# A unified view of linear AR(1) models

G.K. Grunwald*, R.J. Hyndman† and L.M. Tedesco* June 1996

**Summary** We review and synthesize the wide range of non-Gaussian first order linear autoregressive models that have appeared in the literature. Models are organized into broad classes to clarify similarities and differences and facilitate application in particular situations. General properties for process mean, variance and correlation are derived, unifying many separate results appearing in the literature. Examples illustrate the wide range of properties that can appear even under the autoregressive assumption. These results are used in analysing a variety of real data sets, illustrating general methods of estimation, model diagnostics and model selection.

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

## 1 Introduction

Time series in which the observations are of a clearly non-Gaussian nature are very common in many areas of science. Particular forms of non-normality, such as series of counts, proportions, binary outcomes or non-negative observations are all common. Quite a variety of apparently unrelated models have been suggested for non-Gaussian time series, and this diversity makes it difficult to decide how to proceed in a particular modelling situation.

In this paper we focus on first-order linear autoregressive (AR(1)) models. AR(1) structure is simple, useful and interpretable in a wide range of contexts. Many models have been proposed as non-Gaussian analogues of the Gaussian AR(1) model (more than 30 such models are reviewed or discussed in this paper). We explore how this diverse collection of models can be viewed in a more unified way to emphasize similarities and differences, encourage comparisons, direct data analysis, and suggest further extensions.

In Section 2 we give a general formulation of the linear AR(1) model, and in Section 3 we clarify some common structures by describing five classes of models within this general formulation. These classes contain most of the AR(1) models which have been proposed, and in Section 4 we briefly survey many of these models. More limited surveys of some of the non-Gaussian models are found in Lewis (1985), McKenzie (1985a) and Sim (1994). A preliminary version of some of the results in this paper is given in Tedesco (1995).

*Department of Statistics, University of Melbourne, Parkville VIC 3052, Australia.

†Department of Mathematics, Monash University, Clayton VIC 3168, Australia. Please send all correspondence to Dr R.J. Hyndman at this address or by e-mail to Rob.Hyndman@sci.monash.edu.au.

Results concerning moments and correlation structure have often been proven for particular AR(1) models, but in fact under very mild assumptions many properties are a consequence of the general formulation given in Section 2. In Sections 5 and 6 we give some of these general properties. These results are not mathematically difficult, and are very useful in understanding general AR(1) structure and in formulating or selecting models appropriate to given situations.

Section 7 considers the application of these models in data analysis. We discuss parameter estimation, and in particular the issue of selecting the most appropriate model amongst several possible AR(1) models for a given series. 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, 1986) or smoothing (for instance, Altman, 1990 or Hart, 1991). We mention some possible such extensions to non-Gaussian models in Section 8.

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 scope of this paper.

## 2 A general formulation of linear AR(1) models

Let $\{Y_t\}$, $t = 0, 1, \ldots$ be a first order Markov process on sample space $\mathbf{Y} \subseteq \mathbb{R}$ with conditional (transition) density $p(y_t \mid y_{t-1})$. Let $m_t \equiv \mathsf{E}(Y_t \mid Y_{t-1} = y_{t-1}) \in \mathbf{M}$, where $\mathbf{M}$ is the set of values of $m_t$ for which $p$ is a nondegenerate proper probability density. Assume $Y_0$ has the stationary distribution of $Y_t$ when it exists, and otherwise is a given random variable.

We shall consider *linear first order autoregressive (AR(1)) structure* as defined by

$$m_t = \phi y_{t-1} + \lambda \tag{2.1}$$

where $\phi$ and $\lambda$ can take any values such that $m_t \in \mathbf{M}$ for all $y_{t-1} \in \mathbf{Y}$. We refer to such values of $\phi$ and $\lambda$ as *allowable*, and assume throughout this paper that this holds for any model considered. The conditional density $p$ may depend on other parameters besides $\phi$ and $\lambda$, so let $\boldsymbol{\theta}$ be a vector of these parameters ($\boldsymbol{\theta}$ may be null if there are no other parameters). The linear Gaussian AR(1) model is a special case with $p$ a normal density, $\mathbf{Y} = \mathbb{R}$, $\mathbf{M} = \mathbb{R}$, and $\boldsymbol{\theta} = \sigma$.

We take the preceding two paragraphs to define a *linear AR(1) process*. Most linear AR(1) models which have been studied in the literature have this form. Non-linear AR(1) processes, where $m_t$ is a non-linear function of $y_{t-1}$, have also received considerable attention (e.g. Tong, 1990), but will not be considered here.

Alternative definitions of AR(1) structure have been considered (Lewis, 1985). A more restrictive definition than (2.1) is to require the model to satisfy the innovation form

$$Y_t = \phi Y_{t-1} + Z_t$$

where $Z_t$ is an iid sequence with mean $\lambda$. This model is only suitable where $Z_t$ has a continuous sample space. Furthermore, even when $Z_t$ has a continuous sample space, the range of possible models can be limited.

A weaker definition than (2.1) is to require an AR(1) process to have autocorrelation function

$$\rho_k = \mathsf{Corr}[Y_t, Y_{t-k}] = \phi^k, \qquad k = 1, 2, \ldots. \tag{2.2}$$

There are a few processes which have the ACF (2.2) but do not have a linear conditional mean (for example, the minification processes of Tavares (1977, 1980a, 1980b) and Lewis and McKenzie (1991) and the product AR processes of McKenzie (1982)). In this paper, we use the definition (2.1), which better conveys the idea of autoregression—regressing the series on previous values of itself.

## 3 AR(1) model classes

We now consider organizing models in classes with common forms. This helps understand similarities and differences among the various models, often allows general properties to be calculated as in Section 5, and will be useful later in choosing among possible models in practice. We give five classes which subsume almost all of the linear AR(1) models, satisfying (2.1), which have appeared in the literature. The few exceptions which satisfy (2.1) but do not fit in any of these classes are noted in Section 4. Some models may fit into more than one class.

The Gaussian AR(1) process with mean $\mu$ is usually written in terms of a series of white noise innovations $\{E_t\}$:

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + E_t$$

where $E_t \sim \mathrm{N}(0, \sigma^2)$ are iid and $|\phi| < 1$. This can be rewritten in several ways including

$$Y_t = \phi Y_{t-1} + Z_t \qquad \text{where } Z_t \sim \mathrm{N}(\lambda, \sigma^2) \tag{3.1}$$

$$\text{or} \qquad [Y_t \mid Y_{t-1} = y_{t-1}] \sim \mathrm{N}(m_t, \sigma^2), \tag{3.2}$$

with $m_t$ as in (2.1), $\lambda = \mu(1-\phi)$, and $Z_t$ is a sequence of iid innovations. A useful feature of Gaussian AR(1) processes is that the marginal distribution is also normal, hence

$$Y_t \sim \mathrm{N}(\mu, \sigma^2/(1-\phi^2)). \tag{3.3}$$

Various non-Gaussian AR(1) models have been proposed by replacing either the innovation distribution in (3.1) or the conditional distribution in (3.2) by a non-Gaussian distribution. Thus we have the following classes of AR(1) models:

**Innovation class:** $Y_t = \phi Y_{t-1} + Z_t, \qquad Z_t \sim (\lambda, \boldsymbol{\theta})$ (3.4)

**Conditional distribution class:** $[Y_t \mid Y_{t-1} = y_{t-1}] \sim (m_t, \boldsymbol{\theta})$ (3.5)

where $Y \sim (m, \boldsymbol{\theta})$ denotes a random variable with mean $m$ and any other parameters contained in $\boldsymbol{\theta}$. Both classes have linear conditional mean (2.1).

Extensions to the innovations class have been proposed by replacing $\phi Y_{t-1}$ by $X_t$ where $\mathsf{E}[X_t \mid Y_{t-1} = y_{t-1}] = \phi y_{t-1}$. This still retains the property of the linear conditional mean, and allows some more general models to be considered. Three extensions of this kind will be considered here:

**Random coefficient class:** $Y_t = \phi_t Y_{t-1} + Z_t$ (3.6)

**Thinning class:** $Y_t = \phi * Y_{t-1} + Z_t$ (3.7)

**Random coefficient thinning class:** $Y_t = \phi_t * Y_{t-1} + Z_t$ (3.8)

where, in each case, $Z_t \sim (\lambda, \boldsymbol{\theta})$. Here, $\{\phi_t\}$ represents an iid sequence of random coefficients such that $\mathsf{E}\phi_t = \phi$ and $\{\phi_t\}$ is independent of $\{Z_t\}$ unless otherwise stated. The thinning operation denoted by $*$ is defined as

$$\phi * X = \sum_{i=1}^{N(X)} W_i$$

where $N(x)$ is a random variable and $\{W_i\}$ is a sequence of iid random variables, independent of $N(x)$, such that $\mathsf{E}[N(X)W_i \mid X = x] = \phi x$. The most common form of thinning is binomial thinning where $N(X) = X$ and $W_i \sim \text{Bin}(\phi)$, but several other possibilities for $N(X)$ and $W_i$ have been proposed.

There are several common approaches to constructing models in the classes (3.4) – (3.8). One may specify a known distribution for the innovation $Z_t$, and allow marginal $p(y_t)$ and conditional $p(y_t \mid y_{t-1})$ to take whatever form arises (it may not be available analytically). Alternatively, one may give a known marginal for $Y_t$, and attempt to find a distribution for $Z_t$ which gives that marginal (this is not always possible). Or, one may specify the conditional distribution $p(y_t \mid y_{t-1})$ directly. It is rarely possible to have the same marginal and conditional distributions, or even to have marginal, innovation and conditional distributions having relatively simple forms. The Gaussian distribution is a remarkable exception.

## 4 Review of models in literature

Appendix 1 contains tables listing linear AR(1) models which have appeared in the literature. They are listed according to their class and sample space: whole real line, positive real line, non-negative integers, the (0,1) interval and the $(-\pi, \pi)$ interval. This is useful in practice, since given a data series the sample space will be known. A few linear AR(1) models which satisfy (2.1) but which do not fit our class structure are noted below. Some models fit into more than one class. In such cases, we list the model in the class which leads to the simplest form for the distributions involved in the model definition. For instance, any AR(1) model can be written in conditional distribution form by giving $p(y_t \mid y_{t-1})$, but this may be very complicated. For distributions which are invariant (up to a shift in location) when a constant is added, we write them in innovations form.

There is no standard parameterization for these models, and we have reparameterized almost every model so they are of the forms given in (3.4) – (3.8). As will be seen in Section 5, $\mu = \lambda/(1-\phi)$ generally gives the process mean, and this parameter will be used throughout. The common parameterization presented here has several advantages: it enables easier comparisons between models; it emphasizes similarities between models; it allows easier simulation of data from different models with the same mean and/or autocorrelation; it allows parameters of competing models to be directly compared. For most models, some restrictions on the parameters are necessary to obtain well-defined stationary distributions. These restrictions are not specified here.

For each model, a reference is given. In most cases, the reference is to the original paper which proposed the model, but for some models a convenient later reference is given. Distributions are given either by the notation in Appendix 2, or by the probability mass function, or by the density.

Models of the conditional distribution class (3.5) have been considered generally by Zeger and Qaqish (1988), Li (1994) and Shephard (1995), though there remains much to be done in learning about these models. Unlike the other model classes, most individual models within the conditional

distribution class have not been treated separately in the literature, and so are not listed in Appendix 1. Some appear in the examples in Section 7. The model for directional time series is an exception and this is listed in Table 11 of Appendix 1.

### 4.1 Whole real line

For models with sample space consisting of the whole real line, there is no real distinction between the innovation and conditional density classes since distributions defined on the real line are invariant (up to a shift in location) when a constant is added.

Tables 2 – 3 show some of the models that have been studied which have sample space $\mathbb{R}$. For simplicity, we have shown the innovation densities with scale parameters set to one. There is also a large literature on innovation models and random coefficient models with general innovation distributions having mean 0 and variance $\sigma^2$. See Brockwell and Davis (1991) and Nicholls and Quinn (1982) respectively.

### 4.2 Positive real line

Tables 4 – 6 show some of the models defined on the positive real line which have been proposed in the literature. Sim's (1990) GEAR model with exponential marginal distribution is closely related to this class, but is not listed as the innovations are not iid.

### 4.3 Non-negative integers

Tables 7 – 9 show some of the models defined on some or all of the non-negative integers. Many of these are analogues of processes with sample space the positive real line. The Binomial AR(1) process of McKenzie (1985a) and the geometric model of Al-Osh and Aly (1992) are both omitted, having innovations which are not iid. Other models on this sample space which do not fit into any of our classes are the binary AR(1) model of Kanter (1975) and the DAR(1) model of Jacobs and Lewis (1978a,b). Several of the thinning models on this space also appear in the branching processes literature as branching processes with immigration.

### 4.4 (0,1) interval

Table 10 shows two models defined on the unit interval (0,1), one with positive correlation and one with negative correlation.

### 4.5 $(-\pi, \pi)$ interval

Two models for directional time series defined conditionally on the interval $(-\pi, \pi)$ were given by Breckling (1989), and are shown in Table 11.

# 5 Properties of general AR(1) models

In this section we show that under the assumption of AR(1) structure, general results can be obtained for process mean, variance and stationarity properties. These results unify many individual results in the literature, and will be of practical use in model selection in Section 7.2.

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).

## 5.1 Process mean

Under very mild assumptions, the process mean can be easily derived from the linear conditional mean (2.1).

**Proposition 1** *For an AR(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* $t \geq 0$.

**Proof:** 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 $A = \lambda$. □

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

## 5.2 Process variance

In exponential family theory, the class of distributions with quadratic variance function includes the most common models 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 conditional quadratic variance relation includes most of the models in this paper and allows simple and general expressions for process variances, as given in Proposition 2 below.

The conditional distribution $[Y_t \mid Y_{t-1} = y_{t-1}]$ has *quadratic variance relation* if

$$\mathsf{Var}(Y_t \mid Y_{t-1} = y_{t-1}) = am_t^2 + bm_t + c \equiv v(m_t)$$

where, as before, $m_t = \mathsf{E}(Y_t \mid Y_{t-1} = y_{t-1})$ and $a$, $b$ and $c$ are constants possibly depending on $\mu$ or $\boldsymbol{\theta}$ such that $\mathsf{Var}(Y_t \mid Y_{t-1} = y_{t-1}) \geq 0$ for all $y_{t-1} \in \mathbf{Y}$. In particular, note that $a \geq 0$ unless $\mathbf{Y}$ is finite. Values of $a < 0$ do occur, as illustrated in the examples in Section 5.4.

**Proposition 2** *For an AR(1) process, suppose* $[Y_t \mid Y_{t-1} = y_{t-1}]$ *has quadratic variance relation and* $|\phi| < 1$.

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*

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

**Proof:** By the conditional variance formula we have

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

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

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

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

When $|\phi| < 1/|a+1|^{\frac{1}{2}}$ and $a \neq -1$ we will assume $\mathsf{Var}(Y_0) = v(\mu)/(1-(a+1)\phi^2)$ so that $\mathsf{Var}(Y_t) = v(\mu)/(1-(a+1)\phi^2)$ for all $t \geq 1$. Proposition 2 clarifies the close relation between conditional and marginal variances.

For a more general variance relation $\mathsf{Var}(Y_t \mid Y_{t-1} = y_{t-1}) = f(m_t)$, we have the expression

$$\mathsf{Var}(Y_t) = \mathsf{Var}[\mathsf{E}(Y_t \mid Y_{t-1} = y_{t-1})] + \mathsf{E}[\mathsf{Var}(Y_t \mid Y_{t-1} = y_{t-1})] = \phi^2\mathsf{Var}(Y_{t-1}) + \mathsf{E}[f(m_t)]$$

but little more can be said without further assumptions on $f$.

## 5.3 Stationarity

When an AR(1) model is defined to have a specific marginal distribution for $Y_t$, the process is clearly stationary. In other cases when the process is defined by innovations or conditional distributions, the existence or form of a stationary distribution may not be known. We now give conditions under which conditionally linear AR(1) models have an ergodic distribution with moments represented by the limits in Propositions 1 and 2. This situation is usually referred to as stationarity.

The results of Feigin and Tweedie (1985) give the necessary tools for relating marginal to conditional distribution 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). To show the required irreducibility for an AR(1) model, it suffices and is generally easy to show that the support of the conditional density is the entire sample space $\mathbf{Y}$. This holds for all 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 $y_{t-1}$, which is again true for all models in this paper.

The following result gives a sufficient but not necessary condition for existence of an ergodic distribution, in the sense that

$$\|\Pr(Y_t \in A \mid Y_{t-1} = y_{t-1}) - \Pr(Y_t \in A)\| \to 0 \text{ as } t \to \infty \quad \text{for } y_{t-1} \in \mathbf{Y}$$

where $A \subseteq \mathbf{Y}$ and $\|\cdot\|$ denotes total variation. We give a result for the two cases $\mathbf{Y} = \mathbb{R}$ and $\mathbf{Y} \subseteq [0, \infty)$, which include all of the models in this paper. The results do not apply to model g in Table 2, which is defined in terms of Cauchy innovations, or to other models for which the conditional distribution does not have a mean.

**Proposition 3** *Assume $\{Y_t\}$ is a general AR(1) model as defined by (2.1) and also is an irreducible Feller chain. Then consider two cases:*

*Case I: If $\mathbf{Y} = \mathbb{R}$ and $\mathsf{E}[|Y_t| \mid Y_{t-1} = 0] < \infty$ and $|\phi| < 1$ then $\{Y_t\}$ is ergodic.*

*Case II: If $\mathbf{Y} \subseteq \mathbb{R}^+$ and $0 \leq \phi < 1$ then $\{Y_t\}$ is ergodic.*

**Proof:** Using Theorem 1 of Feigin and Tweedie (1985), we take the function $g(y) = |y| + 1$ and must find a $\delta > 0$ and a compact set $A$ such that

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

*Case I:* The model can be stated in innovations form as $Y_t = \phi Y_{t-1} + Z_t$ with $Z_t$ iid, and the assumption is that $\mathsf{E}|Z_t| < \infty$, so

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

Take $k$ so that $|\phi| < k = 1 - \delta < 1$, and then $|\phi y| + \mathsf{E}|Z_t| + 1 \leq kg(y) = k(|y| + 1)$ whenever $|y| \geq \alpha \equiv (\mathsf{E}|Z_t| + 1 - k)/(k - |\phi|)$. So taking $A = [-\alpha, \alpha]$ gives the required condition.

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

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

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

These conditions are sufficient but may not be necessary, and in some cases such as in the conditionally Gamma model in Example 3 in Section 5.4, $\{Y_t\}$ may be ergodic for some $\phi \geq 1$.

It is not necessarily true 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 similar methods to those in Proposition 3, 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.

## 5.4 Examples

The results in Sections 5.1, 5.2 and 5.3 give insight into the process structure, as illustrated in the following examples. Since the conditional variance is typically easier to calculate, Proposition 2 gives an easy way of calculating the process variance even when the marginal distribution is unknown. This is useful in developing model diagnostic and selection methods as discussed in Section 7.

1 If the conditional variance relation is finite and does not depend on $m_{t-1}$ ($a = b = 0$ and $c < \infty$) then $\mathsf{Var}(Y_t) < \infty$ for allowable values of $|\phi| < 1$ and $\mathsf{Var}(Y_t) = c/(1-\phi^2)$. This is the case for Gaussian processes ($c = \sigma^2$ gives the usual result), and also for any model in innovation form.

2 If the conditional variance relation is finite and linear ($a = 0$, $b \neq 0$), then again $\mathsf{Var}(Y_t) < \infty$ for allowable values of $|\phi| < 1$ and $\mathsf{Var}(Y_t) = v(\mu)/(1-\phi^2)$. This is is similar to the previous example but now depends on the process mean $\mu$. An example of such a model is a conditionally Poisson model: $[Y_t \mid Y_{t-1} = y_{t-1}] \sim \mathrm{Pn}(m_t)$, which has $\mathbf{Y} = \{0, 1, \ldots\}$, $\mathbf{M} = (0, \infty)$, $\lambda \geq 0$, $\phi \geq 0$, $\lambda + \phi > 0$ and $\boldsymbol{\theta}$ is null. Then $a = 0$, $b = 1$ and $c = 0$, and

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

This model can also be written as a thinning model or branching process with immigration, by letting $N(x) = x$, $W_i \sim \mathrm{Pn}(\phi)$, and $Z_t \sim \mathrm{Pn}(\lambda)$.

3 If the conditional variance relation is quadratic with $a > 0$, an ergodic distribution exists for allowable values of $|\phi| < 1$, 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 don't know of any such models.) An example of such a model is a conditionally Gamma model: $[Y_t \mid Y_{t-1} = y_{t-1}] \sim \mathrm{G}(r, r/m_t)$. Here, $\mathbf{Y} = (0, \infty)$, $\mathbf{M} = (0, \infty)$, $\lambda \geq 0$, $\phi \geq 0$, $\lambda + \phi > 0$ and $\boldsymbol{\theta} = r$. $\mathsf{Var}(Y_t \mid Y_{t-1} = y_{t-1}) = m_t^2/r$ 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 $r = 1$, $[Y_t \mid Y_{t-1} = y_{t-1}] \sim \mathrm{Exp}(1/m_t)$, $\mathsf{Var}(Y_t \mid Y_{t-1} = y_{t-1}) = m_t^2$ and

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

Grunwald and Feigin (1996) have studied this and similar models, and show that the model is ergodic for some values of $\phi \geq 1$, with a greater region of ergodicity for smaller $r$.

4 Values of $a < 0$ are possible only when $\mathbf{Y}$ is finite. Therefore $\mathsf{Var}(Y_t)$ is also finite and Proposition 2 gives the variance for allowable values of $|\phi| < 1$. An example of such a process is a conditionally Binomial process: $[Y_t \mid Y_{t-1} = y_{t-1}] \sim \mathrm{Bi}(n, p_t)$ with $p_t \equiv m_t/n$, $\mathbf{Y} = \{0, 1, \ldots, n\}$, $\mathbf{M} = (0, n)$, $\mu \in [0, n]$, $0 \leq \phi \leq 1$, $\mu + \phi > 0$, $\mu + \phi < n + 1$ and $\boldsymbol{\theta} = n$. $\mathsf{Var}(Y_t \mid Y_{t-1} = y_{t-1}) = np_t(1-p_t) = -m_t^2/n + m_t$ so $a = -1/n$, $b = 1$ and $c = 0$. Proposition 2 gives

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

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 in the random coefficient class,

$$\mathsf{Var}(Y_t \mid Y_{t-1} = y_{t-1}) = y_{t-1}^2 \mathsf{Var}(\phi_t) + \mathsf{Var}(Z_t) = \frac{\mathsf{Var}(\phi_t)}{\phi^2} m_t^2 - 2\lambda \frac{\mathsf{Var}(\phi_t)}{\phi^2} m_t + \lambda^2 \frac{\mathsf{Var}(\phi_t)}{\phi^2} + \mathsf{Var}(Z_t)$$

so $a = \mathsf{Var}(\phi_t)/\phi^2$ and the condition for finite 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 the innovations class when $\mathsf{Var}(\phi_t) = 0$ (i.e. when $\phi_t = \phi$ is non-random).

6 For models in the thinning class, direct calculation using the conditional variance formula gives

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

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.

## 5.5 Forecasting

For linear AR(1) models with quadratic variance function $v$, $h$-step forecasts and forecast variances are easily derived by the same methods as the recursions for the process mean and variance in Propositions 1 and 2. Let $m_{t+h|t} \equiv \mathsf{E}[Y_{t+h} \mid Y_t = y_t]$ denote the $h$-step forecast and $\mathrm{V}_{t+h|t} \equiv \mathrm{Var}[Y_{t+h} \mid Y_t = y_t]$ denote the forecast variance for $h = 1, 2, \ldots$, using the previous notation $m_{t+1|t} = m_{t+1}$ and $\mathrm{V}_{t+1|t} = v(m_{t+1})$ for $h = 1$.

**Proposition 4** *For a linear AR(1) model with quadratic variance function $v$,*

1 $m_{t+h|t} = \phi m_{t+h-1|t} + \lambda$, *and*
2 $V_{t+h|t} = \phi^2(a+1)\, V_{t+h-1|t} + v(m_{t+h-1|t})$

*for $h = 2, 3, \ldots$.*

These expressions can be expanded into expressions in terms of $y_t$ using the result:
If $x_t$ is defined recursively for $t = 1, 2, \ldots$ by $x_t = ax_{t-1} + b_t$ then $x_{t+h} = a^h x_t + \sum_{i=1}^{h-1} a^i b_{t+h-i}$ for $h = 1, 2, \ldots$.

However, since forecast (conditional) distributions are typically very non-Gaussian, the forecast variances are of little practical use. Prediction regions for non-Gaussian (and non-linear) models are better constructed using highest density regions as described in Hyndman (1995).

# 6 Autocorrelation structure for AR(1) models

Many authors have defined models and proven both linear conditional mean (2.1) and exponentially decaying ACF (2.2). 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 typically (2.1) is much easier to check than (2.2). This and related results clarify the use and interpretation of the ACF as a model diagnostic for general AR(1) structure.

**Proposition 5** *For an AR(1) process with* $|\phi| < 1$ *and* $\mathsf{Var}(Y_t) < \infty$ *constant in time,*

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

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

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 time series literature.

The result is very useful in analyzing data. The sample ACF can be used very generally to determine if AR(1) structure is appropriate, in the same way it is used with Gaussian series. The result also makes it clear that the sample ACF will not help in determining which of several possible AR(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. For $R_k$ the lag $k$ sample autocorrelation, if $Y_1, \ldots Y_n$ are iid with $\mathsf{Var}(Y_t) < \infty$ and constant, then $\mathsf{Var}(R_k) \approx 1/n$. Thus, the usual $\pm 1.96 n^{-\frac{1}{2}}$ 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 (1995). This result can be proven by examining the proof of Bartlett (1946), and is also given in Brockwell and Davis (1991).

# 7 Modelling

## 7.1 Estimation

If the conditional density is known, it is relatively easy to compute maximum likelihood estimates. The likelihood can be calculated using the first-order Markov property of the models:

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

where $p(y_1)$ is the density of $Y_1$. If $p(y_1)$ is unknown, the likelihood conditional on $y_1$ can be calculated by omitting $p(y_1)$ 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 in Tables 4 and 5, 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). Another potential problem with maximum likelihood estimators is the lack of robustness for some of these models. For example, Anděl (1988) gives the maximum likelihood estimator of $\phi$ for model (a) of Table 4 as

$$\hat{\phi} = \min(\phi^\circ, \phi_0)$$

where

$$\phi^\circ = \min_{2\le t\le n} \frac{Y_t}{Y_{t-1}} \quad \text{and} \quad \phi_0 = 1 - \lambda\left(2Y_1 + \sum_{i=2}^{n-1} Y_i\right)^{-1}.$$

Clearly this estimator can break down disastrously with just one outlier.

An alternative, less efficient, but simpler estimation method is to use ordinary least squares regression of $Y_t$ against $Y_{t-1}$. Because the conditional mean given by (2.1) is linear for each of the models considered here, ordinary least squares will give unbiased estimators of $\lambda$ and $\phi$ for all of these models. Although least squares estimators will not generally be the most efficient, they are very easily calculated, are often more robust than maximum likelihood estimators, and avoid the numerical problems associated with discontinuous likelihood functions.

## 7.2 Model selection and diagnostics

Given a data series on a particular sample space, the ACF can be compared with (2.2) to determine if some AR(1) model is appropriate, but there will still typically be several possible AR(1) models on that sample space. By Proposition 5, the ACF cannot be used to select among them. A particular model is often assumed for computational convenience, or standard diagnostics such as residuals are used to show that a given model could be appropriate. Examination of the marginal distribution is also recommended, but even with quantile-quantile 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 AR(1) models with the same form of marginal distribution (see Example 2 below).

To our knowledge there has not been any study of methods for this aspect of model selection, or of the extent to which the various AR(1) models on a given sample space can be distinguished in practice. The problem is challenging because series are often short, distributions non-normal, models are not nested, and all models being considered share some features such as (2.1) and (2.2) so differences may be subtle. In some cases, the physical situation may suggest a model or models, or at least eliminate some models from consideration, and the material in previous sections of this paper is helpful in doing this. In this section we give a discussion and two examples to illustrate these issues, to show how the methods in this paper can be useful, and to encourage further study.

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

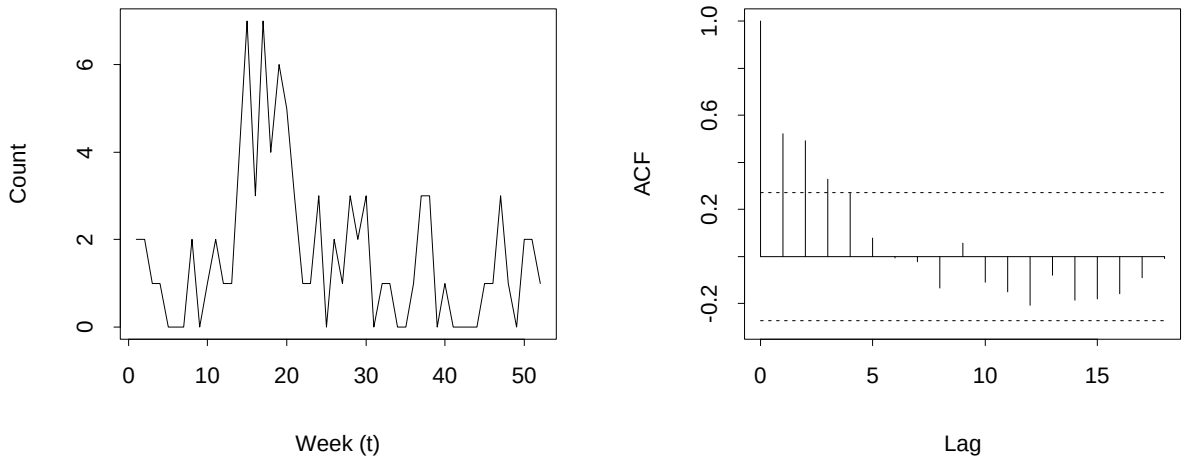


**Figure 1:** *Weekly counts of MCLS in 1982 in Totori-prefecture, Japan, with their sample ACF*

series simulated from the fitted model should share the stochastic properties of the series being modelled. Tsay proposed specifying a particular characteristic or functional, say $\tau$, 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, Hamza and Hyndman (1996) 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) used this method to show that model (a) of Table 6 gave an adequate fit for a given positive series, but in such cases it is unclear which of the other possible models for positive series would also have been adequate.

We illustrate this method and other data-analytic methods in the next section with two real data series and various possible models. In some cases, models are easily distinguishable even with short series and moderate correlation, while in other cases specialized diagnostics based on particular properties of the models under consideration are needed. However, it is clear that different AR(1) models on the same sample space can have very different properties, and selecting a model based on convenience is not sufficient.

## 7.3 Examples

1 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. These authors used a state space model to estimate a postulated underlying smooth disease rate. An alternative analysis, useful for other purposes, is based on the AR(1) models in this paper. Figure 1 shows the series and the sample ACF. These are consistent with (2.2) and AR(1) structure on sample space $\mathbf{Y} = \{0, 1, \ldots\}$. We fitted two models, the INAR(1) model (model (a) of Table 8) of McKenzie (1985a, 1988) and Al-Osh and Alzaid (1987), and the conditionally Poisson model of Example 2, Section 5.4.

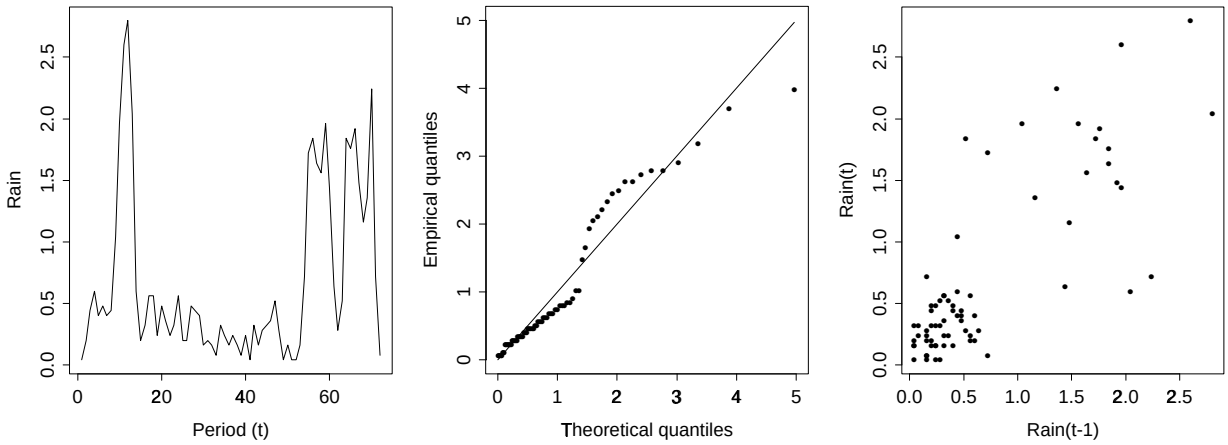


**Figure 2:** *Half-hourly rainfall at river Hirnant at periods 1574–1645, ACF, and exponential QQplot of series.*

Least squares conditional on $y_1$ was used to fit both models, since these models involve only $\phi$ and $\lambda$. The result was $\hat{\phi} = .524$ and $\hat{\lambda} = .802$ (the estimates are the same 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, even though an explicit form for the conditional distribution of latter model is not directly available. So we consider the diagnostic $\tau \equiv \mathsf{Var}(Y_t)/\mathsf{E}(Y_t)$, and estimates of it given by the ratio $\hat{\tau} = s_y^2/\bar{y}$ where $\bar{y}$ 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 is clear that the INAR(1) is not appropriate.

2 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 $\mathbf{Y} = (0, \infty)$. (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, for instance.) Any of the models in Tables 4, 5 and 6 are possible. The left graph in Figure 2 shows the series. The ACF (not shown) is again consistent with AR(1) structure. We consider three models which are based on exponential distributions: the EAR(1) model (Table 4 (c)) of Gaver and Lewis (1980), the thinning model (Table 6 (a)) of Sim (1990) with $\beta = 1/\lambda$ (so having exponential innovation and marginal distributions), and the conditionally Gamma model of Example 3, Section 5.4, with $r = 1$ (so conditionally exponential). 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 straight.
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 1 shows the results of 100 simulations from each of the three models (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 statistic 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 5 shows that the thinning model has conditional variance linear in $m_t$ (and so also in $y_{t-1}$), while the conditional model has conditional variance quadratic in $m_t$ and $y_{t-1}$. This suggested the statistic $\tau$ 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 1 shows the percentile intervals and observed value. The thinning and innovations models now seem unlikely, and the methods in 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 1: *95% intervals for model diagnostics for the Hirnant rainfall series.*

## 8 Conclusions and extensions

We have classified the vast range of non-Gaussian linear AR(1) models into several categories based on model form and sample space. Our aims in doing so have been to highlight similarities and differences between the various models, to derive some more general results for these models than have previously been obtained, and to encourage a more unified view of the area for both research and data analysis. In particular, there is a need for more data-based choice in selecting a model to use. We have given some examples illustrating this need and one approach that might be taken in this area.

Of course, linear AR(1) models provide a limited class with which to model real data. However, they can be used as building blocks for more complex models. For example, covariates can be included by defining a model that satisfies

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

where $\mathbf{x}_t$ denotes a vector of covariates at time $t$. This is particularly easy for innovations or conditional models. Grunwald and Hyndman (1995) show how a smooth mean function can be included by setting

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

where $\mu_t$ represents the smooth mean function. 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 and Shephard, 1995, for instance) have considered using a link function $g(\cdot)$ as in Generalized Linear Models (McCullagh and Nelder, 1989), giving

$$g(m_t) = \lambda + \phi y_{t-1} \, .$$

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. A link function can give a wider range of allowable values of $\phi$ and $\lambda$, 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 $\phi$ 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 $\phi \leq 0$. The same 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. Modifications of this approach include transforming $y_{t-1}$ (Zeger and Qaqish, 1988) or working with linear approximations (Shephard, 1995). Further work in this area, including to what extent data can be used to identify an appropriate link function, is needed.

## 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 helpful discussions.

## Appendix 1: A survey of linear AR(1) models

| | Name | Reference | Innovation distribution | Marginal distribution |
|---|---|---|---|---|
| a | Gaussian | Brockwell & Davis (1991) | $\mathrm{N}(\lambda, 1)$ | $\mathrm{N}\left(\mu, \frac{1}{1-\phi^2}\right)$ |
| b | Uniform | Bell & Smith (1986) | $\mathrm{U}(0, 2\lambda)$ | Unknown |
| c | Logistic | Sim (1993) | $\frac{\sin(\phi\pi)}{2\phi\pi[\cosh(z-\lambda)+\cos(\phi\pi)]}$ $-\infty<z<\infty$ | $\mathrm{Lg}(\mu, 1)$ |
| d | Hyperbolic secant | Rao & Johnson (1988) | $\frac{\cos(\phi\pi/2)\cosh(\pi(z-\lambda)/2)}{\cosh(\pi(z-\lambda))+\cos(\phi\pi)}$ $-\infty<z<\infty$ | $\mathrm{Hs}(\mu, 1)$ |
| e | Laplace | Dewald & Lewis (1985) | $\begin{cases} \lambda & \text{w.p. } \phi^2 \\ \mathrm{La}(\lambda, 1) & \text{w.p. } 1-\phi^2 \end{cases}$ | $\mathrm{La}(\mu,1)$ |
| f | GENTS | Lye & Martin (1994) | $\mathrm{Gt}(\mu, \sigma^2)$ | Unknown |
| g | Cauchy | Brockwell & Davis (1991) | $\mathrm{C}(\lambda, 1)$ | $\mathrm{C}(\mu, \frac{1}{1-\phi})$ |

Table 2: *Innovation models defined on the whole real line.*

| | Reference | Innovation distribution | Marginal distribution |
|---|---|---|---|
| a | Nicholls & Quinn (1982) | $\mathrm{N}(\lambda, 1)$ | Depends on $\phi_t$. |

Table 3: *Random coefficient models defined on the whole real line.*

| | Name | Reference | Innovation distribution | Marginal distribution |
|---|---|---|---|---|
| a | | Bell & Smith (1986)<br>Anděl (1988) | $\text{Exp}(1/\lambda)$ | unknown |
| b | | Hutton (1990) | $\begin{cases} 0 & \text{w.p. } \delta \\ \text{Exp}\left(\frac{1-\delta}{\lambda}\right) & \text{w.p. } 1-\delta \end{cases}$<br>$0 < \delta < 1$ | unknown |
| c | EAR | Gaver & Lewis (1980) | $\begin{cases} 0 & \text{w.p. } \phi \\ \text{Exp}(1/\mu) & \text{w.p. } 1-\phi \end{cases}$ | $\text{Exp}(1/\mu)$ |
| d | GAR | Gaver & Lewis (1980)<br>Lawrance (1982) | $\begin{cases} 0 & \text{w.p. } \phi^k \\ \sum_{i=1}^{N} \phi^{U_i} W_i & \text{w.p. } 1-\phi^k \end{cases}$<br>$U_i \sim U(0,1)$, $N \sim \text{Pn}(-\mu\alpha \log \phi)$<br>$W_i \sim \text{Exp}(\alpha)$ | $\text{G}(\mu\alpha, \alpha)$ |

Table 4: *Innovation models defined on the positive real line.*

| | Name | Reference | Coefficient $\phi_t$ distribution | Innovation distribution | Marginal distribution |
|---|---|---|---|---|---|
| a | TEAR | Lawrance & Lewis (1981) | $\begin{cases} 1 & \text{w.p. } \phi \\ 0 & \text{w.p. } 1-\phi \end{cases}$ | $\text{Exp}(1/\lambda)$ | $\text{Exp}(1/\mu)$ |
| b | NEAR | Lawrance & Lewis (1981) | $\begin{cases} \beta & \text{w.p. } \frac{\phi}{\beta} \\ 0 & \text{w.p. } 1-\frac{\phi}{\beta} \end{cases}$ | $\begin{cases} \text{Exp}(\frac{1}{\mu}) & \text{w.p. } \frac{1-\beta}{1-\beta+\phi} \\ \text{Exp}(\frac{1}{\mu(\beta-\phi)}) & \text{w.p. } \frac{\phi}{1-\beta+\phi} \end{cases}$ | $\text{Exp}(1/\mu)$ |
| c | | Sim (1986) | $\frac{\phi}{1-\phi} v^{(2\phi-1)/(1-\phi)}$<br>$0<v<1$ | $\text{Exp}(1/\lambda)$ | $\text{G}(\frac{1}{1-\phi}, \frac{1}{\lambda})$ |
| d | BGAR | Lewis, McKenzie & Hugus (1989) | $\text{Be}(k\phi, k(1-\phi))$ | $\text{G}(k(1-\phi), k/\mu)$ | $\text{G}(k, k/\mu)$ |

Table 5: *Random coefficient models defined on the positive real line.*

| | Reference | $N(x)$ | $W_i$ | Innovation distribution | Marginal distribution |
|---|---|---|---|---|---|
| a | Sim (1990) | $\text{Pn}(\phi\beta x)$ | $\text{Exp}(\beta)$ | $\text{G}(\beta\lambda, \beta)$ | $\text{G}(\beta\lambda, \beta(1-\phi))$ |

Table 6: *Thinning models defined on the positive real line.*

| | Reference | Coefficient $\phi_t$ distribution | Innovation distribution | Marginal distribution |
|---|---|---|---|---|
| a | McKenzie (1985a) | $A_t - B_t$ where $A_t \sim \text{Bin}(\lambda + \phi)$ and $B_t \sim \text{Bin}(\lambda)$ | $B_t$ | $\text{Bin}(\mu)$ |

Table 7: *Random coefficient models defined on the non-negative integers.*

| | Name | Reference | $N(x)$ | $W_i$ | Innovation distribution | Marginal distribution |
|---|---|---|---|---|---|---|
| a | INAR | McKenzie (1985a, 1988) Al-Osh & Alzaid (1987) | $x$ | $\text{Bin}(\phi)$ | $\text{Pn}(\lambda)$ | $\text{Pn}(\mu)$ |
| b | SINAR | Franke & Seligmann (1993) | $x$ | $\text{Bin}(\phi)$ | $Q_t A_t + (1-Q_t)B_t$ $Q_t \sim \text{Bin}\left(\frac{\lambda-\lambda_2}{\lambda_1-\lambda_2}\right)$ $A_t \sim \text{Pn}(\lambda_1)$, $B_t \sim \text{Pn}(\lambda_2)$ | Unknown |
| c | | Al-Osh & Aly (1992) | $\text{Bi}(x, \beta\phi)$ | $\text{Ge}(\frac{\beta}{1+\beta})$ | $\text{NB}\left(\lambda\beta, \frac{\beta}{1+\beta}\right)$ | $\text{NB}\left(\lambda\beta, \frac{\beta(1-\phi)}{1+\beta(1-\phi)}\right)$ |
| d | | McKenzie (1985a) | $x$ | $\text{Bin}(\phi)$ | $\sum_{i=1}^{N} \phi^{U_i} * G_i$ $U_i \sim U(0,1)$, $N \sim \text{Pn}(-\beta \log \phi)$ $G_i \sim \text{Ge}(\beta/(\mu+\beta))$ | $\text{NB}(\beta, \frac{\beta}{\mu+\beta})$ |
| e | GINAR | McKenzie (1985a) Alzaid & Al-Osh (1988) | $x$ | $\text{Bin}(\phi)$ | $\begin{cases} G_t & \text{w.p. } 1-\phi \\ 0 & \text{w.p. } \phi \end{cases}$ $G_t \sim \text{Ge}(1/(1+\mu))$ | $\text{Ge}(1/(1+\mu))$ |

Table 8: *Thinning models defined on the non-negative integers.*

| | Reference | Coefficient $\phi_t$ distribution | $N(x)$ | $W_i$ | Innovation distribution | Marginal distribution |
|---|---|---|---|---|---|---|
| a | McKenzie (1985a) | $\text{Be}(\phi, 1-\phi)$ | $x$ | $\text{Bin}(\phi)$ | $\text{NB}\left(\frac{\beta(1-\phi)}{\phi}, \frac{\beta}{\phi\mu+\beta}\right)$ | $\text{NB}\left(\frac{\beta}{\phi}, \frac{\beta}{\phi\mu+\beta}\right)$ |
| b | McKenzie (1985a) | $\begin{cases} \beta & \text{w.p. } \frac{\phi}{\beta} \\ 0 & \text{w.p. } 1-\frac{\phi}{\beta} \end{cases}$ | $x$ | $\text{Bin}(\phi)$ | $\{1-(1-\beta+\phi)V_t\} * G_t$<br>$V_t \sim \text{Bin}\left(\frac{\phi}{1-\beta+\phi}\right)$<br>$G_t \sim \text{Ge}\left(\frac{1}{1+\mu}\right)$ | $\text{Ge}\left(\frac{1}{1+\mu}\right)$ |

Table 9: *Random coefficient thinning models defined on the non-negative integers.*

| | Name | Reference | Coefficient $\phi_t$ distribution | Innovation distribution | Marginal distribution |
|---|---|---|---|---|---|
| a | PBAR | McKenzie (1985b) | $A_tB_t$ where $A_t \sim \text{Be}\left(\beta, \frac{\lambda\beta}{1-\lambda}\right)$<br>$B_t \sim \text{Be}\left(\frac{\lambda\beta\phi}{(1-\lambda-\phi)(1-\lambda)}, \frac{\lambda\beta}{1-\lambda}\right)$<br>and $\phi > 0$. | $1-A_t$ | $\text{Be}\left(\frac{\lambda\beta}{1-\lambda-\phi}, \beta\right)$ |
| b | NBAR | McKenzie (1985b) | $-A_tB_t$ where $A_t \sim \text{Be}\left(\frac{\lambda\beta}{1-\lambda-\phi}, \frac{\beta(1-\lambda)}{1-\lambda-\phi}\right)$<br>$B_t \sim \text{Be}\left(\frac{-\beta\phi}{(1-\lambda-\phi)}, \frac{\beta(\lambda+\phi)}{1-\lambda-\phi}\right)$<br>and $\phi < 0$. | $A_t$ | $\text{Be}\left(\frac{\lambda\beta}{1-\lambda-\phi}, \beta\right)$ |

Table 10: *Random coefficient models defined on (0,1).*

| | Name | Reference | Conditional distribution |
|---|---|---|---|
| a | von Mises | Breckling (1989) | $[2\pi I_0(v_t)]^{-1} \exp\{\kappa_1 \cos(y - y_{t-1}) + \kappa_0 \cos Y_t\}$.<br>where $\kappa_0$ and $\kappa_1$ are functions of $\phi$, $\lambda$ and $\sigma$, and $v_t = \|\boldsymbol{v}_t\|$ denotes the length of the concentration vector $\boldsymbol{v}_t = \kappa_1(\cos Y_{t-1}, \sin Y_{t-1})^T + (\kappa_0, 0)^T$. Here $I_0(\kappa)$ is the modified Bessel function of the first kind and order 0. |
| b | Wrapped Normal | Breckling (1989) | $\text{WN}(\phi y_{t-1} + \lambda, \sigma^2)$ |

Table 11: *Conditional distribution models defined on $(-\pi, \pi)$.*

## Appendix 2: Notation for distributions

| | | |
|---|---|---|
| $\mathrm{N}(\mu, \sigma^2)$ | Normal with mean $\mu$ and variance $\sigma^2$ | $\mu_1 = \mu$ |
| | $p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}(x-\mu)^2/\sigma^2), \quad -\infty < x < \infty$ | $\mu_2 = \sigma^2$ |
| $\mathrm{Gt}(\mu, \sigma^2)$ | Generalized $t$ with mean $\mu$ and variance $\sigma^2$ | $\mu_1 = \mu$ |
| | $p(x) = \exp\{\theta_1 \tan^{-1}(\frac{x}{\gamma}) + \theta_2 \log(\gamma^2 + x^2) + \sum_{i=3}^{M} \theta_i x^{i-2} - \eta\}, \quad -\infty < x < \infty$ | $\mu_2 = \sigma^2$ |
| $\mathrm{La}(\mu, \alpha)$ | Laplace mean $\mu$ and scale $\alpha$ | $\mu_1 = \mu$ |
| | $p(x) = \frac{1}{2}\alpha \exp(-\alpha\lvert x-\mu\rvert), \quad -\infty < x < \infty$ | $\mu_2 = 2/\alpha^2$ |
| $\mathrm{C}(\mu, a)$ | Cauchy location $\mu$, scale $a$ | |
| | $p(x) = \frac{1}{\pi}\frac{a}{a^2+(x-\mu)^2}, \quad -\infty < x < \infty$ | |
| $\mathrm{Lg}(\alpha, \beta)$ | Logistic mean $\alpha$, scale $\beta$ | $\mu_1 = \alpha$ |
| | $p(x) = \frac{1}{4\beta}\mathrm{sech}^2\left(\frac{x-\alpha}{2\beta}\right), \quad -\infty < x < \infty$ | $\mu_2 = \frac{4}{3}\pi^2\beta^2$ |
| $\mathrm{Hs}(\alpha, \beta)$ | Hyperbolic secant mean $\alpha$, scale $\beta$ | $\mu_1 = \alpha$ |
| | $p(x) = \frac{1}{2\beta}\mathrm{sech}\left(\frac{\pi(x-\alpha)}{2\beta}\right), \quad -\infty < x < \infty$ | $\mu_2 = \beta^2$ |
| $\mathrm{WN}(\mu, \sigma^2)$ | Wrapped normal mean $\mu$ and concentration $\sigma^2$ | $\mu_1 = \mu$ |
| | $p(x) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu-2\pi k)^2\right\}, -\pi \le x < \pi$ | |
| $\mathrm{Exp}(\alpha)$ | Exponential rate $\alpha$ | $\mu_1 = 1/\alpha$ |
| | $p(x) = \alpha \exp(-\alpha x), \quad x > 0$ | $\mu_2 = 1/\alpha^2$ |
| $\mathrm{G}(r, \alpha)$ | Gamma shape $r$, rate $\alpha$ | $\mu_1 = r/\alpha$ |
| | $p(x) = \frac{\alpha^r e^{-\alpha x} x^{r-1}}{\Gamma(r)}, \quad x > 0$ | $\mu_2 = r/\alpha^2$ |
| $\mathrm{Be}(m, n)$ | Beta parameters $m$ and $n$ | $\mu_1 = \frac{m}{m+n}$ |
| | $p(x) = \frac{x^{m-1}(1-x)^{n-1}\Gamma(m+n)}{\Gamma(m)\Gamma(n)}, \quad 0 < x < 1$ | $\mu_2 = \frac{mn}{(m+n)^2(m+n+1)}$ |
| $\mathrm{U}(a, b)$ | Uniform range $[a, b]$ | $\mu_1 = \frac{1}{2}(a+b)$ |
| | $p(x) = \frac{1}{b-a}, \quad a < x < b$ | $\mu_2 = \frac{1}{12}(b-a)^2$ |
| $\mathrm{Bin}(p)$ | Binary probability $p$ | $\mu_1 = p$ |
| | $\mathsf{Pr}(X = x) = xp + (1-x)(1-p), \quad x = 0, 1$ | $\mu_2 = p(1-p)$ |
| $\mathrm{Bi}(n, p)$ | Binomial sample size $n$, probability $p$ | $\mu_1 = np$ |
| | $\mathsf{Pr}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n$ | $\mu_2 = np(1-p)$ |
| $\mathrm{Pn}(\mu)$ | Poisson rate $\mu$ | $\mu_1 = \mu$ |
| | $\mathsf{Pr}(X = x) = \frac{e^{-\mu}\mu^x}{x!}, \quad x = 0, 1, 2, \ldots$ | $\mu_2 = \mu$ |
| $\mathrm{Ge}(p)$ | Geometric probability $p$ | $\mu_1 = \frac{1-p}{p}$ |
| | $\mathsf{Pr}(X = x) = p(1-p)^x, \quad x = 0, 1, 2, \ldots$ | $\mu_2 = \frac{1-p}{p^2}$ |
| $\mathrm{NB}(r, p)$ | Negative binomial parameters $p$ and $r$ | $\mu_1 = \frac{r(1-p)}{p}$ |
| | $\mathsf{Pr}(X = x) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \ldots$ | $\mu_2 = \frac{r(1-p)}{p^2}$ |

$p(x)$ denotes the probability density function, $\mu_1 = \mathsf{E}(X)$ and $\mu_2 = \mathsf{Var}(X)$.

## 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, Jens (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.

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. and HYNDMAN, R.J. (1995) Smoothing non-Gaussian time series with autoregressive structure, Research Report 18 (revised), Department of Statistics, University of Melbourne.

GRUNWALD, G.K., HAMZA, K. and HYNDMAN, R.J. (1996) Some properties and generalizations of Bayesian time series models, *J. Roy. Stat. Soc. B*, to appear.

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.

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

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

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. (1980) Non-negative time series models for dry riverflow, *J. Appl. Prob.*, **27**, 171–182.

HYNDMAN, R.J. (1995) Highest density forecast regions for non-linear and non-normal time series models, *J. Forecasting*, **14**, 431–441.

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

JUDGE, G.G., HILL, R.C., GRIFFITHS, W.E., LÜTKEPOHL, H. and LEE, T.-C. (1982) *Introduction to the theory and practice of econometrics*, John Wiley & Sons: New York.

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

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

KITAGAWA, G. (1987) Non-Gaussian state-space modeling of nonstationary time series, *J. Amer. Statist. 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 au-

toregressive time series in exponential variables (with discussion), *J. Roy. Stat. Soc. B*, **47**, 165–202.

Lewis, P.A.W. (1980) Simple models for positive-valued and discrete-valued time series with ARMA correlation structure, In *Multivariate analysis V*, ed. P.R. Krishnaiah, pp.151–166, North Holland.

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.

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.

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. Roy. 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.

Tedesco, L.M. (1995) *Non-Gaussian first order autoregressive time series models*, M.Sc. thesis, University of Melbourne.

Tong, H. (1990) *Non-linear time series: a dynamical system approach*, Oxford University Press: Oxford.

Tsay, R.S. (1992) Model checking via parametric bootstraps in time series analysis, *J. R. Statist. Soc. A*, **155**, 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. Amer. 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.

## Résumé

Nous passons en revue et faisons une synthese de lénsemble des modeles lineaires autoregressifs non-Gaussiens du premier ordre qui apparaissent dans la litterature. Les modeles sont classes pour mettre en valeur les similarites et les differences existant entre eux afin de faciliter leur utilisation dans telle ou telle situation particuliere. Les proprietes generales de la moyenne, variance et correlation des processus sont deduites, unifiant ainsi de nombreux resultats obtenus separement dans la litterature. Nous presentons des exemples illustrant létendue des proprietes qui peuvent apparaitre meme dans lh́ypothese autoregressive. Ces resultats sont utilises dans lánalyse dúne serie dénsembles de donnees reelles, illustrant les methodes generales déstimation de diagnostic et de selection des modeles.