Non-Gaussian conditional linear AR(1) models

Gary K. Grunwald[1], Rob J. Hyndman[2], Leanna Tedesco[3] & Richard L. Tweedie[4]

30 November 1999

Abstract: We give a general formulation of a non-Gaussian conditional linear AR(1) model subsuming most of the non-Gaussian AR(1) models that have appeared in the literature. We derive some general results giving properties for the stationary process mean, variance and correlation structure, and conditions for stationarity. These results highlight similarities and differences with the Gaussian AR(1) model, and unify many separate results appearing in the literature. Examples illustrate the wide range of properties that can appear under the conditional linear autoregressive assumption. These results are used in analysing three real data sets, illustrating general methods of estimation, model diagnostics and model selection. In particular, we show that the theoretical results can be used to develop diagnostics for deciding if a time series can be modelled by some linear autoregressive model, and for selecting among several candidate models.

Keywords: autoregression, data analysis, Exponential time series, Gamma time series, non-Gaussian time series, Poisson time series.

[1]Department of Preventive Medicine and Biometrics, Box B-119, University of Colorado Health Sciences Center, Denver, CO 80262, USA.

[2]Department of Econometrics and Business Statistics, Monash University, Clayton VIC 3168, Australia. Please send all correspondence to Dr R.J. Hyndman at this address or to Rob.Hyndman@monash.edu.au.

[3]Tillinghast-Towers Perrin, Level 17, MLC Centre, 19–29 Martin Place, Sydney NSW 2000, Australia.

[4]Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455-0378, USA.

1 Introduction

We frequently encounter time series which are clearly non-Gaussian. Particular forms of non-normality, such as series of counts, proportions, binary outcomes or non-negative or heavy-tailed observations are all common. In the course of analysing various non-Gaussian series, we have encountered more than 30 different models described as having first order autoregressive (AR(1)) structure. This diversity makes it difficult to see just what AR(1) structure is, and to decide how to proceed in a particular modelling situation.

In this paper we discuss first-order conditionally linear autoregressive models. Linear AR(1) structure is simple, useful and interpretable in a wide range of contexts. Our aims are two-fold. We want to derive theoretical results analogous to standard Gaussian results which will allow us to better understand AR(1) structure and to see similarities and differences among the various AR(1) models in the literature. We also want to develop data-analytic methods to aid the practitioner in using these many models in a given situation.

In Section 2 we give a general formulation of linear AR(1) structure, and in Section 3 we review some of the models which fit within our general formulation. More limited surveys of some of the non-Gaussian models are found in Lewis (1985), McKenzie (1985a) and Sim (1994). A more detailed survey is given in Grunwald et al. (1995).

Theoretical results concerning stationarity, moments and correlation structure have been proven for many particular AR(1) models, but in fact under very mild assumptions many of these properties can be derived in much more generality. In Section 4 we derive these new general results which are very useful in understanding AR(1) structure and in formulating or selecting models appropriate to given situations.

Section 5 considers the application of these models in data analysis. We discuss parameter estimation, and in particular the issue of selecting among several possible AR(1) models for a given series. We use a general parametric bootstrap diagnostic proposed by

Tsay (1992) to show that it is possible to distinguish between various AR(1) models on a given sample space.

In analysing Gaussian series, AR(1) models often appear as building blocks in more complex models, for instance as a means of including correlated errors in regression (e.g., Judge et al., 1985) or smoothing (e.g., Altman, 1990 or Hart, 1991). We mention some possible extensions to non-Gaussian models in Section 7.

Various alternative approaches to modelling non-Gaussian time series have been proposed including the Bayesian forecasting models of West, Harrison and Migon (1985) or Harvey and Fernandes (1989), state space models as in Kashiwagi and Yanagimoto (1992), Kitagawa (1987) or Fahrmeir (1992), and the transformation approach of Swift and Janacek (1991). These alternative approaches are outside the range of this paper.

2 The conditional linear AR(1) model

Let $\{Y_t\}$, $t = 0, 1, \ldots$ be a time-homogeneous first order Markov process on sample space $\mathcal{Y} \subseteq \mathbb{R}$ with conditional (transition) distribution function $F(y_t \mid y_{t-1})$. We assume Y_0 has a fixed but (usually) arbitrary distribution, and for some results that Y_0 has the stationary distribution of Y_t when it exists.

Let $m(Y_{t-1}) \equiv \mathsf{E}(Y_t \mid Y_{t-1})$ denote the conditional mean of Y_t. We say $\{Y_t\}$ has *first order conditional linear autoregressive (CLAR(1)) structure* if

$$m(Y_{t-1}) = \phi Y_{t-1} + \lambda \tag{2.1}$$

where ϕ and λ are any real numbers. When the sample space $\mathcal{Y}$ is not the entire real line, restrictions on ϕ and λ may be required to ensure that $m(Y_{t-1})$ remains in the parameter space of $F(y_t \mid y_{t-1})$. We assume appropriate restrictions for any model we consider.

The conditional distribution function $F(y_t \mid y_{t-1})$ may depend on other parameters be-

sides ϕ and λ, and we let $\boldsymbol{\theta}$ be a vector of these parameters ($\boldsymbol{\theta}$ may be null if there are no other parameters). The usual Gaussian AR(1) model

$$(Y_t - \mu) \;=\; \phi(Y_{t-1} - \mu) + Z_t\,, \qquad Z_t \overset{iid}{\sim} N(0, \sigma^2) \tag{2.2}$$

is a special case where F denotes a normal distribution, $\mathcal{Y} = \mathbb{R}$, $\lambda = (1-\phi)\mu$ and $\boldsymbol{\theta} = \sigma$ since $[Y_t \,|\, Y_{t-1}] \sim N(m(Y_{t-1}), \sigma^2)$.

The CLAR(1) class includes nearly all of the non-Gaussian AR(1) models that have been proposed in the literature, and allows derivation of some very general theoretical results as in Section 4. The definition also conveys the original idea of autoregression—regressing the series on previous values of itself.

An alternative definition of first order autoregressive structure is obtained by requiring an exponentially decaying autocorrelation function (ACF)

$$\rho_k = \mathsf{Corr}[Y_t, Y_{t-k}] = \phi^k, \qquad k = 1, 2, \ldots\,. \tag{2.3}$$

In Section 4.5 we show that under very mild conditions, (2.3) is implied by CLAR(1), but the converse is not true. There are a few processes in the literature that have exponentially decaying ACF (2.3) but do not have linear conditional mean (2.1). For example, the minification processes of Tavares (1977, 1980a, 1980b) and Lewis and McKenzie (1991) and the product AR processes of McKenzie (1982). Our general results below do not apply to these models.

3 Literature on non-Gaussian AR(1) models

We give a brief review of models and methods that have appeared in the literature on non-Gaussian AR(1) models, in order to illustrate the number and variety of individual models subsumed under the CLAR(1) class. In our recent review of non-Gaussian AR(1)

Method	Model formulation	
Innovation	$Y_t = \phi Y_{t-1} + Z_t,$	(3.1)
Conditional distribution	$[Y_t \,\vert\, Y_{t-1}] \sim (m(Y_{t-1}), \boldsymbol{\theta})$	(3.2)
Random coefficient	$Y_t = \phi_t Y_{t-1} + Z_t$	(3.3)
Thinning	$Y_t = \phi * Y_{t-1} + Z_t$	(3.4)
Random coefficient thinning	$Y_t = \phi_t * Y_{t-1} + Z_t$	(3.5)

Table 1: *Methods of constructing non-Gaussian linear AR(1) models. Where relevant, $Z_t \overset{iid}{\sim} (\lambda, \boldsymbol{\theta})$, $\{\phi_t\}$ represents an iid sequence of random coefficients such that $\mathsf{E}\phi_t = \phi$ and $\{\phi_t\}$ is independent of $\{Z_t\}$, and $\phi * Y$ represents the "thinning" operation.*

models we summarized more than 30 models (Grunwald et al., 1995). Most models were on the sample spaces $\mathbb{R}$ (eight models), $(0, \infty)$ (nine models) or $\{0, 1, \ldots, \}$ (eight models) but other sample spaces included $(0, 1)$, $\{0, 1, \ldots, N\}$ and $(-\pi, \pi)$. Nearly all of the AR(1) models we found are contained in the CLAR(1) class, but differ in other properties, illustrating the range of CLAR(1) models.

Several general methods have been used to construct non-Gaussian AR(1) models. All of these methods lead to models satisfying CLAR(1), and some models can be equivalently constructed in several different ways. Let $Y \sim (\lambda, \boldsymbol{\theta})$ denote a random variable with mean λ and any other parameters contained in $\boldsymbol{\theta}$, with $m(Y_{t-1})$ as in (2.1). One could replace the innovations form (2.2) by a similar form with non-Gaussian innovations to give models of the form shown in (3.1) of Table 1. Any such model retains the linear conditional mean (2.1) (provided the mean of the innovations exists) and so is CLAR(1). This method has been used to construct models on $\mathbb{R}$ or $(0, \infty)$ using particular innovations distributions by Gaver and Lewis (1980), Lawrance (1982), Dewald and Lewis (1985), Bell and Smith (1986), Anděl (1988), Rao and Johnson (1988), Hutton (1990), Sim (1993), Lye and Martin (1994), and using a general innovation distribution on $\mathbb{R}$ by Brockwell and Davis (1991).

Alternatively, one could specify the conditional distribution associated with (2.1) to be of a particular form, as shown in (3.2) of Table 1. These models are again CLAR(1). Zeger and Qaqish (1988), Li (1994), Diggle et al. (1994), Shephard (1995) and Hyndman (1999)

have used this method to construct models on $\mathbb{R}$, $(0,\infty)$ and $\{0,1,\ldots,N\}$.

Extensions to the innovations method have been proposed by replacing ϕY_{t-1} by random variables X_t where $\mathsf{E}[X_t \,|\, Y_{t-1}] = \phi Y_{t-1}$. The resulting models retain the linear conditional mean and so are CLAR(1). Random coefficient models represent one such approach and have been used by Lawrance and Lewis (1981), Nicholls and Quinn (1982), McKenzie (1985a), McKenzie (1985b), Sim (1986), and Lewis et al. (1989) to construct models on $\mathbb{R}$, $(0,\infty)$, $\{0,1,\ldots,\}$ or $(0,1)$.

The thinning operation in (3.4) and (3.5) denoted by $*$ is defined as $\phi * X = \sum_{i=1}^{N(X)} W_i$ where $N(X)$ is a non-negative integer random variable and $\{W_i\}$ is a sequence of iid random variables, independent of $N(X)$, such that $\mathsf{E}[N(X)W_i \,|\, X] = \phi X$. The most common form of thinning is binomial thinning where $N(X) = X$ and $W_i \stackrel{d}{=} \text{Bin}(1,\phi)$, where $\text{Bin}(n,p)$ denotes a binomial variable, but several other possibilities for $N(X)$ and W_i have been proposed. Thinning has been used by McKenzie (1985a), Al-Osh and Alzaid (1987), Alzaid and Al-Osh (1988), McKenzie (1988), Al-Osh and Aly (1992), Sim (1990), and Franke and Seligmann (1993) to construct AR(1) models on $(0,\infty)$ or $\{0,1,\ldots,\}$. Random coefficients combined with thinning was used by McKenzie (1985a) to obtain models on $\{0,1,\ldots,\}$.

Several other methods have yielded non-Gaussian AR(1) models that are CLAR(1), including those of Kanter (1975), Jacobs and Lewis (1978a,b), McKenzie (1985a), Sim (1990), and Al-Osh and Aly (1992).

4 Stochastic properties CLAR(1) models

We now give general results describing the stochastic structure of CLAR(1) models. These results are similar in form to those for the Gaussian AR(1) model and so facilitate comparisons with that case. They also unify many individual results in the literature, and show the wide variety of stochastic behavior that can be obtained within the

CLAR(1) class. Understanding these stochastic properties will be of practical use in developing methods for analyzing non-Gaussian time series in later sections. Some of the results have been derived in particular cases, but not, to our knowledge, in this general setting. Results are stated below, with proofs given in Appendix 1.

4.1 Stationary process mean

Under very mild assumptions, the stationary process mean can be easily derived from the linear conditional mean (2.1) without any knowledge of the distributions involved.

Proposition 1 *For a CLAR(1) process, if* $|\mathsf{E}(Y_0)| < \infty$ *and* $|\phi| < 1$ *then* $lim_{t\to\infty}\, \mathsf{E}(Y_t) = \lambda/(1-\phi) \equiv \mu$. *If* $\mathsf{E}(Y_0) = \mu$, *then* $\mathsf{E}(Y_t) = \mu$ *for* $t \geq 0$.

When $|\phi| < 1$ and $\mathsf{E}(Y_0) = \mu$, we can rewrite (2.1) as $m(Y_{t-1}) = \phi Y_{t-1} + (1-\phi)\mu$, showing the conditional mean $m(Y_{t-1})$ to be a combination of the previous observation and the stationary process mean.

4.2 Stationary process variance

Further assumptions are needed to obtain results for higher moments. In exponential family theory, the class of distributions with quadratic variance function includes the most common probability distributions and gives a class of distributions for which many theoretical results are available (Morris, 1982, 1983). Even without the exponential family structure, the assumption of quadratic conditional variance includes all of the models in this paper and allows simple and general expressions for stationary process variances, as given in Proposition 2 below. Note, however, that it doesn't include the conditional Cauchy model (e.g., Brockwell and Davis, 1991, section 13.3) which has infinite variance, or the von Mises model (Breckling, 1989) which has a non-quadratic conditional variance relationship.

The conditional distribution $[Y_t \,|\, Y_{t-1}]$ has *quadratic conditional variance* if

$$v(m(Y_{t-1})) \;\equiv\; \mathsf{Var}(Y_t \,|\, Y_{t-1}) \;=\; am(Y_{t-1})^2 + bm(Y_{t-1}) + c \tag{4.1}$$

where a, b and c are constants possibly depending on μ or $\boldsymbol{\theta}$ (we suppress this dependence in the notation). In particular, note that $a \geq 0$ unless the sample space $\mathcal{Y}$ is finite. Values of $a < 0$ can occur for finite sample spaces, as illustrated in the examples below.

Proposition 2 *For a CLAR(1) process, suppose* $[Y_t \,|\, Y_{t-1}]$ *has quadratic conditional variance (4.1), and* $|\phi| < 1$*; and assume* $\mathsf{E}(Y_0) = \mu$*.*

1. *If* $a = -1$ *then* $\mathsf{Var}(Y_t) = v(\mu)$ *for all* $t \geq 1$*.*
2. *If* $a \neq -1$ *and* $|\phi| < 1/|a+1|^{\frac{1}{2}}$ *then*

$$\lim_{t\to\infty} \mathsf{Var}(Y_t) = \frac{v(\mu)}{1-(a+1)\phi^2}\ .$$

When $|\phi| < 1/|a+1|^{\frac{1}{2}}$ or $a \neq 1$, if we assume $\mathsf{Var}(Y_0) = v(\mu)/(1-(a+1)\phi^2)$, then it follows that $\mathsf{Var}(Y_t) = v(\mu)/(1-(a+1)\phi^2)$ for all $t \geq 1$. Proposition 2 clarifies the analogy with the Gaussian stationary marginal variance (see Example 1 below), and the close relation between conditional and stationary marginal variances in CLAR(1) models.

For a conditional variance relation more general than quadratic, say $\mathsf{Var}(Y_t \,|\, Y_{t-1}) = f(m(Y_{t-1}))$, we can obtain the expression

$$\mathsf{Var}(Y_t) = \mathsf{Var}[\mathsf{E}(Y_t \,|\, Y_{t-1})] + \mathsf{E}[\mathsf{Var}(Y_t \,|\, Y_{t-1})] = \phi^2\mathsf{Var}(Y_{t-1}) + \mathsf{E}[f(m(Y_{t-1}))]$$

but little more can be said in general. With further assumptions on f, results for higher moments analogous to Proposition 2 could also be derived.

4.3 Conditions for stationarity

When a CLAR(1) model is defined to have a specific marginal distribution for Y_t for all t, the process is clearly stationary. In other cases the form or even the existence of a stationary distribution may not be known. We now give conditions under which a CLAR(1) model has an ergodic distribution with moments represented by the limits in Propositions 1 and 2.

The results of Feigin and Tweedie (1985) give the necessary tools for finding conditions for stationarity in terms of the conditional distribution, based on methods developed for a Markov chain on a general state space. A few technical assumptions are required, which we give in condensed form—details may be found in Feigin and Tweedie (1985) or in more generality in Meyn and Tweedie (1993). To show the required *irreducibility* for an AR(1) model it suffices (and is generally easy) to show that the support of the conditional distribution $F(y_t \mid y_{t-1})$ is the entire sample space $\mathcal{Y}$ for each $y_{t-1} \in \mathcal{Y}$. This holds for all CLAR(1) models discussed in this paper. We also require that Y_t is a *Feller chain*, and this holds if the transition from $Y_{t-1} = y_{t-1}$ to Y_t is a pointwise continuous function of y_{t-1}; this is again true for all CLAR(1) models discussed in this paper.

The following result gives a sufficient but not necessary condition for existence of an ergodic distribution π on $\mathcal{Y}$, in the sense that for every $y \in \mathcal{Y}$,

$$\|\mathrm{P}^t(y, \cdot) - \pi(\cdot)\| \to 0 \text{ as } t \to \infty \tag{4.2}$$

where $\mathrm{P}^t(y, A) = \Pr(Y_t \in A \mid Y_0 = y)$ and $\|\cdot\|$ denotes the total variation norm. We give a result for the two cases $\mathcal{Y} = \mathbb{R}$ and $\mathcal{Y} \subseteq [0, \infty)$, which include all of the CLAR(1) models discussed in this paper.

Proposition 3 *Assume $\{Y_t\}$ is a CLAR(1) model and that it is also an irreducible Feller chain. Then consider two cases*

Case I: $\mathcal{Y} = \mathbb{R}$. *If* $\mathsf{E}\Big[|Y_t - m(Y_{t-1})| \;\Big|\; Y_{t-1} = y\Big] < B$ *for all* y *and some finite* B, *and if* $|\phi| < 1$, *then* $\{Y_t\}$ *is ergodic, and then convergence in (4.2) is geometrically fast.*

Case II: $\mathcal{Y} \subseteq [0, \infty)$. *If* $0 \leq \phi < 1$ *then* $\{Y_t\}$ *is ergodic, and then convergence in (4.2) is geometrically fast.*

These conditions are sufficient but may not be necessary, and in some cases such as the conditionally Gamma model in Example 3 below, $\{Y_t\}$ may be ergodic even for some $\phi \geq 1$.

It is not obvious that the limits of sequences of moments given in Propositions 1 and 2 correspond to the moments of the ergodic distribution when it exists. Using Theorem 2 of Feigin and Tweedie (1985) and methods similar to those in the proof of Proposition 3 (taking $g(y) = y^k + 1$ for the kth moment), it can be shown that for an AR(1) process which is ergodic, if the limits of the sequences in Propositions 1 and 2 are finite, they do in fact represent the moments of the ergodic distribution. Moreover, from Theorem 14.3.1 of Meyn and Tweedie (1993), we can also see that if the limits in Proposition 2 are infinite then the stationary process has infinite marginal variance. Hence, at least for irreducible Feller models, we do have that the limits in Propositions 1 and 2 give the stationary moments as we would hope.

The boundedness condition in Case I cannot be dispensed with; it is easy to construct examples where the conditional mean $m(Y_{t-1})$ is linear and satisfies $|\phi| < 1$, but where no stationary distribution exists. However, weaker conditions than the boundedness may suffice in some cases.

4.4 Examples

The results above give insight into the stationary process structure, as we illustrate in the following examples. In particular this illustrates that CLAR(1) models on a given sample space can have quite different stochastic properties. This is useful in developing model

diagnostic and selection methods as discussed in Section 5.

1. If the conditional variance (4.1) is finite and does not depend on $m(Y_{t-1})$ ($a = b = 0$ and $c < \infty$) then $\mathsf{Var}(Y_t) = c/(1-\phi^2)$. This is the case for Gaussian stationary processes ($c = \sigma^2$ gives the usual result), and also for any model of the innovations form $Y_t = \phi Y_{t-1} + Z_t$ with Z_t iid with finite variance (then $c = \mathsf{Var}(Z_t)$). For these models the stationary variance is greater than the innovations variance by the factor $1/(1-\phi^2)$.

2. If the conditional variance is finite and linear ($a = 0$, $b \neq 0$), then again $\mathsf{Var}(Y_t) = v(\mu)/(1-\phi^2)$. This is similar in form to the usual Gaussian result and more generally to the previous example but now the stationary variance depends on the stationary process mean μ in the same way that the conditional variance depends on the conditional mean. An example of such a model is a conditionally Poisson model: $[Y_t \,|\, Y_{t-1}] \sim \mathrm{Pn}(m(Y_{t-1}))$ where $\mathrm{Pn}(m)$ denotes a Poisson distribution with mean m. This model has $\mathcal{Y} = \{0, 1, \ldots\}$, $\lambda \geq 0$, $\phi \geq 0$, $\lambda + \phi > 0$ and $\boldsymbol{\theta}$ null. Then $a = 0$, $b = 1$ and $c = 0$, and

$$\mathsf{Var}(Y_t) = \frac{\mu}{1-\phi^2} \qquad \text{for } 0 \leq \phi < 1\,.$$

3. If the conditional variance is quadratic with $a > 0$, an ergodic distribution exists but there will be values of $|\phi| < 1$ for which $\lim_{t\to\infty} \mathsf{Var}(Y_t)$ is infinite. (This will also be true if $a < -2$, but we do not know of any such models in the literature.) An example is a conditionally Gamma model: $[Y_t \,|\, Y_{t-1}] \sim \mathrm{G}(r, r/m(Y_{t-1}))$ where $\mathrm{G}(r, \alpha)$ denotes a Gamma distribution with shape parameter r, rate parameter α, mean r/α and variance r/α^2. Here, $\mathcal{Y} = (0, \infty)$, $\lambda \geq 0$, $\phi \geq 0$, $\lambda + \phi > 0$ and $\boldsymbol{\theta} = r$. This gives $\mathsf{Var}(Y_t \,|\, Y_{t-1}) = m(Y_{t-1})^2/r$ so $a = 1/r$, $b = 0$ and $c = 0$, and

$$\mathsf{Var}(Y_t) = \frac{\mu^2/r}{1-\phi^2(r+1)/r} \qquad \text{for } 0 \leq \phi < \left(\tfrac{r}{r+1}\right)^{\frac{1}{2}} < 1.$$

Here, because of the heavy tails of the conditional distribution, the marginal variance is infinite for some values of $\phi < 1$, with a greater region of infinite variance for smaller r (greater conditional skewness). For instance, if $r = 1$, $[Y_t \,|\, Y_{t-1}]$ is exponentially distribution with mean $m(Y_{t-1})$, then $\mathsf{Var}(Y_t \,|\, Y_{t-1}) = m(Y_{t-1})^2$ and

$$\mathsf{Var}(Y_t) = \frac{\mu^2}{1-2\phi^2} \qquad \text{for } 0 \le \phi < \tfrac{1}{\sqrt{2}} \approx 0.707\,.$$

However, Grunwald and Feigin (1996) have studied this and similar models, and show that the conditionally Gamma model is ergodic for some values of $\phi \ge 1$, with a greater region of ergodicity for smaller r. For example, the conditionally exponential model is ergodic for $\phi < \exp\{-\psi(1)\} \approx 1.77$ ($\psi(r) = \frac{d}{dr}\log\Gamma(r)$ is the digamma function) when $\lambda > 0$. When ϕ satisfies the same condition and $\lambda = 0$, $Y_t \to 0$ almost surely. Meyn and Tweedie (1993, p.226) mention in general the possibility of ergodicity when $\phi > 1$.

4. Values of $a < 0$ are possible only when $\mathcal{Y}$ is finite. Therefore $\mathsf{Var}(Y_t)$ is also finite and Proposition 2 gives the variance for allowable values of $|\phi| < 1$. An example is the conditionally Binomial process: $[Y_t \,|\, Y_{t-1}] \sim \mathsf{Bin}(n, p(Y_{t-1}))$ with $p(Y_{t-1}) \equiv m(Y_{t-1})/n$, $\mathcal{Y} = \{0, 1, \ldots, n\}$, $\mu \in [0, n]$, $0 \le \phi \le 1$, $\mu + \phi > 0$, $\mu + \phi < n + 1$ and $\boldsymbol{\theta} = n$. $\mathsf{Var}(Y_t \,|\, Y_{t-1}) = np(Y_{t-1})(1 - p(Y_{t-1})) = -m^2(Y_{t-1})/n + m(Y_{t-1})$ so $a = -1/n$, $b = 1$ and $c = 0$. Proposition 2 gives

$$\mathsf{Var}(Y_t) = \frac{np(1-p)}{1 - \phi^2(n-1)/n} \qquad \text{for } 0 \le \phi \le 1$$

where $p \equiv \mu/n$. The stationary distribution is a distribution on $\{0, 1, \ldots, n\}$ with variance greater (unless $n = 1$ or $\phi = 0$) than the Binomial.

5. For models with random coefficients and iid innovations, $Y_t = \phi_t Y_{t-1} + Z_t$, Z_t iid,

$$\begin{aligned}
\mathsf{Var}(Y_t \,|\, Y_{t-1}) &= Y_{t-1}^2 \mathsf{Var}(\phi_t) + \mathsf{Var}(Z_t) \\
&= \frac{\mathsf{Var}(\phi_t)}{\phi^2} m^2(Y_{t-1}) - 2\lambda \frac{\mathsf{Var}(\phi_t)}{\phi^2} m(Y_{t-1}) + \lambda^2 \frac{\mathsf{Var}(\phi_t)}{\phi^2} + \mathsf{Var}(Z_t),
\end{aligned}$$

so $a = \mathsf{Var}(\phi_t)/\phi^2$ and the condition for finite stationary marginal variance is

$$|\phi| < \frac{1}{(\mathsf{Var}(\phi_t)/\phi^2 + 1)^{\frac{1}{2}}}$$

i.e. $\phi^2 < 1 - \mathsf{Var}(\phi_t) \leq 1$. This agrees with the eigenvalue condition for vector AR(p) random coefficient models given by Feigin and Tweedie (1985) Theorem 4 when $p = 1$.

Some algebra shows that in this case,

$$\mathsf{Var}(Y_t) = \frac{\mathsf{Var}(\phi_t)\mu^2 + \mathsf{Var}(Z_t)}{1 - [\mathsf{Var}(\phi_t) + \phi^2]} \ .$$

These results reduce to those for iid innovations models with constant coefficients ($\phi_t = \phi$) since then $\mathsf{Var}(\phi_t) = 0$.

6. For models with thinning and iid innovations, direct calculation using the conditional variance formula gives

$$\begin{aligned} \mathsf{Var}(Y_t \,|\, Y_{t-1}) &= \mathsf{Var}(W_i)\mathsf{E}[N(Y_{t-1}) \,|\, Y_{t-1}] + \mathsf{Var}(Z_t) \\ &\quad + [\mathsf{E}(W_i)]^2 \mathsf{Var}(N(Y_{t-1}) \,|\, Y_{t-1}) \ . \end{aligned}$$

In general, nothing more can be said, but special cases can be easily calculated. For instance, if $N(x) = x$,

$$\mathsf{Var}(Y_t) = \frac{\mathsf{Var}(W_i)\mu + \mathsf{Var}(Z_t)}{1 - \phi^2} \ .$$

This result includes several standard results for branching processes with immigration.

4.5 Autocorrelation structure

Many authors have defined specific models and proven both the linear conditional mean (2.1) and the exponentially decaying ACF results (2.3) using special methods. We now show that this is typically unnecessary, since under very mild conditions the exponentially decaying ACF is a consequence of the linear conditional mean (2.1) and holds very generally. This is particularly useful since (2.1) is typically much easier to check than (2.3). This and related results clarify the use and interpretation of the ACF as a model diagnostic for general AR(1) structure.

Proposition 4 *For a CLAR(1) process with* $|\phi| < 1$ *and* $\mathsf{Var}(Y_t) < \infty$ *constant in time,*

$$\rho_k = \mathsf{Corr}[Y_t, Y_{t-k}] = \phi^k, \qquad k = 1, 2, \ldots.$$

This result is mentioned in Heyde and Seneta (1972) in the context of branching processes, but does not seem to have appeared in such generality in the non-Gaussian time series literature.

The result is very useful in analyzing data. If the sample ACF does not appear to be exponentially decaying as in (2.3) then the model does not have CLAR(1) structure. (As mentioned previously, exponentially decaying ACF is also consistent with some models outside the CLAR(1) class.) Proposition 4 also makes it clear that the sample ACF will not help in determining which of several possible CLAR(1) models is most appropriate for a given series.

Standard errors for sample autocorrelations are useful in interpreting the sample ACF, and in fact this standard result holds generally also. Let R_k denote the lag k sample autocorrelation. Under the null hypothesis that $Y_1, \ldots, Y_n$ are iid with constant $\mathsf{Var}(Y_t) < \infty$, then $\rho_k = 0$ for $k \neq 0$ and $R_1, \ldots, R_h$ are approximately iid normal with mean 0 and variance $1/n$. (This result can be proved by examining the proof of Bartlett (1946), and

is also given in Brockwell and Davis (1991), Example 7.2.1.) Thus, the usual $\pm 1.96/\sqrt{n}$ bands for the ACF hold very generally, and there is typically no need to use simulation methods like those used by Sim (1994) and Grunwald and Hyndman (1998).

5 Modelling

5.1 Estimation

If the conditional distribution is known, it is relatively easy to compute maximum likelihood estimates for parameters in CLAR(1) models. The likelihood can be calculated using the first-order Markov property of the models:

$$L(y_1, \dots, y_n; \phi, \lambda, \boldsymbol{\theta}) = \pi(y_1) \prod_{t=2}^{n} p(y_t \,|\, y_{t-1})$$

where $p(y_t \,|\, y_{t-1})$ denotes the conditional density or the conditional probability function and $\pi(y_1)$ denotes the marginal density or marginal probability function. If $\pi(y_1)$ is unknown, the likelihood conditional on y_1 can be calculated by omitting $\pi(y_1)$ in the above expression. Hence, if the conditional distribution is known, maximum likelihood estimates can always be found, at least numerically.

However, several difficulties can arise with maximum likelihood estimators. If the conditional density function has discontinuities (as with many of the models on $(0, \infty)$ for example), then the likelihood will also be discontinuous and numerical optimization is very difficult. See the comments by Raftery in the discussion of Lawrance and Lewis (1985). Maximum likelihood estimators can also be extremely non-robust for some models (e.g., Anděl (1988)).

Ordinary least squares regression of Y_t against Y_{t-1} gives a less efficient but simpler and more robust estimation method. For CLAR(1) models, ordinary least squares gives unbiased estimators of λ and ϕ. Parameters in $\boldsymbol{\theta}$ must be estimated by other means, possibly

including maximum likelihood conditional on $\hat{\phi}$ and $\hat{\lambda}$.

5.2 Model selection and diagnostics

In a given data analysis, the sample space will be known. The ACF can be compared with (2.3) to determine if some CLAR(1) model may be appropriate, but typically there will still be several possible CLAR(1) models on that sample space. By Proposition 4, the ACF cannot be used to select among them. A particular model is often assumed for computational convenience or familiarity, as in Sim (1994). Standard diagnostics such as residuals are usually used to show that a proposed model could be appropriate. Examination of the marginal distribution is also recommended, but even with QQ plots against the given theoretical marginal distribution, there is often too much variability for these to be of much practical use. Additionally, there are often several different CLAR(1) models with the same form of marginal distribution (see Example 3 below), and QQ plots of the series cannot distinguish them.

To our knowledge there has not been any study of methods for model selection among several possible AR(1) models, or of the extent to which the various CLAR(1) models on a given sample space can be distinguished in practice. The problem is challenging because series are often short, distributions are non-Gaussian, models are not nested, and all models being considered share some features such as (2.1) and (2.3) so differences may be subtle. In this section we show how an understanding of the stochastic properties of each model, as given in Section 4, can be used to develop model diagnostics.

Tsay (1992) developed a very general approach to model diagnostics for time series based on using parametric bootstrap samples to assess the adequacy of a fitted model. The premise is that series simulated from the fitted model should share the stochastic properties of the series being modelled. Tsay proposed specifying a particular characteristic or functional, such as τ defined below, of the series or model, obtaining its sampling distribution using the parametric bootstrap on the fitted model, and comparing the observed

value for the series with this distribution. We have found this approach particularly useful in situations like those in this paper. Grunwald et al. (1997) used this approach to discover and study some surprising properties of Bayesian time series models, even in cases when the fit had passed a set of standard residual diagnostics. Sim (1994) also used this method to show that a particular model gave an adequate fit for a given positive series, but did not consider distinguishing among several possible models for the series.

We now illustrate the methods described above with three real data series. In some cases, models are easily distinguishable even with short series and moderate correlation, while in other cases specialized diagnostics based on particular stochastic properties of the models under consideration are needed.

6 Examples

6.1 Example 1: Weekly incidence of MCLS

Consider the series of 52 weekly counts of the incidence of acute febrile muco-cutaneous lymph node syndrome (MCLS) in Totori-prefecture in Japan during 1982, given by Kashiwagi and Yanagimoto (1992). In that year, a nationwide outbreak of MCLS was reported. The authors used a state space model to estimate a postulated underlying smooth disease rate. An alternative analysis, useful for other purposes, is based on CLAR(1) models. Figure 1 shows the series and the sample ACF. These are consistent with (2.3) and a CLAR(1) model on sample space $\mathcal{Y} = \{0, 1, \ldots\}$. (The non-significant negative correlations for lags 10–17 are not unusual for short CLAR(1) series even in the Gaussian case.) We considered two CLAR(1) models:

1. INAR(1) (McKenzie, 1985a, 1988; Al-Osh and Alzaid, 1987):

$$Y_t = \phi * Y_{t-1} + Z_t = \sum_{i=1}^{Y_{t-1}} W_i + Z_t$$

with $Z_t \overset{iid}{\sim} \text{Pn}(\lambda)$ and $W_i \overset{iid}{\sim} \text{Bin}(1, \phi)$.

2. Conditionally Poisson (Example 2, Section 4.4):

$$[Y_t \,|\, Y_{t-1}] \sim \text{Pn}(m(Y_{t-1}))$$

with $m(Y_{t-1}) = \phi Y_{t-1} + \lambda$.

Figure 1 about here

The INAR(1) model also appears in the stochastic processes literature as an infinite server ($M/M/\infty$) queue (Parzen, 1962) and as a Poisson branching process with immigration (Heyde and Seneta, 1972, for instance). Least squares conditional on y_1 was used to estimate ϕ and λ in each model. The result was $\hat{\phi} = 0.524$ and $\hat{\lambda} = 0.802$ (the estimated parameters are the same for the two models since LS does not use distributional information).

Using Proposition 2 the marginal variances of the models are $\mathsf{Var}(Y_t) = \mu$ for INAR(1), and $\mathsf{Var}(Y_t) = \mu/(1-\phi^2)$ for the conditionally Poisson model. (The INAR model has Poisson marginal distribution; an explicit form for the marginal distribution of the latter model is not directly available.) This difference in the theoretical relationship between marginal variance and marginal mean suggests the diagnostic $\tau \equiv \mathsf{Var}(Y_t)/\mathsf{E}(Y_t)$. Estimates of τ are given by the ratio $\hat{\tau} = s_y^2/\overline{y}$ where $\overline{y}$ and s_y^2 are the series sample mean and variance respectively. For the MCLS series, $\hat{\tau} = 1.818$. Simulating 100 series from each fitted model, computing $\hat{\tau}$ for each, and constructing 95% intervals from the 2.5 and 97.5 percentiles of each set of estimates gave $(0.529, 1.437)$ for INAR(1) and $(0.893, 2.180)$ for the conditionally Poisson model. The series is consistent with the conditionally Poisson model, but despite its short length and moderate correlation, it can be seen that the INAR(1) does not give a large enough variance to model this series.

6.2 Example 2: Gold particles

We repeated this procedure for another series on the same sample space, the first 60 values of the series of counts of gold particles as given by Guttorp (1991). Graph similar to those in Figure 1 (not shown) show the series and sample ACF are again consistent with CLAR(1) structure. LS estimates are $\hat{\phi} = 0.583$ and $\hat{\lambda} = 0.636$. Using the same two models and τ as in Example 1, $\hat{\tau} = 0.850$ for the series and the INAR(1) and conditionally Poisson percentile intervals were $(0.583, 1.458)$ and $(0.885, 2.041)$ respectively. This series is consistent with INAR(1), but again despite its short length and moderate correlation, it can be seen to have too small a variance to have arisen from a conditionally Poisson model. Examination of the physical experiment (Chandrasekhar, 1954) shows that the INAR(1) model is expected to be appropriate, though for the full series of 1598 observations the ACF shows substantial additional variation (Grunwald and Hyndman, 1998). Examples 1 and 2 together show that no single CLAR(1) model on a given sample space will suffice.

6.3 Example 3: Rainfall data

Weiss (1985) gave half-hourly riverflow and rainfall data for the Hirnant river, Wales, for November and December, 1972. Here we consider a period of 72 consecutive half-hours with some recorded rainfall (observations #1574–1645) as an example of a series on $\mathcal{Y} = (0, \infty)$. (More sophisticated models could impose a binary Markov chain to allow for periods with no rain, along with a positive AR(1) model for rain amounts—see Stern and Coe (1984), Grunwald and Jones (1999), or Hyndman and Grunwald (2000) for instance.) The left graph in Figure 2 shows the series. The ACF (not shown) is again consistent with CLAR(1) structure.

Figure 2 about here

We consider three models for series on $(0, \infty)$ that are based on exponential distributions. We use the notation $\text{Exp}(\mu)$ to denote exponential random variable with mean $1/\mu$.

1. EAR(1) (Gaver and Lewis, 1980):

$$Y_t = \phi Y_{t-1} + Z_t \qquad \text{where} \qquad Z_t \sim \begin{cases} 0 & \text{w.p. } \phi \\ \text{Exp}(1/\mu) & \text{w.p. } 1-\phi. \end{cases}$$

2. Thinning model of Sim (1990):

$$Y_t = \phi * Y_{t-1} + Z_t = \sum_{i=1}^{N(Y_{t-1})} W_i + Z_t$$

   where $Z_t \overset{iid}{\sim} \text{Exp}(1/\lambda)$, $W_i \overset{iid}{\sim} \text{Exp}(1/\lambda)$ and $N(x) \sim \text{Pn}(\phi x/\lambda)$.

3. Conditionally exponential (Example 3, Section 4.4 with $r = 1$):

$$[Y_t \,|\, Y_{t-1}] \sim \text{Exp}(1/m(Y_{t-1}))$$

   with $m(Y_{t-1}) = \phi Y_{t-1} + \lambda$.

The first two models have exponential marginal distributions. The middle graph in Figure 2 shows an exponential QQplot of the 72 values. Simulations from exponential distributions can show this much deviation from a straight line.

We first considered three test statistics: the positive/negative ratio (pn) for diagnosing time-irreversibility, as discussed by Tsay (1992); the coefficient of skewness; and the ratio of series variance to series mean. The first two were used by Sim (1994). Table 2 shows the results of 100 simulations from each of the three models (models were fitted using least squares). Skewness and variance/mean ratio are inconclusive and do not reject any of the models, despite the marginal distributions having quite different forms. The pn rejects EAR(1). Viewing graphs of simulated series also makes it clear that EAR(1) could not have generated this series. In particular, EAR(1) requires $y_t \geq \phi y_{t-1}$, and on

a graph of y_t versus y_{t-1} this shows up as a lower bound at $y_t = \phi y_{t-1}$, which is not evident in the lagged plot in the right panel of Figure 2. Further examination of model properties as in Section 4 shows that the thinning model has conditional variance linear in $m(Y_{t-1})$ (and so also in Y_{t-1}), while the conditional model has conditional variance quadratic in $m(Y_{t-1})$ and Y_{t-1}. This suggested the statistic τ defined by the ratio of the variance of y_t for periods with y_{t-1} greater than median y_{t-1} to the variance of y_t for periods with y_{t-1} less than median y_{t-1}. The column labelled "condvar" in Table 2 shows the percentile intervals and observed value. The thinning and innovations models now seem unlikely, and the theoretical results in Section 4 of this paper helped develop an appropriate diagnostic.

	Statistic			
Model	pn	skew	ratio	condvar
EAR(1)	(0.127, 0.340)	(0.614, 2.447)	(0.194, 1.191)	(0.494, 9.393)
Thinning	(0.732, 1.367)	(0.430, 2.427)	(0.253, 0.920)	(1.439, 16.031)
Conditional	(0.868, 1.448)	(1.297, 5.630)	(0.276, 13.109)	(2.147, 471.520)
Observed	1.152	1.330	0.722	20.372

Table 2: *95% intervals for model diagnostics for the Hirnant rainfall series.*

7 Conclusions and extensions

We have defined a large class of non-Gaussian first-order autoregressive models with linear conditional mean and we have derived several new theoretical results in this general setting. This work clarifies AR(1) structure for non-Gaussian models, unifies many separate results in the literature on non-Gaussian AR(1) models, and provides a theoretical basis for developing practical methods of model diagnostics and selection.

Despite the more than 25 different CLAR(1) models that have appeared in the literature, the CLAR(1) class still provides a limited selection of models for real data because of its linear conditional mean assumption and relatively simple correlation structure. However, CLAR(1) models can be used as building blocks for more complex models. For

example, a vector $\mathbf{x}_t$ of covariates can be included by defining a model that satisfies

$$\mathsf{E}[Y_t \mid Y_{t-1}, \mathbf{x}_t] = \phi Y_{t-1} + \lambda + \mathbf{x}_t'\boldsymbol{\beta} - \phi \mathbf{x}_{t-1}'\boldsymbol{\beta} \; .$$

This is particularly easy for innovations or conditional models.

Grunwald and Hyndman (1998) show how a smooth mean function μ_t can be included by setting

$$\mathsf{E}[Y_t \mid Y_{t-1}] = \phi Y_{t-1} + \mu_t - \phi \mu_{t-1}.$$

This is analogous to smoothing with correlated errors in the Gaussian case (see Altman, 1990 or Hart, 1991). If the conditional distribution is an exponential family form, these models are closely related to the Generalized Additive Models of Hastie and Tibshirani (1990).

When the conditional distribution is an exponential family form, some authors (Zeger and Qaqish, 1988, Li, 1994, Shephard, 1995, and Diggle et al. (1994) for instance) have considered using a link function $g(\cdot)$ as in Generalized Linear Models (McCullagh and Nelder, 1989), giving

$$g(m(Y_{t-1})) = \lambda + \phi Y_{t-1} \; .$$

Of course, unless the link is the identity, this model does not give a linear dependence between the conditional mean and the previous observation, but this is not necessarily a deficiency and has some advantages. A link function can give a wider range of allowable values of ϕ and λ, and allows the methods and software of Generalized Linear Models to be used. However, the ACF is then somewhat more difficult to relate to the models, and the properties of the models, particularly the range of ϕ which gives stationarity, are affected. For instance, Zeger and Qaqish (1988) show that the conditionally Poisson model with log link function is stationary only for $\phi \leq 0$. A similar effect has been noted in models for spatial correlation, as in the auto-Poisson model of Besag (1974), which is also capable of modelling only negative correlation. As Example 2 in Section 6 above

illustrates, this approach will also not provide the full range of models needed for real data. Modifications of the GLM approach include transforming Y_{t-1} (Zeger and Qaqish, 1988) or working with linear approximations (Shephard, 1995).

Acknowledgment

The work of the first two authors was supported by Australian Research Council grants. The authors are grateful to Paul Feigin and Kais Hamza for several useful discussions and to a referee for some helpful suggestions.

Appendix: Proofs of results in Section 4

We will repeatedly use three standard results:

Convergence of geometric series: If $|k| < 1$ then the recursion $x_t = kx_{t-1} + A$ has limit $A/(1-k)$ as $t \to \infty$, and if $x_0 = A/(1-k)$ then $x_t = A/(1-k)$ for $t \geq 1$.

Double expectation formula: For random variables X and Y with $\mathsf{E}(X) < \infty$, $\mathsf{E}(Y) = \mathsf{E}[\mathsf{E}(Y \mid X)]$ (Bickel and Doksum, 1977, 1.1.20).

Conditional variance formula: For random variables X and Y with $\mathsf{E}(X) < \infty$, $\mathsf{Var}(Y) = \mathsf{Var}[\mathsf{E}(Y \mid X)] + \mathsf{E}[\mathsf{Var}(Y \mid X)]$ (Bickel and Doksum, 1977, 1.6.12).

Proof of proposition 1:

From (2.1) and the double expectation formula, $\mathsf{E}(Y_t) = \phi\mathsf{E}(Y_{t-1}) + \lambda$ so the result follows directly from the convergence of geometric series with $x_t = \mathsf{E}(Y_t)$, $k = \phi$ and $A = \lambda$. □

Proof of proposition 2:

By the conditional variance formula we have

$$\mathsf{Var}(Y_t) = \mathsf{E}[v(m(Y_{t-1}))] + \mathsf{Var}(m(Y_{t-1})) \; .$$

Direct calculation gives $\mathsf{E}(m(Y_{t-1})) = \mu$, $\mathsf{E}[m^2(Y_{t-1})] = \phi^2\mathsf{E}(Y_{t-1}^2) + (1-\phi^2)\mu^2$ and $\mathsf{Var}(m(Y_{t-1})) = \phi^2\mathsf{Var}(Y_{t-1})$. Thus

$$\mathsf{Var}(Y_t) = \mathsf{E}(am^2(Y_{t-1}) + bm(Y_{t-1}) + c) + \phi^2\mathsf{Var}(Y_{t-1}) = \phi^2(a+1)\mathsf{Var}(Y_{t-1}) + v(\mu) \; .$$

By convergence of geometric series with $x_t = \mathsf{Var}(Y_t)$, $k = \phi^2(a+1)$ and $A = v(\mu)$, the limit is as stated. □

Proof of proposition 3:

Using Theorem 1 of Feigin and Tweedie (1985), the results will hold if we take the function $g(y) = |y| + 1$ and can find a $\delta > 0$ and a compact set A such that

$$\mathsf{E}[g(Y_t) \,|\, Y_{t-1} = y] \leq (1-\delta)g(y) \quad \text{for } y \in A^c \,.$$

Case I: The model can be stated in terms of innovations $Z_t \equiv Y_t - m(Y_{t-1})$ (which are not necessarily iid), giving $Y_t = m(Y_{t-1}) + Z_t$. Then by the hypothesis on Z_t,

$$\mathsf{E}[g(Y_t) \,|\, Y_{t-1} = y] \;\leq\; |\phi|\,|y| + |\lambda| + B + 1 \,.$$

Since $|\phi| < 1$, take $\delta > 0$ with $|\phi| < 1 - \delta < 1$, and then

$$|\phi|\,|y| + |\lambda| + B + 1 \;<\; (1-\delta)g(y)$$

whenever

$$|y| \;>\; \frac{|\lambda| + B + \delta}{1 - \delta - |\phi|} \equiv \alpha \,.$$

So taking $A = [-\alpha,\ \alpha]$ gives the required result.

Case II: Directly from the definition, since $Y_t \geq 0$,

$$\mathsf{E}[g(Y_t) \,|\, Y_{t-1} = y] = \phi y + \lambda + 1 \,.$$

A similar argument to that in Case I leads to $\alpha = (\lambda + \delta)/(1 - \delta - \phi)$ and $A = [0, \alpha]$. □

Proof of proposition 4:

Let $X_t = Y_t - \mu$ so that $\mathsf{E}(X_t) = 0$. Induction and the double expectation formula give $\mathsf{E}(X_t \,|\, X_{t-k}) = \phi^k X_{t-k}$, as follows: It is true for $k = 0$ since $\mathsf{E}(X_t \,|\, X_{t-0}) = \phi^0 X_t$. Assum-

ing it to be true for some $k > 0$,

$$\mathsf{E}(X_t \mid X_{t-(k+1)}) = \mathsf{E}[\mathsf{E}(X_t \mid X_{t-k}) \mid X_{t-(k+1)}] = \mathsf{E}(\phi^k X_{t-k} \mid X_{t-(k+1)}) = \phi^{k+1} X_{t-(k+1)}$$

since by (2.1), $\mathsf{E}(X_j \mid X_{j-1}) = \phi X_{j-1}$. Now,

$$\mathsf{Cov}(Y_t, Y_{t-k}) = \mathsf{E}(X_t X_{t-k}) = \mathsf{E}[X_{t-k}\mathsf{E}(X_t \mid X_{t-k})] = \phi^k \mathsf{E}(X_{t-k}^2) = \phi^k \mathsf{Var}(Y_{t-k}).$$

Dividing by $\mathsf{Var}(Y_{t-k})$ gives the result. □

References

AL-OSH, M.A. and ALY, E-E.A.A. (1992) First order autoregressive time series with negative binomial and geometric marginals, *Commun. Statist.-Theory Meth.*, **21**, (9) 2483–2492.

AL-OSH, M.A. and ALZAID, A.A. (1987) First order integer-valued autoregressive (INAR(1)) process, *J. Time Series Anal.*, **8**, 261–275.

ALTMAN, N.S. (1990) Kernel smoothing of data with correlated errors, *J. Amer. Statist. Assoc.*, **85**, 749–759.

ALZAID, A.A. and AL-OSH, M.A. (1988) First order integer-valued autoregressive process: distributional and regression properties, *Statistica Neerlandica*, **42**, 53–61.

ANDĚL, J. (1988) On AR(1) processes with exponential white noise, *Commun. Statist.-Theory Meth.*, **17**, 1481–1495.

BARTLETT, M.S. (1946) On the theoretical specification of sampling properties of autocorrelated time series, *J. Roy. Stat. Soc., B*, **8**, 27–41.

BELL, C.B. and SMITH, E.P. (1986) Inference for non-negative autoregressive schemes, *Commun. Statist.-Theory Meth.*, **15**, 2267–2293.

BESAG, J.E. (1974) Spatial interaction and the statistical analysis of lattice systems, *J. Roy. Stat. Soc., B*, **36**, 192–225.

BICKEL, P.J. and DOKSUM, K.A. (1977) *Mathematical statistics: basic ideas and selected topics*, Holden-Day: San Francisco.

BRECKLING, J. (1989) *The analysis of directional time series: applications to wind speed and direction*, Lecture Notes in Statistics **61**, Springer-Verlag: Berlin.

BROCKWELL, P.J. and DAVIS, R.A. (1991) *Time series: theory and methods,* second edition, Springer-Verlag: New York.

CHANDRASEKHAR, S. (1954) Selected papers on noise and stochastic processes, In *Stochastic problems in physics and astronomy*, ed. N. Wax, Dover: New York.

DEWALD, L.S. and LEWIS, P.A.W. (1985) A new Laplace second-order autoregressive

time series model—NLAR(2), *IEEE Trans. Inf. Theory*, **31**, 645–51.

DIGGLE, P.J., LIANG, K.Y. AND ZEGER, S.L. (1994) *Analysis of Longitudinal Data*, Clarendon Press, Oxford.

FAHRMEIR, L. (1992) State space modeling and conditional mode estimation for categorical time series, In *New directions in time series analysis, part 1*, (eds D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, and M.S. Taqqu), Springer: New York.

FEIGIN, P. and TWEEDIE, R.L. (1985) Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments, *J. Time Series Anal.*, **6**, 1–14.

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 analysis: in honour of Maurice B. Priestley*, ed. T. Subba Rao, Chapman and Hall: New York.

GAVER D.P. and LEWIS, P.A.W. (1980) First-order autoregressive gamma sequences and point processes, *Adv. Appl. Prob*, **12**, 726–745.

GRUNWALD, G.K. and FEIGIN, P. (1996) *First order product autoregressive processes*, Research Report 2, Department of Statistics, University of Melbourne.

GRUNWALD, G.K., HAMZA, K. and HYNDMAN, R.J. (1997) Some properties and generalizations of Bayesian time series models, *J. Roy. Stat. Soc., B*, **59**, 615–626.

GRUNWALD, G.K. and HYNDMAN, R.J. (1998) Smoothing non-Gaussian time series with autoregressive structure, *Computational Statistics and Data Analysis*, **28**, 171–191.

GRUNWALD, G.K., HYNDMAN, R.J., and TEDESCO, L. (1995) *A unified view of linear AR(1) models*, Research Report, Department of Statistics, University of Melbourne. Available at http://www-personal.buseco.monash.edu.au/~hyndman/papers/ar1.htm.

GRUNWALD, G.K. and JONES, R.J. (1999) Markov models for time series with mixed distribution. *Environmetrics*, **10**, in press.

GUTTORP, P. (1991) *Statistical inference for branching processes*, Wiley, New York.

HART, J.D. (1991) Kernel regression estimation with time series errors, *J. Roy. Statist. Soc., B*, **53**, 173–187.

HARVEY, A.C. and FERNANDES, C. (1989) Time series models for count or qualitative observations (with discussion), *J. Bus. Econ. Stat.*, **7**, 407–422.

HASTIE, T.J. AND TIBSHIRANI, R.J. (1990) *Generalized Additive Models*, Chapman and Hall, London.

HEYDE, C.C. and SENETA, E. (1972) Estimation theory for growth and immigration rates in a multiplicative process, *J. Appl. Prob.*, **9**, 235–256.

HUTTON, J.L. (1990) Non-negative time series models for dry riverflow, *J. Appl. Prob.*, **27**, 171–182.

HYNDMAN, R.J. (1999) "Nonparametric additive regression models for binary time series", Proceedings, 1999 Australasian Meeting of the Econometric Society, 7–9 July 1999, University of Technology, Sydney.

HYNDMAN, R.J., and GRUNWALD, G.K. (2000) Generalized additive modelling of mixed distribution Markov models with application to Melbourne's rainfall, *Australian & New Zealand J. Statistics*, to appear.

JACOBS, P.A. and LEWIS, P.A.W. (1978a) Discrete time series generated by mixtures I: correlational and runs properties, *J. Roy. Statist. Soc. Ser. B*, **40**, 94–105.

JACOBS, P.A. and LEWIS, P.A.W. (1978a) Discrete time series generated by mixtures II: asymptotic properties, *J. Roy. Statist. Soc. Ser. B*, **40**, 222–228.

JUDGE, G.G., HILL, R.C., GRIFFITHS, W.E., LÜTKEPOHL, H. and LEE, T.-C. (1985) *The theory and practice of econometrics*, 2nd edition, John Wiley & Sons: New York.

KANTER, M. (1975) Autoregression for discrete processes mod 2, *J. Appl. Prob.*, **12**, 371–375.

KASHIWAGI, N. and YANAGIMOTO, T. (1992) Smoothing serial count data through a state-space model, *Biometrics*, **48**, 1187–1194.

KITAGAWA, G. (1987) Non-Gaussian state-space modeling of nonstationary time series,

J. Amer. Stat. Assoc., **82**, 1032–1063.

LAWRANCE, A.J. (1982) The innovation distribution of a gamma distributed autoregressive process, *Scand. J. Statist.*, **9**, 234–236.

LAWRANCE, A.J. and LEWIS, P.A.W. (1981) A new autoregressive time series model in exponential variables (NEAR(1)), *Adv. Appl. Prob.*, **13**, 826–845.

LAWRANCE, A.J. and LEWIS, P.A.W. (1985) Modelling and residual analysis of nonlinear autoregressive time series in exponential variables (with discussion), *J. Roy. Stat. Soc., B*, **47**, 165–202.

LEWIS, P.A.W. (1985) Some simple models for continuous variate time series, *Water Resources Bulletin*, **21**, 635–644.

LEWIS, P.A.W. and MCKENZIE, E. (1991) Minification processes and their transformations, *J. Appl. Prob.*, **28**, 45–57.

LEWIS, P.A.W., MCKENZIE, E. and HUGUS, D.K. (1989) Gamma processes, *Commun. Statist.-Stochastic Models*, **5**, 1–30.

LI, W.K. (1994) Time series models based on generalized linear models: some further results, *Biometrics*, **50**, 506–511.

LYE, J.N. and MARTIN, V.L. (1994) Non-linear time series modelling and distributional flexibility, *J. Time Series Anal.*, **15**, 65–84.

MCCULLAGH, P., and NELDER, J.A. (1989) *Generalized Linear Models,* second edition, Chapman and Hall, London.

MCKENZIE, E. (1982) Product autoregression: a time series characterization of the gamma distribution, *J. Appl. Prob*, **19**, 463–468.

MCKENZIE, E. (1985a) Some simple models for discrete variate time series, *Water Resources Bulletin*, **21**, 645–650.

MCKENZIE, E. (1985b) An autoregressive process for beta random variables, *Management Science*, **31**, 988–997.

MCKENZIE, E. (1988) Some ARMA models for dependent sequences of Poisson counts,

Adv. Appl. Prob, **20**, 822–835.

MEYN, S.P. and TWEEDIE, R.L. (1993) *Markov Chains and Stochastic Stability*, Springer-Verlag: New York.

MORRIS, C.N. (1982) Natural exponential families with quadratic variance functions, *Ann. Statist.*, **10**, 65–80.

MORRIS, C.N. (1983) Natural exponential families with quadratic variance functions: Statistical theory, *Ann. Statist.*, **11**, 515–529.

NICHOLLS, D.F. and QUINN, B.G. (1982) *Random coefficient autoregressive models: an introduction*, Lectures notes in Statistics **11**, Springer–Verlag: New York.

PARZEN, E. (1962) *Stochastic Processes*, Holden-Day: San Francisco.

RAO, P.S. and JOHNSON, D.H. (1988) A first-order AR model for non-Gaussian time series, *Proceedings of IEEE international conference on ASSP*, **3**, 1534–1537.

SHEPHARD, N. (1995) *Generalized linear autoregressions*, Economics Working paper 8, Nuffield College, Oxford.

SIM, C.H. (1986) Simulation of Weibull and gamma autoregressive stationary processes, *Commun. Statist.-Simula.*, **15**, 1141–1146.

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

SIM, C.H. (1993) First-order autoregressive logistic processes, *J. Appl. Prob.*, **30**, 467–470.

SIM, C.H. (1994) Modelling non-normal first-order autoregressive time series, *J. Forecasting*, **13**, 369–381.

STERN, R.D. and COE, R. (1984) A model fitting analysis of daily rainfall data (with discussion), *J. R. Statist. Soc. A*, **147**, 1–34.

SWIFT, A.L. and JANACEK, G.J. (1991) Forecasting non-normal time series, *J. Forecasting*, **10**, 501–520.

TAVARES, L.V. (1977) The exact distribution of extremes in a non-Gaussian process, *Stoch. Proc. Appl.*, **5**, 151–156.

TAVARES, L.V. (1980a) A non-Gaussian Markovian model to simulate hydrologic processes, *J. Hydrology*, **46**, 281–287.

TAVARES, L.V. (1980b) An exponential Markovian stationary process, *J. Appl. Prob.*, **17**, 1117–1120.

TSAY, R.S. (1992) Model checking via parametric bootstraps in time series analysis, *Applied Statistics*, **41**, 1–15.

WEISS, G. (1985) Half-hourly precipitation and stream flow, river Hirnant, Wales, U.K., November and December, 1972., In *Data: A collection of problems from many fields for the student and research worker*, ed. D.F. Andrews and A.M. Herzberg, Springer-Verlag: New York.

WEST, M., HARRISON, P.J. and MIGON, H.S. (1985) Dynamic generalized linear models and Bayesian forecasting (with discussion), *J. Am. Statist. Assoc.*, **80**, 73–96.

ZEGER, S.L. and QAQISH, B. (1988) Markov regression models for time series: a quasi-likelihood approach, *Biometrics*, **44**, 1019–1031.

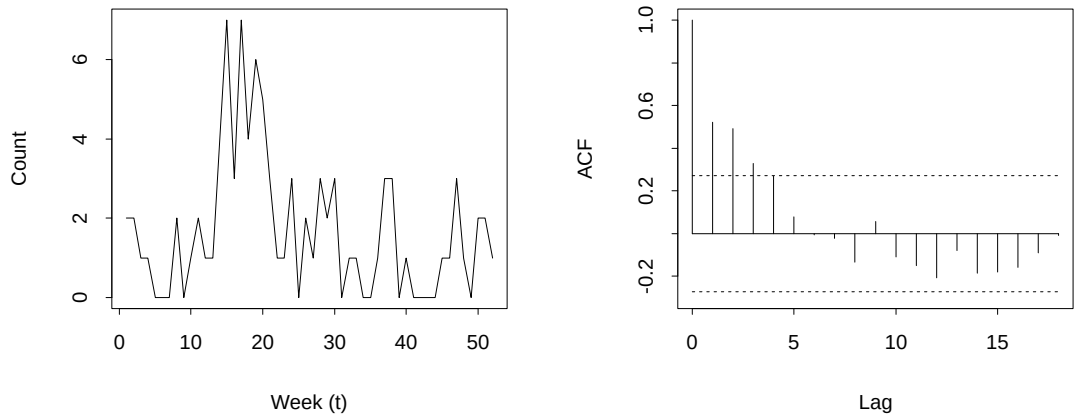

Figure 1: *Weekly counts of MCLS in 1982 in Totori-prefecture, Japan, with the sample ACF.*

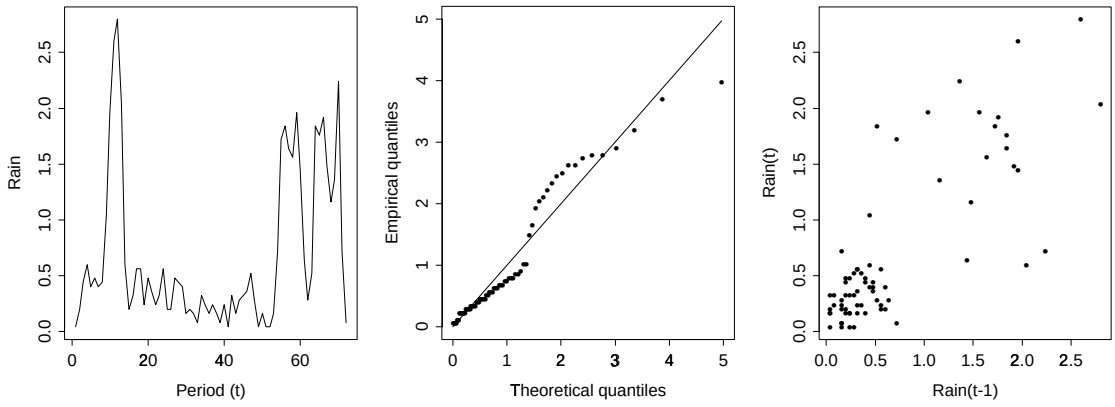

Figure 2: *Half-hourly rainfall at river Hirnant at periods 1574–1645, exponential QQplot and lag one scatterplot of series.*