STATIONARY AUTOREGRESSIVE MODELS VIA A BAYESIAN NONPARAMETRIC APPROACH

By Ramsés H. Mena and Stephen G. Walker

IIMAS–UNAM, México and University of Kent

First Version received May 2004

Abstract. An approach to constructing strictly stationary AR(1)-type models with arbitrary stationary distributions and a flexible dependence structure is introduced. Bayesian nonparametric predictive density functions, based on single observations, are used to construct the one-step ahead predictive density. This is a natural and highly flexible way to model a one-step predictive/transition density.

Keywords. AR model; Bayesian nonparametrics; random probability measure; stationary process.

1. INTRODUCTION

Constructing *strictly stationary* autoregressive type (AR-type) models with arbitrary stationary distributions has been a focus of research for the past three decades. Typically, such models are useful when marginal inspection from the data is feasible. Early accounts of this type of construction, outside the Gaussian framework, can be found in Lawrance and Lewis (1977, 1980), Jacobs and Lewis (1977), Gaver and Lewis (1980) and Lawrance (1982), where positive real-valued AR-type models with marginal distributions falling in the gamma family were proposed. The construction proposed by these authors rely on the notion of a self-decomposable (SD) random variables.

A random variable X that can be decomposed as $X =_{\mathrm{d}} \rho X + W$, where $=_{\mathrm{d}}$ denotes equal in distribution, for all $0 < \rho < 1$ and (*innovation*) variable W, independent of X, is known to be self-decomposable. The SD random variables include many well-known distributions such as the gamma, log-normal and inverse Gaussian families.

It follows that $X_t = \rho X_{t-1} + W_t$, where W_t are independent and identically distributed copies of W, gives a stationary sequence X_t, with the stationary distribution being SD. See Mena and Walker (2004) for a recent study of the innovation distribution, which is generally of a complicated form.

Most of the contributions have focused on constructing integer-valued models. See McKenzie (1986, 1988), Al-Osh and Alzaid (1987), Alzaid and Al-Osh (1990), Du and Li (1991) and Al-Osh and Aly (1992), who used the thinning operator to construct AR-type models with Poisson, negative binomial and geometric

0143-9782/05/06 789–805 JOURNAL OF TIME SERIES ANALYSIS Vol. 26, No. 6

stationary distributions. The thinning operator, defined in its simplest form by $\rho \cdot X \sim \text{Binomial}(X, \rho)$, is useful to construct discrete versions of SD random variables (see Steutel and van Harn, 1979). For example, the thinning operator allows the construction of integer-valued AR-type (INAR) models, $X_t = \rho \cdot X_{t-1} + W_t$. See also Jacobs and Lewis (1978a, 1978b, 1978c) for an alternative approach to constructing integer-valued models.

A unified approach to constructing stationary AR-type models that encompasses many of the approaches has been highlighted by Joe (1996) and Jørgensen and Song (1998). In their approach stationary models of the type $X_t = A_t(X_{t-1}) + W_t$, where $A_t(\cdot)$ denotes a random operator, allow the construction of models with stationary distributions belonging to the infinitely divisible convolution-closed exponential family. This can itself be seen as a particular case of a more recent approach studied by Pitt *et al.* (2002).

The approach of Pitt *et al.* (2002) goes as follows: Suppose the required marginal density for the process is given by $p(x)$. A conditional distribution is introduced $p(y|x)$ and the transition density driving the AR(1)-type model $\{X_t\}$ is obtained as

$$p(x|x_{t-1}) = \int p(x|y)p(y|x_{t-1})\eta_1(\mathrm{d}y) \tag{1}$$

with

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\eta_2(\mathrm{d}x)},$$

where η_1 and η_2 are reference measures, such as the Lebesgue or counting measures. It is easy to show that $p(\cdot)$ constitutes an invariant density for the transition (1); that is

$$p(x) = \int p(x|x_{t-1})p(x_{t-1})\eta_2(\mathrm{d}x_{t-1}). \tag{2}$$

This implies that the resulting AR(1)-type model with transition (1) defines a stationary process with marginal distribution whose density is given by $p(\cdot)$. It is worth mentioning that this approach resembles the construction of a reversible Markov chain in the MCMC literature; in particular, with that of the Gibbs sampler method (see Smith and Roberts, 1993). Therefore, AR(1)-type models constructed using this approach will retain the reversibility property (see Robert and Casella, 2002).

In the construction of Pitt *et al.* (2002), although the stationary density is fixed, the transition density has a number of choices based on the selection of $p(y|x)$. In time series analysis it is this transition which is estimated and so a parametric form for $p(y|x)$ would lead to a parametric approach. In this paper we propose a nonparametric approach for modelling the transition density, while retaining the known stationary density.

The idea is based on a closer inspection of (1). Note that the transition density has the interpretation of

$$p(x|x_{t-1}) = \mathbb{E}\{p(x|y)|x_{t-1}\},$$

where the expectation is with respect to $p(y|x_{t-1})$ which itself can be seen as a Bayes posterior distribution with likelihood $p(x_{t-1}|y)$ and prior $p(y)$. If we now consider $y \equiv f$ to be a density function with prior $\Pi(\mathrm{d}f)$ then we can see the transition density as

$$p(x|x_{t-1}) = \int f(x)\Pi(\mathrm{d}f|x_{t-1}),$$

where $\Pi(\mathrm{d}f|x_{t-1})$ is the Bayes posterior for the density f given observation x_{t-1}. Also, $p(x|y) \equiv f(x)$ and $p(y|x_{t-1}) \equiv \Pi(\mathrm{d}f|x_{t-1})$. In other words, the transition density is the Bayesian predictive density based on the prior $\Pi(\mathrm{d}f)$ and a single observation. This is therefore connected to the area of Bayesian nonparametric methods.

Describing the layout of the paper; in Section 2 we present detailed background to the relevant concepts of Bayesian nonparametrics and define the general form for the model. Section 3 details the specific Bayesian nonparametric model discussed in the paper. Section 4 presents a numerical illustration of our approach and finally Section 6 contains a brief discussion and possible future work.

2. BAYESIAN NONPARAMETRIC CONSTRUCTION

We will be using a Bayesian nonparametric model to construct the transition density. Hence it is prudent here to describe such nonparametric models. The basic idea is a probability measure on spaces of density functions. In practice, the prior distributions are based on stochastic processes, such as Gaussian processes or Lévy processes (independent increment processes). The probability distribution governing the stochastic process acts as the prior distribution. See Walker *et al.* (1999) for a recent review.

For example, the Dirichlet process is based on a Lévy process, specifically, an independent increment gamma process. Let this be denoted by $Z(t)$, for $t \geq 0$. If $\lim_{t\to\infty} Z(t) < +\infty$ with probability one then

$$F(t) = Z(t)/Z(\infty)$$

behaves as a distribution function. The probability governing this process is the prior distribution. The posterior distribution is also a Dirichlet process (Ferguson, 1973). In the case of the Dirichlet process there is no admission of density functions. We are interested in priors which admit densities and so the actual prior process we will use is based on a Gaussian process, suitably normalized, which provides a process behaving almost surely as a density function. A detailed description of this process and the random density functions generated are given in Section 3.

So the prior will be denoted by $\Pi(\mathrm{d}f)$, or equivalently $\Pi(\mathrm{d}F)$ if preference is on the distribution rather than density function, and the posterior, given an observation x from f, is given by

$$\Pi(\mathrm{d}f|x) = \frac{f(x)\Pi(\mathrm{d}f)}{\int f(x)\,\Pi(\mathrm{d}f)}.$$

The predictive density based on the single observation x is then given by

$$p(x'|x) = \mathbb{E}\{f(x')|x\} = \int f(x')\,\Pi(\mathrm{d}f|x).$$

This forms the basis of the transition density.

To illustrate our idea we will use a simple random distribution model with marginal distribution Q. Consider the following joint distribution:

$$P(X \leq x, F \in A) = \mathbb{E}_{\Pi}\{F(x)\mathbb{I}(F \in A)\} = \int_A F(x)\Pi(\mathrm{d}F)$$

for any measureable set of distributions A. Integrating out the random distribution component, we have $Q(\cdot)$ given by

$$Q(x) = \mathbb{E}_{\Pi}\{F(x)\} = \int F(x)\Pi(\mathrm{d}F). \tag{3}$$

In the Bayesian nonparametric literature, Q is known as the centering or expected distribution associated with the prior Π. In our context, the distribution $Q(\cdot)$ will act as the required stationary distribution for the AR(1)-type model.

The posterior of F given $X = x$ is characterized by

$$\mathrm{P}(F \in A|X = x). \tag{4}$$

The conditional distribution (4) provides us with a way to construct the following one-step transition probability:

$$\mathrm{P}(x_{t-1}, x) = \mathrm{P}(X_t \leq x|X_{t-1} = x_{t-1}) = \mathbb{E}\{F(x)|X_{t-1} = x_{t-1}\}, \tag{5}$$

where the expectation is taken with respect to the conditional distribution (4).

The construction of the above one-step transition probability is equivalent to finding predictive densities, based on one observation, within the Bayesian nonparametric framework. The probability measure Π can be seen as the nonparametric prior. It is easy to check that if $X_{t-1} \sim Q$ and $\mathrm{Law}\{X_t|X_{t-1} = x\} = \mathbb{E}\{F|X_{t-1} = x\}$ then marginally $X_t \sim Q$. Note that

$$\begin{aligned} P_{X_t}(x) &= \int \mathrm{P}(x_{t-1}, x)Q(\mathrm{d}x_{t-1}) = \int\int F(x)\mathrm{d}F(x_{t-1})\Pi(\mathrm{d}F) \\ &= \int F(x)\left\{\int \mathrm{d}F(x_{t-1})\right\}\Pi(\mathrm{d}F) = Q(x). \end{aligned}$$

In this case, the constructed AR(1)-type model is also reversible since the balance condition

$$\mathrm{P}(x_{t-1},x)\,Q(x_{t-1}) = \int F(x)\,F(x_{t-1})\,\Pi(\mathrm{d}F) = \mathrm{P}(x,x_{t-1})\,Q(x),$$

is satisfied for any space–time points x and x_{t-1}.

The construction at issue, as an extension of the parametric setting introduced by Pitt *et al.* (2002) was also motivated in their paper. They found that the first order discrete autoregressive, DAR(1), model of Jacobs and Lewis (1978c) can be seen as an AR(1)-type process constructed in this way, when the probability measure Π is set to be the Dirichlet process. More precisely, if we denote by $\mathcal{D}(cQ)$ a Dirichlet process driven by the measure $c\,Q(\cdot)$, where $c > 0$, then a random distribution chosen by $F \sim \mathcal{D}(cQ)$ satisfies

$$\mathbb{E}\{F(x)\} = Q(x)$$

for any $x \in \mathbb{R}$ (see Ferguson, 1973). The parameter $c > 0$ is commonly associated to the variability of the random distributions F about Q.

In this case, the well-known conjugacy property of the Dirichlet process leads to

$$F|X = x \sim \mathcal{D}(cQ + \delta_x),$$

where δ_x denotes the point mass at x. See Theorem 1 in Ferguson (1973). With these assumptions, we can construct the following transition distribution driving the process $\{X_t\}_{t=1}^{\infty}$

$$\begin{aligned}\mathrm{P}(x_{t-1},x) &= \mathbb{E}\{F(x)|X_{t-1} = x_{t-1}\}\\ &= \frac{c}{c+1}\,Q(x) + \frac{1}{c+1}\,\delta_{x_{t-1}}((-\infty,x]),\end{aligned} \tag{6}$$

which remains invariant with respect to Q. Model (6) is known as the DAR(1) model. Take $c = \rho^{-1} - 1$ to obtain the notation of Jacobs and Lewis (1978c).

This model has a tractable dependence structure. It also inherits the discreteness associated with the Dirichlet process, a problem that rules it out as a model in most applications of time series analysis. Therefore, an alternative is to focus on choices of random distributions that lead to models that put probability one on the set of all absolutely continuous distributions. In other words, we look for models where

$$\mathrm{P}(X_t = x|X_{t-1} = x) = 0$$

for all $x \in \mathbb{R}$. In what follows we study one choice of prior Π with such constraint and with the property of having a more flexible dependence structure.

3. AR(1) MODELS BASED ON GENERALIZED LOG-GAUSSIAN PROCESSES

We use a method based on Gaussian processes studied by Lenk (1988) based on ideas of Leonard (1978) and Thorburn (1986). In order to introduce this method first let us assume that we want to construct a probability measure on density functions that have support on the set E, a subset of the real line. In other words, a density f will be modelled by a stochastic process $f = \{f(x);\ x \in E\}$. The 'sample paths' or 'trajectories' of such a process are densities.

Working in the framework described above, we can write the transition density corresponding to the transition distribution (5) and stationary density q, as follows

$$p(x_{t-1}, x_t) = \frac{1}{q(x_{t-1})} \int f(x_t) f(x_{t-1})\, \Pi(\mathrm{d}f), \tag{7}$$

where $q(x) = \int f(x)\Pi(\mathrm{d}f)$.

The logistic normal distribution described in Lenk (1988) is constructed as follows: Let Z be a Gaussian process with mean function μ and covariance function σ, both being continuous so a continuous separable version with integrable sample paths of Z is available. Consider the following logistic transform of Gaussian processes

$$f(x) = \frac{W(x)}{\int_E W(s)\mathrm{d}\lambda(s)}, \tag{8}$$

where λ is a σ-finite measure on E and $W(x) = \exp\{Z(x)\}$. The law, Λ, for the log-normal process W will be denoted by $\mathrm{LN}_E(\mu, \sigma)$. Provided the existence of such process, f has sample paths that are densities, and its support contains the densities on E with respect to λ.

As we saw in Section 2, the transition probability is given in terms of the moments of the random density. Following Lenk (1988), let us notice that the joint moments of W are given by

$$\begin{aligned} \mathrm{M}(\mathbf{x}) &= \mathbb{E}\left[\prod_{k=1}^{K} W(x_k)\right] \\ &= \exp\left\{ \left(\sum_{k=1}^{K} \mu(x_k) + \frac{\sigma(x_k, x_k)}{2} \right) + \sum_{i<j}^{K} \sigma(x_i, x_j) \right\}, \end{aligned} \tag{9}$$

where $\mathbf{x} \in E^K$. The random density f uses the path-integral of W, for which the positive moments are given by

$$\begin{aligned} \mathrm{C}(K, \mu) &= \mathbb{E}\left[\left(\int_E W \mathrm{d}\lambda\right)^K\right] \\ &= \int_{E^K} \mathbb{E}\left[\prod_{k=1}^{K} W(s_k)\right] \mathrm{d}\lambda(\mathbf{s}) \\ &= \int_{E^K} \exp\left\{\sum_{i<j}^{K} \sigma(s_i, s_j)\right\} \prod_{k=1}^{K} W_0(s_k) \mathrm{d}\lambda(\mathbf{s}), \end{aligned} \tag{10}$$

where $d\lambda(\mathbf{s}) = d\lambda(s_1)d\lambda(s_2)\cdots d\lambda(s_K)$ and $W_0(s) = \mathbb{E}[W(s)] = \exp\{\mu(s) + \sigma(s,s)/2\}$.

It can be proved that the use of process (8) as nonparametric prior leads to nonconjugate posteriors (4). In order to circumvent this issue Lenk (1988) considered a generalization of the random density (8) characterized by the following distribution:

$$\Lambda_\xi(A) = \frac{\int_A \{\int_E W(s)d\lambda(s)\}^\xi d\Lambda(W)}{C(\xi,\mu)}, \tag{11}$$

where $\xi = 1,2,\ldots$ and $\Lambda \sim W \sim \mathrm{LN}_E(\mu, \sigma)$. The generalized distribution (11) is indicated by $\mathrm{LN}_E(\mu, \sigma, \xi)$.

The *logistic normal process*, denoted by $\mathrm{LNS}_E(\mu, \sigma, \xi)$, is defined as the random density on $\mathcal{F}_d$ defined by the logistic transformation (8) with $W \sim \mathrm{LN}_E(\mu, \sigma, \xi)$.

Proposition 1. *Let* $f \sim \mathrm{LNS}_E(\mu, \sigma, \xi)$, *then*

$$\mathbb{E}\left[\prod_{i=1}^K f(x_i)\right] = M(\mathbf{x})\frac{C(\xi - K, \mu^*)}{C(\xi,\mu)},$$

where

$$\mu^*(s) = \mu(s) + \sum_{i=1}^K \sigma(s, x_i),$$

$M(\cdot)$ *is given by* (9) *and* $C(i, \mu)$ *is the ith moment, defined by* (10).

Proof. For a proof of this proposition we refer to Corollary 3 in Lenk (1988). □

Hence, modelling the random densities as $f \sim \mathrm{LNS}_E(\mu, \sigma, \xi)$ the transition density (7) is given by

$$\begin{aligned} p(x_{t-1}, x_t) &= \frac{\mathbb{E}[f(x_t)f(x_{t-1})]}{q(x_{t-1})} \\ &= \frac{1}{q(x_{t-1})} M((x_t, x_{t-1})')\frac{C(\xi - 2, \mu^\diamond)}{C(\xi,\mu)}, \end{aligned} \tag{12}$$

where $q(x) = \mathbb{E}[f(x)]$ and $\mu^\diamond(s) = \mu(s) + \sigma(s, x_t) + \sigma(s, x_{t-1})$. Using Proposition (1), the stationary density q can be represented as follows

$$q(x) = \mathbb{E}[f(x)] = \exp\{\mu(x) + \sigma(x,x)/2\}\frac{C(\xi - 1, \mu^x)}{C(\xi,\mu)},$$

where $\mu^x(s) = \mu(s) + \sigma(s, x)$. We thus have a very general method for constructing a stationary AR(1)-type models with stationary density q and transition density given by (12). The functions μ and σ present in (12) and q will allow us to model both marginal and dependence structure in a highly flexible way.

3.1. *Case $\xi = 1$*

As pointed out in Lenk (1988), the dependence structure of the process $\{X_t\}_{t=1}^{\infty}$ is driven by the covariance function σ chosen for the required Gaussian process. If $\sigma \to 0$ then the model approaches independence, meaning that not lag-dependence is captured.

Hence choosing a model translates into choosing ξ, μ and σ adequately. In order to easily represent the stationary distribution let us assume that $\xi = 1$, so that the random density is modeled by $f \sim \mathrm{LNS}_E(\mu, \sigma, 1)$. Under this assumption we have

$$q(x) \propto \exp\{\mu(x) + \sigma(x,x)/2\}. \tag{13}$$

Therefore, if we want to construct a AR(1)-type model with stationary density being q, we take the mean function in the required Gaussian process to be equal to

$$\mu(x) = \ln q(x) - \frac{\sigma(x,x)}{2}. \tag{14}$$

We are left with σ modelling the dependence structure.

PROPOSITION 2. *Assume that q and σ are given and that $f \sim \mathrm{LNS}_X(\mu, \sigma, 1)$, with μ given by* (14), *then*

$$p(x_{t-1}, x_t) \propto q(x_t) \exp\{\sigma(x_t, x_{t-1})\} C(-1, \mu^{\diamond}). \tag{15}$$

PROOF. The proof is given by direct substitution of (14) in (12) with M given by expression (9). □

Hence, for an arbitrary choice of q we can construct an AR(1)-type model with transition (15).

The complicated part of the transition (15) lies in the negative moment $C(-1,\mu^{\diamond})$. Such moments are not very tractable. In Lenk (1988) a Monte Carlo approximation scheme was proposed to compute such negative moments, which results in the following estimator,

$$C(-1, \mu^*) \propto \exp\{-\sigma(x_t, +)\},$$

where

$$\sigma(x, +) = \int_E \sigma(x, s) \mathrm{d}\lambda(s)/\lambda(E).$$

Hence, the transition density (15) can be approximated by

$$p(x_{t-1}, x_t) \propto q(x_t)\exp\{\sigma(x_t, x_{t-1}) - \sigma(x_t, +)\}. \tag{16}$$

For illustration purposes we consider a particular choice of covariance function, through which the dependence, in the AR(1)-type model, will be included. Natural choices of valid covariance functions are given by those corresponding to stationary Gaussian process, namely $\sigma(x, y) = \sigma^*(|x - y|)$. A general form within this class is the Matérn covariance function given by

$$\sigma^*(\tau) = \frac{\alpha}{\Gamma(v)2^{v-1}}\left(\frac{2\sqrt{v}\tau}{\kappa}\right)^v K_v\left(\frac{2\sqrt{v}\tau}{\kappa}\right), \quad \alpha, \kappa, v > 0, \tag{17}$$

where $\tau = |x - y|$ and K_v denotes the modified Bessel function of the second kind of order v (see Abramowitz and Stegun, 1992, Sec. 9.6). This covariance function leads to Gaussian processes with $(v - 1)$-times diferenciable paths. Therefore, v forms a smoothing parameter, for the paths of the Gaussian process Z, required in the construction of the random density. The parameter κ controls the dependency strength and α is a scaling parameter. For more on covariance functions we refer to Abrahamsen (1997).

Figure 1 shows some sample paths and densities of stationary AR(1)-type models driven by transition (16) with normal and Student-t stationary distributions. The stationary distributions were also plotted (dotted lines). From the density plots we can observe some skewness and kurtosis, this

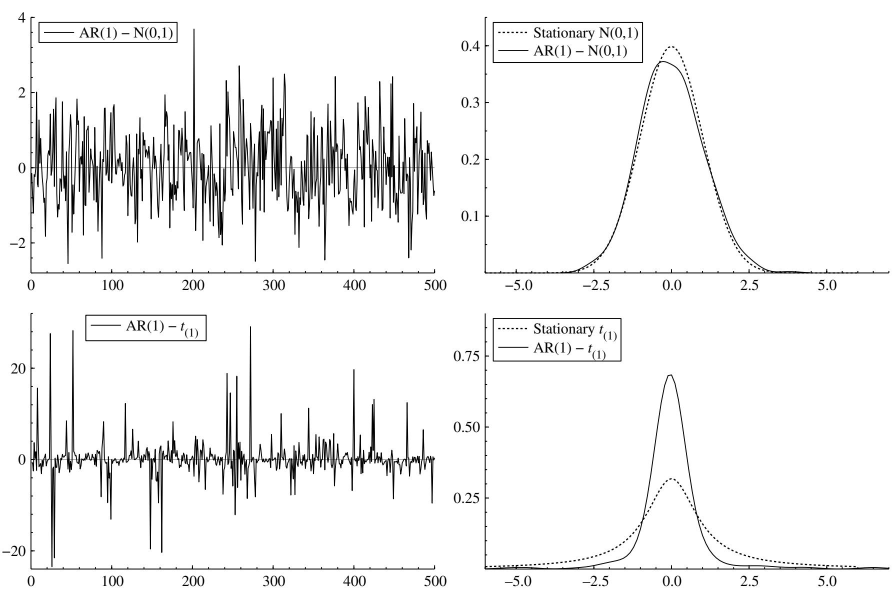

FIGURE 1. Data simulated from stationary AR(1)-type models using the approach described in Section 3.1. The Matérn covariance function with parameters $\alpha = \kappa = v = 1$ was used.

suggests that the constructed models are able to represent some of these features. The simulations were done using the inverse CDF method. The required normalization was done with numerical integration.

The Matérn covariance function is quite general among stationary covariances. However, the resulting dependence in the AR(1)-type model may be limited to these stationary behaviour. In Section 4 we give an example with a more general covariance function, leading to a wide flexibility in the dependence structure.

3.2. *Case* $\xi = 2$

The difficulty of working with negative moments leads us to consider the case when $\xi = 2$. From the transition density (12) is clear that any choice of $\xi \geq 2$ leads to a transition that does not depend on the negative moments. In particular for the case $\xi = 2$, we have the following simple form for the transition density,

PROPOSITION 3. *Let us assume that q and σ are given and that* $f \sim \mathrm{LNS}_X(\mu, \sigma, 2)$ *then*

$$p(x_{t-1}, x_t) \propto q(x_t) \exp\{\sigma(x_t, x_{t-1})\}. \tag{18}$$

PROOF. The proof follows the same argument as the one given for Proposition 2. □

In this case the form of the marginal density is not straightforward to write down (in terms of μ and σ) because it will involve the term $C(1, \mu^x)$. However, this is more than compensated by the simple form of the transition density.

4. A STATIONARY AR(1) MODEL WITH BETA MARGINALS

As mentioned before, Pitt *et al.* (2002) introduced a technique to construct stationary AR(1) models with a given stationary distribution. For the sake of illustration, let us consider the case, where the stationary distribution belongs to the beta family with density function denoted by $q(x) = \mathrm{Beta}(x; \alpha, \beta)$. Following the approach in Pitt *et al.* (2002) a way to construct an AR(1)-type model with such stationary distribution can be done by imposing the parametric family $f_{Y|X}(y|x) = \mathrm{Binomial}(y; r, x)$. Hence, the other component required to get the transition density (1) is given by $f_{X|Y}(x|y) = \mathrm{Beta}(x; \alpha + y, \beta + r - y)$. It can be easily seen that, the transition probability of this model has linear expectation given by $\mathbb{E}(X_t|X_{t-1}) = \rho X_{t-1} + (1 - \rho)\mu$, where $\mu = \mathbb{E}(X) = \alpha/(\alpha + \beta)$ and $\rho = r/(r + \alpha + \beta)$, the later also denoting the one-lag autocorrelation in the model.

The dependence structure in the model is determined by the form assumed for $f_{Y|X}$. However, changing such specification with another suitable family of distributions might lead to a completely different model with the same marginal distributions. In other words, given that the phenomenon under study has a beta distribution as the marginal, there is still a wide choice for the dependence structure. Hence our model is useful in such contexts. As an example we will try to capture the dependence in two simulated data sets.

Let us consider two data sets of size 200, simulated from the beta stationary AR(1)-type model described at the beginning of this section. For the first data set, we have chosen $\alpha = \beta = 9$ and $r = 2$ implying $\rho = 0.1$, whereas for the second data set, we choose $\alpha = \beta = 4$ and $r = 24$ implying $\rho = 0.75$. Namely, low and high autocorrelation respectively. It is worth noting that simulating from such a model is an easy task if one uses the representation (1). To model such data sets we fit the model described in Section 3.2, that is the specification corresponding to $\xi = 2$. In this case we do not restrict ourselves in the dependence structure and we assume a general form of the covariance function given by the orthonormal cosine basis representation

$$\sigma(x, y) = \sum_{k=1}^{\infty} V_k^2 \cos(k\pi x) \cos(k\pi y). \tag{19}$$

This choice give us a wide flexibility in the dependence structure. For more on this sort of covariance functions we refer to Grenander (1981).

For the estimation of the parameters we have truncated the series representation (19), that is we consider the estimation of the vector $V = (V_1, \ldots, V_M)$. Specifically we took $M = 4$. The numerical integration was done using the globally adaptive Gauss–Kronrod-based integrator and for the maximum likelihood estimation the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization algorithm was used. See Press *et al.* (1992) for more on this algorithm. All the routines were implemented in OX, see Doornik (2002). Figures 2 and 3 show the results from this estimation procedure. As is evident from these figures, the joint bivariate density estimators are very good.

5. REAL DATA EXAMPLE

We fit the AR(1)-type model with beta marginals to real data. The data set comes from daily average wind speeds during the year of 1978 at Rosslare's meteorological station in the Republic of Ireland. The data consist of 365 observations, measured in knots and belong to a bigger data set analysed in Haslett and Raftery (1989) to quantify the power resources of wind energy in the Republic of Ireland. The data can be obtained in Raftery's web page http://www.stat.washington.edu/raftery/index.html. Figure 4 shows the data together with its histogram, autocorrelation and partial autocorrelation function.

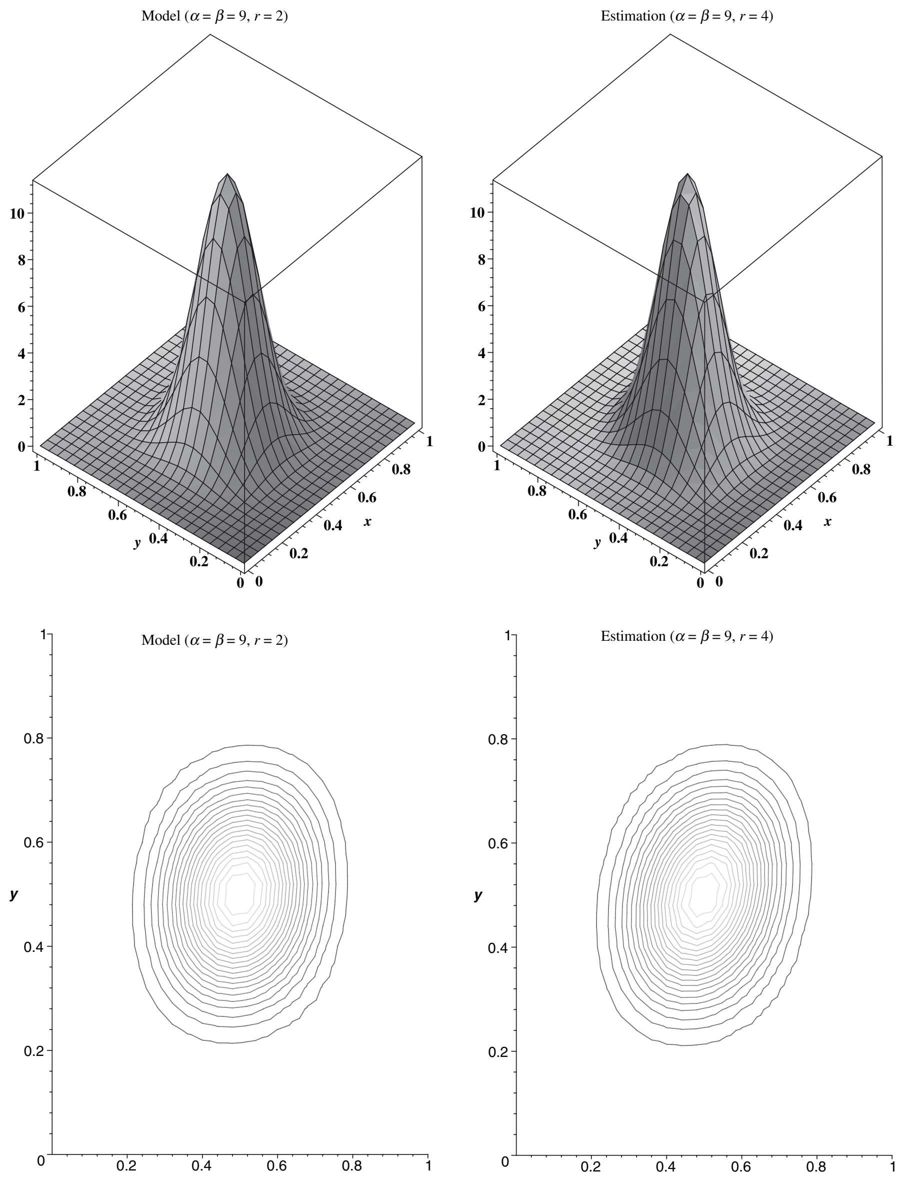

FIGURE 2. Bivariate densities $f_{X_t,X_{t-1}}(x,y)$ and their contours for the stationary beta model with correlation $\rho = 0.1$. The estimation was done assuming the cosine representation (19) truncated at $M = 4$.

Following the same line as in Section (4), we fit our model using the covariance function (19) truncated at $M = 15$. A way to visualize the fitting of such an order one model, can be done by comparing the estimated bivariate distribution $f_{X_t,X_{t-1}}$

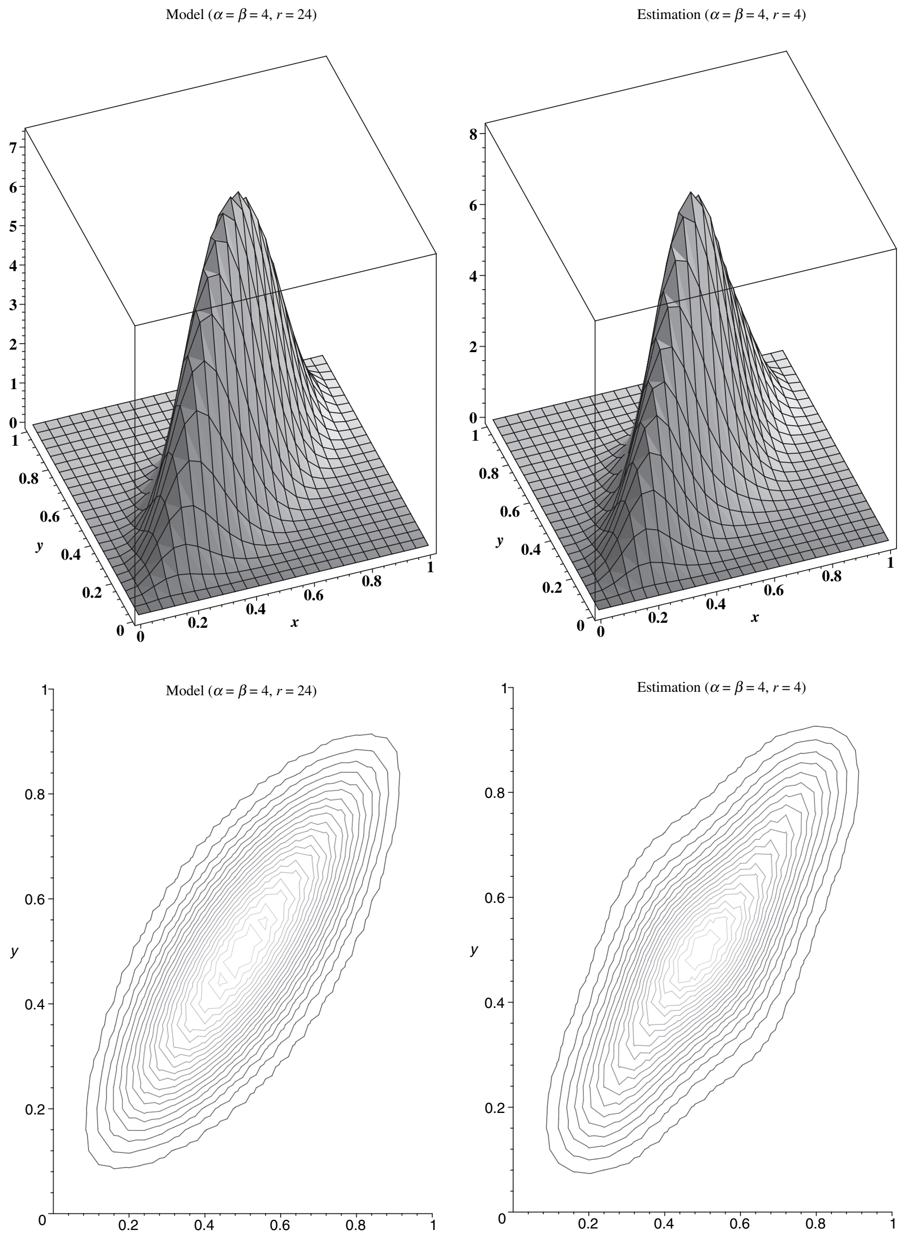

FIGURE 3. Bivariate densities $f_{X_t, X_{t-1}}(x, y)$ and their contours for the stationary beta model with correlation $\rho = 0.75$. The estimation was done assuming the cosine representation (19) truncated at $M = 4$.

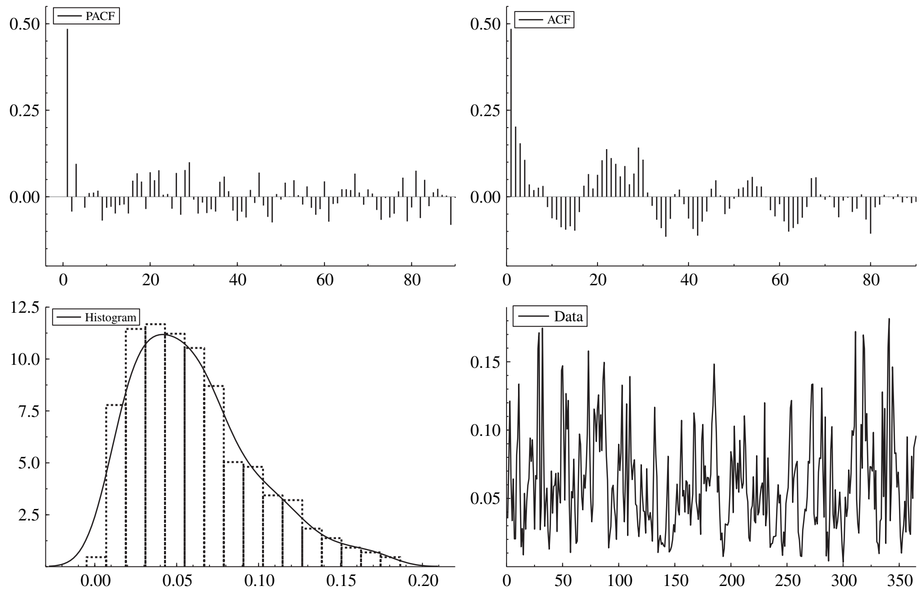

FIGURE 4. Daily average wind speeds for 1978 at Rosslare's meteorological station in the Republic of Ireland.

with the one corresponding to the data. In Figure 5, this two distributions are superimposed. In here, it is remarkable the ability in the model to capture high-moment dependence structures, such as the noticeable hump in the bivariate density corresponding to the data.

6. DISCUSSION

We have introduced a highly flexible way to model the dependence structure for an AR(1)-type stationary model. Moreover, this is achieved in a natural way using a Bayesian nonparametric predictive density, based on a single observation, to model the one-step transition density.

We have focused on the Gaussian process prior of Lenk (1988). Clearly alternative nonparametric priors are feasible. Here we mention the possibility of a mixture model; such as

$$f_P(x) = \int K(x;\theta)\mathrm{d}P(\theta),$$

where K is a kernel density and P is a random distribution function, such as a Dirichlet process. This is the well known and widely used Mixture of Dirichlet

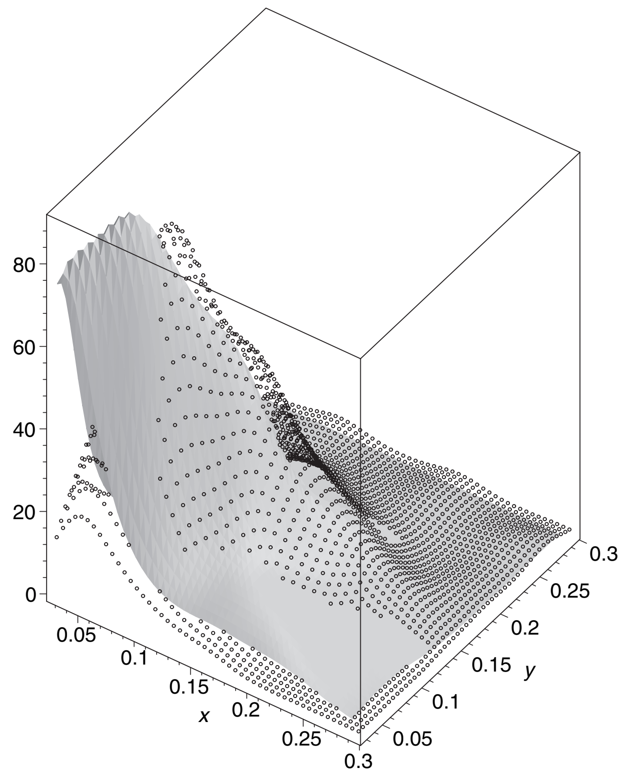

FIGURE 5. Comparison of bivariate densities $f_{X_t,X_{t-1}}(x,y)$ corresponding to wind speeds data (in circles) and fitted stationary AR(1) beta model with cosine representation (19) truncated at $M = 15$ (in patch). The surface estimation for the data was done using the sm.density function of Bowman and Azzalini (1997).

Process model. If $\mathbb{E}(P) = Q$ then the marginal density is $q(x) = \int K(x;\, \theta)\mathrm{d}Q(\theta)$ and the transition density is given by

$$p(x, y) = \mathbb{E}\{f_P(x)f_P(y)\}/q(x).$$

This is quite easy to simplify. Other possible priors include Pólya trees, Lavine (1992) and mixtures of Bernstein polynomial priors, Petrone (1999). These priors will form the basis of future research.

ACKNOWLEDGEMENTS

The research of the first author was partially supported by the National Council for Science and Technology of México (CONACYT) and of the second author by an EPSRC Advanced Research Fellowship.

NOTE

Corresponding author: Ramsés H. Mena, Departamento de probabilidad y estadística, IIMAS-UNAM, A.P. 20-726, 01000 D.F., México. E-mail: ramses@sigma.iimas.unam.mx

REFERENCES

Abrahamsen, P. (1997) A review of Gaussian random fields and correlation functions. Technical Report 917, Norwegian Computing Center., Box 114, Blindern, N-0314 Oslo, Norway.

Abramowitz, M. and Stegun, I. A., eds (1992) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications Inc. (Reprint of the 1972 edition.)

Al-Osh, M. A. and Alzaid, A. A. (1987) First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* 8, 261–75.

Al-Osh, M. A. and Aly E. A. A. (1992) First order autoregressive time series with negative binomial and geometric marginals. *Communication of Statistical Theory Methods* 21, 2483–92.

Alzaid, A. A. and Al-Osh, M. A. (1990) An integer-valued *p*th order autoregressive structure (INAR(*p*)) process. *Journal of Applied Probability* 27, 314–24.

Bowman, A. W. and Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Oxford: Oxford University Press.

Doornik, J. A. (2002) *Object-Oriented Matrix Programming Using Ox*, 3rd edn. Oxford: Timberlake Consultants Press.

Du, J. G. and Li, Y. (1991) The integer-valued autoregressive (INAR(*p*)) model. *Journal of Time Series Analysis* 12, 129–42.

Ferguson, T. (1973) A Bayesian analysis of some nonparametric problems. *Annals of Statistics* 1, 209–30.

Gaver, D. P. and Lewis, P. A. W. (1980) First-order autoregressive gamma sequences and point processes. *Advance in Applied Probability* 12, 727–45.

Grenander, U. (1981) *Abstract Inference*. New York: Wiley.

Haslett, J. and Raftery, A. E. (1989) Space-time modelling with long-memory dependence: assesing Ireland's wind power resource. *Applied Statistics* 38, 1–50.

Jacobs, P. A. and Lewis, P. A. W. (1977) A mixed autoregressive-moving average exponential sequence and point process (EARMA(1, 1)). *Advances in Applied Probability* 9, 87–104.

Jacobs, P. A. and Lewis, P. A. W. (1978a) Discrete time series generated by mixtures. I: correlational and runs properties. *Journal of Royal Statistical Society* B 40, 94–105.

Jacobs, P. A. and Lewis, P. A. W. (1978b) Discrete time series generated by mixtures. II: asymptotic properties. *Journal of Royal Statistical Society* B 40, 222–8.

Jacobs, P. A. and Lewis, P. A. W. (1978c) Discrete time series generated by mixtures. III: autoregressive processes (DAR(p)). Technical Report NPS55-78-022, Naval Postgraduate School, Monterey, California.

Joe, H. (1996) Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability* 33, 664–77.

Jørgensen, B. and Song, P. (1998) Stationary time series models with exponentials dispersion model margins. *Journal of Applied Probability* 35, 78–92.

Lavine, M. (1992) Some aspects of Pólya tree distributions for statistical modelling. *Annals of Statistics* 20, 1203–21.

Lawrance, A. J. (1982) The innovation distribution of a gamma distributed autoregressive process. *Scandinavian Journal of Statistics* 9, 234–6.

Lawrance, A. J. and Lewis P. A. W. (1977). An exponential moving-average sequence and point process (EMA1). *Journal of Applied Probability* 14, 98–113.

Lawrance, A. J. and Lewis P. A. W. (1980) The exponential autoregressive-moving average EARMA(p, q) process. *Journal of Royal Statistical Society Series* B 42, 150–61.

Lenk, P. J. (1988) The logistic normal distribution for Bayesian nonparametric predictive densities. *Journal of American Statistical Association* 83(402), 509–16.

Leonard, T. (1978) Density estimation stochastic processes and prior information. *Journal of Royal Statistical Society* B 40, 113–46.

McKenzie, E. (1986) Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advanced in Applied Probability* 18, 679–705.

McKenzie, E. (1988) Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability* 20, 822–35.

Mena, R. H. and Walker, S. G. (2004) Distribution of the innovation part of a self-decomposable random variable. *Journal of Statistical Computation and Simulation* 74, 765–75.

Petrone, S. (1999) Random bernstein polynomials. *Scandinavian Journal of Statistics* 26, 373–93.

Pitt, M. K., Chatfield, C. and Walker S. G. (2002) Constructing first order autoregressive models via latent processes. *Scandinavian Journal of Statistics* 29, 657–63.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (1992) *Numerical Recipes in C*. Cambridge: Cambridge University Press.

Robert C. P. and Casella, G. (2002) *Monte Carlo statistical methods*, 2nd edn. New york: Springer-Verlag.

Smith, A. and Roberts, G. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of Royal Statistical Socity* B 55, 3–24.

Steutel, F. W. and van Harn, K. (1979) Discrete analogues of self-decomposability and stability. *Annals of Probability* 7, 893–99.

Thorburn, D. (1986) A Bayesian approach to density estimation. *Biometrika* 73, 65–76.

Walker, S., Damien, P., Laud, P. and Smith, A. (1999) Bayesian nonparametric inference for random distributions and related functions. *Journal of Royal Statistical Society* B 61, 485–527.