 Vol 29: 657–663, 2002

Constructing First Order Stationary Autoregressive Models via Latent Processes

MICHAEL K. PITT

University of Warwick

CHRIS CHATFIELD and STEPHEN G. WALKER

University of Bath

ABSTRACT. First order stationary autoregressive (AR(1)) models are introduced for which there exists a linear relation between the expectations of the observations, and where it is readily possible to arrange the marginal distributions to be other than normal.

Key words: autocorrelation function, autoregressive process, EM algorithm, exponential family, latent process, stationary time series

1. Introduction

The usual zero-mean AR(1) model can be written in the form

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \tag{1}$$

where the $\{\varepsilon_t\}$ are i.i.d innovation variables with zero mean and constant variance σ^2 such that ε_t is independent of $\{Y_1, \ldots, Y_{t-1}\}$, and $-1 < \rho < 1$ for stationarity. It is easy to generalize (1) to have non-zero mean μ by rewriting it in the form

$$(Y_t - \mu) = \rho(Y_{t-1} - \mu) + \varepsilon_t, \tag{2}$$

or by arranging the $\{\varepsilon_t\}$ to have mean $(1-\rho)\mu$ in (1).

The recursive nature of (1) makes it convenient to use in many applications, especially when ε_t, and hence Y_t, has a normal distribution. However, it is a difficult task to establish the distribution of ε_t so as to ensure that $\{Y_t\}$ is strictly stationary with a particular non-normal density, say $f_Y(y)$. The first to tackle such problems were Lawrance & Lewis (1977) and Jacobs & Lewis (1977) who restricted their attention to exponential marginal distributions. Gaver & Lewis (1980) considered processes with gamma marginals. However, even in the apparently straightforward case when Y_t is required to have a gamma density, the distribution of ε_t is complicated (Lawrance, 1982; Walker, 2000). Moreover, it is found that $\mathrm{pr}\{\varepsilon_t = 0\} > 0$ which renders the model deficient in the sense that the experimenter would usually not realistically desire $\mathrm{pr}\{Y_t = \rho Y_{t-1}\}$ to be greater than zero.

Recent advances have been made by Joe (1996). Joe (1996) produced a means by which one can generate stationary first-order processes with members of the infinitely divisible (i.d.) convolution-closed exponential family as marginal distributions.

The aim of this paper is to introduce an alternative class of models, which are first-order Markov. In general this can be done by specifying the joint distribution of (Y_t, Y_{t-1}), or equivalently in the stationary case by specifying the marginal distribution of Y_t and the conditional (or transition) density of $(Y_t|Y_{t-1})$. By a partial analogy with (2), we aim to do this by requiring that there is a linear relation with respect to the mean, i.e.

$$\mathrm{E}(Y_t|Y_{t-1}) = \rho Y_{t-1} + (1-\rho)\mu, \tag{3}$$

where $\mu = \int y f_Y(y)\, dy$ and $f_Y(y)$ is the required stationary density of Y_t. As for the standard AR(1) model, it can be shown straightforwardly that the autocorrelation function must be of the form ρ^r if (3) holds. We demonstrate that it is usually possible to define a transition density $f(y_t|y_{t-1})$, temporarily expressed here as $f_{Y|Z}(y|z)$ (i.e. $Y = Y_t$ and $Z = Y_{t-1}$), for which the above properties hold, i.e.

$$f_Y(y) = \int f_{Y|Z}(y|z) f_Y(z)\, dz \tag{4}$$

and

$$\int y f_{Y|Z}(y|z)\, dy = \rho z + (1-\rho)\mu. \tag{5}$$

We achieve this by introducing a third (latent) variable for mathematical convenience, say x, and considering transition densities of the form

$$f_{Y|Z}(y|z) = \int f_1(y|x) f_2(x|z)\, d\lambda(x)$$

where $d\lambda$ represents either the counting measure (x discrete) or the Lebesgue measure (x continuous).

The key point is, if we construct a joint density $f_{Y,X}(y,x)$ such that $f_1(y|x) = f_{Y|X}(y|x)$ and $f_2(x|z) = f_{X|Y}(x|z)$ are the two conditional densities and $f_Y(y) = \int f_{Y,X}(y,x)\, d\lambda(x)$ then $f_Y(y) = \int f_{Y|Z}(y|z) f_Y(z)\, dz$ follows trivially. Hence, to ensure that (4) holds, we only need to introduce a joint density $f_{Y,X}(y,x)$ with marginal $f_Y(y)$. Note that the transition density can be sampled via simulating a latent process $\{X_t\}$, i.e. $Y_{t+1}|X_t \sim f_{Y|X}(\cdot|X_t)$ and $X_t|Y_t \sim f_{X|Y}(\cdot|Y_t)$. Since the two densities sampled are the two conditionals from a joint density, there is an obvious connection with the Gibbs sampler (Smith & Roberts, 1993).

In order to achieve the linear relation between expectations in (5), more work is required. This will need careful choice of $f_{X|Y}$, ensuring that $f_{Y|X}$ has a simple form, in fact the same form as f_Y, and that $\mathrm{E}(X|Y) \propto Y$.

Example 1. As an illustration, consider the case when $f_Y(y)$ is required to have a gamma density with shape parameter $a > 0$ and scale parameter 1, i.e. $f_Y(y) \propto y^{a-1}\exp(-y)$. We write this distribution as $\mathrm{ga}(y|a,1)$ and note that it has mean a. Let $f_{X|Y}(x|y)$ be a Poisson probability mass function with mean $y\phi$ with $\phi \geqslant 0$. Marginally $Y_t \sim \mathrm{ga}(a,1)$; if $Y_{t+1}|[X_t = x] \sim \mathrm{ga}(a+x, 1+\phi)$ and $X_t|[Y_t = y] \sim \mathrm{po}(y\phi)$, then marginally X_t is negative-binomial, i.e. a Poisson–gamma mixture. Then $\mathrm{E}(Y_t|Y_{t-1})$ may be derived via the intermediate calculation

$$\mathrm{E}[\mathrm{E}(Y_t|X_{t-1})|Y_{t-1}].$$

Of course, if $\phi = 0$, then $X = 0$ with probability 1, and $\{Y_t\}$ is a i.i.d. sequence. As $\phi \to \infty$, the autocorrelation function of $\{Y_t\}$ converges to 1.

Since $f_{Y|X}(y|x)$ is $\mathrm{ga}(y|a+x, 1+\phi)$ we obtain the required linear relation with respect to the mean, namely

$$\mathrm{E}(Y_t|Y_{t-1}) = \frac{a + \phi Y_{t-1}}{1+\phi}.$$

Comparing this with (3), we see that $\rho = \phi/(1+\phi)$ and $\mu = a$. Note that ρ is restricted to the range $(0,1)$, though this is not a serious restriction in practice. The autocorrelation function is given by $\phi^r/(1+\phi)^r$. Note that marginal aspects of the series are carried in $f_Y(y)$ only and that the dependence properties of the sequence are carried solely by $f_{X|Y}(x|y)$.

In this example, the density function $f_{Y_t|Y_{t-1}}(y_t|y_{t-1})$ is a mixture of gamma densities with Poisson weights:

$$f_{Y_t|Y_{t-1}}(y_t|y_{t-1}) = \sum_{x=0}^{\infty} w(y_{t-1}, x)\mathrm{ga}(y_t|a+x, 1+\phi),$$

where

$$w(y, x) = \frac{(y\phi)^x}{x!}\exp(-y\phi).$$

The conditional variance is given by

$$\mathrm{var}(Y_t|Y_{t-1}) = \frac{\mu + 2\phi Y_{t-1}}{(1+\phi)^2}. \tag{6}$$

More generally, the above example suggests that, in order to be able to achieve the linear relation in the mean, we need to specify $f_{Y,X}(y,x)$ in a particular way. This boils down to specifying $f_{X|Y}(x|y)$ and from the above gamma illustration it appears that this should be done to ensure that the $f_{Y|X}(y|x) \propto f_{X|Y}(x|y)f_Y(y)$ belongs to the same family as $f_Y(y)$.

A referee has drawn our attention to Barndorff-Nielsen (1997) who generates a Markov chain $\{Y_t\}$ via a sequence of unknown latent states $\{X_t\}$ and the transition densities $f(y_t|x_t)$ and $f(x_t|y_{t-1})$. The former of these is a normal distribution and the latter a generalized inverse Gaussian distribution, not arising from a single joint density. Consequently, in this case, the marginal and stationary distribution of the $\{Y_t\}$ are not known.

2. General models

We would like to be able to understand how to construct a suitable $f_{X|Y}(x|y)$ given $f_Y(y)$. In section 2.1 we do this when f_Y is a member of the i.d. convolution-closed exponential family and in section 2.2 for a new family of density functions which includes the beta, inverse gamma and normal, among others.

2.1. ID convolution-closed exponential family

If $f_Y(y)$ is a member of the i.d. convolution-closed exponential family then the density function takes the form

$$f(y; \theta, \tau) = c(y; \tau)\exp\{y\theta - \tau M(\theta)\},$$

where $\tau > 0$. Recently, Joe (1996) developed and constructed an AR(1) model whose marginal distribution is an i.d. convolution-closed exponential model (see also Jørgensen, 1986, 1987; Jørgensen & Song, 1998). Here we present the work of Joe (1996) using our ideas and then extend to a new family of densities.

Consider

$$f_{X|Y}(x|y, \tau_1, \tau_2) = \frac{c(x; \tau_1)c(y-x; \tau_2)}{c(y; \tau)},$$

where $\tau_1 = \rho\tau$ and $\tau_2 = (1-\rho)\tau$ with $0 < \rho < 1$. Note that

$$f_X(x) = c(x; \tau_1)\exp\{x\theta - \tau_1 M(\theta)\}$$

and so $\mathrm{E}(X) = \rho\tau m(\theta)$, where $m = M'$. Since $f_Y(y)$ is convolution-closed, i.e.

$$\int f(x; \theta, \tau_1)f(y-x; \theta, \tau_2)\, d\lambda(x) = f(y; \theta, \tau_1 + \tau_2)$$

then $f_{X|Y}(x|y, \tau_1, \tau_2)$ is a density and $\mathrm{E}(X|Y) = \rho Y$. That $f_{X|Y}(x|y, \tau_1, \tau_2)$ is a density follows from the fact that

$$\frac{c(x; \tau_1)c(y - x; \tau_2)}{c(y; \tau)} = \frac{f(x; \theta, \tau_1)f(y - x; \theta, \tau_2)}{f(y; \theta, \tau)}.$$

Now suppose $\mathrm{E}(X|Y) = \alpha(Y)$ so for all θ,

$$\mathrm{E}_\theta\{\alpha(Y) - \rho Y\} = \int \{\alpha(y) - \rho y\} f(y; \theta, \tau)\, dy \equiv 0,$$

since $\mathrm{E}(X) = \int \alpha(y) f(y; \theta, \tau)\, dy = \rho\tau m(\theta)$. From completeness arguments, it follows that $\alpha(y) = \rho y$. The conditional density $f_{Y|X}(y|x)$ is given by

$$f_{Y|X}(y|x) \propto c(y - x; \tau_2) \exp\{(y - x)\theta\},$$

meaning that conditional on $X = x$, $Y = x + U$ where the density of U is

$$f_U(u) = c(u; \tau_2) \exp\{u\theta - \tau_2 M(\theta)\}$$

so that $\mathrm{E}(Y|X) = X + \tau_2 m(\theta)$. Putting these together, we have

$$\mathrm{E}(Y_t|Y_{t-1}) = \mathrm{E}(X|Y_{t-1}) + \tau_2 m(\theta) = \rho Y_{t-1} + (1 - \rho)\mathrm{E}(Y)$$

as required.

Example 2. Let $\mathrm{po}(y|\theta)$ denote a Poisson density with mean θ. Consider the joint density

$$f_{Y,X}(y, x) = \mathrm{po}(y|\theta)\, \mathrm{bn}(x|y, \rho),$$

where bn denotes the binomial distribution. The conditional distribution of X given $Y = y$ is $\mathrm{bn}(\cdot|y, \rho)$ and for Y given $X = x$ is, by construction, $Y = x + \Pi(\theta\bar{\rho})$, where $\bar{\rho} = 1 - \rho$, and $\Pi(\theta\bar{\rho})$ is a Poisson random variable with mean $\theta\bar{\rho}$. Consequently, we can construct a stationary first-order process with Poisson marginals via

$$Y_{t+1} = \rho * Y_t + \Pi_t,$$

where $\{\Pi_t\}$ is an i.i.d. sequence of $\mathrm{po}(\theta\bar{\rho})$ random variables and $\rho * Y$ is a binomial random variable with parameters $0 < \rho < 1$ and Y.

This process is known to the applied probability literature, in fact it is a simple birth with immigration process. The Markov chain in example 1 is also known (see Feller, 1971). However, it is a hard task to construct Markov chains with known stationary distribution. We provide a general construction of stationary first-order processes with known marginal distributions. Moreover, our constructions lead to a means by which to make inference about the relevant unknowns since we have a complete data likelihood, based on treating the $\{X_t\}$ as missing data (see section 3).

2.2. *A new family of densities*

We start this section with an illustration.

Example 3. Suppose $f_Y(y)$ is a beta density, denoted by $\mathrm{be}(\alpha, \beta)$, and not a member of the convolution-closed exponential family. Take $f_{X|Y}(x|y)$ to be binomial with parameters r and y, ensuring that $f_{Y|X}(y|x) = \mathrm{be}(\alpha + x, \beta + r - x)$. Then

$$\mathrm{E}(Y_t|Y_{t-1}) = \mathrm{E}\{(X + \alpha)/(r + \alpha + \beta)|Y_{t-1}\} = (rY_{t-1} + \alpha)/(r + \alpha + \beta)$$

and hence

$$\mathrm{E}(Y_t|Y_{t-1}) = \rho Y_{t-1} + (1 - \rho)\mu,$$

where $\mu = \mathrm{E}(Y) = \alpha/(\alpha + \beta)$ and $\rho = r/(r + \alpha + \beta)$.

This beta model can be put into a general framework. Suppose

$$f_Y(y) = c\exp\{ag(y) - bh(y)\}g'(y),$$

where $h'/g' = y$ and a, b and c are constants. The beta case, with parameters $(a, b-a)$, arises when $g(y) = \log\{y/(1-y)\}$ and $h(y) = -\log(1-y)$, so that $h'/g' = y$.

We can show that $\mathrm{E}(Y) = a/b$ provided $f_Y(y)/g'(y) \to 0$ as y tends to the limits of the support. This follows since $yg'(y) = h'(y)$ and so

$$\mathrm{E}(Y) = c\int h'(y)\exp\{ag(y) - bh(y)\}\,dy.$$

The result follows via integration by parts.

We now introduce the conditional density

$$f_{X|Y}(x|y) = d(x,\lambda)\exp\{xg(y) - \lambda h(y)\},$$

which is a member of the exponential family. Consequently, assuming the regularity conditions hold for allowing differentiation under the integral sign;

$$0 \equiv \frac{\partial}{\partial y}\int f_{X|Y}(x|y)\,dx = \int \frac{\partial}{\partial y} f_{X|Y}(x|y)\,dx,$$

the standard result

$$\mathrm{E}(X|Y=y) = \lambda h'(y)/g'(y) = \lambda y$$

follows. Note that d is not necessarily unique but can be chosen so that $f_{X|Y}$ is familiar. Then

$$f_{Y|X}(y|x) \propto \exp\{(a+x)g(y) - (b+\lambda)h(y)\}g'(y),$$

and hence

$$\mathrm{E}(Y|X) = \frac{a+X}{b+\lambda},$$

which all leads to

$$\mathrm{E}(Y_t|Y_{t-1}) = \frac{a+\lambda Y_{t-1}}{b+\lambda} = \rho Y_{t-1} + (1-\rho)\mu,$$

where $\rho = \lambda/(b+\lambda)$ and $\mu = a/b$.

Example 4. Here we note that the inverse-gamma density belongs to this family of densities, i.e. $Y \sim \mathrm{iga}(a,1)$ with density

$$f_Y(y) \propto y^{-a-1}\exp(-1/y) = \exp\{-1/y - (a-1)\log y\}y^{-2}.$$

It is seen that $g(y) = -1/y$, $g'(y) = 1/y^2$, $h(y) = \log y$ and $h'(y) = 1/y$. Then $f_{X|Y}(x|y) \propto c(x)\exp(-x/y)$, indicating that the appropriate form for $f_{X|Y}(x|y)$ is $\mathrm{ga}(\lambda, 1/y)$. Then $f_{Y|X}(y|x)$ is inverse gamma, i.e. $\mathrm{iga}(a+\lambda, 1+x)$, so

$$\mathrm{E}(Y|X) = \frac{1+X}{a+\lambda-1}$$

leading to

$$\mathrm{E}(Y_t|Y_{t-1}) = \rho Y_t + (1-\rho)\mu,$$

where $\rho = \lambda/(\lambda+a-1)$ and $\mu = 1/(a-1)$.

Note also that we obtain the gamma family when $g(y) = \log y$ and $h(y) = y$ and we obtain the normal family when $g(y) = y$ and $h(y) = y^2/2$.

2.3. Dirichlet process

Finally, we note the use of the Dirichlet process for generating a first-order process with arbitrary marginal distribution. Let $\mathrm{DP}(c, g)$ denote the Dirichlet process probability (Ferguson, 1973); essentially DP generates random distributions, say F, such that $\mathrm{E}F = G$, where G is the distribution function corresponding to the density g, and $c > 0$ represents the variability of the F about G. If Y, given F, is a random draw from F, then marginally Y is from g. Additionally, the conditional probability $\mathrm{DP}(F|Y)$ is $\mathrm{DP}(c+1, g_Y)$ where $g_y = (cg + \delta_y)/(c+1)$, and δ_y is the discrete density mass function with unit mass at y. Consequently, the sequence $\{Y_t\}$ is given by simulation as $Y_{t+1} \sim F$ and $F \sim \mathrm{DP}(c+1, g_{Y_t})$. Marginally, $Y_t \sim g$ and $\mathrm{E}(Y_{t+1}|Y_t) = (c\mu + Y_t)/(c+1)$, where μ is the mean of g. Integrating over F in each step, we have

$$Y_{t+1} \sim \frac{cg + \delta_{Y_t}}{c+1}$$

and so $\mathrm{pr}(Y_{t+1} = Y_t) = 1/(c+1)$. In fact, this is the DAR(1) model (see McDonald & Zucchini, 1997).

More generally, if $\pi(dF)$ is a probability on a set of distributions, then the Markov chain with transition probability $\mathrm{P}(Y_{t+1} \in A|Y_t) = \int F(A)\pi(dF|Y_t)$ has stationary distribution $G(A) = \int F(A)\pi(dF)$.

3. Estimation

Although the construction of the likelihood function is straightforward, estimating parameters via maximum likelihood, for example, is not straightforward. However, we can regard the $\{X_t\}$ as latent or missing data and construct a complete data likelihood function in the form

$$L = f(y_0) \prod_{t \geqslant 0} f_{Y|X}(y_t|x_{t-1}) f_{X|Y}(x_{t-1}|y_{t-1}).$$

Then maximum likelihood estimation via the EM algorithm (Dempster *et al.*, 1977) is possible.

The mechanics, for the gamma example detailed in example 1, involves obtaining the mean from $f(x_{t-1}|y_{t-1}, y_t, \theta)$, where $\theta = (a, \phi)$ and

$$f(x|y_{t-1}, y_t, \theta) \propto \frac{\{y_{t-1} y_t \phi(1+\phi)\}^x}{z!\Gamma(a+x)},$$

where Γ denotes the usual gamma function. Denote this mean by $\tilde{x}_{t-1}$. The next step involves maximizing

$$\tilde{L}(\theta) = f(y_0) \prod_{t > 0} f_{Y|X}(y_t|\tilde{x}_{t-1}) f_{X|Y}(\tilde{x}_{t-1}|y_{t-1})$$

over θ. These two steps combine into an iterative algorithm producing $\{\theta_k\}_{k \geqslant 1}$ in which $\{\tilde{x}_t\}$ is obtained using θ_k and then θ_{k+1} is obtained via maximizing $\tilde{L}(\theta)$. While the expectation step can be done simply enough, the maximization step may need to be done using numerical methods or an iterative routine such as Newton–Raphson. The latter is easy to construct.

The point is that obtaining estimates via the EM algorithm is not difficult. Suppose instead we have $f_Y(y)$ as a Poisson mass function with mean parameter a. The appropriate conjugate style $f_{X|Y}(x|y)$ in this case is a binomial distribution with parameters y and $0 < \rho < 1$. Then $f_{Y|X}(y|x)$ can be sampled by taking $y = x + w$ where w is Poisson with mean $(1-\rho)a$. Now the mean $\tilde{x}_{t-1}$ is again easy to obtain and in this case the maximization step is available explicitly.

An alternative estimation procedure, in a Bayesian context, is to use Markov chain Monte Carlo methods. This is straightforward to implement in all the examples we have highlighted in the paper.

Acknowledgements

We are grateful for the comments and suggestions of two referees.

References

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.* **24**, 1–13.

Dempster, A. P., Laird, N. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1–38.

Feller, W. (1971). *An introduction to probability theory and its applications*. 2nd edn. Wiley, New York.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

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

Jacobs, P. A. & Lewis, P. A. W. (1977). A mixed autoregressive moving average exponential sequence and point process (EARMA 1,1). *Adv. Appl. Probab.* **12**, 87–104.

Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *J. Appl. Probab.* **33**, 664–677.

Jørgensen, B. (1986). Some properties of exponential dispersion models. *Scand. J. Statist.* **13**, 187–198.

Jørgensen, B. (1987). Exponential dispersion models (with discussion). *J. Roy. Statist. Soc. Ser. B.* **49**, 127–162.

Jørgensen, B. & Song, P. X.-K. (1998). Stationary time series models with exponential dispersion model margins. *J. Appl. Probab.* **35**, 78–92.

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

Lawrance, A. J. & Lewis, P. A. W. (1977). An exponential moving average sequence and point process (EMA 1). *J. Appl. Probab.* **14**, 98–113.

McDonald, I. L. & Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*. Chapman & Hall/CRC Press, London/Baton Rouge.

Smith, A. F. M. & Roberts, G. O. (1993). Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B.* **55**, 3–23.

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

Received December 1999, in final form October 2001

S. G. Walker. Department of Mathematical Sciences, University of Bath, BA2 7AY, UK.
E-mail: s.g.walker@bath.ac.uk