

DIFFERENCE EQUATIONS FOR THE HIGHER-ORDER MOMENTS AND CUMULANTS OF THE INAR(1) MODEL

BY MARIA EDUARDA DA SILVA AND VERA LÚCIA OLIVEIRA

Faculdade de Ciências Universidade do Porto

First Version received March 2000

Abstract. Recently, as a result of the growing interest in modelling stationary processes with discrete marginal distributions, several models for integer value time series have been proposed in the literature. One of these models is the INteger-AutoRegressive (INAR) model. Here we consider the higher-order moments and cumulants of the INAR(1) process and show that they satisfy a set of Yule–Walker type difference equations. We also obtain the spectral and bispectral density functions, thus characterizing the INAR(1) process in the frequency domain. We use a frequency domain approach, namely the Whittle criterion, to estimate the parameters of the model. The estimation theory and associated asymptotic theory of this estimation method are illustrated numerically.

Keywords. INAR models; moments; cumulants; Yule–Walker equations; estimation; frequency domain; Whittle’s criterion.

1. INTRODUCTION

Recently, there has been a growing interest in modelling discrete time stationary processes with discrete marginal distributions. The usual linear models for time series, the well-known ARMA models, are suitable for modelling stationary dependent sequences under the assumption of Gaussianity. However, this assumption is inappropriate for modelling counting processes. One of the first models for integer-valued time series proposed in the literature were the discrete autoregressive-moving average (DARMA) models (Jacobs and Lewis, 1978a, 1978b, 1983). These models are obtained by a probabilistic mixture of a sequence of i.i.d. (independent identically distributed) discrete random variables. The Jacobs–Lewis approach can be viewed as analogous to the Box–Jenkins approach in which the linear combination of the continuous-valued case is replaced with a probabilistic mixture. As a result, a realization of the process will generally, as the authors remarked, contain many runs of a single value. Later, McKenzie (1985, 1986, 1988) introduced a different class of models for integer-valued processes, with specific marginal distributions (geometric and negative binomial) as analogues of continuous variate models for gamma and negative exponential marginal distributions. These models are a particular case of the general integer-valued autoregressive, (INAR) models proposed and studied by Al-Osh and Alzaid (1987). The main motivation for these models is the need to model series

with correlated observations expressed as counts. Examples of these series are the number of patients in a hospital at a specific time, the number of persons in a queue waiting for a service at a certain moment, daily counts of epileptic seizure in one patient (Franke and Seligmann, 1993).

Our concern here are the higher-order moments and cumulants of the INAR(1) process. Al-Osh and Alzaid (1987) show that the autocorrelation function of the INAR(1) process is the same as that of an AR(1) process. We show that the higher-order moments and cumulants of the INAR(1) process satisfy a set of Yule–Walker type difference equations. We also obtain the spectral and bispectral density functions, thus characterizing the INAR(1) process in the frequency domain. The difference equations for the higher-order moments and cumulants may be used to estimate the parameters of the model. This will be considered in the near future. Here we use a frequency domain approach, namely the Whittle criterion, to estimate the parameters of the model, then compare the estimates thus obtained with the conditional maximum likelihood estimates, proposed by Al-Osh and Alzaid (1987). We find that although their sampling properties are similar, the Whittle estimates are computationally easier to implement and more efficient.

2. INTEGER AUTOREGRESSIONS OF ORDER 1

We begin by introducing the *thinning* operation necessary to the definition of the INAR models.

2.1. The thinning operation

Let X be a non negative integer-valued random variable. Then for any $\alpha \in [0, 1]$ and according to Steutel and Van Harn (1979) the operator ‘*’ is defined by

$$\alpha * X = \sum_{i=1}^X Y_i \quad (1)$$

where $\{Y_i\}_i$ is a sequence of i.i.d. random variables, independent of X such that

$$P(Y_i = 1) = 1 - P(Y_i = 0) = \alpha$$

The sequence Y_1, Y_2, \dots is called the counting series of $\alpha * X$ and we remark that, given X , $\alpha * X$ has a binomial distribution with parameters (X, α) . The properties of the thinning operation are stated in Lemma 1.

LEMMA 1. *Thinning Operation Properties*

- (i) $0 * X = 0$
- (ii) $1 * X = X$

- (iii) $\alpha * (\beta * X) \stackrel{d}{=} (\alpha\beta) * X$ where $\stackrel{d}{=}$ stands for equal in distribution.
- (iv) $E(\alpha * X) = \alpha E(X)$
- (v) $E(\alpha * X)^2 = \alpha^2 E(X^2) + \alpha(1 - \alpha)E(X)$
- (vi) $E(\alpha * X)^3 = \alpha^3 E(X^3) + 3\alpha^2(1 - \alpha)E(X^2) + \alpha(1 - \alpha)(1 - 2\alpha)E(X)$
- (vii) $E(X(\alpha * Y)) = \alpha E(XY)$
- (viii) $E(X(\alpha * Y)^2) = \alpha^2 E(XY^2) + \alpha(1 - \alpha)E(XY)$
- (ix) If X and Y are independent, $E[(\alpha * X)(\beta * Y)] = \alpha\beta E(X)E(Y)$
- (x) $E((\alpha * X)(\beta * Y)) = \alpha\beta E(XY)$ if the counting series of $\alpha * X$ and $\beta * Y$ are independent, and independent of X and Y .
- (xi) $E((\alpha * X)^2(\beta * Y)) = \alpha^2\beta E(X^2Y) + \alpha(1 - \alpha)\beta E(XY)$ if the counting series of $\alpha * X$ and $\beta * Y$ are independent, and independent of X and Y .
- (xii) $E(XY(\beta * Z)) = \beta E(XYZ)$
- (xiii) $E(X(\beta * Y)(\gamma * Z)) = \beta\gamma E(XYZ)$ if the counting series of $\beta * Y$ and $\gamma * Z$ are independent, and independent of X , Y and Z .
- (xiv) $E((\alpha * X)(\beta * Y)(\gamma * Z)) = \alpha\beta\gamma E(XYZ)$ if the counting series of $\alpha * X$, $\beta * Y$ and $\gamma * Z$ are independent, and independent of X , Y and Z .

PROOF. Properties (i), (ii) and (iii) follow from the definition; properties (iv), (v), (ix) and (x) are due to Al-Osh and Alzaid (1987), Du and Li (1991) and Franke and Subba Rao (1995); properties (vi), (vii), (viii), (xi), (xii) follow from the definition of $*$ and the conditional expectation.

2.2. The INAR(1) process

The INAR(1) process has been introduced by Al-Osh and Alzaid (1987) as a tool for modelling and generating sequences of dependent counting processes. A discrete time stochastic process, $\{X_t\}$, is said to be an INAR(1) process if it satisfies the equation

$$X_t = \alpha * X_{t-1} + e_t \tag{2}$$

where $\alpha \in [0, 1]$ and $\{e_t\}$ is a sequence of i.i.d. non-negative integer-valued random variables, independent of X_t , with mean μ_e and finite variance σ_e^2 . The process $\{X_t\}$ satisfying (2) is second-order stationary if $0 \leq \alpha < 1$ (Al-Osh and Alzaid, 1987, Du and Li, 1991). It follows that $\{X_t\}$ is a Markov process. We can view a realization of X_t as having two components: the survivors of elements of X_{t-1} , each with probability of survival α , and the elements which entered the system in the interval $[t - 1, t]$ (the innovation term e_t). The INAR(1) process as defined by (2) can be viewed as a Galton–Watson branching process with immigration. The offspring in this process has a Bernoulli distribution with probability generating function, $f(s)$, given by

$$f(s) = 1 - \alpha + \alpha s \quad \text{for } |s| \leq 1$$

The probability generating function of the random variable X_t , $\phi_{X_t}(s)$, is given by Alzaid and Al-Osh (1988)

$$\phi_{X_t}(s) = \phi_{X_0}(1 - \alpha^t + \alpha^t s) \prod_{k=0}^{t-1} \phi_e(1 - \alpha^k - \alpha^k s) \quad |s| \leq 1$$

where $\phi_e(s)$ is the probability generating function of e_1 . Since $E(e_1) < \infty$, the limit $\lim_{n \rightarrow \infty} \phi_{X_t}(s)$ exists and is the probability generating function of some random variable X , for which

$$\phi_X(s) = \phi_e(s) \phi_X(1 - \alpha + \alpha s) \quad (3)$$

(Alzaid and Al-Osh, 1988). Equation (3) is related to the definition of a self-decomposable distribution on the set of non-negative integers as introduced by Steutel and Van Harn (1979). In fact, assuming that the INAR(1) process is stationary, the probability generating function of X_t satisfies (3). Thus, one can choose any member of the class of discrete self-decomposable distributions as the marginal distribution of a stationary INAR(1) model.

Al-Osh and Alzaid (1987) provide the following representation for the marginal distribution of the INAR(1) model expressed in terms of the innovation sequence $\{e_t\}$

$$X_t \stackrel{d}{=} \sum_{j=0}^{\infty} \alpha^j * e_{t-j} \quad (4)$$

Poisson INAR(1)

We may mention here that the class of discrete self-decomposable distributions is a wide subclass of discrete infinitely divisible distributions and that it includes many of the well-known distributions such as the Poisson and the negative binomial. In particular, it contains the class of discrete stable distributions. A property of this class is that, apart from the Poisson distribution, none of its members has a finite mean (Van Harn, 1978, Cor. 3.3.15). Also an interesting property of self-decomposable distributions is that they are unimodal (Steutel and Van Harn, 1979). It is, perhaps, these properties that make the Poisson distribution the most commonly used distribution in modelling counting processes.

In the INAR(1) process, the Poisson distribution plays a role similar to that of the Gaussian distribution in the AR(1) process. Specifically, it has the property that if the innovation sequence $\{e_t\}$ and the initial distribution are Poisson, the marginal distribution of $\{X_t\}$ is also Poisson.

In this paper, our main emphasis is the study of the Poisson INAR(1) process. Other possible marginal distributions for the stationary INAR(1) processes are the negative binomial (Mckenzie, 1985; Al-Osh and Alzaid, 1987), which is also a discrete self-decomposable distribution and the geometric distribution (Mckenzie, 1986; Alzaid and Al-Osh, 1988, GINAR(1) processes).

3. MOMENTS AND CUMULANTS OF THE INAR(1)

3.1. Preliminaries

Let $\{X_t\}$ be a stationary time series for which moments up to the order k exist. If the time series $\{X_t\}$ is k th-order stationary then the k th-order joint moment of $X_t, X_{t+s_1}, \dots, X_{t+s_{k-1}}$, is a function of $k - 1$ variables which we denote by

$$\mu(s_1, \dots, s_{k-1}) = E(X_t X_{t+s_1} \dots X_{t+s_{k-1}})$$

and $\mu = E(X_t)$. Also the corresponding moments about the mean will be denoted by $R(s_1, \dots, s_{k-1})$. Under the same conditions, let $\text{cum}_k(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ denote the k th-order joint cumulant which can be written as a function of $k - 1$ variables, too, which we, henceforth, shall denote by $C_k(s_1, \dots, s_{k-1})$. Leonov and Shiryaev (1959) derived relations between joint moments and joint cumulants of a stationary time series. In this section, we give, without proof, explicit relations between moments and cumulants that will be needed later for the derivation of the difference equations. It is easily shown that

$$\text{cum}(X_t) = E(X_t) = \mu \tag{5}$$

$$C_2(s) = \mu(s) - \mu^2 = R(s) \tag{6}$$

$$C_3(s_1, s_2) = R(s_1, s_2) \tag{7}$$

From (6) and (7), we see that the second- and third-order cumulants are the same as the second- and third-order moments about the mean. If the time series $\{X_t\}$ is a real-valued stationary time series, the third-order cumulants satisfy the following set of symmetry relations over and above the well-known conditions $C_2(s) = C_2(-s)$. For $s \geq 0$ and $\tau \geq 0$ we have,

$$\begin{aligned} C_3(s, s + \tau) &= C_3(s + \tau, s) = C_3(-s, \tau) = C_3(\tau, -s) \\ &= C_3(-\tau, -\tau - s) = C_3(-s - \tau, -\tau) \end{aligned} \tag{8}$$

In view of the above symmetry relations, $C_3(s_1, s_2)$ is completely determined over the entire plane by its values for $s_1 \geq 0$ and $0 \leq s_2 \leq s_1$.

3.2. Second- and third-order moments of INAR(1)

In this section, we begin by obtaining the mean, second- and third-order moments about the origin for the Poisson INAR(1) process. The mean, covariance, autocovariance and autocorrelation function of an INAR(1) process have been obtained by Al-Osh and Alzaid (1987). Thus, some of the results presented in this section are a particular case of the above reference.

THEOREM 1. (FIRST-, SECOND- AND THIRD-ORDER MOMENTS). *Let $\{X_t\}$ be a stationary process satisfying (2) where $\{e_t\}$ are i.i.d. random variables, independent of $\{X_t\}$, with Poisson distribution with parameter λ . Then*

$$E(X_t) = \mu = \frac{\lambda}{1 - \alpha} \quad (9)$$

$$\mu(0) = \frac{\lambda(\lambda + 1 - \alpha)}{(1 - \alpha)^2} \quad (10)$$

$$\mu(s) = \alpha\mu(s - 1) + \lambda\mu \quad s > 0 \quad (11)$$

$$\mu(0, 0) = \frac{\lambda^3 + \lambda(1 - \alpha)(3\lambda + 1 - \alpha)}{(1 - \alpha)^3} \quad (12)$$

$$\mu(0, s) = \alpha\mu(0, s - 1) + \lambda\mu(0) \quad s > 0 \quad (13)$$

$$\mu(s, s) = \alpha^2\mu(s - 1, s - 1) + [\alpha(1 - \alpha) + 2\alpha\lambda]\mu(s - 1) + \lambda(\lambda + 1)\mu \quad s > 0 \quad (14)$$

$$\mu(s, \tau) = \alpha\mu(s, \tau - 1) + \lambda\mu(s) \quad s > 0, \tau > s \quad (15)$$

PROOF. Property (9) is due to Al-Osh and Alzaid (1987); (10) and (12) follow from the fact that $\{X_t\}$ has a Poisson distribution with parameter $\lambda/(1 - \alpha)$. Now, (11) and (13)–(15) follow from the definitions and the properties of the thinning operation. Here we exemplify the calculations for (14) and omit the other ones.

$$\begin{aligned} \mu(s, s) &= E(X_t X_{t+s} X_{t+s}) \\ &= \alpha^2 E(X_t X_{t+s-1} X_{t+s-1}) + \alpha(1 - \alpha) E(X_t X_{t+s-1}) \\ &\quad + 2\alpha E(X_t X_{t+s-1}) E(e_{t+s}) + E(X_t) E(e_{t+s} e_{t+s}) \\ &= \alpha^2 \mu(s - 1, s - 1) + [\alpha(1 - \alpha) + 2\alpha\lambda] \mu(s - 1) + \lambda(\lambda + 1) \mu \quad s > 0 \end{aligned}$$

3.3. Cumulants

Expressions for the cumulants of the Poisson INAR(1) process can be obtained using either a direct method of evaluation or the method of Leonov and Shiryaev (1959). However, since we are computing relations for the cumulants up to the third order and in view of (6) and (7), we adopt the direct method of evaluation.

THEOREM 2 (FIRST-, SECOND- AND THIRD-ORDER CUMULANTS). *Let $\{X_t\}$ be a stationary process satisfying (2) where $\{e_t\}$ are i.i.d. random variables, independent of $\{X_t\}$, with Poisson distribution with parameter λ . Then*

$$\text{cum}(X_t) = \mu \quad (16)$$

$$C_2(0) = \frac{\lambda}{1 - \alpha} \quad (17)$$

$$C_2(s) = \alpha C_2(0) \quad s > 0 \quad (18)$$

$$C_3(0, 0) = \frac{\lambda}{1 - \alpha} \tag{19}$$

$$C_3(0, s) = \alpha C_3(0, s - 1) \quad s > 0 \tag{20}$$

$$C_3(s, s) = \alpha^2 C_3(s - 1, s - 1) + \alpha(1 - \alpha)C_2(s - 1) \quad s > 0 \tag{21}$$

$$C_3(s, \tau) = \alpha C_3(s, \tau - 1) \quad s > 0, \tau > s \tag{22}$$

PROOF. The equations (16), (17) and (19) follow from the fact that the Poisson INAR(1) process has a Poisson marginal distribution. Equations (18) and (20)–(22) follow from the relations between the moments and the cumulants. Here, we exemplify for (22).

$$\begin{aligned} C_3(s, \tau) &= E[(X_t - \mu)(X_{t+s} - \mu)(X_{t+\tau} - \mu)] \\ &= \mu(s, \tau) - \mu\mu(s) - \mu\mu(\tau) - \mu E(X_{t+s}X_{t+\tau}) + 2\mu^3 \\ &= \alpha[\mu(s, \tau - 1) - \mu\mu(s) - \mu\mu(\tau - 1) - \mu E(X_{t+s}X_{t+\tau-1}) + 2\mu^3] \\ &\quad - (1 - \alpha)\mu\mu(s) + 2(1 - \alpha)\mu^3 + \lambda\mu(s) - 2\lambda\mu^2 \\ &= \alpha C_3(s, \tau - 1) \quad s > 0, \tau > s \end{aligned}$$

From (17) and (18), we obtain

$$\rho(k) = \alpha^{|k|} \quad k \in Z \tag{23}$$

$$\phi_{11} = \alpha \quad \phi_{kk} = 0 \quad k \geq 2 \tag{24}$$

where ϕ_{kk} is the partial autocorrelation of function at lag k .

The above results show that the second-order properties of the INAR(1) process are similar to those of the standard AR(1) process. These results regarding the second-order properties of the INAR(1) process may be easily extended to INAR(p) processes. Note that the partial autocorrelations are then useful to identify the order of the process. Moreover, the above equations show that the third-order cumulants of the INAR(1) process satisfy a set Yule–Walker type difference equations. We note that the third-order structure of the INAR(1) is similar to that of well-known nonlinear processes, such as the BL(1,0,1,1) processes. These equations may be used to estimate the parameters of the model by the method of moments.

4. THE SPECTRAL AND BISPECTRAL DENSITY FUNCTIONS

Let $f(\omega)$ denote the spectral density function of a stationary process $\{X_t\}$, defined as the Fourier transform of the autocovariance function of the process,

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R(k)e^{-i\omega k} \quad -\pi \leq \omega \leq \pi$$

Also, let $f(\omega_1, \omega_2)$ denote the bispectral density function of a stationary process $\{X_t\}$, defined as the Fourier transform of the third-order cumulants of the process,

$$f(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_3(j, k) e^{-i\omega_1 j - i\omega_2 k} \quad -\pi \leq \omega_1, \omega_2 \leq \pi$$

Then the spectral and bispectral density functions of a Poisson INAR(1) process are given in Theorem 3.

THEOREM 3 SPECTRUM AND BISPECTRUM OF POISSON INAR(1). *Let $\{X_t\}$ be a stationary process, satisfying (2) where $\{e_t\}$ are i.i.d. random variables, independent of $\{X_t\}$ with Poisson distribution with parameter λ . Then*

$$f_X(\omega) = \frac{\lambda(\alpha + 1)}{2\pi(1 - 2\alpha \cos \omega + \alpha^2)} \quad -\pi \leq \omega \leq \pi \tag{25}$$

$$\begin{aligned} f_X(\omega_1, \omega_2) = & \frac{1}{(2\pi)^2} \{C_3 + (g(-\omega_1) + g(-\omega_2) + g(\omega_1 + \omega_2))C_3 \\ & + (h(\omega_1) + h(\omega_2) + h(-\omega_1 - \omega_2))(C_3 - C_2) \\ & + (g(\omega_1) + g(\omega_2) + g(-\omega_1 - \omega_2))C_2 \\ & + g(-\omega_1)[(h(\omega_2) + h(-\omega_1 - \omega_2))(C_3 - C_2) \\ & + (g(\omega_2) + g(-\omega_1 - \omega_2))C_2] \\ & + g(-\omega_2)[(h(\omega_1) + h(-\omega_1 - \omega_2))(C_3 - C_2) \\ & + (g(\omega_1) + g(-\omega_1 - \omega_2))C_2] \\ & + g(\omega_1 + \omega_2)[(h(\omega_1) + h(\omega_2))(C_3 - C_2) \\ & + (g(\omega_1) + g(\omega_2))C_2]\} \tag{26} \end{aligned}$$

where

$$\begin{aligned} C_2 &= C_2(0) \\ C_3 &= C_3(0, 0) \\ g(\omega_j) &= \frac{\alpha z_j}{1 - \alpha z_j} \end{aligned}$$

and

$$h(\omega_j) = \frac{\alpha^2 z_j}{1 - \alpha^2 z_j}$$

with $z_j = e^{i\omega_j}, j = 1, 2$.

PROOF. The spectral density function of the INAR(1) is the same, up to the constant, as that of an AR(1) process, since the autocorrelations are identical. The two spectra differ on the scaling factor which, being the variance of the one-step-ahead prediction error, for the Poisson INAR(1) process is based on the Poisson innovations.

To obtain the expression (26), we use the symmetry relations, given by (8), satisfied by the cumulants. Thus,

$$\begin{aligned}
 f(\omega_1, \omega_2) = & \frac{1}{(2\pi)^2} \left\{ C_3(0, 0) + \sum_{\tau=1}^{\infty} C_3(0, \tau) [e^{-i\tau\omega_1} + e^{-i\tau\omega_2} + e^{-i\tau(\omega_1+\omega_2)}] \right. \\
 & + \sum_{\tau=1}^{\infty} C_3(\tau, \tau) [e^{-i\tau(\omega_1+\omega_2)} + e^{i\tau\omega_1} + e^{i\tau\omega_2}] \\
 & + \sum_{\tau=1}^{\infty} \sum_{s=1}^{\infty} C_3(s, s + \tau) \left[e^{-is\omega_1 - i(s+\tau)\omega_2} + e^{-i(s+\tau)\omega_1 - is\omega_2} \right. \\
 & \left. \left. + e^{is\omega_1 - i\tau\omega_2} + e^{-i\tau\omega_1 + is\omega_2} + e^{i(s+\tau)\omega_1 + is\omega_2} + e^{i\tau\omega_1 + i(s+\tau)\omega_2} \right] \right\}
 \end{aligned}$$

Substituting for $C(0, \tau)$, $C(\tau, \tau)$ and $C(s, s + \tau)$ using the expressions (20), (21) and (22) and we may write

$$\begin{aligned}
 f(\omega_1, \omega_2) = & \frac{1}{(2\pi)^2} \left\{ C_3(0, 0) + \sum_{\tau=1}^{\infty} \alpha^\tau C_3(0, 0) [e^{-i\tau\omega_1} + e^{-i\tau\omega_2} + e^{i\tau(\omega_1+\omega_2)}] \right. \\
 & + \sum_{\tau=1}^{\infty} (\alpha^{2\tau} C_3(0, 0) + \alpha^\tau (1 - \alpha)(\alpha^{\tau-1} + \alpha^{\tau-2} + \dots + \alpha + 1) C_2(0)) \\
 & \left[e^{-i\tau(\omega_1+\omega_2)} + e^{i\tau\omega_1} + e^{i\tau\omega_2} \right] \\
 & + \sum_{\tau=1}^{\infty} \sum_{s=1}^{\infty} \alpha^\tau C_3(\tau, \tau) \left[e^{-is\omega_1 - i(s+\tau)\omega_2} + e^{-i(s+\tau)\omega_1 - is\omega_2} \right. \\
 & \left. + e^{is\omega_1 - i\tau\omega_2} + e^{-i\tau\omega_1 + is\omega_2} + e^{i(s+\tau)\omega_1 + is\omega_2} + e^{i\tau\omega_1 + i(s+\tau)\omega_2} \right] \left. \right\} \\
 = & \frac{1}{(2\pi)^2} \left\{ C_3(0, 0) + \sum_{\tau=1}^{\infty} \alpha^\tau C_3(0, 0) [e^{-i\tau\omega_1} + e^{-i\tau\omega_2} + e^{i\tau(\omega_1+\omega_2)}] \right. \\
 & + \sum_{\tau=1}^{\infty} (\alpha^{2\tau} C_3(0, 0) + \alpha^\tau (1 - \alpha^\tau) C_2(0)) \left[e^{-i\tau(\omega_1+\omega_2)} + e^{i\tau\omega_1} + e^{i\tau\omega_2} \right] \\
 & + \sum_{\tau=1}^{\infty} \sum_{s=1}^{\infty} (\alpha^{2s+\tau} C_3(0, 0) + \alpha^{s+\tau} (1 - \alpha^s) C_2(0)) \left[e^{-is\omega_1 - i(s+\tau)\omega_2} \right. \\
 & + e^{-i(s+\tau)\omega_1 - is\omega_2} \\
 & \left. \left. + e^{is\omega_1 - i\tau\omega_2} + e^{-i\tau\omega_1 + is\omega_2} + e^{i(s+\tau)\omega_1 + is\omega_2} + e^{i\tau\omega_1 + i(s+\tau)\omega_2} \right] \right\}
 \end{aligned}$$

Now setting

$$\begin{aligned}
 C_2 &= C_2(0) \\
 C_3 &= C_3(0, 0) \\
 g(\omega_j) &= \frac{\alpha z_j}{1 - \alpha z_j}
 \end{aligned}$$

and

$$h(\omega_j) = \frac{\alpha^2 z_j}{1 - \alpha^2 z_j}$$

with $z_j = e^{i\omega_j}$, $j = 1, 2$, we obtain the desired expression, (26).

Since $\alpha > 0$, most of the power is concentrated at the low frequency end of the spectrum. The bispectral modulus of INAR(1) process has a peak in the low frequencies.

In Figure 1, we illustrate a realization of an INAR(1) process with parameters $\alpha = 0.5$ and $\lambda = 1$, together with the spectral and bispectral density functions of the process.

5. FREQUENCY DOMAIN ESTIMATION OF THE INAR(1) MODEL

Let X_1, \dots, X_n be n observations from the INAR(1) model given by (2), denote the parameters of the model by θ and their true values by $\theta_0 = (\alpha, \lambda)$. We now consider the problem of estimating the parameters of the model. There are two approaches for the estimation of the parameters of finite-parameter time series models:

- (i) the frequency domain approach, and
- (ii) the time domain approach.

The Whittle criterion comes under the former approach. This approach has been originally proposed by Whittle (1953) for Gaussian processes, further investigated by several authors (Walker, 1964; Hannan, 1973; Rice, 1979; Dzhaparidze and Yaglom, 1983) and since used in many situations (Fox and Taquq, 1986; Sesay and Subba Rao, 1992; Subba Rao and Chandler, 1996). The motivation is that the spectral density function of a model may be easy to obtain whereas an exact likelihood is not. In the following section, we review the Whittle criterion.

5.1. Whittle criterion

The idea of representing the likelihood of a stochastic process via its spectral properties rather than directly was first suggested by Whittle (1953) for Gaussian models. If the time series is Gaussian, the maximization of the logarithm of the likelihood function of the sample is asymptotically equivalent to minimization of Whittle criterion given by

$$\hat{L}_N(\theta) = \frac{N}{4\pi} \int_{-\pi}^{\pi} \left\{ \log f(\omega, \theta) + \frac{I_N(\omega)}{f(\omega, \theta)} \right\} d\omega \quad (27)$$

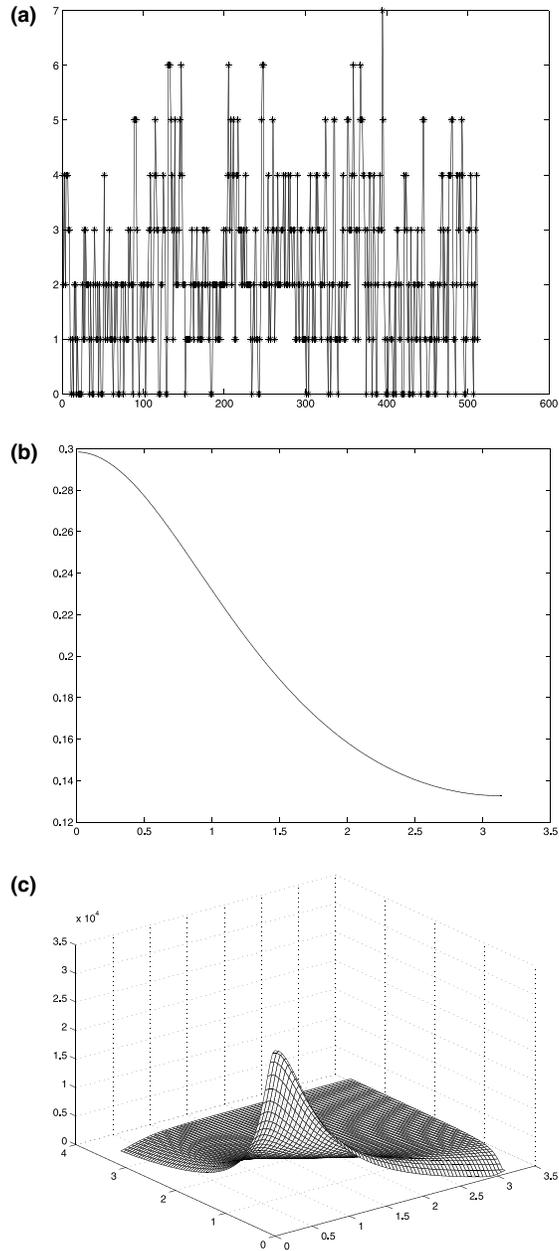


FIGURE 1. A realization (a), spectral (b) and bispectral (c) density functions of the INAR(1) process with parameters $\alpha = 0.5, \lambda = 1$

where $f(\omega, \theta)$ denotes the non-normalized spectral density function of the process and $I_N(\omega)$ is given by

$$I_N(\omega) = (2\pi N)^{-1} \left| \sum_{t=1}^N X_t e^{i\omega t} \right|^2$$

Let $\hat{\theta}$ represent the estimator obtained by minimizing (27), the so-called Whittle estimator. Under the Gaussianity assumption, Whittle (1954) and Walker (1964) have shown that the estimator obtained by minimizing (27) is weakly consistent and has an asymptotic normal distribution. If the series is not Gaussian, Chandler (1995) has shown that the use of Whittle's criterion is justified because this is, in fact, the criterion one would arrive at if one considers approximate likelihood functions of collections of sample Fourier coefficients. In this case, the Whittle criterion has a quasi-likelihood interpretation. Results for the consistency and asymptotic normality of the Whittle estimators when the series is not Gaussian, were provided by Rice (1979). However, the asymptotic variance of $(\hat{\theta} - \theta_0)$ depends on the fourth-order spectra of the process, the trispectrum. Thus the distributional results of the estimators are not very useful in practice, since it is very difficult to obtain expressions for the trispectrum.

5.2. Numerical illustrations and discussion

In practice, the minimization of Whittle criterion is achieved iteratively, replacing (27) by

$$\hat{I}_N(\theta) = \frac{1}{N} \sum_{j=1}^{[N/2]} \left\{ \log f(\omega_j, \theta) + \frac{I_N(\omega_j)}{f(\omega_j, \theta)} \right\} \quad (28)$$

Here, $f(\omega_j, \theta)$ is the spectral density function at the frequency point $\omega_j = 2\pi j/N$, $I_N(\omega_j)$ is the periodogram ordinate at the same frequency point and \sum is the summation over $j = \pm 1, \dots, \pm[N/2]$, the point $j = 0$ being excluded as a mean correlation for X_t .

The numerical minimization of (28) is achieved using numerical algorithms. Due to the nature of the model parameters, α is a probability, we consider both unconstrained minimization and constrained minimization, imposing that $\alpha < 1$. We use the minimization algorithms implemented in the Matlab functions, *fmins* and *constr*. The function *fmins* used the simplex search algorithm of Nelder and Mead which is an algorithm generally referred to as unconstrained nonlinear minimization. The algorithm used by *constr* finds the constrained minimum of a scalar function of several variables and uses a sequential quadratic programming method. Both functions start from an initial estimate for (α, λ) which is obtained by the method of moments from the Yule-Walker equations.

To study the properties of the estimates, we use Monte Carlo methods. We consider the following set of INAR(1) parameter values: $\alpha = 0.1, 0.5, 0.9$ and

$\lambda = 1, 3$. For each INAR(1) model considered, 200 realizations of $n = 64, 128, 512, 1024$ observations are generated. The average bias, mean square error, sample measures of skewness and sample measures of kurtosis are computed for each of the estimated parameters from the 200 realizations. The sample multivariate measures of skewness and kurtosis (Mardia, 1970) are also calculated to study the joint distributions of the parameter estimates. The numerical results obtained using unconstrained minimization are presented in Tables I and II. The results obtained using constrained minimization are not

TABLE I
 PROPERTIES OF THE WHITTLE'S ESTIMATES AND CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATES (INSIDE BRACKETS) FOR THE INAR(1) MODEL WITH PARAMETERS α AND $\lambda = 1$, OBTAINED FROM 200 REALIZATIONS OF EACH MODEL

		<i>n</i>			
		64	128	512	1024
$\alpha = 0.1$	<i>Bias</i> ($\hat{\alpha}$)	-0.0015 (0.0207)	-0.0005 (-0.0009)	0.0008 (0.0027)	0.0059 (0.0053)
	<i>RMSE</i> ($\hat{\alpha}$)	0.0180 (0.0253)	0.0076 (0.0087)	0.0022 (0.0021)	0.0010 (0.0011)
	<i>Skew</i> ($\hat{\alpha}$)	0.0062 (-0.7842)	-0.0728 (0.2446)	-0.3395 (-0.2441)	0.2073 (0.2215)
	<i>Kurt</i> ($\hat{\alpha}$)	-0.3064 (1.6876)	0.6364 (0.6006)	0.1651 (-0.0766)	-0.2397 (-0.3335)
$\lambda = 1$	<i>Bias</i> ($\hat{\lambda}$)	0.0119 (-0.0235)	-0.0121 (-0.0005)	0.0009 (-0.0014)	-0.0017 (-0.0058)
	<i>RMSE</i> ($\hat{\lambda}$)	0.0641 (0.0478)	0.0264 (0.0177)	0.0070 (0.0041)	0.0047 (0.0024)
	<i>Skew</i> ($\hat{\lambda}$)	0.8870 (0.7794)	0.5149 (0.2581)	0.3171 (0.4385)	0.3438 (0.1663)
	<i>Kurt</i> ($\hat{\lambda}$)	2.2123 (1.3967)	0.4990 (-0.1863)	-0.1622 (1.0231)	-0.0048 (0.1412)
Multivariate					
	Skewness	39.7992 (19.4611)	9.3024 (9.2480)	7.1884 (3.5271)	6.6616 (5.0978)
	Kurtosis	4.1540 (-0.2958)	2.3939 (0.2059)	0.2361 (-0.3277)	-0.7083 (-0.4442)
$\alpha = 0.5$	<i>Bias</i> ($\hat{\alpha}$)	0.0257 (0.0164)	0.0173 (0.0203)	-0.0001 (-0.0013)	0.0029 (0.0040)
	<i>RMSE</i> ($\hat{\alpha}$)	0.0145 (0.0109)	0.0077 (0.0045)	0.0016 (0.0010)	0.0007 (0.0006)
	<i>Skew</i> ($\hat{\alpha}$)	-0.2252 (-0.6079)	0.0529 (-0.2846)	-0.0979 (-0.3107)	-0.2624 (-0.0233)
	<i>Kurt</i> ($\hat{\alpha}$)	-0.0342 (0.2475)	-0.4226 (0.4518)	-0.2122 (0.2741)	-0.3940 (0.1685)
$\lambda = 1$	<i>Bias</i> ($\hat{\lambda}$)	-0.0058 (-0.0252)	0.0049 (-0.0257)	0.0065 (-0.0050)	-0.0001 (-0.0046)
	<i>RMSE</i> ($\hat{\lambda}$)	0.0536 (0.0422)	0.0243 (0.0202)	0.0062 (0.0060)	0.0037 (0.0029)
	<i>Skew</i> ($\hat{\lambda}$)	0.9907 (0.4459)	0.3189 (0.3562)	0.1886 (0.3954)	0.4680 (0.3929)
	<i>Kurt</i> ($\hat{\lambda}$)	2.1519 (-0.1574)	-0.3832 (0.0290)	-0.1935 (0.1900)	1.1608 (0.5536)
Multivariate					
	Skewness	39.9566 (32.4521)	4.9456 (15.4782)	1.8430 (9.6861)	12.3867 (6.9978)
	Kurtosis	3.2768 (1.0975)	-0.1397 (0.8239)	-1.3047 (1.6209)	0.9815 (0.4037)
$\alpha = 0.9$	<i>Bias</i> ($\hat{\alpha}$)	0.0326 (0.0011)	0.0222 (0.0019)	0.0060 (0.0015)	0.0027
	<i>RMSE</i> ($\hat{\alpha}$)	0.0134 (0.0013)	0.0031 (0.0002)	0.0005 (0.0001)	0.0002
	<i>Skew</i> ($\hat{\alpha}$)	-0.0862 (3.2411)	-0.1394 (-0.7555)	-0.3748 (-0.0139)	-0.0161
	<i>Kurt</i> ($\hat{\alpha}$)	-0.0785 (16.2815)	0.6059 (0.8002)	-0.1088 (-0.5248)	-0.6202
$\lambda = 1$	<i>Bias</i> ($\hat{\lambda}$)	-0.0871 (-0.0016)	-0.0778 (-0.0157)	-0.0123 (-0.0137)	-0.0135
	<i>RMSE</i> ($\hat{\lambda}$)	0.1034 (0.0961)	0.0518 (0.0215)	0.0070 (0.0056)	0.0033
	<i>Skew</i> ($\hat{\lambda}$)	1.2503 (-1.9351)	0.7796 (0.5971)	0.5630 (0.3876)	0.6283
	<i>Kurt</i> ($\hat{\lambda}$)	3.6984 (10.0548)	0.9052 (0.2692)	0.1525 (0.1427)	1.3543
Multivariate					
	Skewness	57.3971 (96.2967)	21.4786 (72.9137)	17.3023 (22.9597)	12.3867
	Kurtosis	5.9025 (12.1638)	2.3687 (10.6125)	-0.2832 (10.2782)	0.7947

TABLE II
 PROPERTIES OF THE WHITTLE'S ESTIMATES AND CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATES (INSIDE BRACKETS) FOR THE INAR(1) MODEL WITH PARAMETERS α AND $\lambda = 3$, OBTAINED FROM 200 REALIZATIONS OF EACH MODEL

		<i>n</i>		
		64	128	512
$\alpha = 0.1$	<i>Bias</i> ($\hat{\alpha}$)	-0.0153 (0.0153)	-0.0062 (0.0129)	-0.0022 (0.0035)
	<i>MSE</i> ($\hat{\alpha}$)	0.0165 (0.1348)	0.0084 (0.0089)	0.0017 (0.0016)
	<i>Skew</i> ($\hat{\alpha}$)	-0.1055 (-10.8762)	0.0236 (-0.1397)	-0.1072 (0.2517)
	<i>Kurt</i> ($\hat{\alpha}$)	-0.5671 (138.3768)	0.1789 (-0.6253)	-0.1908 (-0.3134)
$\lambda = 3$	<i>Bias</i> ($\hat{\lambda}$)	0.1846 (0.0256)	-0.0005 (-0.0567)	0.0062 (-0.0163)
	<i>RMSE</i> ($\hat{\lambda}$)	0.4813 (0.4299)	0.2163 (0.1141)	0.0461 (0.0264)
	<i>Skew</i> ($\hat{\lambda}$)	0.6571 (4.3413)	0.3563 (0.1856)	0.7128 (-0.0652)
	<i>Kurt</i> ($\hat{\lambda}$)	0.6302 (40.5262)	0.1973 (-0.2533)	0.7957 (-0.1212)
Multivariate				
	Skewness	17.0489 (5179.8000)	4.5808 (3.0121)	20.9501 (5.5333)
	Kurtosis	-0.4113 (290.5987)	0.3933 (-1.2316)	0.5282 (-0.0083)
$\alpha = 0.5$	<i>Bias</i> ($\hat{\alpha}$)	0.0288 (0.0163)	-0.0033 (0.0041)	0.0070 (0.0036)
	<i>MSE</i> ($\hat{\alpha}$)	0.0156 (0.0098)	0.0067 (0.0035)	0.0014 (0.0009)
	<i>Skew</i> ($\hat{\alpha}$)	-0.1786 (-0.4214)	-0.4677 (-0.2334)	-0.0688 (-0.4301)
	<i>Kurt</i> ($\hat{\alpha}$)	-0.7219 (-0.3822)	0.2348 (0.1647)	-0.1712 (0.7871)
$\lambda = 3$	<i>Bias</i> ($\hat{\lambda}$)	-0.0749 (-0.0972)	-0.0138 (-0.0200)	0.0028 (-0.0236)
	<i>MSE</i> ($\hat{\lambda}$)	0.4191 (0.3390)	0.1791 (0.1435)	0.0523 (0.0431)
	<i>Skew</i> ($\hat{\lambda}$)	0.5765 (0.4144)	0.3860 (0.3666)	0.1144 (0.6457)
	<i>Kurt</i> ($\hat{\lambda}$)	-0.0494 (-0.3658)	0.0243 (0.4831)	-0.1343 (1.3987)
Multivariate				
	Skewness	12.4326 (10.7616)	13.4777 (6.9590)	2.8770 (18.4420)
	Kurtosis	-1.7127 (-0.7739)	0.5338 (0.5315)	-0.5455 (1.8734)
$\alpha = 0.9$	<i>Bias</i> ($\hat{\alpha}$)	0.0358	0.0219	0.0043
	<i>RMSE</i> ($\hat{\alpha}$)	0.0134	0.0037	0.0004
	<i>Skew</i> ($\hat{\alpha}$)	-0.4637	0.0481	-0.6392
	<i>Kurt</i> ($\hat{\alpha}$)	0.8384	0.4915	0.5583
$\lambda = 3$	<i>Bias</i> ($\hat{\lambda}$)	-0.3826	-0.1898	-0.0193
	<i>RMSE</i> ($\hat{\lambda}$)	0.9321	0.3139	0.0490
	<i>Skew</i> ($\hat{\lambda}$)	1.2093	0.7496	0.4014
	<i>Kurt</i> ($\hat{\lambda}$)	2.8291	1.4428	0.6882
Multivariate				
	Skewness	57.1508	18.8074	25.6986
	Kurtosis	7.5367	3.1660	2.0408

presented here for brevity but may be obtained from the authors. The following conclusions can be drawn from the tables.

The estimates for α present small biases which decrease with n . However, the absolute biases are larger for larger values of α and λ . The estimates for the parameter λ present negative bias as n increases, indicating that the parameter seems to be underestimated. Also, the bias of the estimates for λ increase in absolute value as α approaches the nonstationarity region. Small samples and larger values of λ , also lead to an increase in the bias of the estimates for λ . This is

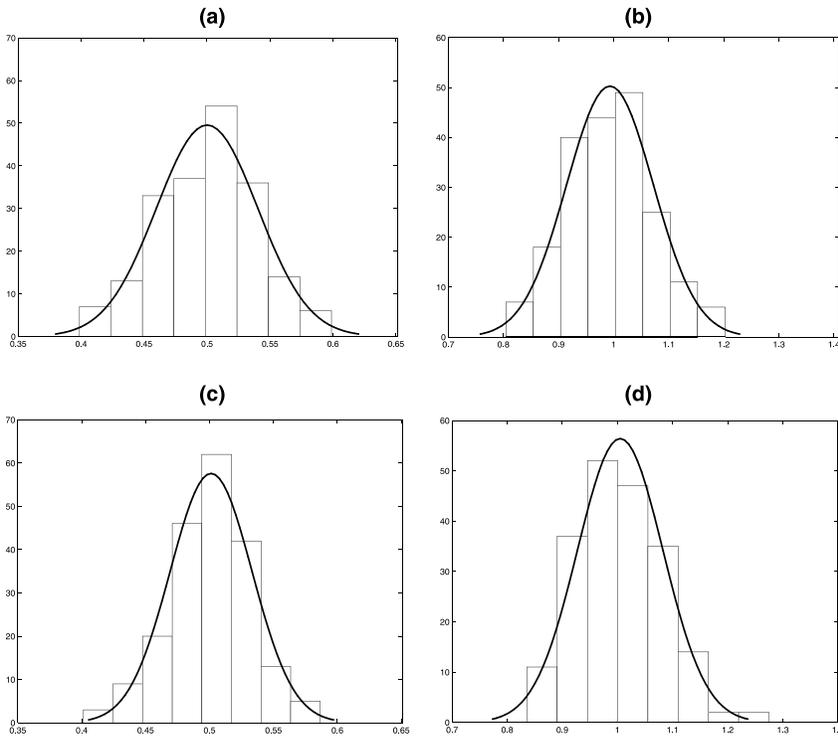


FIGURE 2. Histogram of the estimates, $\hat{\alpha}$ of $\alpha = 0.5$: **(a)** Whittle's criterion and **(c)** conditional maximum likelihood. Histogram of the estimates $\hat{\lambda}$ of $\lambda = 1$: **(b)** Whittle's criterion and **(d)** conditional maximum likelihood

due to the fact that an increase in the parameter λ leads to an increase in the variance of the process. Then, the Whittle criterion becomes a function with many peaks and troughs and numerical problems usually arise.

The root mean square error of the estimates of both α and λ decreases as n increases but the estimates of λ , for larger values of the parameter, present larger root mean square error. However, the root mean square error of the estimates of either parameter does not seem to depend on the value of the other.

The univariate measures of skewness and kurtosis do not indicate departure from Gaussianity in the marginal distributions. The value of the multivariate coefficient of skewness decreases as n increases, indicating asymptotic symmetry of the joint distribution of the estimates. However, for small sample sizes, there is evidence of departure from symmetry in both the marginal and the joint distributions, especially for larger values of the parameters.

We note that the estimates obtained with constrained estimation present the same properties as those obtained with unconstrained estimation. However, the constrained estimation is essential when the parameter α is near 0 or 1. When the parameter α approaches the nonstationarity region, the estimates of λ present

very large biases and mean square errors and both the marginal and joint distributions are not symmetric.

Thus, the numerical results obtained agree with the asymptotic properties of the Whittle estimates (WE) presented in the literature (Rice, 1979). It is, however, important to compare the behaviour of the WEs with that of the conditional maximum likelihood estimated (CMLE). Therefore, we have also computed, for the same set of realizations, the conditional maximum likelihood estimates of the parameters. However, for certain values of the parameters and sample size, it was not possible to obtain the conditional maximum likelihood estimates, namely for $(\alpha = 0.9, \lambda = 1, n = 1024)$, $(\alpha = 0.9, \lambda = 3, n = 64, 512, 1024)$ and $(\alpha = 0.1, 0.5, 0.9, \lambda = 3, n = 1024)$. The CMLEs are given in Tables I and II inside brackets. An analysis of the tables show that the properties of the two types of estimates are similar. This is clearly seen from the histograms of the estimates, $\hat{\alpha}$ and $\hat{\lambda}$, represented in Figure 2 and obtained from 200 realizations of size 512 from the INAR(1) model with parameters $\alpha = 0.5$ and $\lambda = 1$. Moreover, further simulation studies show that for large sample sizes, $n > 500$, the two estimates (WE and CMLE) are equally efficient. We note, however, that, the actual computation of the Whittle estimates is simpler and faster (in cpu time) than the computation of the conditional maximum likelihood estimates.

ACKNOWLEDGEMENTS

The authors thank the anonymous referee for the comments that helped to improve the paper. This work has been partially funded by Centro de Matemática Aplicada do Porto, Praxis PCEX/Mat/47/96 and PRODEP - Medida 5.

REFERENCES

- AL-OSH, M. A. and ALZAID, A. A. (1987) First order integer-valued autoregressive (INAR(1)) processes. *J. Time Ser. Anal.* 8, 261–75.
- ALZAID, A. A. and AL-OSH, M. A. (1988) First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties. *Statistica Neerlandica* 42, 53–61.
- CHANDLER, R. (1995) A spectral method for estimating parameters in rainfall models. Research Report 142, Department of Statistical Science, UCL.
- DU, J.-Guan and LI, Y. (1991) The integer-valued autoregressive (INAR(p)) model. *J. Time Series Analysis* 12, 129–42.
- DZHAPARIDZE, K. O. and YAGLOM, A. M. (1983) Spectrum parameter estimation in time series analysis. In *Developments in Statistics* vol. 4 (ed. P. R. Krishnaiah). London: Academic Press, pp. 1–98.
- FOX, R. and TAQQU, M. S. (1986) Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals Statist.* 14, 517–32.
- FRANKE, J. and SELIGMANN, T. (1993) Conditional maximum-likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In *Developments in Time Series* (ed. T. Subba Rao). London: Chapman & Hall, pp. 310–30.

- FRANKE, J. and SUBBA RAO, T. (1995) Multivariate first order integer valued autoregressions. Technical Report, Maths Dep., UMIST.
- HANNAN, E. J. (1973) The asymptotic theory of linear time series models. *J. Appl. Probab.* 10, 130–45.
- JACOBS, P. A. and LEWIS, P. A. W. (1978a) Discrete time series generated by mixture I: Correlational and runs properties. *J. Royal Statist. Soc. B* 40, 94–105.
- JACOBS, P. A. and LEWIS, P. A. W. (1978b) Discrete time series generated by mixture II: Asymptotic properties. *J. Royal Statist. Soc. B* 40, 222–8.
- JACOBS, P. A. and LEWIS, P. A. W. (1983) Stationary discrete autoregressive moving average time series generate by mixtures. *J. Time Ser. Anal.* 4, 19–36.
- LEONOV, V. P. and SHIRYAEV, A. N. (1959) On a method of calculation of semi-invariants. *Theory Probab. Its Appl.* (URSS) 4, 319–29.
- MARDIA, K. V. (1970) Measures of multivariate skewness and kurtosis with applications *Biometrika*. 57, 519–29.
- MCKENZIE, E. (1985) Some simple models for discrete variate series. *Water Resources Bulletin* 21, 645–50.
- MCKENZIE, E. (1986) Autoregressive moving-average processes with negative binomial and geometric marginal distributions. *Adv. Appl. Prob.* 18, 679–705.
- MCKENZIE, E. (1988) Some ARMA models for dependent sequence of Poisson counts. *Adv. Appl. Prob.* 20, 822–35.
- RICE, J. (1979) On the estimation of the parameters of a power spectrum. *J. Mult. Analysis* 9, 378–92.
- SESAY, S. A. O. and SUBBA RAO, T. (1992) Frequency domain estimation of bilinear time series models. *J. Time Series Anal.* 13, 521–45.
- STEUTEL, F. W. and VAN HARN, K. (1979) Discrete analogues of self-decomposability and stability. *Ann. Probab.* 7, 893–99.
- SUBBA RAO, T. and CHANDLER, R. E. (1996) A frequency domain approach for estimating parameters in point process models. In *Athens Conference on Applied Probability and Time Series* Vol. II, Lecture Notes in Statistics, 115 (eds P. Robinson and M. Rosenblatt). New York: Springer-Verlag, pp. 392–405.
- VAN HARN, K. (1978) *Classifying Infinitely Divisible Distributions by Functional Equations*. Amsterdam: Mathematisch Centrum.
- WALKER, A. M. (1964) Asymptotic properties of least squares estimates of parameters of the spectrum of a stationary non-deterministic series. *J. Austl. Math. Soc.* 4, 107–28.
- WHITTLE, P. (1953) Estimation and information in stationary time series. *Ark. Mat.* 2, 423–34.
- WHITTLE, P. (1954) Some recent contributions to the theory of stationary processes. In *A Study in the Analysis of Stationary Time Series*, 2nd edn (ed. H. Wold). Stockholm: Almqvist and Wiksell. Appendix 2, pp. 196–228.