ISSN 1645-6726

Instituto Nacional de Estatística
Statistics Portugal

REVSTAT

Statistical Journal

Volume 8, No.2
November 2010

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

- **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 21 842 6364
    Web site: http://www.ine.pt
    Customer Support Service
    (National network) : 808 201 808
    Other networks: + 351 22 605 07 48
- **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**
  - *350 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT 5% included]

- Single issue ……. **€ 8**
- Annual subscription (No. 1, No. 2) ……. **€ 13**

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 8, Number 2, November 2010, 81–103

AUTOREGRESSIVE SEQUENCES VIA LÉVY PROCESSES

Author: Nadjib Bouzar
– Department of Mathematics and Computer Science,
University of Indianapolis, U.S.A.
nbouzar@uindy.edu

Received: November 2009 Revised: May 2010 Accepted: May 2010

Abstract:

- We use Lévy processes to develop a family of first-order autoregressive sequences of random variables with values in $\mathbf{R}_+$, called C-AR(1) processes. We obtain various distributional and regression properties for these processes and we establish a limit theorem that leads to the property of stationarity. A connection between stationarity of C-AR(1) processes and the notion of C-self-decomposability of van Harn and Steutel (1993) is discussed. A number of stationary C-AR(1) processes with specific marginals are presented and are shown to generalize several existing $\mathbf{R}_+$-valued AR(1) models. The question of time reversibility is addressed and some examples are discussed.

Key-Words:

- *stationarity; semigroup of cumulant generating functions; self-decomposability; stability; time-reversibility.*

AMS Subject Classification:

- 60G10, 62M10.

1. INTRODUCTION

In recent years, several authors proposed generalized first-order autoregressive (or AR(1)) models with marginal distributions on $\mathbf{R}_+ := [0, \infty)$. Lewis *et al.* (1989) constructed gamma AR(1) processes with random coefficients based on the beta-gamma transformation. As an application, they analyzed inter-failure times of a computer system. Sim (1990) introduced a generalized multiplication based on a conditional compound Poisson distribution to construct a gamma AR(1) processes. Al-Osh and Alzaid (1993) provided extensions of the gamma models in Lewis *et al.* (1989) via the Gamma-Dirichlet transformation. Grunwald *et al.* (2000) introduced the family of conditional linear AR(1) (CLAR(1)) models. A CLAR(1) process is a Markov process $(X_n,\, n \geq 0)$ such that the conditional expectation $E(X_n|X_{n-1})$ is linear affine. The authors fitted a CLAR(1) model to rainfall data. Zhu (2002) introduced a class of generalized AR(1) (GAR(1)) processes with marginal distributions on $\mathbf{R}_+$. As an application, Zhu (2002) fitted a GAR(1) process with a gamma marginal distribution to ozone data. Darolles *et al.* (2006) introduced a general class of compound autoregressive AR(1) (CAR(1)) processes for non-Gaussian time series. CAR(1) processes are specified in terms of their conditional Laplace transforms.

The aim of this paper is to develop a class of AR(1) sequences of random variables (rv's) with values in $\mathbf{R}_+$ by way of Lévy processes, or processes with stationary independent increments. Our starting point is a continuous convolution semigroup of cumulant generating functions denoted by $C = (C_t,\, t \geq 0)$ and its related operator $\odot_C$ (the definition is recalled below) introduced by van Harn and Steutel (1993). The equation governing our models (equation (2.1)) is analogous to the one describing the standard AR(1) process, with the operator $\odot_C$ replacing the standard multiplication. We obtain various distributional and regression properties for these models and we discuss conditions that lead to stationarity and time reversibility. A number of stationary C-AR(1) processes with specific marginals are presented and are shown to generalize several existing models. The paper is organized as follows. In Section 2 we introduce C-AR(1) processes and give their representation in terms of independent sequences of $\mathbf{R}_+$-valued Lévy processes. We describe their various properties and obtain a limit theorem that leads to the property of stationarity for C-AR(1) processes. We also discuss a connection between the concept of C-self-decomposability of van Harn and Steutel (1993) and stationarity of C-AR(1) processes. In Section 3 we present a number of specific stationary solutions for C-AR(1) processes. Characterizations of their marginal distributions are obtained and some examples are discussed. The question of time reversibility of C-AR(1) processes is addressed in Section 4.

In the rest of this section we recall some definitions and results that are needed in the sequel. For proofs and further details we refer to Hansen (1989), van Harn and Steutel (1993), and Steutel and van Harn (2004).

The Laplace–Stieltjes transform (LST) of an $\mathbf{R}_+$-valued rv X with distribution function $F(x)$ is denoted by ϕ_X:

$$\phi_X(\tau) = \int_0^\infty e^{-\tau x}\, dF(x) \qquad (\tau \geq 0)\ .$$

$C = (C_t;\, t \geq 0)$ will denote a continuous composition semigroup of cumulant generating functions (cgf's): for every $t \geq 0$, $C_t = -\ln L_t$ for some infinitely divisible LST L_t, $C_t \not\equiv 1$, and $\delta_C = -\ln(-L_1'(0)) > 0$. For any $\tau \geq 0$,

$$C_0(\tau) = \tau;\ \ C_s \circ C_t(\tau) = C_{s+t}(\tau)\ \ (s,t \geq 0);\ \ \lim_{t \downarrow 0} C_t(\tau) = \tau;\ \ \lim_{t \to \infty} C_t(\tau) = 0\,. \tag{1.1}$$

The infinitesimal generator U of the semigroup C is defined by

$$U(\tau) = \lim_{t \downarrow 0} \big(C_t(\tau) - \tau\big)/t \qquad (\tau \geq 0)\ ,$$

and satisfies $U(0) = 0$ and $U(\tau) < 0$ for $\tau > 0$. U admits the representation

$$U(\tau) \ = \ a\tau - \frac{1}{2}\,\sigma^2\tau^2 - \int_0^\infty \big(e^{-\tau x} - 1 + \tau x/(1+x^2)\big)\, dN(x)\ ,$$

where a is real, $\sigma \geq 0$, and $N(dx)$ is a Lévy spectral function such that $\int_0^y x^2\, dN(x) < \infty$ for every $y > 0$. Moreover, the following non-explosion condition holds:

$$\left| \int_{0^+}^{y} U(x)^{-1}\, dx \right| = \infty \qquad \text{for sufficiently small } \ y > 0\ .$$

A related function, called the A-function, is defined by

$$A(\tau) = \exp\left\{ \int_\tau^1 \big(U(x)\big)^{-1}\, dx \right\} \qquad \big(\tau \geq 0;\ \ A(0) = 0\big)\,. \tag{1.2}$$

The functions $U(\tau)$ and $A(\tau)$ satisfy for every $t \geq 0$ and $\tau \geq 0$,

$$\frac{\partial}{\partial t} C_t(\tau) = U\big(C_t(\tau)\big) = U(\tau)\, C_t'(\tau) \qquad \text{and} \qquad A\big(C_t(\tau)\big) = e^{-t} A(\tau)\ . \tag{1.3}$$

Moreover,

$$\delta_C = -U'(0) \qquad \text{and} \qquad C_t'(0) = e^{-\delta_C t} \quad (t \geq 0)\ . \tag{1.4}$$

The infinite divisibility of $L_t(\tau)$ and the second part of (1.4) imply that for any $\tau > 0$ and $t > 0$,

$$C_t(\tau) < \tau\ . \tag{1.5}$$

For an $\mathbf{R}_+$-valued rv X and $\alpha \in (0,1)$, the generalized multiplication $\alpha \odot_C X$ is defined by

$$\alpha \odot_C X = Y(X) \ , \tag{1.6}$$

where $(Y(s), s \geq 0)$ is an $\mathbf{R}_+$-valued Lévy process, independent of X, such that $\phi_{Y(1)}(\tau) = \exp(-C_t(\tau))$, $t = -\ln \alpha$. The LST of $\alpha \odot_C X$ is given by

$$\phi_{\alpha \odot_C X}(\tau) = \phi_X\big(C_t(\tau)\big) \ , \qquad t = -\ln \alpha \ . \tag{1.7}$$

If $E(X) < \infty$, then

$$E(\alpha \odot_C X) = \alpha^{\delta_C} E(X) \ . \tag{1.8}$$

2. *C*-AR(1) PROCESSES

Definition 2.1. A sequence $(X_n, n \in \mathbf{Z})$ of $\mathbf{R}_+$-valued rv's is said to be a C-AR(1) process if for any $n \in \mathbf{Z}$,

$$X_n = \alpha \odot_C X_{n-1} + \epsilon_n \ , \tag{2.1}$$

where $0 < \alpha < 1$ and $(\epsilon_n, n \in \mathbf{Z})$ is an iid sequence of $\mathbf{R}_+$-valued rv's that is assumed independent of the Y variables that define the operator $\odot_C$ (see below). $(\epsilon_n, n \in Z)$ is called the innovation sequence of $(X_n, n \in \mathbf{Z})$.

In the remainder of this paper we will at times refer to the single-ended C-AR(1) processes $(X_n, n \geq 0)$ that arises when equation (2.1) is assumed to hold only for $n \geq 0$.

The generalized multiplication $\alpha \odot_C X_{n-1}$ in (2.1) is performed independently for each n. More precisely, we assume the existence of a sequence $(Y^{(n)}(\cdot), n \in \mathbf{Z})$ of iid $\mathbf{R}_+$-valued Lévy processes, independent of $(\epsilon_n, n \in \mathbf{Z})$, such that the LST of $Y^{(n)}(1)$ is

$$\phi_{Y^{(n)}(1)}(\tau) \ = \ \exp\big(-C_t(\tau)\big) \ , \qquad \tau \geq 0 \ , \tag{2.2}$$

where $t = -\ln \alpha$, and (see (1.6))

$$\alpha \odot_C X_{n-1} = Y^{(n-1)}(X_{n-1}) \ , \qquad n \in \mathbf{Z} \ .$$

In terms of LST's, equation (2.1) translates, by way of (1.7), into

$$\phi_{X_n}(\tau) \ = \ \phi_{X_{n-1}}\big(C_t(\tau)\big)\, \phi_\epsilon(\tau) \ , \qquad \tau \geq 0 \ , \tag{2.3}$$

where $\phi_\epsilon(\tau)$ is the marginal LST of $(\epsilon_n, n \in \mathbf{Z})$ and $t = -\ln \alpha$.

Some results on conditional and joint distributions of a C-AR(1) process are given next.

Proposition 2.1. *Let $(X_n, n \geq 0)$ be a C-AR(1) process for some $\alpha \in (0,1)$. Let $t = -\ln \alpha$. The following assertions hold for any $n \geq 1$.*

(**i**) *The conditional LST of X_n given $X_{n-1} = x$, $x \geq 0$, is*

$$\phi_{X_n|X_{n-1}=x}(\tau) = \exp\big(-x\, C_t(\tau)\big)\, \phi_\epsilon(\tau)\ , \qquad \tau \geq 0\ . \tag{2.4}$$

(**ii**) *The bivariate joint LST of (X_{n-1}, X_n) is given by*

$$\phi_{(X_{n-1},X_n)}(\tau_1,\tau_2) = \phi_\epsilon(\tau_2)\, \phi_{X_{n-1}}\big(\tau_1 + C_t(\tau_2)\big)\,. \tag{2.5}$$

(**iii**) *More generally, the joint LST of $(X_1, X_2, ..., X_n)$ can be found recursively by*

$$\phi_{(X_1,...,X_n)}(\tau_1,...,\tau_n) = \phi_\epsilon(\tau_n)\, \phi_{(X_1,...,X_{n-1})}\big(\tau_1,...,\tau_{n-2},\tau_{n-1}+C_t(\tau_n)\big)\,. \tag{2.6}$$

Proof: (i) follows from (2.1) and the fact that $\phi_{Y^{(n)}(x)}(\tau) = \exp(-xC_t(\tau))$. To show (ii), we recall that the joint LST $\phi_{(X_{n-1},X_n)}(\tau_1,\tau_2)$ of (X_{n-1}, X_n) is defined by

$$\phi_{(X_{n-1},X_n)}(\tau_1,\tau_2) = E\Big(e^{-(\tau_1 X_{n-1}+\tau_2 X_n)}\Big)\,, \qquad \tau_1 \geq 0,\ \ \tau_2 \geq 0\ .$$

It can be rewritten as

$$\phi_{(X_{n-1},X_n)}(\tau_1,\tau_2) = E\Big(e^{-\tau_1 X_{n-1}}\, E\big(e^{-\tau_2 X_n}|X_{n-1}\big)\Big)\,,$$

which, combined with (2.4), yields

$$\phi_{(X_{n-1},X_n)}(\tau_1,\tau_2) = E\Big(e^{-(\tau_1+C_t(\tau_2))X_{n-1}}\, \phi_\epsilon(\tau_2)\Big)\,,$$

which, in turn, implies (2.5). The exact same argument establishes (2.6). The details are omitted. □

We note, by definition, that any C-AR(1) process is necessarily a Markov process. Moreover, by using (2.3) recursively (and the fact that $(C_t, t \geq 0)$ is a semigroup), it can be shown that a C-AR(1) process $(X_n, n \in \mathbf{Z})$ admits the following representation for any $k \geq 1$,

$$X_n \stackrel{d}{=} \alpha^k \odot_C X_{n-k} + \sum_{i=0}^{k-1} \alpha^i \odot_C \epsilon_{n-i}\ , \qquad n \in \mathbf{Z}\ . \tag{2.7}$$

Basic regression properties of C-AR(1) processes are gathered in the following proposition.

Proposition 2.2. *Assume $U''(0) < \infty$. Let $(X_n, n \in \mathbf{Z})$ be a C-AR(1) process (for some $0 < \alpha < 1$) such that $E(X_n) < \infty$ and $E(X_n^2) < \infty$ for any $n \in \mathbf{Z}$, $\mu_\epsilon = E(\epsilon_0) < \infty$ and $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_0) < \infty$.*

(**i**) *The regression of X_n on X_{n-1} is linear:*

$$E(X_n|X_{n-1}) = \alpha^{\delta_C} X_{n-1} + \mu_\epsilon \ , \qquad n \in \mathbf{Z} \ . \tag{2.8}$$

(**ii**) *The conditional variance of X_n given X_{n-1} is linear:*

$$\mathrm{Var}(X_n|X_{n-1}) = BX_{n-1} + \sigma_\epsilon^2 \ , \qquad n \in \mathbf{Z} \ . \tag{2.9}$$

where $B = \left(1 - \frac{U''(0)}{U'(0)}\right) \alpha^{\delta_C}(1 - \alpha^{\delta_C})$.

(**iii**) *For any $n \in \mathbf{Z}$ and $k \geq 0$, the covariance at lag k, $\Gamma_n(k) = \mathrm{cov}(X_{n-k}, X_n)$ of $(X_n, n \in \mathbf{Z})$ is*

$$\Gamma_n(k) = \alpha^{k\delta_C} \mathrm{Var}(X_{n-k}) \ . \tag{2.10}$$

(**iv**) *For any $n \in \mathbf{Z}$ and $k \geq 0$,*

$$E(X_n) = \alpha^{k\delta_C} E(X_{n-k}) + \mu_\epsilon \sum_{i=o}^{k-1} \alpha^{i\delta_C} \ , \tag{2.11}$$

and

$$\mathrm{Var}(X_n) = \alpha^{2k\delta_C} \mathrm{Var}(X_{n-k}) + B \sum_{i=1}^{k} \alpha^{2(i-1)\delta_C} E(X_{n-i}) + \sigma_\epsilon^2 \sum_{i=1}^{k} \alpha^{2(i-1)\delta_C}, \tag{2.12}$$

where the constant B is as in (2.9) above.

Proof: We note that for $x \geq 0$

$$E(X_n|X_{n-1} = x) = -\phi'_{X_n|X_{n-1}=x}(0) \ ,$$

where $\phi_{X_n|X_{n-1}=x}(\tau)$ is given by (2.4). By differentiating (2.4) and using (1.4) we obtain (2.8). By differentiating twice (w.r.t. τ) the expression $U(C_t(\tau)) = C_t'(\tau)\, U(\tau)$ $(t = -\ln \alpha)$ and letting $\tau \downarrow 0$, we obtain via (1.4), $C_t''(0) = \alpha^{\delta_C}(\alpha^{\delta_C} - 1) \cdot$ $\cdot U''(0)/U'(0)$. Moreover,

$$E(X_n^2|X_{n-1} = x) = \phi''_{X_n|X_{n-1}=x}(0) \ ,$$

and

$$\mathrm{Var}(X_n|X_{n-1} = x) = \phi''_{X_n|X_{n-1}=x}(0) - \left(\phi'_{X_n|X_{n-1}=x}(0)\right)^2 .$$

Direct calculations, along with (1.4) and the formula for $C_t''(0)$ found above, leads to (2.9). Equation (2.10) is obtained by applying a conditioning argument to (2.7). Finally, (2.11) and (2.12) are easily derived from (2.8) and (2.9). □

The following result demonstrates the existence of a stationary C-AR(1) process.

Theorem 2.1. *Let $(X_n, n \geq 0)$ be a single-sided C-AR(1) process with coefficient $\alpha \in (0,1)$. Then $(X_n, n \geq 0)$ admits a proper limit distribution as $n \to \infty$ if and only if*

$$\int_0^y \frac{1-\phi_\epsilon(x)}{x - C_t(x)}\, dx < \infty\ , \qquad t = -\ln \alpha\ , \tag{2.13}$$

for some $y > 0$, and therefore for all $y > 0$.

Proof: We combine a Poisson mixture argument due to van Harn and Steutel (1993) and a convergence result for branching processes with immigration due to Foster and Williamson (1971). An induction argument based on (2.3) leads to

$$\phi_{X_n}(\tau) \,=\, \phi_{X_0}\big(C_{nt}(\tau)\big) \prod_{j=0}^{n-1} \phi_\epsilon\big(C_{jt}(\tau)\big)\,, \qquad t = -\ln \alpha, \quad \tau \geq 0, \quad n \geq 1\ .$$

Therefore, the sequence $(\phi_{X_n}(\tau), n \geq 0)$ is decreasing for every $\tau \geq 0$. It follows that

$$\phi(\tau) = \lim_{n\to\infty} \phi_{X_n}(\tau) \tag{2.14}$$

exists for each $\tau \geq 0$.

Let $\lambda > 0$. By van Harn and Steutel (1993), the functions $F^{(\lambda)} = (F_t^{(\lambda)};\, t \geq 0)$ defined by

$$F_t^{(\lambda)}(z) \,=\, 1 - \frac{1}{\lambda}\, C_t\big(\lambda(1-z)\big) \qquad (z \in [0,1])\ . \tag{2.15}$$

form a continuous composition semigroup of probability generating functions (pgf's), with

$$\frac{\partial}{\partial z} F_t^{(\lambda)}(z)\bigg|_{z=1} \,=\, e^{-\delta_c t} \tag{2.16}$$

for each $t > 0$.

Consider the branching process with immigration $\big(Y_n^{(\lambda)}, n \geq 0\big)$

$$Y_n^{(\lambda)} \,=\, \sum_{i=1}^{Y_{n-1}^{(\lambda)}} W_i^{(\lambda)} + \epsilon_n^{(\lambda)}\ , \tag{2.17}$$

where $(W_n^{(\lambda)}, n \geq 1)$ and $(\epsilon_n^{(\lambda)}, n \geq 0)$ are independent sequences of iid $\mathbf{Z}_+$-valued rv's with respective marginal pgf's $F_t^{(\lambda)}(z)$ and $P_{\epsilon^{(\lambda)}}(z) = \phi_\epsilon(\lambda(1-z))$, $0 \leq z \leq 1$

and $t = -\ln\alpha$. Moreover, $Y_0^{(\lambda)}$ has pgf $P_{Y_0^{(\lambda)}}(z) = \phi_{X_0}(\lambda(1-z))$ and is independent of $(W_n^{(\lambda)}, n \geq 1)$ and $(\epsilon_n^{(\lambda)}, n \geq 0)$. By (2.3), 2.15), (2.17), and an induction argument, we have

$$P_{Y_n^{(\lambda)}}(z) = \phi_{X_n}\big(\lambda(1-z)\big)\,, \qquad n \geq 0,\ \ 0 \leq z \leq 1\,. \tag{2.18}$$

By (2.16), $(Y_n^{(\lambda)}, n \geq 0)$ is a sub-critical branching process.

Let's now assume that (2.13) holds. Simple calculations show that

$$\int_0^1 \frac{1 - P_{\epsilon^{(\lambda)}}(x)}{F_t^{(\lambda)}(x) - x}\, dx = \int_0^\lambda \frac{1-\phi_\epsilon(x)}{x - C_t(x)}\, dx < \infty\,, \qquad t = -\ln\alpha\,. \tag{2.19}$$

By the main Theorem of Foster and Williamson (1971), case (iii), $(Y_n^{(\lambda)}, n \geq 0)$ has a proper limit distribution, as $n \to \infty$, whose pgf is (by (2.18))

$$P^{(\lambda)}(z) = \lim_{n\to\infty} \phi_{X_n}\big(\lambda(1-z)\big)\,, \qquad 0 \leq z \leq 1\,. \tag{2.20}$$

It follows by (2.14) that for every $\lambda > 0$, $\phi(\lambda(1-z)) = P^{(\lambda)}(z)$, $0 \leq z \leq 1$. Therefore, by Lemma A.6 in van Harn and Steutel (1993), $\phi(\tau)$ is the LST of a distribution on $\mathbf{R}_+$.

Conversely, assume that $(X_n, n \geq 0)$ admits a proper limit distribution as $n \to \infty$. The limit LST $\phi(\tau)$ is given by (2.14). Hence, for every $\lambda > 0$, $(Y_n^{(\lambda)}, n \geq 0)$ has a proper limit distribution whose pgf is $P^{(\lambda)}(z)$ of (2.20). We deduce by the converse of the Theorem in Foster and Williamson (1971), case (iii), that (2.19) holds for every $\lambda > 0$, which in turn implies (2.13). □

Since a single-sided C-AR(1) process $(X_n, n \geq 0)$ is Markovian, it is stationary if and only if it is started with its limit distribution. By Theorem 2.1, such a limit distribution exists if condition (2.13) holds. We note that a single-sided process can be extended to a doubly-infinite stationary process (see the proof of Theorem 2.2 below).

Next, we explore the relationship between self-decomposability and stationary C-AR(1) processes. A distribution on $\mathbf{R}_+$ with LST $\phi(\tau)$ is said to be C-self-decomposable (van Harn and Steutel, 1993) if for any $t > 0$, there exists an LST $\phi_t(\tau)$ such that

$$\phi(\tau) = \phi\big(C_t(\tau)\big)\,\phi_t(\tau)\,, \qquad \tau \geq 0\,. \tag{2.21}$$

Any C-self-decomposable distribution can arise as the marginal distribution of a stationary C-AR(1) process. More precisely, we have the following result.

Theorem 2.2. *Let $\phi(\tau)$ be the LST of a C-self-decomposable distribution. For any $\alpha \in (0,1)$, there exists a stationary C-AR(1) process $(X_n, n \in \mathbf{Z})$ whose marginal distribution has LST $\phi(\tau)$.*

Proof: Let $\alpha \in (0,1)$ and $t = -\ln \alpha$. By the Kolmogorov extension theorem (Breiman, 1968), there exists a probability space $(\Omega, \mathcal{F}, \mu)$ on which one can define an array $(Y^{(n)}(\cdot),\ n \geq 0)$ of iid $\mathbf{R}_+$-valued Lévy processes such that $Y^{(n)}(1)$ has LST (2.2), a sequence of iid rv's $(\epsilon_n,\ n \geq 0)$ with common LST $\phi_\epsilon(\tau) = \phi_t(\tau)$ of (2.21), and a rv X_0 with LST $\phi(\tau)$, with the further property that $(Y^{(n)}(\cdot),\ n \geq 0)$, $(\epsilon_n,\ n \geq 0)$, and X_0 are independent. We then construct a single-ended INAR1 process $(X_n,\ n \geq 0)$ via equation (2.1). This implies that for every $n \geq 1$, the LST $\phi_{X_n}(\tau)$ of X_n satisfies (2.3), with $\phi_{X_0}(\tau) = \phi(\tau)$. It follows by (2.3) and (2.21) that $\phi_{X_n}(\tau) = \phi(\tau)$ for every $n \geq 0$. Therefore, the X_n's are identically distributed. Since $(X_n,\ n \geq 0)$ is a Markov process, its stationarity ensues. The existence of the doubly infinite extension $(X_n, n \in \mathbf{Z})$ follows from Proposition 6.5, page 105, in Breiman (1968). □

Next, we state a representation theorem for stationary C-AR(1) processes. The proof follows easily from (2.7) and is omitted.

Theorem 2.3. *Any stationary C-AR(1) process $(X_n, n \in \mathbf{Z})$ admits the following (infinite order) moving average representation for some $0 < \alpha < 1$:*

$$X_n \stackrel{d}{=} \sum_{i=0}^{\infty} \alpha^i \odot_C \epsilon_{n-i}\ , \qquad n \in \mathbf{Z}\ , \tag{2.22}$$

where the convergence of the series is in the weak sense.

The mean, variance, and autocorrelation function (ACRF) of a stationary C-AR(1) process follow straightforwardly from Proposition 2.2.

Proposition 2.3. *Assume $U''(0) < \infty$. Let $(X_n, n \in \mathbf{Z})$ be a stationary C-AR(1) process (for some $0 < \alpha < 1$) such that $E(X_0) < \infty$, $E(X_0^2) < \infty$, $\mu_\epsilon = E(\epsilon_0) < \infty$ and $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_0) < \infty$. Then*

(**i**) *For any $n \in \mathbf{Z}$,*

$$E(X_n) = \mu_\epsilon (1 - \alpha^{\delta_C})^{-1}\ ,$$

and

$$\mathrm{Var}(X_n) = \frac{\left(1 - \frac{U''(0)}{U'(0)}\right)\alpha^{\delta_C}\mu_\epsilon + \sigma_\epsilon^2}{1 - \alpha^{2\delta_C}}\ .$$

(**ii**) *For any $k \geq 0$ and $n \in \mathbf{Z}$, the correlation coefficient of (X_{n-k}, X_n) is*

$$\rho(k) = \alpha^{k\delta_C}\ . \tag{2.23}$$

We note that the $ACRF$ of a stationary C-AR(1) process, as given by (2.23), has the same form as that of the standard $AR(1)$ process. It decays exponentially at lag k. However, unlike the standard $AR(1)$ case, $\rho(k)$ remains always positive.

3. STATIONARY C-AR(1) PROCESSES WITH SPECIFIC MARGINAL DISTRIBUTIONS

In this section we present several stationary solutions for C-AR(1) processes.

An $\mathbf{R}_+$-valued rv X is said to have a C-stable distribution with exponent $\gamma > 0$ if there exists a sequence of iid $\mathbf{R}_+$-valued rv's $(X_i, i \geq 0)$, $X_i \stackrel{d}{=} X$ for all i, such that for any $n > 0$,

$$X \stackrel{d}{=} n^{-1/\gamma} \odot_C \sum_{i=1}^{n} X_i \ .$$

C-stable distributions are C-self-decomposable and exist only when $0 < \gamma \leq \delta_C$ (van Harn and Steutel, 1993). Moreover, the LST $\phi(\tau)$ of a C-stable distribution with exponent $\gamma \in (0, \delta_C]$ admits the canonical representation

$$\phi(\tau) = \exp\big[-\lambda A(\tau)^\gamma\big] \ , \qquad \tau \geq 0 \tag{3.1}$$

for some $\lambda > 0$, where $A(\tau)$ is given in (1.2).

It follows by Theorem 2.2 that for every $0 < \alpha < 1$, there exists a stationary C-AR(1) process $(X_n, n \in \mathbf{Z})$ with a C-stable marginal distribution with exponent γ $(0 < \gamma \leq \delta_C)$. The marginal distribution of the innovation sequence $(\epsilon_n, n \in \mathbf{Z})$, obtained by solving for ϕ_ϵ in (2.3) and by using (1.3), is also C-stable with exponent γ and has LST

$$\phi_\epsilon(\tau) = \exp\big[-\lambda(1-\alpha^\gamma) A(\tau)^\gamma\big] \ . \tag{3.2}$$

Moreover, it can be shown (see van Harn and Steutel, 1993) that stationary C-AR(1) processes whose marginal is C-stable with finite mean arise only in the case $\gamma = \delta_C$ and $A'(0) < \infty$. The process has finite variance if $A''(0) < \infty$.

We have shown via (3.2) (by letting $\alpha = e^{-t}$) that the LST $\phi(\tau)$ of the marginal distribution of a stationary C-stable C-AR(1) process satisfies the following property: for any $t > 0$, there exist $\lambda(t) > 0$ such that

$$\ln \phi(\tau) = \lambda(t) \ln \phi[C_t(\tau)] \ , \qquad \tau \geq 0 \ . \tag{3.3}$$

It turns out that this property characterizes such processes.

Theorem 3.1. *A function $\phi(\tau)$ on $\mathbf{R}_+$ is the LST of a C-stable distribution with some exponent $\gamma \in (0, \delta_C]$ if and only if for any $t > 0$, there exists $\lambda(t) > 0$ such that (3.3) holds for every $\tau \geq 0$. The function $\lambda(t)$ is necessarily of the form $\lambda(t) = e^{\gamma t}$.*

Proof: The 'only if' part follows from the preceding discussion. We prove only the 'if' part. Let $\psi(\tau) = \ln \phi(\tau)/\ln \phi(1)$. By (3.3), we have for any $t > 0$ and $\tau \geq 0$ (note $\lambda(t) = 1/\psi(C_t(1))$),

$$\psi\big(C_t(\tau)\big) = \psi\big(C_t(1)\big)\,\psi(\tau)\,, \qquad \tau \geq 0\,. \tag{3.4}$$

By differentiating (3.4) w.r.t. t, we obtain

$$\frac{\partial}{\partial t} C_t(\tau)\,\psi'\big(C_t(\tau)\big) = \frac{\partial}{\partial t} C_t(1)\,\psi'\big(C_t(1)\big)\,\psi(\tau)\,, \qquad \tau \geq 0\,.$$

Using $\frac{\partial}{\partial t} C_t(\tau) = U(C_t(\tau))$ and letting $t \downarrow 0$, it follows by (1.1) that

$$\frac{\psi'(\tau)}{\psi(\tau)} = \frac{U(1)}{U(\tau)}\,\psi'(1)\,, \qquad \tau \geq 0\,,$$

whose solution is $\psi(\tau) = A(\tau)^\gamma$ where $\gamma = -\psi'(1)\,U(1) > 0$. Hence, $\phi(\tau)$ has the form (3.1). Since $\phi(\tau)$ is an LST, γ must satisfy $\gamma \leq \delta_F$ (it follows by adapting to our case the argument in the proof of Lemma 4.2. in van Harn and Steutel (1993)). The form of $\lambda(t)$ results from its uniqueness and the 'only if' part. □

Next, we present a stationary C-AR(1) process with a C-geometric stable marginal distribution.

An $\mathbf{R}_+$-valued rv X is said to have a C-geometric stable distribution if for any $p \in (0,1)$, there exists $\alpha(p) \in (0,1)$ such that

$$X \stackrel{d}{=} \alpha(p) \odot_C \sum_{i=1}^{N_p} X_i\,,$$

where $(X_i, i \geq 1)$ is a sequence of iid $\mathbf{R}_+$-valued rv's, $X_i \stackrel{d}{=} X$, N_p has the geometric distribution with parameter p, and $(X_i, i \geq 1)$ and N_p are independent (Bouzar, 1999). C-geometric stable distributions are C-self-decomposable and their LST's admit the canonical representation

$$\phi(\tau) = \big(1 + \lambda\,A(\tau)^\gamma\big)^{-1}, \qquad \tau \geq 0\,, \tag{3.5}$$

for $0 < \gamma \leq \delta_C$ and $\lambda > 0$. We will refer to a distribution with LST (3.5) as C-geometric stable with exponent γ.

By Theorem 2.2, for every $\alpha \in (0,1)$, there exists a stationary C-AR(1) process $(X_n, n \in \mathbf{Z})$ with a C-geometric stable marginal distribution with LST

(3.5). Its innovation sequence $(\epsilon_n, n \in \mathbf{Z})$ has marginal LST (obtained by solving for $\phi_\epsilon(\tau)$ in (2.3) and by using (1.3))

$$\phi_\epsilon(\tau) = \alpha^\gamma + (1-\alpha^\gamma)\big(1+\lambda\, A(\tau)^\gamma\big)^{-1}, \qquad \tau \geq 0\ , \tag{3.6}$$

where $0 < \gamma \leq \delta_C$ and $\lambda > 0$.

It follows from (3.6) that a stationary C-AR(1) process $(X_n, n \in \mathbf{Z})$ with a C-geometric stable marginal distribution can be written as

$$X_n = \alpha \odot_C X_{n-1} + I_n E_n\ , \qquad n \in \mathbf{Z}\ , \tag{3.7}$$

where $(I_n, n \in \mathbf{Z})$ and $(E_n, n \in \mathbf{Z})$ are independent sequences of iid rv's such that I_n is Bernoulli$(1-\alpha^\gamma)$ and E_n has the same distribution as X_n.

A stationary C-AR(1) process with a C-geometric stable marginal distribution has finite mean only if $\gamma = \delta_F$ and $A'(0) < \infty$. It has a finite variance if $A''(0) < \infty$.

We have in fact shown by the above argument (and by letting $\alpha = e^{-t}$) that the LST $\phi(\tau)$ of the marginal distribution of a stationary C-geometric stable C-AR(1) process satisfies the following property: for any $t > 0$ there exists $c(t) \in (0,1)$ such that

$$\phi(\tau) \,=\, \phi\big(C_t(\tau)\big)\,\Big(c(t) + \big(1-c(t)\big)\,\phi(\tau)\Big)\,, \qquad \tau \geq 0\ . \tag{3.8}$$

We show next that the converse is true.

Theorem 3.2. *A function $\phi(\tau)$ on $\mathbf{R}_+$ is the LST of a C-geometric stable distribution with some exponent $\gamma \in (0, \delta_C]$ if and only if for any $t > 0$ there exists $c(t) \in (0,1)$ such that (3.8) holds. The function $c(t)$ is necessarily of the form $c(t) = e^{-\gamma t}$.*

Proof: The 'only if' part was established in the preceding discussion. We show the 'if' part. Rewriting $\phi(\tau) = (1+\psi(\tau))^{-1}$, it follows by (3.8) that for any $t > 0$, there exists $c(t) \in (0,1)$ such that

$$\psi\big(C_t(\tau)\big) \,=\, c(t)\,\psi(\tau)\ , \qquad \tau \geq 0\ . \tag{3.9}$$

Using the exact same argument as the one in the proof of Theorem 3.1 (following (3.4)), we have $\psi(\tau) = \lambda\, A(\tau)^\gamma$ for some $0 < \gamma \leq \delta_C$ and $\lambda > 0$. The form of $c(t)$ follows from its uniqueness and the 'only if' part. □

We define next a compound gamma distribution and construct the corresponding stationary C-AR(1) process.

Let $0 < \gamma \leq \delta_C$, $\lambda > 0$, and $r > 0$. An $\mathbf{R}_+$-valued rv X is said to have a C-compound gamma (γ, λ, r) distribution if its LST has the form

$$\phi(\tau) = \left(1 + \lambda\, A(\tau)^\gamma\right)^{-r}, \qquad \tau \geq 0\ . \tag{3.10}$$

Note that $\phi(\tau)$ indeed results from the compounding of C-stable distributions (with LST (3.1)) by a gamma distribution (with LST $\phi_1(\tau) = (1+\tau)^{-r}$). The special case $r = 1$ in (3.10) gives the C-geometric stable distribution. van Harn and Steutel (1993) showed that C-compound gamma distributions are C-self-decomposable (see also Proposition 3.1 below) and arise as solutions to stability equations for $\mathbf{R}_+$-valued processes with stationary independent increments.

Let $0 < \gamma \leq \delta_C$, $\lambda > 0$, and $r > 0$. By Theorem 2.2, for every $\alpha \in (0,1)$, there exists a stationary C-AR(1) process $(X_n, n \in \mathbf{Z})$ with a C-compound gamma (γ, λ, r) marginal distribution. Its innovation sequence $(\epsilon_n, n \in \mathbf{Z})$ has LST

$$\phi_\epsilon(\tau) = \left(\frac{1 + \lambda\, \alpha^\gamma A(\tau)^\gamma}{1 + \lambda\, A(\tau)^\gamma}\right)^r, \qquad \tau \geq 0\ . \tag{3.11}$$

It can be shown by a straightforward calculations that ϵ_n with LST (3.11) has the representation

$$\epsilon \stackrel{d}{=} \sum_{i=1}^{N} (\alpha^{U_i}) \odot_C W_i\ , \tag{3.12}$$

where $(W_i, i \geq 0)$ is a sequence of iid C-geometric stable rv's (with LST (3.5)), $(U_i, i \geq 0)$ are iid uniform $(0,1)$ rv's, and N is Poisson with mean $-r\,\gamma \ln \alpha$, with all these variables being independent. This allows for a shot-noise interpretation of the process that is similar to the one given by Lawrance (1982) for the gamma AR(1) process. A shot-noise process is defined by

$$X(t) \;=\; \sum_{m=N(-\infty)}^{N(t)} \alpha^{t-\tau_m} \odot_C W_m\ , \tag{3.13}$$

where $(W_m, m \geq 0)$ are $\mathbf{R}_+$-valued iid rv's (amplitudes of the shots) and $(N(t), t \geq 0)$ is a Poisson process with occurrence times at τ_m. If the W_m's have their common LST given by (3.5) and $N(t)$ has rate $-r\,\gamma \ln \alpha$, then $X(t)$ of (3.13) sampled at $n = 0, \pm 1, \pm 2, ...$ gives another representation of the stationary C-AR(1) process (2.1) with a C-compound gamma (γ, λ, r) marginal distribution. The proof of this fact is an adaptation of Lawrance's (1982) argument and the details are omitted.

Other representations of the innovation variable ϵ_n for a C-AR(1) process with a C-compound gamma (γ, λ, r) marginal distribution can be obtained by adapting the ones derived by McKenzie (1987) for an integer-valued AR(1) process and by Walker (2000) for the gamma AR(1) process of Gaver and Lewis (1980).

As in the previously seen models, a stationary C-AR(1) process with a C-compound gamma (γ, λ, r) marginal distribution has finite mean only if $\gamma = \delta_C$ and $A'(0) < \infty$. It has a finite variance if $A''(0) < \infty$.

The C-self-decomposability of the C-geometric stable distributions (with LST (3.5)) and the C-compound gamma distributions (with LST (3.10)) can be derived from the following, more general, result.

Proposition 3.1. *Let $\varphi(\tau)$ be the LST of a self-decomposable distribution on $\mathbf{R}_+$ with respect to the usual multiplication. Then the compound distribution on $\mathbf{R}_+$ with LST*

$$\phi(\tau) = \varphi\big(\lambda A(\tau)^\gamma\big) , \qquad \tau \geq 0 , \tag{3.14}$$

for some $0 < \gamma \leq \delta_C$ and $\lambda > 0$, is C-self-decomposable.

Proof: We note first that $\phi(\tau)$ is indeed an LST. Specifically, it is the LST of the $\mathbf{R}_+$-valued rv $Y = X(T)$ where $(X(t), t \geq 0)$ is an $\mathbf{R}_+$-valued Lévy process such that $X(1)$ has LST (3.1) and T is a rv (independent of $(X(t), t \geq 0)$) with LST $\varphi(\tau)$ (see Steutel and van Harn (2004), Chapter I, Section 3, for a discussion on compound distributions of the type (3.14)). By self-decomposability with respect to the usual multiplication, we have for every $\tau \geq 0$ and $t > 0$

$$\varphi(\tau) = \varphi(e^{-\gamma t}\tau)\, \varphi_{\gamma,t}(\tau) , \tag{3.15}$$

for some LST $\varphi_{\gamma,t}(\tau)$. Combining equations (1.3), (3.14) and (3.15), yields for every $\tau \geq 0$ and $t > 0$,

$$\phi(\tau) = \varphi\big(e^{-\gamma t}\lambda\, A(\tau)^\gamma\big)\, \varphi_{\gamma,t}\big(\lambda\, A(\tau)^\gamma\big) = \varphi\big(\lambda\, A\big(C_t(\tau)\big)\big)\, \varphi_{\gamma,t}\big(\lambda\, A(\tau)^\gamma\big) .$$

Therefore, (2.21) holds for $\phi(\tau)$, with $\phi_t(\tau) = \varphi_{\gamma,t}(\lambda A(\tau)^\gamma)$. The same argument we used above to show that $\phi(\tau)$ is an LST can be repeated to conclude $\phi_t(\tau)$ is also an LST. □

The LST's described by (3.5) and (3.10) are special cases of (3.14). In this case, $\varphi(\tau) = (1+\tau)^{-r}$ (with $r = 1$ for (3.5)). Steutel and van Harn (2004), Chapter 5, Section 9, offer a multitude of examples of LST's $\varphi(\tau)$ from which one can construct stationary C-AR(1) processes (by combining Proposition 3.1 and Theorem 2.2).

Next, we present a random coefficient stationary C-AR(1) process with the C-compound gamma marginal distribution of (LST) (3.10) and with an innovation sequence that is simpler than (3.12) or (3.13).

Let B be a rv taking values in $(0, 1)$ and X an $\mathbf{R}_+$-valued rv independent of B. The random coefficient operator $B \odot_C X$ is defined via its LST by the

equation

$$\phi_{B\odot_C X}(\tau) = \int_0^1 \phi_X\big(C_{-\ln b/\delta_C}(\tau)\big)\, dF(b)\,, \tag{3.16}$$

where $F(b)$ is the distribution function of B.

Lemma 3.1. *Let $0 < s < r$, $0 < \gamma \leq \delta_C$, and $\lambda > 0$. Define $\gamma_1 = \frac{\gamma}{\delta_C}$ and note $\gamma_1 \in (0, 1]$. Assume that B has the probability density function*

$$f(b) = \frac{\gamma_1 \Gamma(r)}{\Gamma(s)\,\Gamma(r-s)}\; b^{\gamma_1 s - 1}(1 - b^{\gamma_1})^{r-s-1}\,, \qquad 0 < b < 1\,, \tag{3.17}$$

and that X has the C-compound gamma (γ, λ, r) distribution. Then $B \odot_C X$ has a C-compound gamma (γ, λ, s) distribution.

Proof: Using (3.16), (3.17), and the change of variable $w^{\gamma} = \dfrac{1 - b^{\gamma_1}}{1 + \lambda A(\tau)^{\gamma} b^{\gamma_1}}$, we obtain

$$\phi_{B\odot_C X}(\tau) = \left[\frac{\gamma_1 \Gamma(r)}{\Gamma(s)\,\Gamma(r-s)} \int_0^1 (1 - w^{\gamma_1})^{s-1}\; w^{\gamma_1(r-s)-1}\, dw\right] \big(1 + \lambda\, A(\tau)^{\gamma}\big)^{-s}.$$

Since $\dfrac{\gamma_1 \Gamma(r)}{\Gamma(s)\,\Gamma(r-s)} \displaystyle\int_0^1 (1-w^{\gamma_1})^{s-1} w^{\gamma_1(r-s)-1}\, dw = 1$, the conclusion follows. □

An $\mathbf{R}_+$-valued stochastic process $(X_n, n \geq 0)$ is said to be a random coefficient C-AR(1) process if it satisfies the equation

$$X_n = B_n \odot_C X_{n-1} + \epsilon_n\,, \tag{3.18}$$

where $(B_n, n \geq 1)$ is an iid sequence of rv's with $0 < B_n < 1$ and $(\epsilon_n, n \geq 1)$ is an iid sequence of $\mathbf{R}_+$-valued rv's. Moreover, it is assumed that B_n, X_{n-1}, and ϵ_n are mutually independent.

Theorem 3.3. *Let $0 < s < r$, $0 < \gamma \leq \delta_C$, and $\lambda > 0$. Let $(X_n, n \geq 0)$ be the random coefficient C-AR(1) process of (3.18) such that B_n has probability density function (3.17) and ϵ_n has a C-compound gamma $(\gamma, \lambda, r - s)$ distribution. If X_0 has a C-compound gamma (γ, λ, r) distribution, then $(X_n, n \geq 0)$ is stationary with a C-compound gamma (γ, λ, r) marginal distribution.*

Proof: We have by (3.18) and Lemma 3.1,

$$\phi_{X_1}(\tau) = \phi_{B_1\odot_C X_0}(\tau)\,\phi_{\epsilon}(\tau) = \big(1+\lambda A(\tau)^{\gamma}\big)^{-s}\big(1+\lambda A(\tau)^{\gamma}\big)^{s-r} = \big(1+\lambda A(\tau)^{\gamma}\big)^{-r}.$$

An induction argument shows that X_n has a C-compound gamma (γ, λ, r) for all $n \geq 1$. Since $(X_n, n \geq 1)$ is a Markov process, stationarity ensues. □

We conclude the section by mentioning a family of semigroups of cgf's. For $\theta \in [0,1)$, let

$$C_t^{(\theta)}(\tau) = \frac{\overline{\theta}\, e^{-\overline{\theta}t}\tau}{\overline{\theta} + \theta(1 - e^{-\overline{\theta}t})\tau}\ , \qquad t, \tau \geq 0, \quad \overline{\theta} = 1 - \theta\ . \tag{3.19}$$

It is easy to verify that $C_t^{(\theta)}(\tau)$ has a completely monotone derivative and hence is a cgf. Moreover, straightforward calculations show that the properties in (1.1) hold. Therefore, $C^{(\theta)} = (C_t^{(\theta)}, t \geq 0)$ is a continuous semigroup of cgf's. In this case

$$U^{(\theta)}(\tau) = -\tau(\overline{\theta} + \theta\tau)\,, \qquad A^{(\theta)}(\tau) = \left(\frac{\tau}{\overline{\theta} + \theta\tau}\right)^{1/\overline{\theta}}, \qquad \delta_C^{(\theta)} = \overline{\theta}\ . \tag{3.20}$$

The special case $\theta = 0$ corresponds to the ordinary multiplication. The stationary $C^{(0)}$-AR(1) process with a $C^{(0)}$-stable marginal distribution corresponds to the AR(1) process with the standard stable distribution on $\mathbf{R}_+$ as its marginal. The stationary $C^{(0)}$-AR(1) process with a $C^{(0)}$-geometric stable marginal distribution reduces to the Mittag–Leffler AR(1) process of Jayakumar and Pillai (1993). The stationary $C^{(0)}$-AR(1) process with a $C^{(0)}$-compound gamma (with LST (3.10)) becomes the gamma AR(1) process of Gaver and Lewis (1980).

4. TIME-REVERSIBILITY OF STATIONARY C-AR(1) PROCESSES

A stochastic process $(X_n, n \in \mathbf{Z})$ is said to be time-reversible if for any $n \in \mathbf{Z}$ and $k \geq 0$, $(X_n, X_{n+1}, ..., X_{n+k})$ and $(X_{n+k}, X_{n+k-1}, ..., X_n)$ have the same joint distribution.

Let $(X_n, n \in \mathbf{Z})$ be a C-AR(1) process. By the Markov property, $(X_n, n \in \mathbf{Z})$ is time-reversible if and only if for any $n \in \mathbf{Z}$, (X_{n-1}, X_n) and (X_n, X_{n-1}) have the same joint distribution. $(X_n, n \in \mathbf{Z})$ is time-reversible if and only if for every $n \in \mathbf{Z}$,

$$\phi_{(X_{n-1},X_n)}(\tau_1, \tau_2) = \phi_{(X_{n-1},X_n)}(\tau_2, \tau_1)\ , \qquad \tau_1 \geq 0, \quad \tau_2 \geq 0\ , \tag{4.1}$$

where $\phi_{(X_{n-1},X_n)}(\tau_1, \tau_2)$ is the joint LST of (X_{n-1}, X_n).

By Proposition 2.2-(i), a time-reversible C-AR(1) process $(X_n, n \in \mathbf{Z})$ (such that $E(X_n) < \infty$ and $E(\epsilon_n) < \infty$) possesses the property of linear backward regression. That is, there exist $c > 0$ and $d \geq 0$ such that for any $n \in \mathbf{Z}$,

$$E(X_{n-1}|X_n) = d + cX_n\ . \tag{4.2}$$

We show next that a stationary C-AR(1) process with finite mean and finite variance has the property of linear backward regression only if its LST admits a certain form.

Theorem 4.1. *Let $(X_n, n \in \mathbf{Z})$ be a stationary C-AR(1) process with finite mean and finite variance with the property of linear backward regression (4.2). Further, assume*

$$C_t(1) \sim a e^{-\delta_C t} \qquad (t \to \infty) \,, \tag{4.3}$$

for some constant $a > 0$. Then the marginal distribution of $(X_n, n \in \mathbf{Z})$ is infinitely divisible with LST $\phi(\tau)$ of the form

$$\phi(\tau) = \exp\Big\{ -\int_0^\tau \big(b - \lambda A(x)^{\delta_C}\big)\, dx \Big\} , \tag{4.4}$$

for some $b > 0$ and $\lambda > 0$.

Proof: Let $n \in \mathbf{Z}$ and let $\phi(\tau)$, $\tau \geq 0$, and $g(\tau_1, \tau_2)$, $\tau_1, \tau_2 \geq 0$, be the LST of X_n and joint LST of (X_{n-1}, X_n), respectively. Recall that by the stationarity assumption, both $\phi(\tau)$ and $g(\tau_1, \tau_2)$ are independent of n. By Proposition 2.1-(ii) and equation (2.3), we have for any $\tau_1,\ \tau_2 \geq 0$

$$g(\tau_1, \tau_2) = \phi_\epsilon(\tau_2)\, \phi\big(\tau_1 + C_t(\tau_2)\big) = \frac{\phi\big(\tau_1 + C_t(\tau_2)\big)\, \phi(\tau_2)}{\phi\big(C_t(\tau_2)\big)} \,. \tag{4.5}$$

Differentiating g with respect to τ_1, then setting $\tau_1 = 0$ and $\tau_2 = \tau$, it follows that for any $n \in \mathbf{Z}$,

$$E(X_{n-1} e^{-\tau X_n}) = -\frac{\phi(\tau)}{\phi\big(C_t(\tau)\big)}\, \phi'\big(C_t(\tau)\big) \,, \qquad \tau \geq 0 \,. \tag{4.6}$$

By the property of linear backward regression (see equation (4.2)), we have for some $c > 0$ and $d \geq 0$,

$$E(X_{n-1} e^{-\tau X_n}) = E\big(e^{-\tau X_n} E(X_{n-1}|X_n)\big) = c\, E(X_n e^{-\tau X_n}) + d\, E(e^{-\tau X_n}) \,,$$

for any $n \in \mathbf{Z}$ and $\tau \geq 0$. Noting that $E(X_n e^{-\tau X_n}) = -\phi'(\tau)$, it follows that

$$E(X_{n-1} e^{-\tau X_n}) = d\phi(\tau) - c\, \phi'(\tau) \,, \qquad \tau \geq 0 \,. \tag{4.7}$$

Letting $h(\tau) = \phi'(\tau)/\phi(\tau)$ and combining (4.6) and (4.7) , we obtain

$$c\, h(\tau) - d = h\big(C_t(\tau)\big) \,, \qquad \tau \geq 0 \,.$$

It follows by differentiation that $c\, h'(\tau) = C_t'(\tau)\, h'(C_t(\tau))$. Noting that $h'(0) = \mathrm{Var}(X_n) \neq 0$ (and recalling that $C_t(0) = 0$), it follows that $c = C_t'(0) = e^{-\delta_C t}$, with the second equation following from (1.4). This implies

$$h'(\tau) = e^{\delta_C t} C_t'(\tau)\, h'\big(C_t(\tau)\big) \,, \qquad \tau \geq 0 \,.$$

An induction argument yields for any $n \geq 1$,

$$h'(\tau) = e^{n\delta_C t} h'\big(C_{nt}(\tau)\big) \prod_{j=0}^{n-1} C_t'\big(C_{jt}(\tau)\big) , \qquad \tau \geq 0 .$$

By the semigroup properties (1.1) and (1.3), we have

$$C_t'\big(C_{jt}(\tau)\big) = U\big(C_{(j+1)t}(\tau)\big)/U\big(C_{jt}(\tau)\big) , \qquad j = 0, ..., n-1 .$$

Therefore,

$$h'(\tau) = e^{n\delta_C t} \frac{U\big(C_{nt}(\tau)\big)}{U(\tau)} h'\big(C_{nt}(\tau)\big) , \qquad \tau \geq 0 . \tag{4.8}$$

Calling again on the semigroup properties (1.1) and (1.3), we have for any $\tau \geq 0$,

$$\lim_{n\to\infty} C_{nt}(\tau) = 0 , \qquad \lim_{n\to\infty} \frac{U\big(C_{nt}(\tau)\big)}{C_{nt}(\tau)} = U'(0) = -\delta_C .$$

By Lemma 3.2-(i) in Hansen (1989),

$$\lim_{n\to\infty} \frac{C_{nt}(\tau)}{C_{nt}(1)} = A(\tau)^{\delta_C} , \qquad \tau \geq 0 .$$

Moreover, (4.3) implies

$$\lim_{n\to\infty} e^{n\delta_C t} C_{nt}(1) = a .$$

Therefore, by letting $n \to \infty$ in (4.8), we obtain

$$h'(\tau) = -a\,\delta_C\, h'(0) \frac{A(\tau)^{\delta_C}}{U(\tau)} , \qquad \tau \geq 0 .$$

Since by (1.2) $1/U(\tau) = -A'(\tau)/A(\tau)$, we have

$$h(\tau) - h(0) = \int_0^\tau h'(x)\, dx = a h'(0) \int_0^\tau \delta_C\, A'(x)\, A(x)^{\delta_C - 1} , \qquad \tau \geq 0 ,$$

which implies (note $A(0) = 0$)

$$h(t) = h(0) + a h'(0)\, A(\tau)^{\delta_C} , \qquad \tau \geq 0 ,$$

or $\phi'(\tau)/\phi(\tau) = h(0) + a h'(0)\, A(\tau)^{\delta_C}$. It follows

$$\ln \phi(\tau) = h(0)\tau + a h'(0) \int_0^\tau A(x)^{\delta_C}\, dx , \qquad \tau \geq 0 .$$

The representation (4.4) follows by letting $b = -h(0)$ and $\lambda = a h'(0)$. To show that $\phi(\tau)$ of (4.4) is indeed the LST of an infinitely divisible distribution, let $\psi(\tau) = b - \lambda A(\tau)^{\delta_C}$, $\tau \geq 0$. By Theorem 4.2, Chapter III, Section 4, in Steutel and van Harn (2004), it is enough to establish that $\psi(\tau)$ is completely monotone on $(0, \infty)$. Since $\phi_1(\tau) = \exp(-\lambda A(\tau)^{\delta_C})$ is the LST of a C-stable distribution,

it is infinitely divisible (van Harn and Steutel, 1993). It follows that the function $\psi_1(\tau) = -\ln \phi_1(\tau) = \lambda\, A(\tau)^{\delta_C}$ has a completely monotone derivative on $(0,\infty)$ (again by Theorem 4.2 in Steutel and van Harn, 2004, quoted above). Since $\psi'(\tau) \le 0$ and for any $n \ge 2$,

$$(-1)^n\, \psi^{(n)}(\tau) = (-1)^{n-1} (\psi_1')^{(n-1)}(\tau)\ , \qquad \tau > 0\ ,$$

it follows that $\psi(\tau)$ is completely monotone on $(0,\infty)$. □

We note that Theorem 4.1 remains valid if the property of linear backward regression is replaced by the (stronger) assumption of time-reversibility.

For the family of semigroups $(C^{(\theta)}, \theta \in [0,1))$ of (3.19), the condition (4.3) is easily seen to be satisfied (by (3.19)–(3.20)) as $C_t^{(\theta)}(1) \sim \overline{\theta}\, e^{-\overline{\theta} t}$ $(t \to \infty)$. Applying Theorem 4.1 to the semigroup $C^{(\theta)}$ $(\theta \in [0,1))$, we obtain (via (3.20)) the LST $\phi^{(\theta)}(\tau)$ of (4.4) to be, in the case $\theta = 0$,

$$\phi^{(0)}(\tau) = \exp\left\{-b\,\tau + \frac{\lambda}{2}\,\tau^2\right\}, \qquad \tau \ge 0\ , \tag{4.9}$$

for some $b > 0$ and $\lambda > 0$, and in the case $0 < \theta < 1$,

$$\phi^{(\theta)}(\tau) = e^{-c\tau}\left(1 + \frac{\theta}{\overline{\theta}}\,\tau\right)^{-r}, \qquad \tau \ge 0\ , \tag{4.10}$$

for some $c \ge 0$ and $r > 0$. We note that if a rv X has LST $\phi^{(\theta)}(\tau)$ given by (4.10), for $0 < \theta < 1$, then $X \stackrel{d}{=} c + Y$, where Y admits a gamma distribution with LST

$$\varphi^{(\theta)}(\tau) = \left(1 + \frac{\theta}{\overline{\theta}}\,\tau\right)^{-r}, \qquad \tau \ge 0\ . \tag{4.11}$$

It is a simple exercise to verify that $\phi^{(0)}(\tau)$ of (4.9) is the LST of a $C^{(0)}$-self-decomposable distribution. In this case, the LST $\phi_t^{(0)}(\tau)$ in equation (2.21) is

$$\phi_t^{(0)}(\tau) = \exp\left\{-b\,(1 - e^{-t})\,\tau + \frac{\lambda}{2}(1 - e^{-2t})\tau^2\right\}, \qquad \tau \ge 0\ . \tag{4.12}$$

Assume that $(X_n, n \in \mathbf{Z})$ is a stationary $C^{(0)}$-AR(1) process with marginal LST $\phi^{(0)}(\tau)$. The marginal LST of the innovation sequence $(\epsilon_n, n \in \mathbf{Z})$ of $(X_n, n \in \mathbf{Z})$ is given by (4.12). Combining (4.5) with (4.9) and (4.12), we obtain the joint LST of (X_{n-1}, X_n) to be

$$g_0(\tau_1, \tau_2) = \exp\left\{-b(t_1 + \tau_2) + \frac{\lambda}{2}\,(\tau_1^2 + 2\,e^{-t}\tau_1\tau_2 + \tau_2^2)\right\}, \qquad \tau_1 \ge 0,\ \ \tau_2 \ge 0\ .$$

Since $g_0(\tau_1, \tau_2) = g_0(\tau_2, \tau_1)$, it follows that $(X_{n-1}, X_n) \stackrel{d}{=} (X_n, X_{n-1})$. Therefore, $(X_n, n \in \mathbf{Z})$ is time-reversible (and hence, has the property of linear backward regression).

The case $0 < \theta < 1$ is slightly more involved. We need a lemma.

Lemma 4.1. *Let $0 < \theta < 1$. A distribution μ on $\mathbf{R}_+$ with LST $\phi^{(\theta)}(\tau)$ of (4.10) is $C^{(\theta)}$-self-decomposable if and only if $c = 0$ or, equivalently, if and only if μ is a gamma distribution with LST $\varphi^{(\theta)}(\tau)$ of (4.11).*

Proof: Assume that μ has LST $\varphi^{(\theta)}(\tau)$ of (4.11). Straightforward calculations show that for any $t > 0$ and $\tau \geq 0$,

$$\varphi_t^{(\theta)}(\tau) = \frac{\varphi^{(\theta)}(\tau)}{\varphi^{(\theta)}\big(C_t^{(\theta)}(\tau)\big)} = \left(1 + \frac{\theta}{\overline{\theta}}\big(1 - e^{-\overline{\theta}t}\big)\tau\right)^{-r}. \tag{4.13}$$

Clearly, $\varphi_t^{(\theta)}(\tau)$ is the LST of a gamma distribution. Therefore, μ is $C^{(\theta)}$-self-decomposable. Conversely, assume that μ is $C^{(\theta)}$-self-decomposable with LST $\phi^{(\theta)}(\tau)$ of (4.10). By Theorem 5.4 in van Harn and Steutel (1993),

$$\ln \phi^{(\theta)}(\tau) = -\int_0^\tau \frac{\ln f(x)}{U^{(\theta)}(x)}\, dx\,, \qquad \tau \geq 0\,, \tag{4.14}$$

where $f(\tau)$ is the LST of an infinitely divisible distribution on $\mathbf{R}_+$. By differentiating both sides of (4.14) and using (3.20), we deduce that for every $\tau \geq 0$,

$$-\ln f(\tau) = U^{(\theta)}(\tau)\, \frac{d}{d\tau} \ln \phi^{(\theta)}(\tau) = c\,\theta\tau^2 + (c\,\overline{\theta} + r\,\theta)\tau\,.$$

By Theorem 4.2, Chapter III, Section 4, in Steutel and van Harn (2004), the function

$$-\frac{d}{d\tau} \ln f(\tau) = 2\,c\,\theta\tau + c\,\overline{\theta} + r\,\theta$$

must be a completely monotone function on $(0, \infty)$. This can only hold if $c = 0$. □

Assume now that $(X_n, n \in \mathbf{Z})$ is a stationary $C^{(\theta)}$-AR(1) process with marginal LST $\varphi^{(\theta)}(\tau)$ of (4.11). The marginal LST of the innovation sequence $(\epsilon_n, n \in \mathbf{Z})$ of $(X_n, n \in \mathbf{Z})$ is given by (4.13). Using (4.5), along with (4.11) and (4.13), we find the joint LST of (X_{n-1}, X_n) to be

$$g_\theta(\tau_1, \tau_2) = \left(1 + \frac{\theta}{\overline{\theta}}(\tau_1 + \tau_2) + \frac{\theta^2}{\overline{\theta}^2}\,\tau_1\tau_2\right)^{-r}, \qquad \tau_1 \geq 0\,, \quad \tau_2 \geq 0\,.$$

Since $g_\theta(\tau_1, \tau_2) = g_\theta(\tau_2, \tau_1)$, it follows that $(X_{n-1}, X_n) \stackrel{d}{=} (X_n, X_{n-1})$. Therefore, $(X_n, n \in \mathbf{Z})$ is time-reversible (and hence, has the property of linear backward regression).

We summarize our discussion in the following proposition.

Proposition 4.1. *Let $\theta \in [0,1)$. A stationary $C^{(\theta)}$-AR(1) process with a $C^{(\theta)}$-self-decomposable marginal distribution has the property of linear backward regression if and only if its marginal LST is given by (4.9), if $\theta = 0$, or by (4.11), if $0 < \theta < 1$.*

The stationary $C^{(\theta)}$-AR(1) process with the gamma marginal distribution with LST (4.11) is equivalent to the gamma model developed by Sim (1990). Sim makes use of a generalized multiplication based on a conditional compound Poisson distribution. Sim's operator is tailored to lead to a stationary gamma AR(1) process and offers no other stationary solutions. On the other hand, and as seen in Section 3, the $\odot_{C^{(\theta)}}$-multiplication leads to a variety of stationary models.

We conclude by noting that the extension of Proposition 4.1 to an arbitrary semigroup of cgf's C is an open question. Specifically, what kind of semigroups will give rise to a C-self-decomposable distribution with LST of the form (4.4)?

ACKNOWLEDGMENTS

This work was supported by a University of Indianapolis Summer Research Grant. The author is grateful to a referee for helpful comments.

REFERENCES

[1] Al-Osh, M.A. and Alzaid, A.A. (1993). Some Gamma processes based on the Gamma-Dirichlet transformation, *Commun. Stat., Stochastic Models*, **9**, 123–143.

[2] Bouzar, N. (1999). On geometric stability and Poisson mixtures, *Illinois J. of Math*, **43**, 520–527.

[3] Darolles, S.; Gourieroux, C. and Jasiak, J. (2006). Structural Laplace transform and compound autoregressive models, *J. Time Series Anal.*, **27**, 477–503.

[4] Foster, J.H. and Williamson, J.A. (1971). Limit theorems for the Galton–Watson process with time-dependent immigration, *Z. Wahrsch. Verw. Gebiete*, **20**, 227–235.

[5] Gaver, D.P. and Lewis, P.A.W. (1980). First-order autoregressive gamma sequences and point processes, *Adv. Appl. Prob.*, **12**, 727–745.

[6] Grunwald, G.K.; Hyndman, R.J.; Tedesco, L. and Tweedie, R.L. (2000). Non-Gaussian conditional linear AR(1) models, *Aust. N. Z. J. Stat.*, **42**, 479–495.

[7] Hansen, B.G. (1989). Self-decomposable distributions and branching processes, *Memoir COSOR 89-06*, Eindhoven Univ. of Technology (Eindhoven, Netherlands).

[8] Harn, K. van and Steutel, F.W. (1993). Stability equations for processes with stationary independent increments using branching processes and Poisson mixtures, *Stochastic Processes Appl.*, **45**, 209–230.

[9] Jayakumar, K. and Pillai, R.N. (1993). The first-order autoregressive Mittag–Leffler process, *J. Appl. Prob.*, **30**, 462–466.

[10] Lawrance, A.J. (1982). The innovation distribution of a gamma distributed autoregressive process, *Scan. J. Statist.*, **9**, 234–236.

[11] Lewis, P.A.W.; McKenzie, E. and Hugus, D.K. (1989). Gamma processes, *Commun. Stat., Stochastic Models*, **5**, 1–30.

[12] McKenzie, E. (1987). Innovation distributions for gamma and negative binomial autoregressions, *Scan. J. Statist.*, **14**, 79–85.

[13] Sim, C.H. (1990). First-order autoregressive models for gamma and exponential processes, *J. Appl. Prob.*, **27**, 325–332.

[14] Steutel, F.W. and van Harn, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line*, Marcel Dekker, Inc., New York-Basel.

[15] Walker, S.G. (2000). A note on the innovation distribution of a gamma distributed autoregressive process, *Scand. J. Statist.*, **27**, 575–576.

[16] Zhu, R. (2002). *On Generalized continuous-time AR(1) processes: Models, statistical inference, and applications to non-normal times series*, Ph.D. Thesis, University of British Columbia.

REVSTAT – Statistical Journal
Volume 8, Number 2, November 2010, 105–124

OPTIMIZED CLUSTERS FOR DISAGGREGATED ELECTRICITY LOAD FORECASTING

Authors: Michel Misiti
– Laboratoire de Mathématique, Orsay, France
and Ecole Centrale de Lyon, France
Michel.Misiti@ec-lyon.fr

Yves Misiti
– Laboratoire de Mathématique, Orsay, France
Yves.Misiti@math.u-psud.fr

Georges Oppenheim
– Laboratoire de Mathématique, Orsay, France
and Université de Marne-la-Vallée, France
Georges.Oppenheim@math.u-psud.fr

Jean-Michel Poggi
– Laboratoire de Mathématique, Orsay, France
and Université Paris Descartes, France
Jean-Michel.Poggi@math.u-psud.fr

Received: October 2009 Revised: May 2010 Accepted: June 2010

Abstract:

- To account for the variation of EDF's (the French electrical company) portfolio following the liberalization of the electrical market, it is essential to disaggregate the global load curve. The idea is to disaggregate the global signal in such a way that the sum of disaggregated forecasts significantly improves the prediction of the whole global signal. The strategy is to optimize, a preliminary clustering of individual load curves with respect to a predictability index. The optimized clustering procedure is controlled by a forecasting performance via a cross-prediction dissimilarity index. It can be assimilated to a discrete gradient type algorithm.

Key-Words:

- *clustering; disaggregation; forecasting; optimization; wavelets.*

AMS Subject Classification:

- 62P30, 62H30, 62-07, 62M20.

1. INTRODUCTION

This paper is devoted to electricity load forecasting via the disaggregation of the global signal. This disaggregation is based on customer clustering. To clarify, let us split the introduction in two parts. The first paragraph is dedicated to the general context and the second focuses on the specific application.

1.1. General context

Regular forecasting of the electrical load demand arises from a multiplicity of sources such as consumer behavior linked to social activities, government regulations and conventions and meteorological factors. Forecasting demand using statistical models with different types of explanatory variables provides accurate results. Those models include different components. There is a trend as well as daily, weekly and annual seasonal components. Calendar affects help take into account public holidays. An additional term may account for the effect of meteorological variables on the electricity load.

A recent special issue of the International Journal of Forecasting, devoted to energy forecasting, presents domain-related papers (see the editorial presentation by Taylor, Espasa [22]). Six papers in this journal provide an overview of the recent strategies for short-term or very short-term electricity load forecasting. The following methodologies: multi-equation model, neural networks or switching models are applied at national level in France, Spain, Australia, Brazil and Great Britain. This paper presents another approach to energy forecasting: a forecasting method based on signal disaggregation via the clustering of individual load curves.

Our goal is twofold. We aim to improve the accuracy of the electricity load forecast and to account for variability in the EDF's customer portfolio. This variability is due to the opening up of the previously nationalized electricity market. One way to deal with this difficulty is to disaggregate the global signal to improve the forecasting performance. Therefore, we need to create customer clusters such that the sum of disaggregated forecasts significantly improves the forecast of the whole global signal. In this paper we propose an optimized clustering scheme controlled through a cross-prediction dissimilarity index and based on a discrete gradient type algorithm.

Clustering has already been used for forecasting in similar electricity cases. Let us briefly present three examples.

The first example uses clustering for short-term peak load forecasting, proposed in Goia *et al.* ([10]). For a given load curve, forecasting is based on a

two-stage strategy. A functional clustering is created to classify the daily load curves and then a functional linear regression model is used on each cluster. Next, a new load curve is assigned to the clusters thanks to a functional discriminant analysis.

The second example, proposed by Piao *et al.* ([20]), deals with the prediction of customer load pattern in long duration load profiles. It also starts with clustering based on three daily profiles characteristics and aims at creating classes of load pattern and extracting representative load profiles for each class. Supervised learning methods can possibly be used when a new load curve is treated.

The third example, provided by Espinoza *et al.* ([9]), uses the forecasting step before clustering. Each individual load curve is first modeled using parametric periodic time series. Then, a typical daily profile is extracted from this parametric model for each individual customer. Finally, customer segmentation is obtained from the clustering these typical daily profiles.

The originality of our approach is the inclusion of an optimization step supervised by the forecasting procedure combined with a specific clustering strategy. To complete this introduction, let us present the industrial context of our work.

1.2. Industrial context

Load forecasting is a critical task for a company like EDF since it contributes to production planning. Engineers provide at noon, on a daily basis, the next day's consumption forecast. Forecasting is not only useful for short term decisions but also for optimizing production in dams or plants. Models similar to the following simplified version are built. The load P_t is decomposed into three components:

$$P_t = Pi_t + Pd_t + \varepsilon_t \tag{1.1}$$

where Pi_t is a weather independent component containing trend, seasonality and calendar effects, Pd_t a weather dependent component and ε_t an error term.

The parameters involved in the two first terms depend on the hour h, on the position of the day d within the year and on the day-type. The weather independent part is a linear combination of four sine and cosine terms whose coefficients are mainly functions of d, the type of the day (7 types), and h:

$$Pi_t = \Pi_{h,y} \sum_{m=1}^{4} a_{h,m} \cos\left(\frac{2\pi\, m\, d}{c}\right) + b_{h,m} \sin\left(\frac{2\pi\, m\, d}{c}\right) \tag{1.2}$$

where $c = 326.25$ and $\Pi_{h,y}$ is the load shape at the hour h for the year y, which

depends on the day-type. The coefficients a and b also depend on the day-type. The weather dependent component is composed of two parts. The first part involves a cooling gradient and a smoothed summer temperature while the second part involves a heating gradient and a smoothed winter temperature. This smoothed temperature can be assimilated to the indoor temperature.

After some fine tuning of the parameters, the quality of the EDF model measured by the Mean Absolute Percentage Error or MAPE, is considered as good. While the quality is satisfactory, it is never good enough during holidays such as Christmas. For instance, an estimation based on a five-year period such as [2000–2005] gives an hour-MAPE around 1.2%. For that same period, the one year forecast hour-MAPE is almost the same value. However, because of the deregulation of the French electricity market the situation has changed since 2007. EDF customers can now switch from one electricity provider to another, bringing instability to the market. As a consequence, the data available for forecasting is evolving. Before deregulation, a 5-year database was trademark for quality forecasting. Because of the new legal and commercial context, only one to two years of good enough quality data is now available to researchers.

1.3. Outline

This paper is organized as follows. After this introductory section, Section 2 is devoted to the problem and the data. Section 3 briefly recalls a wavelet based procedure for clustering load curves. Section 4 proposes the optimized clustering for forecasting by disaggregation. Section 5 contains experimental results on real world data and Section 6 presents some perspectives for future work.

2. THE DATA AND THE PROBLEM

2.1. The data

The data considered in this paper is not the French half-an-hour load consumption but individual commercial customer data. For obvious confidentiality and industrial reasons the French database is partially undisclosed. Moreover, the most recent data is not available. We worked on the [2000–2001] electricity consumption period. Individual power electricity demand curves, anonymous for confidentiality reasons are available for 2309 industrial customers during this period. The sampling period is one hour, leading to 17520 samples.

To highlight the differences among individual curves, let us examine four customer load curves during [2000–2001] time period (see Figure 1).

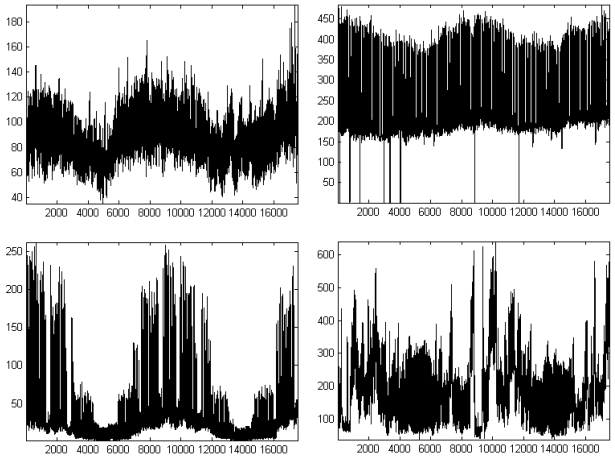

Figure 1: Raw data: 4 customer load curves during the 2000–2001 period (load in kW versus time in hours).

The long term shape of the curves differs a lot. It looks climate free for the customer at the bottom right while we can see three different climatic sensitivities on the other graphics. The same 4-customer load curves for one particular week in 2000 are displayed in Figure 2.

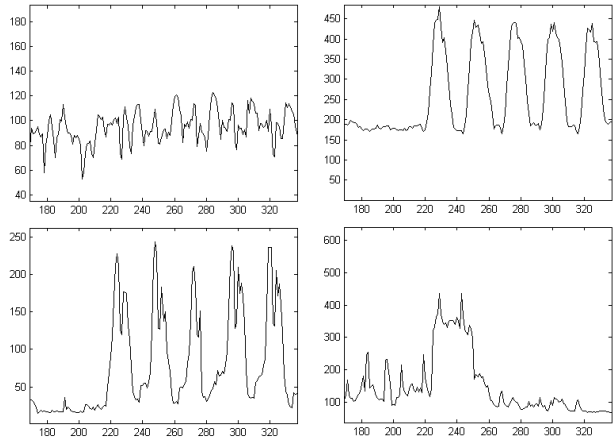

Figure 2: Raw data: 4 customer load curves during one week of 2000 (load in kW versus time in hours).

The two customers on the main diagonal of the plot are different from the others. Their graphics do not show any clear 'social rhythm', whereas the global shape of the two other customers are similar and display a 'social rhythm'.

Weekends are easy to detect and the main differences appear in the middle of the work days. The bottom left graphic, displays a bimodal shape instead of a single peaked curve.

2.2. Aggregated versus disaggregated

The disaggregation based forecasting problem goes as follows. Let us denote by $X_i(t)$ the value of the load curve of the ith customer at time t and we consider the aggregated electricity consumption signal:

$$S(t) = \sum X_i(t) \ . \tag{2.1}$$

The aggregated forecast is obtained by modeling and forecasting the signal:

$$\widehat{S}_{\text{aggr}}(t) = \widehat{S}(t) \ . \tag{2.2}$$

Associated with any partition of clustered individuals, we can define the consumption of each cluster g:

$$S_g(t) = \sum_{i \in g} X_i(t) \ . \tag{2.3}$$

Then, the disaggregated forecast is obtained by modeling and forecasting the signal within each cluster $\widehat{S}_g(t)$ and then summing over all clusters:

$$\widehat{S}_{\text{dis}}(t) = \sum_{g} \widehat{S}_g(t) \ . \tag{2.4}$$

We restrict our attention to this particular form of aggregation (2.4) which may not be the optimal combination. It could be interesting to consider a weighted sum and optimize the weights. However, in this study we specifically focus on the clustering issue and preserve the percentage of the load curve associated with each cluster.

Our challenge is to find the best partition of individuals. This partition has to be as accurate as possible from a forecasting perspective so that the latter will perform significantly better than the aggregated forecast.

Why can we expect such a result? Let us present two useful results already validated in narrower context.

A first indication is provided by the simplest statistical inference problem: the estimation of the mean μ of a variable Y on a given population using the random sample mean $\overline{Y}$. It is an unbiased estimator of μ of variance $\sigma^2(Y)/n$ where n is the sample size. Using stratified representative sampling with respect

to a given partition, the associated stratified estimator's variance (which is the disaggregated one) is reduced to the within variance over n: $\sigma_w^2(Y)/n$, which is always smaller than the variance of $\overline{Y}$.

A second indication comes from a simple result stated for two clusters and true for more clusters. Let us denote X_t and Y_t, two sequences corresponding to generic signal pairs $X_{i,t}$ and $X_{j,t}$ associated with two different clusters. Assume that X_t and Y_t are two sequences of stationary square integrable random variables and define

$$S_t = X_t + Y_t \ . \tag{2.5}$$

Then denoting by

$$\widehat{Z}_t = E(Z_t \mid Z_{t-1}, Z_{t-2}, ..., Z_1) \tag{2.6}$$

the conditional mean of Z_t given its own past. Let us define the two error indices

$$Err_{\text{aggr}} = E(S_t - \widehat{S}_t)^2 \tag{2.7}$$

and

$$Err_{\text{dis}} = E(S_t - \widehat{X}_t - \widehat{Y}_t)^2 \, . \tag{2.8}$$

So, if X_t and Y_t are independent then

$$Err_{\text{dis}} \leq Err_{\text{aggr}} \ . \tag{2.9}$$

If signals corresponding to two different clusters are independent and the conditional mean (in fact actually an accurate estimation) is used to predict, then the disaggregated forecast is of better quality than the forecast on the whole global signal.

Let us give a proof of that result. Starting from the definition of Err_{aggr} and since $S_t = X_t + Y_t$, we get

$$Err_{\text{aggr}} \geq E\big(S_t - E(S_t \mid X_{t-1}, Y_{t-1}, ..., X_1, Y_1)\big)^2 \, . \tag{2.10}$$

Independence between the X's and Y's leads to

$$E\big(X_t - E(X_t \mid X_{t-1}, Y_{t-1}, ..., X_1, Y_1)\big) = E\big(X_t - E(X_t \mid X_{t-1}, ..., X_1)\big) \tag{2.11}$$

as well as the equation obtained by permuting X and Y in (2.11). Adding up these two equations and taking squares of both sides, we obtain

$$E\big(S_t - E(S_t \mid X_{t-1}, Y_{t-1}, ..., X_1, Y_1)\big)^2 = E\big(S_t - \widehat{X}_t - \widehat{Y}_t\big)^2 = Err_{\text{dis}} \ .$$

Therefore, with (2.10) we obtain inequality (2.9).

As a conclusion, the two previously stated results suggest that it may be useful to disaggregate the global signal to significantly improve forecasting. Our idea is to find a good tradeoff between homogeneity within clusters and quality of the model's estimation. Homogeneity increases while the quality decreases with a higher number of clusters. Hence a three-step strategy:

1. Preprocessing individual customer data using wavelets;

2. Primary customer clustering with numerous homogeneous clusters;

3. Aggregation using stepwise optimization algorithm based on a dissimilarity index linked to a cross-prediction error and a discrete gradient type algorithm.

First, let us provide some additional information on the basic forecasting model. Then, we will develop the three-step strategy.

2.3. Eventail-like forecasting model

The aim of this paragraph is to clarify the internal forecasting procedure to the non-initiated reader while avoiding detailed information. Let us emphasize the fact that the error reduction via the new scheme is solely due to clustering optimization. Indeed, we do not perform ad-hoc adaptation of the model design strategy to the obtained clusters.

We circumscribe this paper to a single 'black-box' method used to design the forecasting model, starting from a given time series. Let us explain that we will take full advantage of a fully automatic version of EDF operational model called Eventail. Eventail is designed to predict the aggregated electricity consumption. Bruhns *et al.* ([6]) give a detailed description of a non-linear forecasting model of French electricity load in use at EDF. This model allows for different levels of seasonality and weather dependence. As previously stated, daily, weekly and annual components of the endogenous variable are considered, along with exogenous variables such as temperature, cloud cover, calendar events as well as a long-term trend. The mid-term model is a highly parameterized climate-free SARIMA model additively corrected with a weather dependent term. This model delivers an accurate forecast.

Let us note some results. The forecasting performance on the sample of 2309 customers, measured by the long-term MAPE (for Mean Absolute Percentage Error) is about 4.06% for the global aggregated signal. Meanwhile, on the same sample, the completely disaggregated forecasting performance reaches 2.94%.

Remark 2.1. We will not provide comparisons of Eventail with other forecasting methods. This would be interesting (see for example Hippert *et al.* ([14]) for recent statistical time series tools for load forecasting), however since Eventail is the current operational tool, it is regularly improved in order to take into account the new characteristics of the load curve. Let us mention that Bruhns *et al.* ([6]) describe the forecasting model already used at EDF for mid-term load forecasting and provide a comparative study of various alternatives. Also, a more recent discussion on how to handle changes in customer behavior in a similar context can be found in Dordonnat *et al.* ([8]) who describe a forecasting model based on time-varying processes, specifically a periodic state space model.

3. CLUSTERING USING WAVELETS

The aim of the preliminary step is to build basic clusters (often called super customers hereafter) based on our sensibly assembled customers. The key idea is to take advantage of the hierarchical multiresolution structure of wavelet decomposition (see Misiti *et al.* [18]) for clustering signals. Simply put, wavelets allow us to write each individual signal as the sum of orthogonal signals: a coarser approximation at large scale (low frequency) and additional details at different resolutions, of decreasing scales. The approximation at level j roughly represents the local mean signal on intervals of length 2^j while the detail at level j contains fluctuations around this local mean on the same corresponding intervals. Let n be the common length of the p series individually denoted by $X^{(i)}$. Then, for a given orthogonal wavelet ψ, each time series can be decomposed at level J (which is at most the integer part of $\log_2(n)$). This leads to:

$$X^{(i)} = A_J^{(i)} + \sum_{j=1}^{J} D_j^{(i)}\,, \tag{3.1}$$

where $A_k^{(i)}$ and $D_k^{(i)}$ denote respectively the approximation and the detail, at level k, of the signal $X^{(i)}$.

The procedure, described in Misiti *et al.* ([19]), is a hybrid scheme mixing regularization and filtering approaches, according to James and Sugar's ([15]) terminology. Let us describe this scheme. First, there is individual denoising using a signal-adapted wavelet basis, then a projection on a one common wavelet basis to get a huge dimensionality reduction effect (see Biau *et al.* [5]). Then each customer is characterized by coefficients. The last step of the process is the clustering of the customers using Ward's method with squared Euclidean distances, in order to preserve distances between signals through wavelet coefficients encoding. We generate hierarchies of partitions corresponding to different numbers of clusters and various wavelet representations, that are typically approximations of decreasing resolution level.

For the final step, considering any partition P obtained by clustering data Z and for a given number of clusters, we can compute the following usual variance ratio quality index:

$$I_Z(P) = \frac{\mathrm{Var}_b(Z,P)}{\mathrm{Var}_w(Z,P)} \ , \tag{3.2}$$

where $\mathrm{Var}_b(Z,P)$ and $\mathrm{Var}_w(Z,P)$ denote respectively the variance between clusters and within clusters. This quality index allows us to compare two partitions based on two different signal representations but it depends heavily on the number of clusters. For instance, let us say P' is a finer partition obtained from P. Then $I_Z(P') \geq I_Z(P)$. Since we have to compare partitions with different number of clusters, we will choose the one leading to the best normalized variance ratio index:

$$I_Z^N(P) = \frac{\mathrm{Var}_b(Z,P)}{C(P) \cdot \mathrm{Var}_w(Z,P)} \ , \tag{3.3}$$

where $C(P)$ is the number of clusters within partition P.

This index is similar to the statistic of Calinski and Harabasz ([7]), considered as a 'good competitor' (see for example Tibshirani *et al.* [23]). It allows us to select a convenient number of clusters as well as a critical level of wavelet decomposition (simulated examples, electricity data processing and further details can be found in Misiti *et al.* [19]).

In our electrical context, in an earlier study we obtained various partitions using this clustering scheme but without taking into account the forecasting objective. The most interesting partitions are made of 15 to 19 clusters and highlight wavelet approximation coefficients at level 6 (around 2 coefficients a week) and detail coefficients at level 2 (around 5 coefficients a day). These partitions reached a forecasting performance of 2.75% long-term MAPE which is better than the fully aggregated or the fully disaggregated forecasts. However, partitions describes in this paragraph cannot be improved with the optimization process described in the next section.

Therefore, hereafter we will work from this initial pre-processing. We will select wavelet approximation coefficients at level 6 in order to get the load curve's global shape. We will relax unsupervised clusters constraints. This means that we will start with a large number of clusters and step by step aggregate them with an optimization criterion supervised by predictability. According to the variance ratio, 90 clusters are sufficient to assume strong homogeneity.

4. OPTIMIZED CLUSTERING DIRECTED BY FORECASTING

4.1. A multistage procedure

The proposed optimized clustering scheme is as follows:

1. *Wavelet preprocessing.*
Customer characterization through wavelet representation of each signal after standardization using approximation coefficients at level 6.

2. *First clustering* around numerous centroids.
A minimum of 90 clusters regrouping homogeneous customers. Each cluster is represented by its aggregated signal.

3. *Iterative optimization.*
The starting point being the described initial partition, an optimization process supervised through cross-prediction dissimilarity index is run. A discrete gradient type procedure based on D matrix (defined in the next section) explores the set of partitions.

4.2. Cross-prediction dissimilarity

To qualify a specific aggregation we use cross-prediction dissimilarity between elements. Those elements can be either individual or aggregated signals. This dissimilarity index between X_k and X_j is based on the following idea. The model fitted on past observations of $X_j(t)$ is used to predict the future of $X_k(t)$ and vice-versa. In our specific electrical context, let us denote by

$$forec_{k|j}^{2001} = forecast\big(X_j^{2000}, X_k^{2001}\big)\,, \tag{4.1}$$

the forecasts of X_k on the year 2001 (the test period) obtained from the model fitted on X_j on the year 2000 (the learning period). The fitted model is based on the Eventail-like design tool. Then, the associated error is defined by:

$$E_{k|j} = error\big(X_k^{2001}, forec_{k|j}^{2001}\big)\,. \tag{4.2}$$

Then a natural symmetric measure of dissimilarity is given:

$$D = (D_{j,k}) = \big((E_{k|j} + E_{j|k})/2\big)\,. \tag{4.3}$$

To fairly rescale the X_k and X_j load curves for testing and for estimating the index D is based on errors obtained from the *l1*-normalized versions (*i.e.* signals summing to 1) .

4.3. Zooming in on the optimization step

The iterative optimization of the initial partition is supervised through the cross-prediction dissimilarity. It can be adapted to the forecast horizon as well as to the error criterion. The iterative optimization is based on discrete gradient via a neighborhood definition through dissimilarity between an element and a cluster induced by the matrix D. The basic step is an iterative exploration of elements. These elements are always candidates for cluster change, using nearest D-neighbors. It should be noted that the partition evolves and that the basic step consists of moving an element from one cluster to another. Therefore, this process generates a non monotonic sequence of partitions, which is not a hierarchical approach. This sequence of partitions evolves through element assignment modifications. The number of clusters decreases slowly along the iterations. A cluster disappears only when it is empty. The optimization scheme goes as follows:

1. *Compute matrix* D of dissimilarities between elements;

2. *Compute dissimilarities* between each element and the current clusters using D and a linkage function (the minimum for example);

3. *Select* a neighbor: a couple (E,C), an element E candidate to move to a cluster C;

4. *Test* the new affectation gain for the disaggregated forecast associated to the resulting partition

- *if* the error does not decrease *then*
  - *if* there are candidates *then* select the next one and *go to* step 4
  - *else end* (no improvement by moving an element from a cluster to another)
- *if* the error decreases *then* modify partition and *go to* step 2.

This scheme can be adapted to parallel computations simply through a better organization of the candidates' examination in the more internal loop. Parallel capacities could also be used to explore multistart versions of the algorithm. However, these aspects are out of the scope of this paper which focuses on the question of the possible usefulness of disaggregation.

5. EXPERIMENTAL RESULTS

5.1. Performance results

Starting from 90 clusters, the optimized partition reaches the performances measured by long-term and short-term MAPE, given by Table 1.

Table 1: Performances of optimized partition starting from 90 clusters.

	Aggregated	Disaggregated	Gain
MAPE long-term (LT)	4.06%	2.39% with 19 clusters	41.13%
MAPE short-term (ST)	2.47%	1.51% with 28 clusters	38.86%

The procedure can be stopped at any step of the optimization process, therefore, improving the previous acceptable solution. The 195 step process with an error rate gain of 41%, is illustrated on Figure 3. This error reduction largely and obviously improves the optimization process, which starts with 90 clusters and ends with 19 clusters.

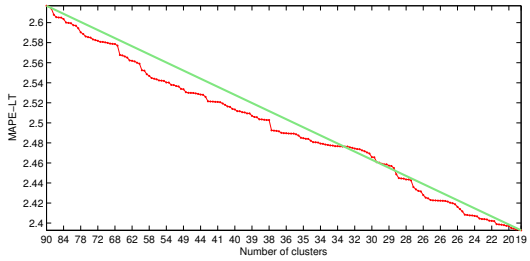

Figure 3: Optimization process: from 90 to 19 clusters leading to a gain of 41%.

5.1.1. About wavelet preprocessing

The first step of the global procedure (wavelet preprocessing and initial clustering using wavelets) is important. Indeed if one performs directly a hierarchical clustering of the original 2309 customers using the dissimilarity matrix D and then optimizes the associated 90 clusters partition, the MAPE-LT error criterion stabilizes around 2.7% instead of 2.5%.

5.1.2. About the optimization step

The optimization step is also important. Indeed, starting from the 90 cluster partition, if one constructs the hierarchy of partitions (by hierarchical clustering using D), it is difficult to select a critical number of clusters (see Figure 4) and the MAPE-LT error criterion remains about 2.6% instead of 2.5%.

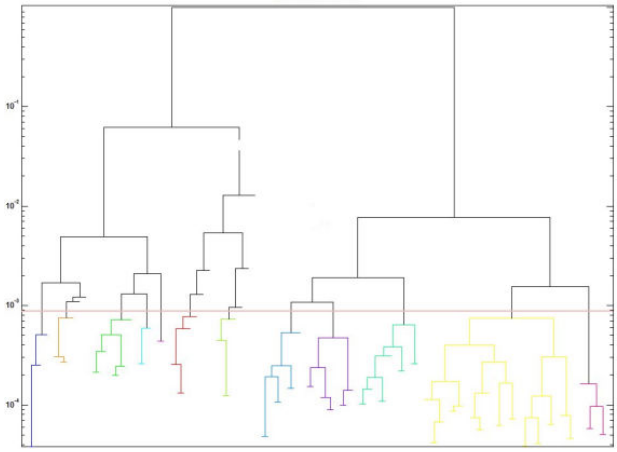

Figure 4: Dendrogram: hierarchical clustering using D.

5.2. About the number of basic customers

Finally, let us mention that the number of initial clusters (taken here to be equal to 90) is an important parameter, especially when the method is used for a significantly large number of customers. Indeed, the actual performance is slightly improved by increasing the number of clusters. The initial 2.39% performance on 90 clusters reaches 2.31% with 200 clusters and even 2.26% for 500 clusters, therefore increasing the reduction rate from 41.1 to 44.3%.

5.3. Clusters interpretation

In this paragraph, we will focus on the 19 clusters resulting from the final optimized partition. For example, Figure 5 presents cluster 1 made up of 10 super customers. It superimposes the 10 super customers consumptions with the average consumption of the cluster. The top graphic represents the year 2000 while the bottom graphic zooms in on the first quarter of that year (January, February and March 2000).

Let us note that the extreme regularity and homogeneity of the final 19 average cluster curves is remarkable. This can be explained by the fact that those curves are perfectly suited for forecast using the Eventail model. In other words, the optimization algorithm produces curves well adapted to Eventail black box forecast method.

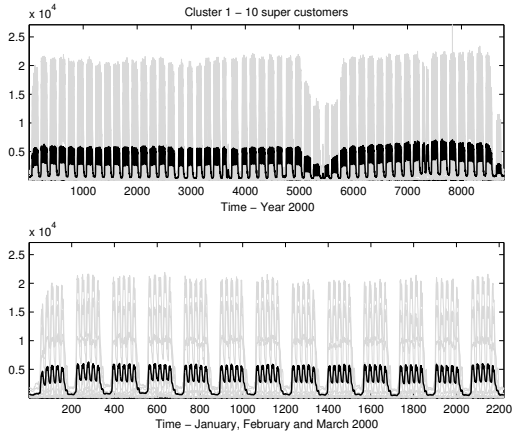

Figure 5: Cluster 1: average consumption (dark curve) and individual consumption of the super customers (light curves). Top: the year 2000. Bottom: zoom in on the first three months of 2000.

To get extra information on cluster 1, let us look at Figure 6 and its 11 graphics. It displays the 198 individual customer consumptions leading to the 10 super customers of the optimized partition. Each one of the ten first plots displays a super customer consumption together with the average consumption. The last plot displays the aggregated signal.

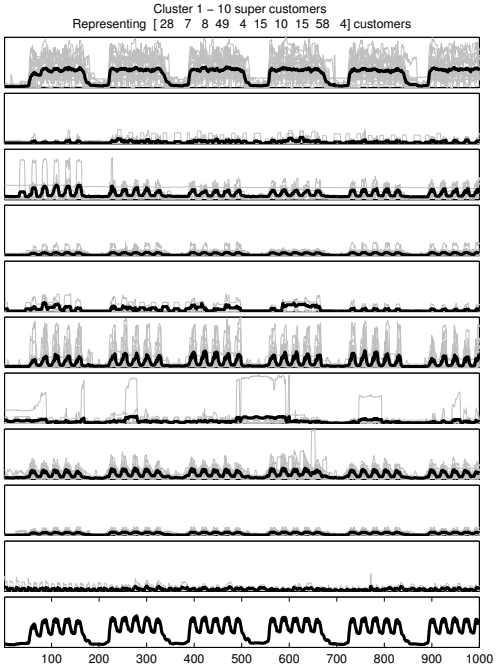

Figure 6: For cluster 1, ten top plots representing individual consumptions (light curves), leading to 10 super customers (dark curves). Bottom plot: total cluster consumption.

So, despite a great heterogeneity of the customers within each cluster, the signals associated with the final clusters are very stable and easy to predict using Eventail.

5.4. Validation results

The optimization procedure can be modeled using the three following steps:

1. Starting from the N individual customers, performing a discrete wavelet transform DWT at a given level j of the $N = 2309$ signals S, normalizing in $l1$-norm and clustering the resulting signals. As a result:

$$(K, P_K) = MSC(N, j; S) \ ,$$

which leads to K super customers associated with a partition P_K. This multiscale clustering (MSC) step involves selecting level of decomposition j (typically $j = 6$ or $j = 4$) as well as choosing K, the number of clusters (usually $K = 60, 90, 200, 500$).

2. Computation of $D(e, e')$ (for elements e and e') then computation of $D_c(e, c)$ (for an element e and a cluster c) and the optimization leads to:

$$(k, \widetilde{P_k}) = Opt(K, P_K) \ .$$

3. Expansion of the k clusters of the K super customers over the N initial individuals to produce:

$$(k, P_k) = Exp(k, \widetilde{P_k}, S) \ .$$

Suppose now that, three years of observations are available for a subset of customers. Then, we can select the parameters (choice of K in particular) and estimate the quality according to the following validation principle. The first two years are used to design the clusters and the last year is used as the test sample.

Unfortunately, a subset of only 1482 customers is available during the three considered years. Quality estimations are as follows. On the learning set (year 2001), quality gain is about 23% since we go from 3.14% to 2.31%. On the test set (year 2002) there is an 8% reduction, from 6.82% to 6.32%. Nonetheless, the disaggregation procedure still provides significant gain. Of course this must be considered with caution since the year 2002 seems to be much more difficult to predict using Eventail: the MAPE forecast error of the aggregated predictor increases from 3.14% to 6.82% and is perhaps not representative.

6. FUTURE WORK

Let us briefly present some future possible developments.

First of all, alternatives to the current algorithm could be studied, the main difficulty being to cope with the computational burden. A scheme better suited for parallelism could be developed. A divisive strategy instead of data aggregation could be used to optimize the forecasting objective. It would start with the whole population and iteratively segment the current subgroup. Segmentation could be completed according to a 2 or 3-means clustering using approximation coefficients.

Another line of work on electrical data could be to further develop forecasting with wavelet methods (see Antoniadis *et al.* ([2]), Amin Ghafari, Poggi ([3])). The aim would be to adapt the models to the clusters using a similar method to the one described in Hathaway, Bezdek ([13]) or more recently clusterwise linear models proposed in Gruen, Leisch ([11]).

Also, we could take advantage of external meteorological and economical information as diagnostic and performance measurement tools. Eventually, the whole procedure should integrate parameters' data-driven choices: the wavelet and the representation basis, the obtained partition and the adaptation of the model to cluster specificities.

Last but not least, theoretically we could explore how to maximize the profits of the disaggregation method in general conditions.

ACKNOWLEDGMENTS

This work is part of a scientific collaboration between EDF Clamart R&D, OSIRIS department and Orsay University. The authors thank EDF for the problem as well as the data. The authors also thank Sophie Lamarche and Alain Dessertaine for helpful discussions, and Chris Roth and Dominique Sorbier for careful reading. Lastly, we thank the anonymous referees for their useful suggestions and meaningful comments.

REFERENCES

[1] Abraham, C.; Cornillon, P.; Matzner-Lober, E. and Molinari, N. (2003). Unsupervised curve clustering using B-splines, *Scand. J. Stat.*, **30**(3), 581–595.

[2] Antoniadis, A.; Paparoditis, E. and Sapatinas, T. (2006). A functional wavelet-kernel approach for time series prediction, *J. of the Royal Stat. Soc., Series B*, **68**, 837–857.

[3] Aminghafari, M. and Poggi, J.M. (2007). Forecasting time series using wavelets, *Int. Journ. of Wavelets, Multiresolution and Inf. Proc.*, **5**(5), 709–724.

[4] Biau, G.; Bunea, F. and Wegkamp, M. (2005). Functional classification in Hilbert Spaces, *IEEE Trans. Inf. Theory*, **51**(6), 2163–2172.

[5] Biau, G.; Devroye, L. and Lugosi, G. (2008). On the performance of clustering in Hilbert spaces, *IEEE Trans. on Inf. Theory*, **54**, 781–790.

[6] Bruhns, A.; Deurveilher, G. and Roy, J.S. (2005). A non linear regression model for mid-term load forecasting and improvements in seasonality, *Proceedings of the 15th Power Systems Computation Conference 2005, Liege Belgium.*

[7] Calinski, R.B. and Harabasz, J. (1974). A dendrite method for cluster analysis, *Comm. Stat.*, **3**, 1–27.

[8] Dordonnat, V.; Koopman, S.J.; Ooms, M.; Dessertaine, A. and Collet, J. (2008). An hourly periodic state space model for modelling French national electricity load, *International Journal of Forecasting*, **24**, 566–587.

[9] Espinoza, M.; Joye, C.; Belmans, R. and De Moor B. (2005). Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series, *IEEE Transactions on Power Systems*, **20**(3), 1622–1630.

[10] Goia, A.; May, C. and Fusai, G. (2009). Functional clustering and linear regression for peak load forecasting, *Int. J. of Forecasting*, to appear.

[11] Gruen, B. and Leisch, F. (2007). Fitting finite mixtures of generalized linear regressions in R, *Computational Statistics and Data Analysis*, **51**(11), 5247–5252.

[12] Hastie, T.; Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning*, Springer.

[13] Hathaway, R.J. and Bezdek, J.C. (1993). Switching regression models and fuzzy clustering, *IEEE Trans. Fuzzy Systems*, **1**(3), 195–203.

[14] Hippert, H.S.; Pedreira, C.E. and Souza, R.C. (2001). Neural networks for short-term load forecasting: A review and evaluation, *IEEE Transactions on Power Systems*, **16**(1), 44–55.

[15] James, G. and Sugar, C. (2003). Clustering for sparsely sampled functional data, *JASA*, **98**, 397–408.

[16] Kaufman, L. and Rousseeuw, P. (2005). *Finding groups in Data, an introduction to Cluster Analysis*, Wiley.

[17] Mallat, S. (1998). *A Wavelet Tour of Signal Processing*, Academic Press.

[18] Misiti, M.; Misiti, Y.; Oppenheim, G. and Poggi, J.-M. (2007). *Wavelets and Their Applications*, Hermes Lavoisier, ISTE Publishing Knowledge.

[19] Misiti, M.; Misiti, Y.; Oppenheim, G. and Poggi, J.-M. (2007). Clustering signals using wavelets, F. Sandoval *et al.* (Eds.): *IWANN 2007*, Lecture Notes in Computer Science, 4507, 514–521, Springer.

[20] Piao, M.; Lee, H.G.; Park, J.H. and Ryu, K.H. (2008). Application of Classification Methods for Forecasting Mid-Term Power Load Patterns in Advanced Intelligent Computing Theories and Applications, D.-S. Huang *et al.* (Eds.): *ICIC 2008, CCIS 15*, Springer, 47–54.

[21] Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*, Springer.

[22] Taylor, J.W. and Espasa, A. (2008). Editorial Energy forecasting, *Int. J. of Forecasting*, **16**, 561–565.

[23] Tibshirani, R.; Walther, G. and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic, *J. of the Royal Stat. Soc., Series B*, **63**(2), 411–423.

REVSTAT – Statistical Journal
Volume 8, Number 2, November 2010, 125–138

IMPROVING ON MINIMUM RISK EQUIVARIANT AND LINEAR MINIMAX ESTIMATORS OF BOUNDED MULTIVARIATE LOCATION PARAMETERS

Authors: Éric Marchand
– Département de Mathématiques, Université de Sherbrooke, Sherbrooke, Canada
eric.marchand@usherbrooke.ca

Amir T. Payandeh Najafabadi
– Department of Mathematical Sciences, Shahid Beheshti University, G.C. Evin, 1983963113, Tehran, Iran
amirtpayandeh@sbu.ac.ir

Received: February 2010 Revised: June 2010 Accepted: June 2010

Abstract:

- We propose improvements under squared error loss of the minimum risk equivariant and the linear minimax estimators for estimating the location parameter θ of a p-variate spherically symmetric distribution, with θ restricted to a ball of radius m centered at the origin. Our construction of explicit improvements relies on a multivariate version of Kubokawa's Integral Expression of Risk Difference (IERD) method. Applications are given for univariate distributions, for the multivariate normal, and for scale mixture of multivariate normal distributions.

Key-Words:

- *decision theory; spherical symmetric distribution; restricted parameter; minimum risk equivariant estimator; linear minimax estimator; dominating estimators; squared error loss.*

AMS Subject Classification:

- 62F10, 62F30, 62H12.

1. INTRODUCTION

We consider the problem of estimating, under squared error loss, the location parameter θ of a p-variate spherically symmetric distribution under the constraint $\|\theta\| \leq m$, with $m > 0$ known. With several authors having obtained interesting results relative to this problem, and more generally for restricted parameter space problems (see Marchand and Strawderman, 2004; van Eeden, 2006 for useful reviews), we focus on the determination of benchmark estimators such as the maximum likelihood estimator (MLE), the minimum risk equivariant estimator (MRE), and the linear minimax estimator (LMX). In this regard, Marchand and Perron (2001) provide for the multivariate normal case improvements on the (always) inadmissible MLE for all (m, p). These include Bayesian improvements, but conditions are then required on (m, p). Complementary findings for the multivariate normal and parallel findings for other spherically symmetric distributions, including in particular multivariate student distributions, were obtained respectively by Fourdrinier and Marchand (2010) and Marchand and Perron (2005); but again conditions for the studied priors π (typically boundary uniform, uniform on spheres, and fully uniform) of the form $m \leq c_\pi(p)$ for the Bayes estimator δ_π to dominate the MLE are necessitated. Hence, the problem of finding a Bayesian or an admissible improvement for any (m, p), for any given spherically symmetric distribution remains unsolved (even for $p = 1$ or the multivariate normal distribution).

Alternatively, for the objective of passing the minimum test of improving upon the minimum risk equivariant estimator, positive findings for the univariate case ($p = 1$) were obtained by Marchand and Strawderman (2005), as well as by Kubokawa (2005). The former establish a general dominance result for the fully uniform prior Bayes estimator, which actually applies more generally for a wider not necessarily symmetric class of location model densities and location invariant losses. The latter provides on the other hand a large class of priors which lead to Bayesian improvements for the univariate version of our problem of symmetric densities and squared error loss. A key feature of these dominance results is the use of Kubokawa's (1994) Integral Expression of Risk Difference (IERD) technique.

For multivariate settings, a lovely result by Hartigan (2004) tells us that for multivariate normal distributions, the fully uniform Bayes procedure improves upon the minimum risk equivariant estimator. The result is actually more general and applies for convex restricted parameter spaces with non-empty interiors. However, Hartigan's result does require normality and hence a spherically symmetric analog remains an open question. Moreover, Hartigan's result does not apply to the benchmark linear minimax estimator, which represents itself a simple improvement on the minimum risk equivariant estimator.

With the above background, our motivation here resides in extending the univariate dominance results to the multivariate case, extending Hartigan's result for balls to spherically symmetric distributions, and considering improvements upon the linear minimax procedure as well. We provide preliminary results in this direction in terms of sufficient conditions for dominating either the minimum risk equivariant estimator, the linear minimax estimator, or both. Our treatment possesses the interesting feature of being unified with respect to the dimension p and the given spherically symmetric distribution. Moreover, we arrive at our dominance results through a novel multivariate variant of Kubokawa's IERD technique. The main dominance results are presented in Section 2, and various examples or illustrations are pursued in Section 3. These include univariate distributions, the multivariate normal distribution, and scale mixture of multivariate normal distributions.

2. MAIN RESULTS

Let X be a p-variate random vector with spherically symmetric density

$$f\big(\|x-\theta\|^2\big)\,, \tag{2.1}$$

where the location parameter θ is constrained to a ball centered at the origin and of radius m, say Θ_m. We seek improvements on the minimum risk equivariant (MRE) estimator $\delta_0(X)=X$, and the linear minimax estimator $\delta_{\rm LMX}(X)=\frac{m^2}{m^2+p\,\sigma^2}\,X$ under squared error loss $L(\theta,d)=\|d-\theta\|^2$, where $E_\theta(\|X-\theta\|^2)=p\,\sigma^2<\infty$. Hereafter, we denote the norms of X, x, and θ by R, r, and λ respectively. Our results bring into play the orthogonally invariant in θ and nonnegative quantities $H(t,\lambda)=\frac{E_\theta\big(\theta^TX|\ \|X\|\geq t\big)}{E_\theta\big(X^TX|\ \|X\|\geq t\big)}$ and $H^*(t,\lambda)=\frac{\lambda\,E_\theta\big(\|X\|\,\|X\|\geq t\big)}{E_\theta\big(|X^TX\|X\|\geq t\big)}$, $t\geq 0,\ \lambda\geq 0$. We will make use of the inequality $H(t,\lambda)\leq H^*(t,\lambda)$ for all $t\geq 0$, $\lambda\geq 0$, which follows as a simple application of the Cauchy–Schwartz inequality. Now, we present the main dominance results of this paper.

Theorem 2.1. *For a model as in (2.1), $\delta_g(X)=g(\|X\|)X$ dominates $g(0)X$, whenever:*

- (**i**) *g is absolutely continuous, nonconstant, and nonincreasing;*
- (**ii**) *and $g(r)\geq \sup_{\lambda\in[0,m]}H(r,\lambda)$ for all $r\geq 0$.*

Moreover, if conditions (i) and (ii) are satisfied, and

- (**iii**) $g(0)\in\Big[\frac{m^2-p\sigma^2}{m^2+p\sigma^2}\,,\,1\Big)$,

then $\delta_g(X)=g(\|X\|)X$ also dominates $\delta_0(X)=X$.

Remark 2.1. By virtue of the inequality $H(t,\lambda) \le H^*(t,\lambda)$ for all $t \ge 0$, $\lambda \ge 0$, condition (ii) of Theorem 2.1 can be replaced by the weaker, but nevertheless useful, condition

(**ii′**) and $g(r) \ge \sup_{\lambda\in[0,m]} H^*(r,\lambda)$ for all $r \ge 0$.

Proof of the Theorem: It is straightforward to verify that $g(0)X$ dominates X under condition (iii), so that conditions for which $\delta_g(X)$ dominates $g(0)X$, such as (i) and (ii), are necessarily conditions for which $\delta_g(X) = g(\|X\|)X$ also dominates $\delta_0(X) = X$. Now, using Kubokawa's IERD technique, the risk difference between the estimators $\delta_g(X)$ and $g(0)X$ can be written as

$$\begin{aligned}
\frac{1}{2}\,\Delta(\theta) &= \frac{1}{2}\Big[R\big(\theta, g(\|X\|)X\big) - R\big(\theta, g(0)X\big)\Big] \\
&= \frac{1}{2}\Big[E_\theta\big\|g(\|X\|)X-\theta\big\|^2 - \big\|g(0)X-\theta\big\|^2\Big] \\
&= \frac{1}{2}\,E_\theta\bigg(\int_0^{\|X\|}\frac{\partial}{\partial t}\big\|g(t)X-\theta\big\|^2\,dt\bigg) \\
&= \int_{\mathbb{R}^p}\int_0^{\|x\|} g'(t)\,\big[g(t)\,x-\theta\big]^T x\,f\big(\|x-\theta\|^2\big)\;dt\,dx \\
&= \int_0^\infty g'(t)\int_{\{x\in\mathbb{R}^p:\,\|x\|\ge t\}}\big[g(t)\,x^Tx-\theta^Tx\big]\,f\big(\|x-\theta\|^2\big)\;dx\,dt\,.
\end{aligned}$$

Now, observe that conditions (i) and (ii) imply that $\Delta(\theta) \le 0$ for all $\theta \in \Theta_m$, establishing the result. □

Here are some further remarks and observations in relationship to Theorem 2.1.

The nonincreasing property of condition (i) is not necessarily restrictive. Indeed, for the multivariate normal case, Marchand and Perron (2001, theorem 5) establish that the nonincreasing property holds for all Bayesian estimators associated with symmetric, logconcave prior densities on $[-m,m]$. The conditions of Theorem 2.1 suggest the bounds (ii) and (ii′) themselves $\sup_{\lambda\in[0,m]} H(r,\lambda)$ and $\sup_{\lambda\in[0,m]} H^*(r,\lambda)$ as candidate g functions. These functions are of the form $H(r,\lambda(r))$ and $H^*(r,\lambda(r))$, where $\lambda(\cdot)$ is some function taking values on $[0,m]$. All such functions lead to range preserving estimators δ_g; i.e., $\|\delta_g(x)\| \le m$ for all $x \in \mathbb{R}^p$; since for all $r \ge 0$ and $\|\theta\| = \lambda(r)$:

$$0 \;\le\; H(r,\lambda(r)) \;\le\; H^*(r,\lambda(r)) \;=\; \frac{\lambda(r)\,E_\theta\big(\|X\|\,\|X\|\ge r\big)}{E_\theta\big(|X^TX\,\|X\|\ge r\big)} \;\le\; \frac{\lambda(r)}{r} \;\le\; \frac{m}{r}\;,$$

and since $\|\delta_g(x)\| \le m$ for all $x \in \mathbb{R}^p$ whenever $0 \le g(r) \le \frac{m}{r}$ for all $r > 0$. Finally, as a consequence of the above, observe that the projection of $\delta_0(X)$ onto Θ_m, given by δ_{g_p} with $g_p(r) = \frac{m}{r} \wedge 1$, satisfies the conditions of Theorem 2.1.

We now focus on related implications for the estimators $\delta_H(X) = H(\|X\|,m)X$ and $\delta_{H^*}(X) = H^*(\|X\|,m)X$, which will turn out in several cases to be the smallest possible g's satisfying respectively conditions (ii) and (ii′) of Theorem 2.1.

Corollary 2.1.

(**a**) *If $H(r,\lambda)$ increases in $\lambda \in [0,m]$ for all $r \geq 0$, and decreases in $r \in [0,\infty]$ for all $\lambda \in [0,m]$, then $\delta_H(X) = H(\|X\|,m)X$ dominates both the linear minimax estimator $\delta_{\mathrm{LMX}}(X)$ and the MRE estimator $\delta_0(X)$;*

(**b**) *If $H^*(r,\lambda)$ increases in $\lambda \in [0,m]$ for all $r \geq 0$, then $\delta_{H^*}(X) = H^*(\|X\|,m)X$ dominates the MRE estimator $\delta_0(X)$.*

Proof: Part (a) follows as a direct application of Theorem 2.1 as $H(0,m) = \frac{E_\theta(\theta^T X)}{E_\theta(X^T X)} = \frac{m^2}{m^2+p\sigma^2} \in \left[\frac{m^2-p\sigma^2}{m^2+p\sigma^2}, 1\right)$, for $\|\theta\| = m$. Part (b) follows for two reasons. First, for any positive random variable Y with density g_Y, and its biased version W with density proportional to $w g_Y(w)$, the ratio $\frac{E(Y^2 | Y>t)}{E(Y|Y>t)} = E(W|W>t)$ is increasing in t, which implies that $H^*(\cdot,m)$ is a decreasing function on $[0,\infty)$. Secondly, for $\|\theta\| = m$, $H^*(0,m) = m\,\frac{E_\theta(\|X\|)}{E_\theta(\|X\|^2)} = \frac{E_\theta(\|X/m\|)}{E_\theta(\|X/m\|^2)} < \frac{E_\theta(\|X/m\|)^2}{E_\theta(\|X/m\|^2)} < 1$, since $E_\theta(\|X\|) > \|E_\theta(X)\| = m$. □

3. EXAMPLES

The following subsections are devoted to applications of Corollary 2.1, with the key difficulty arising in checking the monotonicity conditions relative to H and H^*. We focus on general univariate cases (subsection 3.1), the multivariate normal distribution (subsection 3.2.), and scale mixtures of multivariate normal distributions (subsection 3.3).

3.1. Univariate spherically symmetric distributions

We express the symmetric univariate densities in (2.1) as

$$f_\theta(x) = e^{-q(x-\theta)}\,, \tag{3.1}$$

and restrict ourselves to cases where

$$q \in Q^* = \Big\{ q\colon\ q(\cdot) \text{ is increasing and convex on } (0,\infty), \text{ and } q'(\cdot) \text{ is concave on } (0,\infty) \Big\}.$$

Examples of such distributions include normal, Laplace, exponential power densities with $q(y) = \alpha\, y^\beta + c$, $\alpha > 0$, $1 \leq \beta \leq 2$; Hyperbolic Secant, Logistic, Generalized logistic densities with $q(y) = -y + \frac{2}{\alpha}\log(1+e^{\alpha y}) + c$, $\alpha > 0$; and Champernowne densities with $q(y) = \log(\cosh(y) + \beta)$, $\beta \in [0,2]$, (also see Marchand

and Perron, 2009; Marchand *et al.*, 2008). The next theorem establishes for such densities the applicability of part (a) of Corollary 2.1 and dominance of $\delta_H(X)$ over both the linear minimax estimator, $\delta_{\mathrm{LMX}}(X)$, and the MRE estimator $\delta_0(X)$.

Theorem 3.1. *For model densities as in (3.1) with $q \in Q^*$, the estimator $\delta_H(X) = H(\|X\|, m)X$ dominates both the linear minimax estimator $\delta_{\mathrm{LMX}}(X)$ and the MRE estimator $\delta_0(X)$.*

Proof: By virtue of Corollary 2.1, it suffices to show that $H(r,\lambda)$ decreases in $r \in [0,\infty)$ for all $\lambda \in [0,m]$, and increases in $\lambda \in [0,m]$ for all $r \geq 0$. First, $H(r,\lambda)$ can be written as

$$\begin{aligned} H(r,\lambda) &= \lambda \frac{\int_r^\infty x\big(f_0(x-\lambda) - f_0(x+\lambda)\big)\,dx}{\int_r^\infty x^2\big(f_0(x-\lambda) + f_0(x+\lambda)\big)\,dx} \\ &= \lambda^2 E_\lambda\left(\frac{\tanh\Big(\big(q(Y+\lambda) - q(Y-\lambda)\big)/2\Big)}{\lambda Y}\right), \end{aligned}$$

where Y is a random variable with density proportional to $y^2(f_0(y-\lambda)+f_0(y+\lambda))\cdot$ $\cdot 1_{[r,\infty)}(y)$. Such a family of densities with parameter r has increasing monotone likelihood ratio in Y. Furthermore, since $q \in Q^*$, a result of Marchand *et al.* (2008) (Lemma 1, part e) tells us that the inner function of the above expectation in Y is nonincreasing. Hence, we conclude that, for all $\lambda \in [0,m]$, $H(\lambda,\cdot)$ decreases on $[0,\infty)$. Turning to the monotonicity of $H(\cdot,r)$, begin by writing

$$\begin{aligned} H(r,\lambda) &= \lambda \frac{\int_r^\infty x\big(f_0(x-\lambda) - f_0(x+\lambda)\big)\,dx}{\int_r^\infty x^2\big(f_0(x-\lambda) + f_0(x+\lambda)\big)\,dx} \\ &= \lambda \frac{\int_{r-\lambda}^\infty (y+\lambda)\,f_0(y)\,dy \;-\; \int_{r+\lambda}^\infty (y-\lambda)\,f_0(y)\,dy}{\int_{r-\lambda}^\infty (y+\lambda)^2\,f_0(y)\,dy \;+\; \int_{r+\lambda}^\infty (y-\lambda)^2\,f_0(y)\,dy} \\ &= \lambda \frac{A(r,\lambda)}{B(r,\lambda)}\,, \end{aligned}$$

where $A(r,\lambda)$ and $B(r,\lambda)$ are the numerator and denominator of the above fraction, respectively. Manipulations yield:

$$\begin{aligned} B^2(r,\lambda)\frac{\partial H(r,\lambda)}{\partial \lambda} &= A(r,\lambda)\,B(r,\lambda) + \lambda A'(r,\lambda)\,B(r,\lambda) - \lambda A(r,\lambda)\,B'(r,\lambda) \\ &= \Big[l(r,\lambda) + A_1(r,\lambda)\Big] \\ &\quad \cdot \Big[B_1(r,\lambda) + r\lambda\big(\lambda f_0(r-\lambda) + \lambda f_0(r+\lambda) - r f_0(r-\lambda) + r f_0(r+\lambda)\big)\Big] \\ &\quad + \Big[r\lambda\big(f_0(r-\lambda) + f_0(r+\lambda)\big) + A_1(r,\lambda)\Big]\Big[B_1(r,\lambda) + \lambda\, l(r,\lambda)\Big] \\ &= r\lambda\, G(r,\lambda)\, f_0(r-\lambda) + 2\,A_1(r,\lambda)\,B_1(r,\lambda) + r\lambda^2 f_0(r+\lambda)\,l(r,\lambda) \\ &\quad + r^2\lambda\, f_0(r+\lambda)\,l(r,\lambda) + r\lambda\, f_0(r+\lambda)\,B_1(r,\lambda) + r\lambda^2 f_0(r+\lambda)\,l(r,\lambda)\,, \end{aligned}$$

where

$$G(r,\lambda) = 2\,\lambda \int_{r-\lambda}^{r+\lambda} y f_0(y)\,dy - r\int_{r-\lambda}^{r+\lambda} y f_0(y)\,dy + \int_{r-\lambda}^{\infty} y^2 f_0(y)\,dy + \int_{r+\lambda}^{\infty} y^2 f_0(y)\,dy\ ,$$

$$l(r,\lambda) = \int_{r-\lambda}^{r+\lambda} y f_0(y)\,dy\ ,$$

$$A_1(r,\lambda) = \lambda\Big(\int_{r-\lambda}^{\infty} f_0(y)\,dy + \int_{r+\lambda}^{\infty} f_0(y)\,dy\Big)\,,$$

$$B_1(r,\lambda) = \int_{r-\lambda}^{\infty} y^2 f_0(y)\,dy + \int_{r+\lambda}^{\infty} y^2 f_0(y)\,dy\ .$$

Now, observe that for all $r \geq 0$, $\lambda \in [0,m]$, the quantities $B_1(r,\lambda)$, $A_1(r,\lambda)$, and (r,λ) are nonnegative. Hence, to show the positivity of $\frac{\partial H(r,\lambda)}{\partial \lambda}$, it will suffice to show the positivity of $G(r,\lambda)$. But, we have

$$\begin{aligned} G(r,\lambda) &\geq \int_{r-\lambda}^{r+\lambda} y\, f_0(y)\,(2\,\lambda - r + y)\,dy \\ &\geq \int_{0}^{r+\lambda} \lambda\, y\, f_0(y)\,dy\ 1_{[\lambda,\infty)}(r) + \int_{r-\lambda}^{\lambda-r} y\, f_0(y)\,(2\,\lambda - r + y)\,dy\ 1_{[0,\lambda)}(r) \\ &\geq \int_{0}^{\lambda-r} 2\,y^2 f_0(y)\,dy\ 1_{[0,\lambda)}(r)\ \geq\ 0\ , \end{aligned}$$

which completes the proof. □

3.2. Multivariate normal distributions

We consider here multivariate normal models in (2.1) $X \sim N_p(\theta,\sigma^2)$ with $\|\theta\| \leq m$. We take $\sigma^2 = 1$ without loss of generality (since $\frac{X}{\sigma} \sim N_p(\theta' = \frac{\theta}{\sigma}, I_p)$ with $\|\theta'\| \leq m' = \frac{m}{\sigma}$). We require the following key properties relative to $\rho(\lambda,r) = E_\theta\Big(\frac{\theta^T X}{\|X\|}\,|\ \|X\| = r\Big)$, where $\lambda = \|\theta\|$. These properties involve modified Bessel functions $\mathbb{I}_v$ of order v, and more specifically ratios of the form $\rho_v(t) = \mathbb{I}_{v+1}(t)/\mathbb{I}_v(t)$, $t > 0$.

Lemma 3.1 (Watson, 1983; Marchand and Perron, 2001)**.**

- (**i**) *We have* $\rho(\lambda,r) = \lambda\rho_{p/2-1}(\lambda\,r)$*;*
- (**ii**) $\rho_{p/2-1}(\cdot)$ *is increasing and concave on* $[0,\infty)$*, with* $\rho_{p/2-1}(0) = 0$ *and* $\lim_{t\to\infty} \rho_{p/2-1}(t) = 1$*;*
- (**iii**) $\rho_{p/2-1}(t)/t$ *is decreasing in* t *with* $\lim_{t\to 0^+} \rho_{p/2-1}(t)/t = 1/p$*;*
- (**iv**) $\rho_{p/2}(t) = \rho^{-1}_{p/2-1}(t) - p/t$*.*

Denoting $f_p(\cdot,\lambda)$ and $\bar{F}_p(\cdot,\lambda)$ as the probability density and survival functions of $R = \|X\| \sim \sqrt{\chi^2_p(\lambda^2)}$, we will also require the following useful properties.

Lemma 3.2.

(**i**) *We have* $f_p(r,\lambda) = r\big(\frac{r}{\lambda}\big)^{p/2-1}\,\mathbb{I}_{p/2-1}(r\lambda)\,\exp\big\{-\frac{r^2+\lambda^2}{2}\big\}$;

(**ii**) $r^2 f_p(r,\lambda) = \lambda^2 f_{p+4}(r,\lambda) + p\, f_{p+2}(r,\lambda)$;

(**iii**) $r\, f_p(r,\lambda)\,\rho_{p/2-1}(\lambda r) \,=\, \lambda\, f_{p+2}(r,\lambda)$;

(**iv**) *the ratio* $\frac{\bar{F}_{p+2}(r,\lambda)}{\bar{F}_p(r,\lambda)}$ *decreases in* $\lambda \in [0,\infty)$, *for all* $p \geq 1$ *and* $r > 0$.

Proof: Parts (ii) and (iii) follow directly from (i), while (i) consists of a well known Bessel function representation of the noncentral chi-square distribution. Part (iv) follows from the identity $2\,\frac{\partial}{\partial\lambda}\bar{F}_p(r,\lambda) = \bar{F}_{p+2}(r,\lambda) - \bar{F}_p(r,\lambda)$, and the logconcavity of $\bar{F}_p(r,\cdot)$ on $[0,\infty)$ (see Das Gupta and Sarkar, 1984; Finner and Roters, 1997). □

We now seek to apply part (a) of Corollary 2.1.

Theorem 3.2. *For multivariate normal densities, the estimator* $\delta_H(X) = H(\|X\|,m)X$ *dominates both the linear minimax estimator* $\delta_{\mathrm{LMX}}(X)$ *and the MRE estimator* $\delta_0(X)$.

Proof: By virtue of Corollary 2.1, it suffices to show that $H(r,\lambda)$ decreases in $r \in [0,\infty)$ for all $\lambda \in [0,m]$, and increases in $\lambda \in [0,m]$ for all $r \geq 0$. Making use of Lemmas 3.1 and 3.2, we obtain

$$
\begin{aligned}
H(r,\lambda) \;&=\; \frac{E_\theta\Big(\|X\|\,E_\theta\Big(\frac{\theta^T X}{\|X\|}\,\big|\;\|X\| \geq r\Big)\Big)}{E_\theta\big(\|X\|^2\,\big|\;\|X\| \geq r\big)} \\
&=\; \frac{\int_\infty^r y\,E_\theta\Big(\frac{\theta^T X}{\|X\|}\,\big|\;\|X\| = y\Big)\,f_p(y,\lambda)\,dy}{\int_r^\infty y^2 f_p(y,\lambda)\,dy} \\
&=\; \frac{\int_\infty^r y\,\lambda\rho_{p/2-1}(\lambda y)\,f_p(y,\lambda)\,dy}{\int_r^\infty y^2\,f_p(y,\lambda)\,dy} \qquad (3.2) \\
&=\; \frac{\int_\infty^r \lambda^2 f_{p+2}(y,\lambda)\,dy}{\int_r^\infty y^2\,f_p(y,\lambda)\,dy} \\
&=\; \left\{\frac{p}{\lambda^2} + \frac{\int_r^\infty f_{p+4}(y,\lambda)\,dy}{\int_r^\infty f_{p+2}(y,\lambda)\,dy}\right\}^{-1} \\
&=\; \left\{\frac{p}{\lambda^2} + \frac{\bar{F}_{p+4}(r,\lambda)}{\bar{F}_{p+2}(r,\lambda)}\right\}^{-1}.
\end{aligned}
$$

The monotonicity property of $H(r,\cdot)$ on $[0,m]$ for all $r \geq 0$ now follows from the above expression and part (iv) of Lemma 3.2.

Now, to show that $H(r,\lambda)$ decreases in r, make use of (3.2) to write

$$\begin{aligned} H(r,\lambda) &= \lambda\, E_r\left(\frac{E_Y\Big(\frac{\theta^T X}{\|X\|}\,|\,\|X\| = Y\Big)}{Y}\right) \\ &= \lambda\, E_r\left(\frac{\rho_{p/2-1}(\lambda Y)}{Y}\right), \end{aligned}$$

where Y has density proportional to $y\, f_p(y,\lambda) 1_{[r,\infty)}(y)$. Since this family of densities with parameter r has increasing monotone likelihood ratio in Y, we conclude indeed that $H(r,\lambda)$ decreases for $r \geq 0$ for all $\lambda \in [0,m]$ by making use of part (iii) of Lemma 3.1. □

3.3. Scale mixtures of multivariate normal distributions

We consider here in this subsection scale mixtures of multivariate normal distributions where X admits the representation: $X|Z = z \sim N_p(\theta, z I_p)$, Z having Lebesgue density g on $\mathbb{R}^+$. The corresponding density in (2.1) is of the form

$$\int_0^\infty (2\pi z)^{-p/2} \exp\left\{-\frac{\|x-\theta\|^2}{2z}\right\} g(z)\, dz\,; \tag{3.3}$$

and we further assume that g is logconcave on either $\mathbb{R}^+$ or some open interval (a,b) of $\mathbb{R}^+$. Uniform densities on (a,b) are included. With such a representation, since $X/\sqrt{Z}|Z = z \sim N_p\big(\theta/\sqrt{z}, I_p\big)$, we infer from part (i) of Lemma 3.2 that the density function of $R = \|X\|$ is given by

$$\int_0^\infty \frac{y}{z}\left(\frac{y}{\lambda}\right)^{p/2-1} \mathbb{I}_{p/2-1}\left(\frac{\lambda y}{z}\right) \exp\left\{-\frac{y^2+\lambda^2}{2z}\right\} g(z)\, dz\,. \tag{3.4}$$

We now seek to apply part (a) of Corollary 2.1.

Theorem 3.3. *For scale mixtures of multivariate normal densities as in (3.3) with g logconcave, the estimator $\delta_{H^*}(X) = H^*(\|X\|, m)X$ dominates the MRE estimator $\delta_0(X)$.*

Proof: By virtue of Corollary 2.1, it suffices to show that $H^*(r,\cdot)$ is non-decreasing on $[0,m]$ for all $r \geq 0$ under the given logconcave assumption on g.

Starting from the definition of H^* and making use of 3.4, we obtain

$$\begin{aligned} H^*(r,\lambda) &= \frac{\lambda\, E_\theta\big(R\,|\,R\ge r\big)}{E_\theta\big(R^2|\,R\ge r\big)} \\ &= \frac{\lambda\int_r^\infty\!\!\int_0^\infty y^{\frac{p}{2}+1}\,\frac{g(z)}{z}\,\mathbb{I}_{\frac{p}{2}-1}\big(\frac{y\lambda}{z}\big)\,e^{-\frac{y^2+\lambda^2}{2z}}\,dz\,dy}{\int_r^\infty\!\!\int_0^\infty y^{\frac{p}{2}+2}\,\frac{g(z)}{z}\,\mathbb{I}_{\frac{p}{2}-1}\big(\frac{y\lambda}{z}\big)\,e^{-\frac{y^2+\lambda^2}{2z}}\,dz\,dy} \\ &= \frac{\int_{r/\lambda}^\infty\!\!\int_0^\infty x^{\frac{p}{2}+1}\,\frac{g(\lambda^2 t)}{t}\,\mathbb{I}_{\frac{p}{2}-1}\big(\frac{x}{t}\big)\,e^{-\frac{1+x^2}{2t}}\,dt\,dx}{\int_{r/\lambda}^\infty\!\!\int_0^\infty x^{\frac{p}{2}+2}\,\frac{g(\lambda^2 t)}{t}\,\mathbb{I}_{\frac{p}{2}-1}\big(\frac{x}{t}\big)\,e^{-\frac{1+x^2}{2t}}\,dt\,dx}\,, \end{aligned}$$

with the change of variables $(y,z)=(\lambda x,\lambda^2 t)$. Simple differentiation leads to $\frac{\partial}{\partial\lambda}H^*(r,\lambda)=\frac{1}{B^2}\{A_1-A_2+A_3-A_4\}$, where B is the above denominator of H^*,

$$\begin{aligned} A_1 &= 2\,\lambda\int_{r/\lambda}^\infty\!\!\int_0^\infty x\,M(x,t)\,dt\,dx\int_{r/\lambda}^\infty\!\!\int_0^\infty x^{\frac{p}{2}+1}\,g'(\lambda^2 t)\,\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x}{t}\right)e^{-\frac{1+x^2}{2t}}\,dt\,dx\,, \\ A_2 &= 2\,\lambda\int_{r/\lambda}^\infty\!\!\int_0^\infty M(x,t)\,dt\,dx\int_{r/\lambda}^\infty\!\!\int_0^\infty x^{\frac{p}{2}+2}\,g'(\lambda^2 t)\,\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x}{t}\right)e^{-\frac{1+x^2}{2t}}\,dt\,dx\,, \\ A_3 &= \frac{r}{\lambda^2}\int_{r/\lambda}^\infty\!\!\int_0^\infty x\,M(x,t)\,dt\,dx\int_0^\infty\frac{g(\lambda^2 t)}{t}\left(\frac{r}{\lambda}\right)^{\frac{p}{2}+1}\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{r}{\lambda t}\right)e^{-\frac{\lambda^2+r^2}{2\lambda^2 t}}\,dt\,, \\ A_4 &= \frac{r}{\lambda^2}\int_{r/\lambda}^\infty\!\!\int_0^\infty M(x,t)\,dt\,dx\int_0^\infty\frac{g(\lambda^2 t)}{t}\left(\frac{r}{\lambda}\right)^{\frac{p}{2}+2}\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{r}{\lambda t}\right)e^{-\frac{\lambda^2+r^2}{2\lambda^2 t}}\,dt\,, \end{aligned}$$

with $M(x,t)=\frac{g(\lambda^2 t)}{t}\,x^{\frac{p}{2}+1}\,\mathbb{I}_{\frac{p}{2}-1}\big(\frac{x}{t}\big)\,e^{-\frac{1+x^2}{2t}}$. Obviously, $A_3-A_4\ge 0$, because $x\ge\frac{r}{\lambda}$ on the domain of integration. Furthermore, by setting $h(z)=\big(-g'(z)/g(z)\big)\cdot$ $\cdot 1_{\{z:g(z)>0\}}(z)$, we have

$$\begin{aligned} A_1-A_2 &= 2\,\lambda\int_{r/\lambda}^\infty\!\!\int_0^\infty M(x,t)\,dt\,dx\int_{r/\lambda}^\infty\!\!\int_0^\infty h(\lambda^2 t)\,xt\,M(x,t)\,dt\,dx \\ &\quad -2\,\lambda\int_{r/\lambda}^\infty\!\!\int_0^\infty x\,M(x,t)\,dt\,dx\int_{r/\lambda}^\infty\!\!\int_0^\infty h(\lambda^2 t)\,t\,M(x,t)\,dt\,dx\,. \end{aligned}$$

Now, since h is increasing with the logconcavity of g, the FKG's inequality (see Lemma A.1 in the Appendix) implies that A_1-A_2 is nonnegative whenever $M(x_1,t_2)M(x_2,t_1)-M(x_1,t_1)M(x_2,t_2)\le 0$, for $0\le x_1\le x_2$ and $0\le t_1\le t_2$. From the definition of M, manipulations yield for non-zero values of $M(x_1,t_2)\cdot$

$$\cdot M(x_2,t_1) - M(x_1,t_1)\,M(x_2,t_2)\text{:}$$

$$\begin{aligned}
&\frac{t_1 t_2\, e^{(1/t_1+1/t_2)}}{(x_1x_2)^{p/2+1}\, g(\lambda^2 t_1)\, g(\lambda^2 t_2)} \Big\{ M(x_1,t_2)\,M(x_2,t_1) - M(x_1,t_1)\,M(x_2,t_2) \Big\} = \\
&= \mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_1}{t_2}\right)\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_2}{t_1}\right) - \mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_1}{t_1}\right)\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_2}{t_2}\right) \exp\Big\{-(x_1^2-x_2^2)\,(1/t_1-1/t_2)\Big\} \\
&= \mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_1}{t_2}\right)\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_2}{t_2}\right)\left[\frac{\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_2}{t_1}\right)}{\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_2}{t_2}\right)} - \frac{\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_1}{t_1}\right)}{\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_1}{t_2}\right)} \exp\Big\{-(x_1^2-x_2^2)\,(1/t_1-1/t_2)\Big\}\right] \\
&\le \mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_1}{t_2}\right)\mathbb{I}_{\frac{p}{2}-1}\!\left(\frac{x_2}{t_1}\right)\left(\frac{t_2}{t_1}\right)^{p/2-1}\left[1-\exp\Big\{(x_2^2-x_1^2+x_1)\,(1/t_1-1/t_2)\Big\}\right] \\
&\le 0\,,
\end{aligned}$$

where the former inequality follows from the Ross inequality applications (see Lemma A.2 in Appendix): $\frac{\mathbb{I}_{p/2-1}(x_2/t_1)}{\mathbb{I}_{p/2-1}(x_2/t_2)} \le (t_2/t_1)^{p/2-1}$ and $\frac{\mathbb{I}_{p/2-1}(x_1/t_1)}{\mathbb{I}_{p/2-1}(x_1/t_2)} \ge (t_2/t_1)^{p/2-1}\exp\{x_1/t_1 - x_1/t_2\}$, and where the latter inequality follows from the fact that $(x_2^2 - x_1^2 + x_1)\,(1/t_1 - 1/t_2) \ge 0$, for $0 \le x_1 \le x_2$ and $0 \le t_1 \le t_2$. □

APPENDIX

The FKG inequality due to Fortuin, Kasteleyn, and Ginibre (1971) is useful for Theorem 3.3.

Lemma A.1 (**FKG inequality**). *Suppose a p-variate random variable X is distributed with probability density function ξ and with positive measure ν. For two points $y = (y_1, ..., y_p)$ and $z = (z_1, ..., z_p)$, in the sample space of X, we define $y \wedge z = (y_1 \wedge z_1, ..., y_p \wedge z_p)$ and $y \vee z = (y_1 \vee z_1, ..., y_p \vee z_p)$, where $a \wedge b = \min(a,b)$, $a \vee b = \max(a,b)$. Suppose that ξ satisfies $\xi(y)\,\xi(z) \le \xi(y \vee z)\,\xi(y \wedge z)$ and that $\alpha(y)$, $\beta(y)$ are nondecreasing in each argument and α, β and $\alpha\,\beta$ are integrable with respect to ξ. Then $\int \alpha\beta\xi\, d\nu \ge \int \alpha\xi\, d\nu \int \beta\xi\, d\nu$.*

The following lemma, referred to as the Ross inequality is due to Joshi and Bissu (1991) and establishes useful bounds for a ratio of modified Bessel functions.

Lemma A.2. *Suppose $\mathbb{I}_v(x)$ and $\mathbb{I}_v(y)$ are two modified Bessel functions. Moreover, suppose that $y \ge x$ and $v \ge -\frac{1}{2}$. Then*

$$e^{x-y}\left(\frac{x}{y}\right)^v \le \frac{\mathbb{I}_v(x)}{\mathbb{I}_v(y)} \le \left(\frac{x}{y}\right)^v.$$

ACKNOWLEDGMENTS

Éric Marchand acknowledges the partial research support from NSERC of Canada. We are grateful to Dan Kucerovsky and Reinaldo Arellano-Valle for useful suggestions on the Ross inequality and properties of the H function in Theorem 2.1. Thanks to an anonymous reviewer for constructive comments.

REFERENCES

[1] Das Gupta, S. and Sarkar, S.K. (1984). On TP_2 and log-concavity, *Inequalities in Statistics and Probability, IMS Lecture Notes Monogr. Ser.*, **5**, 54–58.

[2] Fortuin, C.M.; Kasteleyn, P.W. and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets, *Comm. Math. Phys.*, **22**, 89–103.

[3] Fourdrinier, D. and Marchand, É. (2010). On Bayes estimators with uniform priors on spheres and their comparative performance with maximum likelihood estimators for estimating bounded multivariate normal means, *Journal of Multivariate Analysis*, **101**, 1390–1399.

[4] Hartigan, J. (2004). Uniform priors on convex improve risk, *Statistics and Probability Letters*, **67**, 285–288.

[5] Finner, H. and Roters, M. (1997). Log-concavity and inequalities for Chi-square, F and Beta distributions with applications in multiple comparisons, *Statistica Sinica*, **7**, 771–787.

[6] Joshi, C.M. and Bissu, S.K. (1991). Some inequalities of Bessel and modified Bessel functions, *Australian Mathematical Society Journal, Series A, Pure Mathematics and Statistics*, **50**, 333–342.

[7] Kubokawa, T. (1994). A unified approach to improving equivariant estimators, *Annals of Statistics*, **22**, 290–299.

[8] Kubokawa, T. (2005). Estimation of bounded location and scale parameters, *Journal of the Japan Statistical Society*, **35**, 221–249.

[9] Marchand, 'E.; Ouassou, I.; Payandeh, A.T. and Perron, F. (2008). On the estimation of a restricted location parameter for symmetric distributions, *Journal of the Japan Statistical Society*, **38**, 1–17.

[10] Marchand, É. and Perron, F. (2001). Improving on the MLE of a bounded normal mean, *The Annals of Statistics*, **29**, 1078–1093.

[11] Marchand, É. and Perron, F. (2005). Improving on the MLE of a bounded mean for spherical distributions, *Journal of Multivariate Analysis*, **92**, 227–238.

[12] Marchand, É. and Strawderman, W.E. (2004). Estimation in restricted parameter spaces: a review, *A Festschrift for Herman Rubin, IMS Lecture Notes Monogr. Ser.*, **45**, 21–44.

[13] MARCHAND, É. and STRAWDERMAN, W.E. (2005). Improving on the minimum risk equivariant estimator for a location parameter which is constrained to an interval or a half-interval, *Annals of the Institute of Statistical Mathematics*, **57**, 129–143.

[14] VAN EEDEN, C. (2006). *Restricted parameter space estimation problems. Admissibility and minimaxity properties*, Lecture Notes in Statistics, **188**, Springer, New York.

[15] WATSON, G.S. (1983). *Statistics on Spheres*, John Wiley and Sons, Inc., New York.

REVSTAT – Statistical Journal
Volume 8, Number 2, November 2010, 139–149

ESTIMATION OF THE PARAMETER OF A *pARMAX* MODEL

Author: Marta Ferreira
– Department of Mathematics, University of Minho, Portugal
msferreira@math.uminho.pt

Received: November 2009 Revised: July 2010 Accepted: September 2010

Abstract:

- Max-autoregressive models for time series data are useful when we want to make inference about rare events, mainly in areas like hydrology, geophysics and finance. In fact, they are more convenient for analysis than heavy-tailed ARMA, as their finite-dimensional distributions can easily be written explicitly. The recent power max-autoregressive model ($pARMAX$) has the interesting feature of describing an asymptotic independent tail behavior, a property that can be observed in various data series. An estimator of the model parameter c $(0 < c < 1)$ is already available in the literature, but only in the restrictive case $c > 1/2$. Here it is presented an estimator for all $c \in (0, 1)$. Consistency and asymptotic normality are also stated.

Key-Words:

- *extreme value theory; max-autoregressive processes.*

AMS Subject Classification:

- 60G70, 60J10.

1. INTRODUCTION

Extreme Value Theory (EVT) is an important tool for many applied sciences whenever we are faced with modeling high values of certain phenomena. Ocean wave modeling, wind engineering, thermodynamics of earthquakes, risk assessment on financial markets are some examples. The first results were developed considering independent observations but, more recently, models for extreme values have been constructed under the more realistic assumption of temporal dependence. Among these models, stationary Markov chains are very interesting, specially because they may have a somewhat simple treatment in what concerns extremal properties. The max-autoregressive moving average processes MARMA (Davis and Resnick [7]), and also the particular case MAR(1) or ARMAX, given by,

$$X_i \,=\, k\, X_{i-1} \vee W_i \ ,$$

with $0 < k < 1$ and $\{W_i\}_{i\in\mathbb{Z}}$ i.i.d. (Alpuim [2]; Canto e Castro [6]; Ancona-Navarrete and Tawn [3]; Beirlant *et al.* [4]; Lebedev [12]) are some examples. Heavy tailed MARMA and ARMA are both good choices for modeling time series data with sudden large peaks, although the former are more convenient for analysis as their finite-dimensional distributions can easily be written explicitly.

More recently, some careful attention has been given to the statistical modeling of the tail dependence between consecutive pairs from a stationary first-order Markov chain, since it is important to distinguish asymptotic dependence from asymptotic independence. More precisely, according to Bortot and Tawn ([5]), a Markov chain $\{Y_i\}$ is said to be asymptotically tail dependent or independent, whenever $b > 0$ or $b = 0$, respectively, in the limit below:

$$\lim_{y\to y^*} P\big(Y_2 > y \,|\, Y_1 > y\big) \,=\, b \ ,$$

where y^* is the right-endpoint of Y_1, i.e., $y^* = \sup\{y : P(Y_1 \le y) < 1\}$. For asymptotically tail independent Markov chains, the dependence between exceedances of y gradually decreases as $y \to y^*$, which leads to an extremal feature increasingly resembling an i.i.d. sequence at high levels. As pointed out in Bortot and Tawn ([5]), this phenomenon has been noticed in a number of data and theoretical applications. In these cases, procedures as in Smith *et al.* ([16]) assuming that the limiting behavior of the chain is exact above a fixed high threshold, and hence the dependence structure between consecutive random variables (r.v.'s) above the threshold can be modeled through a bivariate extreme value distribution, are not suitable. This problem is overcome by setting the way how $P\big(Y_2 > y \,|\, Y_1 > y\big)$ converges to zero, as $y \to y^*$, which involves the coefficient of asymptotic tail dependence η (Ledford and Tawn [13], [14]). This is a nontrivial class, including many commonly studied processes, such as Gaussian Markov chains (Sibuya [15]).

Coefficient η characterizes the asymptotic tail dependence behavior, i.e., $\eta = 1$ corresponds to tail dependence whilst $\eta < 1$ means asymptotic tail independence, with $\eta = 1/2$ occurring for the (almost) independent case. The *ARMAX* process, which has unit η, is in the group of tail dependent Markov chains (Ferreira and Canto e Castro [8]) and hence is not suitable to model data series expressing the described phenomenon.

Ferreira and Canto e Castro ([8]) introduced the power max-autoregressive process (in short, *pARMAX*), defined as,

$$X_i = X_{i-1}^c \vee Z_i \ , \qquad 0 < c < 1, \quad i \in \mathbb{Z} \ ,$$

with $\{Z_i\}$ i.i.d., for which η is a function of the model parameter c, under the very mild assumption of heavy tailed innovations. More precisely, we have $\eta = \max(1/2, c)$ and hence *pARMAX* is an asymptotic tail independent process, even almost independent in cases $c \leq 1/2$. Hence, it is a suitable model to describe the above mentioned phenomenon of time series exhibiting asymptotic tail independence. In Figure 1, the similarity between the sample paths of heavy tailed *pARMAX* and AR(1) processes, in this case based on marginal d.f.'s Pareto$(1/\gamma)$, with shape parameter $\gamma > 0$, given by

$$K(x) = 1 - x^{-1/\gamma} \ , \qquad x \geq 1 \ , \tag{1.1}$$

indicate that the former can be considered as an alternative for data modeling, particularly with respect to extreme values. The *pARMAX* process has easily derived extremal properties and also easily explicited finite-dimensional distributions (Ferreira and Canto e Castro [8], [9]). Moreover, a generalization of *pARMAX* has also been applied in modeling financial data (Ferreira and Canto e Castro [10]). Based on the estimation procedure for the Ledford and Tawn coefficient η, Ferreira and Canto e Castro ([9]) presented consistent and asymptotically normal estimators for the process parameter c, which applies only in cases where $c > 1/2$. Following a similar procedure to that of Lebedev ([12]) to estimate the parameter of unit Fréchet ARMAX, an estimator for the *pARMAX* parameter c is derived, this time covering all values of $c \in (0, 1)$. From a Klotz's result (Klotz [11], Theorem 1), consistence and asymptotic normality are easily stated.

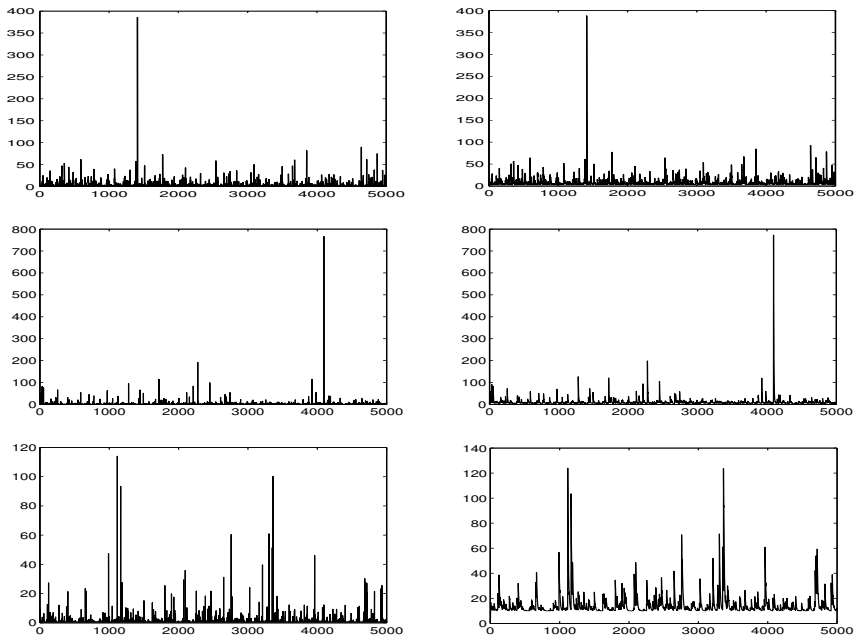

Figure 1: 5000 realizations of *pARMAX*, $X_i = X_{i-1}^c \vee Z_i$, on the left, and of AR(1), $X_i = c\, X_{i-1} + Z_i$, on the right, with, from top to bottom, $c = 0.7$, 0.8, 0.9, respectively, and with marginal Pareto(0.7).

2. THE *pARMAX* PROCESS

Consider $\{Z_i\}$ a sequence of i.i.d. copies of a r.v., Z, having real nonnegative support and marginal d.f. F_Z. A sequence $\{X_i\}$ is said to be a *pARMAX* process if,

$$X_i = X_{i-1}^c \vee Z_i \ , \qquad 0 < c < 1, \quad i = 0, \pm 1, \pm 2, ... \tag{2.1}$$

with X_i independent of Z_j, for all integer $i < j$. The sequence $\{Z_i\}$ is also known as the innovations sequence of the process.

In the sequel we consider that $\{Z_i\}$ has support in $[1, \infty[$, a necessary condition for stationarity.

Let K be the marginal distribution function (d.f.) of the process. Hence K is a solution of the equation

$$K(x) = K(x^{1/c})\, F_Z(x) \ . \tag{2.2}$$

(See Ferreira and Canto e Castro [8], [9] for details). An example of a stationary *pARMAX* process is given below.

Example 2.1. Consider $\{Z_i\}$ with common d.f.,

$$F_Z(x) = c\,\mathbf{1}_{\{x=1\}} + \frac{1-x^{-1/\gamma}}{1-x^{-1/(c\gamma)}}\,\mathbf{1}_{\{x>1\}}\ ,$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. Hence, the Pareto$(1/\gamma)$ d.f. given in (1.1) satisfies (2.2), being, therefore, a stationary distribution for X_i.

The *k-step* transition probability function (t.p.f.) from x to $]-\infty, y]$, given by,

$$Q^k\big(x,]-\infty, y]\big) := P\big(X_{n+k} \le y \,|\, X_n = x\big) = \frac{K(y)}{K(y^{1/c^k})}\,\mathbf{1}_{\left\{x \le y^{1/c^k}\right\}}\ , \tag{2.3}$$

where the last step is due to (2.2), will be used in the forward results.

2.1. Parameter estimation

Now we will present an estimator for the pARMAX parameter (c) based on a similar procedure as in Lebedev ([12]) for unit Fréchet max-autoregressive, $X_i = \max\big(cX_{i-1}, (1-c)Z_i\big)$. In the pARMAX case, Pareto marginals will be considered.

Set, for each $k \ge 1$,

$$p_k = P\big(X_{k+1} \le X_1\big)\,. \tag{2.4}$$

The following result states a relation between p_k and parameter c, more precisely, c^k. For sake of simplicity, from now on consider $a_k := c^k$.

Proposition 2.1. *Let $\{X_i\}$ be a stationary pARMAX process as defined in (2.1) with marginal d.f. K satisfying (1.1). Then the equality,*

$$p_k = a_k\big(\psi(2\,a_k) - \psi(a_k)\big)\,, \tag{2.5}$$

holds where ψ is the well-known digamma function, i.e., $\psi(z) = \Gamma'(z)/\Gamma(z)$ with $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ the Euler Gamma function.

Proof: Just observe that, using (2.3), we have,

$$\begin{aligned} p_k &= \int P\big(X_{k+1} \le x \,|\, X_1 = x\big)\, dK(x) \\ &= \int Q^k\big(x,]-\infty, x]\big)\, dK(x) \qquad (2.6) \\ &= \int \frac{K(x)}{K(x^{1/c^k})}\, dK(x)\ , \qquad k \ge 1\ . \end{aligned}$$

where after some algebra (see for instance Abramowitz and Stegun [1]) and notation $a_k = c^k$, expression (2.5) can be derived. □

Note that p_k does not depend on the marginal d.f. parameter γ. There exist simple estimates for the above probabilities:

$$\widehat{p}_k = \frac{1}{n-k} \sum_{j=k+1}^{n} \mathbf{1}_{\{X_j \le X_{j-k}\}} , \qquad k \ge 1 . \tag{2.7}$$

The next result states consistency and asymptotic normality for estimators $\widehat{a_k}\,(=\widehat{c^k})$, obtained from equation (2.5) by plugging in the empirical estimates $\widehat{p}_k$. More precisely, we have the following result.

Proposition 2.2. *Let $\{X_i\}$ be a stationary pARMAX process as defined in Proposition 2.1. Then, for each positive integer k,*

$$n^{1/2}(\widehat{a_k} - a_k) \xrightarrow{D} N\big(0, \sigma_k^2/g'(a_k)^2\big) \tag{2.8}$$

where $g(x) = x\big(\psi(2\,x) - \psi(x)\big)$ and

$$\sigma_k^2 = p_k(1-p_k)\,(1 - 2\,p_k + \lambda_k)/(1-\lambda_k) , \tag{2.9}$$

with p_k given in (2.5) and $\lambda_k = p_k^{-1}\int_1^\infty \frac{1}{2}\Big[\frac{K(x)\,K(x^{1/a_{k-1}})}{K^2(x^{1/a_k})} + \frac{K^2(x^{a_1})}{K(x)\,K(x^{1/a_{k-1}})}\Big]\,K(dx)$.

Proof: Observe that $\widehat{p}_k$ is the mean of Bernoulli trials with Markov dependence. From Theorem 1 in Klotz ([11]), convergence $n^{1/2}(\widehat{p}_k - p_k) \xrightarrow{D} N(0, \sigma_k^2)$ holds for σ_k^2 given in (2.9), where $\lambda_k = P\big(X_j \le X_{j-k} \,|\, X_{j-1} \le X_{j-k-1}\big)$ with $\max\big(0, (2\,p_k - 1)/p_k\big) \le \lambda_k \le 1$. Hence, the result (2.8) is straightforward by the Delta Method.

In order to obtain the variance in (2.9) we must compute λ_k. First note that,

$$\lambda_k = \frac{P\big(X_j \le X_{j-k},\ X_{j-1} \le X_{j-k-1}\big)}{p_k} , \tag{2.10}$$

in which, using successive conditioning on the numerator lead us to,

$$P\big(X_j \le X_{j-k},\ X_{j-1} \le X_{j-k-1}\big) = \\ = \int_1^\infty \int_1^\infty \int_1^x Q\big(w,]-\infty, y]\big)\ Q^{k-1}(y, dw)\, Q(x, dy)\, K(dx) .$$

Now considering (2.3), the following development holds:

$$P\big(X_j \le X_{j-k},\ X_{j-1} \le X_{j-k-1}\big) =$$

$$= \int_1^\infty \int_1^\infty \int_1^{\min(x, y^{1/c})} F_Z(y)\ Q^{k-1}(y, dw)\, Q(x, dy)\, K(dx)$$

$$= \int_1^\infty \left[\int_1^{x^c} F_Z(y)\, Q^{k-1}\big(y,]-\infty, y^{1/c}]\big) + \int_{x^c}^\infty F_Z(y)\, Q^{k-1}\big(y,]-\infty, x]\big) \right] Q(x, dy)\, K(dx)$$

$$= \int_1^\infty \left[\int_1^{x^c} F_Z(y)\, \frac{K(y^{1/c})}{K(y^{1/c^k})} + \int_{x^c}^{x^{1/c^{k-1}}} F_Z(y)\, \frac{K(x)}{K(x^{1/c^{k-1}})} \right] Q(x, dy)\, K(dx) .$$

If d.f. F_Z admits density f_Z, the transition density of (2.3) is given by $q(x,y) = f_Z(y)\,\mathbf{1}_{\{x^c<y\}} + F_Z(x^c)\,\mathbf{1}_{\{x^c=y\}}$. Thus, the first term in the last integral is null and hence,

$$P\big(X_j \le X_{j-k},\ X_{j-1} \le X_{j-k-1}\big) = \tag{2.11}$$
$$= \int_1^\infty \frac{K(x)}{K\left(x^{1/c^{k-1}}\right)} \frac{F_Z^2\left(x^{1/c^{k-1}}\right) - F_Z^2(x^c)}{2}\, K(dx) + \int_1^\infty \frac{K(x)}{K\left(x^{1/c^{k-1}}\right)} F_Z^2(x^c)\, K(dx)\ .$$

Now the result follows from equation (2.2) and notation $a_k = c^k$. □

Note that $p_k \in (1/2, 1)$ (see Figure 2 and Table 1) and no definite results can be obtained for $\widehat{p}_k < 1/2$.

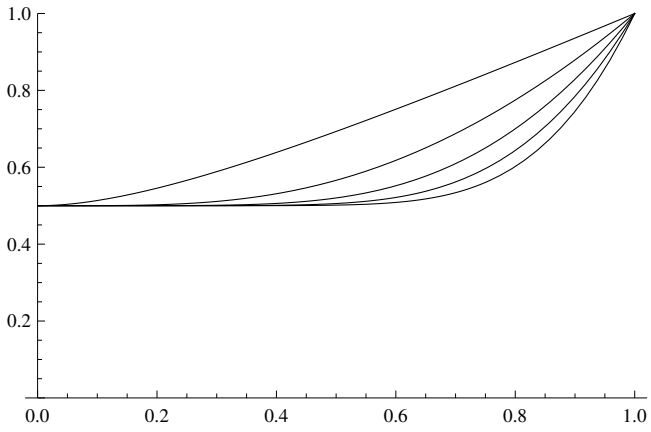

Figure 2: Plot of (from top to bottom) $p_1, ..., p_5$ given in (2.5).

Table 1: Values of p_k computed from (2.5), for Pareto marginal *pARMAX* processes with parameter values: $c = 0.1, 0.2, ..., 0.9$.

c	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
0.1	0.513472	0.500161	0.500002	0.5	0.5
0.2	0.545531	0.502419	0.500103	0.500004	0.5
0.3	0.588572	0.511114	0.501132	0.500106	0.50001
0.4	0.63855	0.531074	0.505905	0.501021	0.500169
0.5	0.693147	0.565986	0.52013	0.505648	0.501503
0.6	0.750948	0.617901	0.551814	0.521466	0.50849
0.7	0.811047	0.687525	0.609375	0.561751	0.533835
0.8	0.872845	0.77475	0.699936	0.643619	0.601832
0.9	0.935927	0.879101	0.828815	0.784424	0.74534

However, the probability of such events goes to zero as $n \to \infty$ and hence, this may be an indication of an inconsistency in our choice of the model. In what concerns the lag k, it can be chosen in order to obtain the smallest variance (σ_k^2) provided that the estimate, $\widehat{p}_k$, takes value in $(1/2, 1)$, which means as small as possible (see, for instance, Table 2).

Table 2: True values of λ_k and respective estimates, $\widehat{\lambda_k}$ in (2.12) and $\widetilde{\lambda_k}$ in (2.13), considering $n = 5000$ realizations of process $pARMAX$ for cases $c = 0.3, 0.5, 0.7, 0.9$, with marginal Pareto(1).

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
			$c = 0.3$		
λ_k	0.4805	0.5744	0.6070	0.6122	0.6128
$\widehat{\lambda_k}$	0.4778	0.5751	0.5991	0.6032	0.6071
$\widetilde{\lambda_k}$	0.4782	0.5754	0.5995	0.6035	0.6075
			$c = 0.5$		
λ_k	0.6393	0.6756	0.7047	0.7195	0.7250
$\widehat{\lambda_k}$	0.6461	0.6686	0.7057	0.7238	0.7236
$\widetilde{\lambda_k}$	0.6461	0.6687	0.7058	0.7239	0.7237
			$c = 0.7$		
λ_k	0.7930	0.8021	0.8119	0.8210	0.8283
$\widehat{\lambda_k}$	0.7973	0.8083	0.8114	0.8269	0.8277
$\widetilde{\lambda_k}$	0.7975	0.8083	0.8116	0.8270	0.8277
			$c = 0.9$		
λ_k	0.9341	0.9348	0.9356	0.9364	0.9373
$\widehat{\lambda_k}$	0.9334	0.9334	0.9342	0.9363	0.9362
$\widetilde{\lambda_k}$	0.9334	0.9334	0.9341	0.9362	0.9361

We remark that this procedure allows to estimate any value of $c \in (0, 1)$, and not only the case $c \in (1/2, 1)$ as in the method considered in Ferreira and Canto e Castro ([9]), which is based on the estimation of Ledford and Tawn tail dependence coefficient η. On the other hand, there is no explicit form for a_k in (2.5) and so it must be obtained numerically. Table 1 presents some computed values.

2.2. An illustrative example

An illustration is now presented. We consider 5000 realizations from $pARMAX$ process in (2.1), for cases $c = 0.3, 0.5, 0.7, 0.9$, with marginal distribution Pareto(1).

In order to obtain an estimate for the variance, we can replace in (2.9), p by $\widehat{p}_k$ stated in (2.7) and λ_k by the empirical counterpart

$$\widehat{\lambda}_k = \frac{1}{n-k-1} \sum_{k+2}^{n} \mathbf{1}_{\{X_j \le X_{j-k}\,,\, X_{j-1} \le X_{j-k-1}\}} \Big/ \widehat{p}_k \tag{2.12}$$

or alternatively, use the estimator proposed by Klotz ([11]),

$$\widetilde{\lambda}_k = \frac{r - \widehat{q}_k(2s-t) + (n-1)\,\widehat{p}_k + \Big(\big(r - \widehat{q}_k(2s-t) + (n-1)\,\widehat{p}_k\big)^2 + 4\,r\,(1-2\widehat{p}_k)\,(n-1)\,\widehat{p}_k \Big)^{1/2}}{2\,(n-1)\,\widehat{p}_k} \tag{2.13}$$

where $\widehat{q}_k = 1 - \widehat{p}_k$, $r = \sum_{i=2}^{n} x_i\, x_{i-1}$, $s = \sum_{i=1}^{n} x_i$ and $t = x_1 + x_n$, which is asymptotically equivalent to the maximum likelihood estimator. Again by Theorem 1 in Klotz ([11]), we have that $\widetilde{\lambda}_k$ is consistent, more precisely, $\sqrt{n}\,(\lambda_k - \widetilde{\lambda}_k) \xrightarrow{D} N\big(0, \lambda(1-\lambda)/p\big)$. See the very close estimates obtained for λ_k in Table 2. Results of estimation are summarized in Table 3.

Table 3: True values of a_k $(= c^k)$ and estimates obtained from (2.5), considering $n = 5000$ realizations of process *pARMAX* in (2.1), with marginal Pareto(1), for cases $c = 0.3, 0.5, 0.7, 0.9$; estimates $\widehat{c}$ were obtained by taking $\widehat{a_k}^{1/k}$; IC(λ),IC($\widehat{\lambda}$) and IC($\widetilde{\lambda}$) are 95% confidence intervals obtained, respectively, with true σ^2 and estimated σ^2 using $\widehat{\lambda}$ given in (2.12) and $\widetilde{\lambda}$ given in (2.13); non filled cells mean that a $\widehat{p}_k$ less than 0.5 was obtained.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
a_k	0.3	0.09	0.027	0.0081	0.00243
IC(λ)	(0.2778, 0.3222)	(0.0203, 0.1598)	(−0.1840,0.2380)	(−0.6627, 0.6789)	(−2.1955, 2.2004)
$\widehat{a_k}$	0.295616	0.093871	—	—	—
IC($\widehat{\lambda}$)	(0.2734,0.3178)	(0.0262,0.1615)	—	—	—
IC($\widetilde{\lambda}$)	(0.2734, 0.3179)	(0.0262, 0.1616)	—	—	—
$\widehat{c}$	0.295616	0.306384	—	—	—
a_k	0.5	0.25	0.125	0.0625	0.03125
IC(λ)	(0.4810, 0.5190)	(0.2088, 0.2912)	(0.0526,0.1974)	(−0.0672, 0.1922)	(−0.2100, 0.2725)
$\widehat{a_k}$	0.500694	0.258363	0.137758	0.062246	0.081734
IC($\widehat{\lambda}$)	(0.4814, 0.5200)	(0.2184, 0.2983)	(0.0701,0.2054)	(−0.0678, 0.1923)	(−0.0244, 0.1880)
IC($\widetilde{\lambda}$)	(0.4814, 0.5200)	(0.2184, 0.2983)	(0.0701,0.2054)	(−0.0678, 0.1923)	(−0.0245, 0.1879)
$\widehat{c}$	0.500694	0.508294	0.516463	0.499491	0.606011
a_k	0.7	0.49	0.343	0.2401	0.16807
IC(λ)	(0.6838, 0.7162)	(0.4563, 0.5237)	(0.2947,0.3913)	(0.1760, 0.3042)	(0.0843, 0.2519)
$\widehat{a_k}$	0.682445	0.469803	0.334072	0.222017	0.149248
IC($\widehat{\lambda}$)	(0.6661, 0.6989)	(0.4355, 0.5041)	(0.2850,0.3831)	(0.1543, 0.2897)	(0.0589, 0.2396)
IC($\widetilde{\lambda}$)	(0.6660, 0.6990)	(0.4355, 0.5042)	(0.2850,0.3832)	(0.1543, 0.2897)	(0.0589, 0.2396)
$\widehat{c}$	0.682445	0.685422	0.693873	0.686430	0.683568
a_k	0.9	0.81	0.729	0.6561	0.59049
IC(λ)	(0.8896, 0.9104)	(0.7862, 0.8338)	(0.6937,0.7643)	(0.6106, 0.7016)	(0.5357, 0.6452)
$\widehat{a_k}$	0.896950	0.803367	0.721969	0.653779	0.583686
IC($\widehat{\lambda}$)	(0.8862, 0.9077)	(0.7790, 0.8277)	(0.6862,0.7577)	(0.6080, 0.6995)	(0.5286, 0.6388)
IC($\widetilde{\lambda}$)	(0.8862, 0.9077)	(0.7790, 0.8277)	(0.6863,0.7577)	(0.6081, 0.6995)	(0.5286, 0.6388)
$\widehat{c}$	0.896950	0.896307	0.897097	0.899203	0.897916

ACKNOWLEDGMENTS

This work has been supported by the grant number FCT/SFRH/BD/38867/2007 from FCT. I also acknowledge the valuable suggestions from the referee.

REFERENCES

[1] Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions: with formulas, graphs and mathematical tables*, New York.

[2] Alpuim, M.T. (1989). An extremal markovian sequence, *J. Appl. Probab.*, **26**, 219–232.

[3] Ancona-Navarrete, M.A. and Tawn, J.A. (2000). A comparison of methods for estimating the extremal index, *Extremes*, **3**, 5–38.

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

[5] Bortot, P. and Tawn, J. (1998). Models for the extremes of Markov Chains, *Biometrika*, **85**, 851–867.

[6] Canto e Castro, L. (1992). *Sobre a Teoria Assintótica de Extremos*, PhD Thesis, FCUL.

[7] Davis, R. and Resnick, S. (1989). Basic properties and prediction of max-ARMA processes, *Adv. Appl. Probab.*, **21**, 781–803.

[8] Ferreira, M. and Canto e Castro, L. (2008). Tail and dependence behavior of levels that persist for a fixed period of time, *Extremes*, **11**, 113–133.

[9] Ferreira, M. and Canto e Castro, L. (2010a). Asymptotic and pre-asymptotic tail behavior of a power max-autoregressive model, *ProbStat Forum*, **3**, 91–107.

[10] Ferreira, M. and Canto e Castro, L. (2010b). Modeling rare events through a pRARMAX process, *J. Statist. Plann. Inference*, **140**(11), 3552–3566.

[11] Klotz, J. (1973). Statistical inference in Bernoulli trials with dependence, *Ann. Statist.*, **1**, 373–379.

[12] Lebedev, A.V. (2008). Statistical analysis of first-order MARMA Processes, *Math. Notes*, **83**, 506–511.

[13] Ledford, A. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values, *Biometrika*, **83**, 169–187.

[14] Ledford, A. and Tawn, J.A. (1997). Modelling Dependence within joint tail regions, *J. Royal Statist. Soc. Ser. B*, **59**, 475–499.

[15] Sibuya, M. (1960). Bivariate extreme statistics, *Ann. Inst. Statist. Math.*, **11**, 195–210.

[16] Smith, R.L.; Tawn, J.A. and Coles, S.G. (1997). Markov chain models for threshold exceedances, *Biometrika*, **84**, 249–68.

REVSTAT – Statistical Journal

Background

Statistical Institute of 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.

— Three volumes are scheduled for publication, one in April, one in June and the other in November.

— On average, four 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 / indexed in

REVSTAT is expected to be abstracted / indexed at least in *Current Index to Statistics*, *Statistical Theory and Method Abstracts* and *Zentralblatt für Mathematik*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: 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.

Submitted 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 not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, 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), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's 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: liliana.martins@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.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística, I.P.
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright