TIME SERIES MODELS FOR NON-GAUSSIAN PROCESSES

BY H.W. BLOCK[1], N.A. LANGBERG[1], D.S. STOFFER[1,2]

University of Pittsburgh, Haifa University, Israel, and University of Pittsburgh

In this paper we present univariate and multivariate time series models for processes with non-Gaussian marginal distributions. These include bivariate autoregressive-type models for processes with bivariate exponential marginals, nonlinear autoregressive-type models for processes with Dirichlet marginals, and nonlinear models for univariate time series with arbitrary marginal distributions. Examples of applications to real data sets are given for some of the models discussed. When applicable, the theory of positive dependence is used to establish the association of the processes.

1. Introduction. The classical model in multivariate time series analysis is the $m \times 1$ vector linear process given by

$$\mathbf{X}(n) = \Sigma_{j=-\infty}^{\infty} A(j)\epsilon(n-j), \quad n \in \mathcal{Z} \tag{1}$$

where $\mathcal{Z} = \{0, \pm 1, \pm 2, \ldots\}$, $\{\epsilon(n), n \in \mathcal{Z}\}$ is a sequence of iid $m \times 1$ random vectors with mean zero and unknown covariance matrix, and $\{A(n), n \in \mathcal{Z}\}$ is a sequence of unknown $m \times m$ matrices such that $\Sigma_{j=-\infty}^{\infty} \parallel A(j) \parallel < \infty$ where $\parallel \cdot \parallel$ denotes the usual eigenvalue norm. Note that autoregressive (AR), moving average (MA), and mixed autoregressive-moving average (ARMA) models are important particular cases of the classical linear process (1).

If the $\epsilon(n)$'s are Gaussian then clearly so are the $\mathbf{X}(n)$'s in (1). Furthermore, if the $\mathbf{X}(n)$'s are Gaussian with mean zero and absolutely continuous spectrum, then there is a sequence of iid normal mean-zero random vectors $\epsilon(n)$, $n \in \mathcal{Z}$, and a sequence of matrices $A(n), n \in \mathcal{Z}$, such that the two processes $\mathbf{X}(n)$ and

[1]Supported in part by the Air Force Office of Scientific Research grant AFOSR–84–0113.

[2] Supported in part by National Science Foundation Grant DMS–9000522 and by a grant from the Centers for Disease Control through a cooperative agreement with the Association of Schools of Public Health.

AMS 1980 subject classifications. Primary 62M10; secondary 62E99, 62H20.

Key words and phrases. Bivariate exponential distributions, Dirichlet distribution, nonlinear time series models, positive dependence, time series with arbitrary marginals.

The authors are grateful to Robert H. Shumway, University of California, Davis for supplying the data sets used in the examples. The authors would also like to thank a referee and the editors for comments and suggestions that improved the presentation.

$\Sigma_{j=-\infty}^{\infty} A(n)\epsilon(n-j), n \in \mathcal{Z}$, are stochastically equal (see Hannan, 1970, p. 221). If the $\mathbf{X}(n)$ process is non-Gaussian, then a decomposition (1) may not exist, and the statistical inference procedures developed for processes satisfying (1) do not apply.

There are, however, many physical situations in which time series are patently non-normal. It has been suggested that time series that depart slightly from normality be handled by data transformations. For other cases, where the departure from normality is more substantial, it has been suggested that new time series models be developed (see, for example, Lewis, 1980).

Over the past decade, there has been a considerable amount of research on modeling time series with exponential, gamma, geometric, or general discrete marginal distributions. For example, Lawrance and Lewis (1980, 1981, 1985) and Jacobs and Lewis (1977) present univariate ARMA-type models with exponential marginals. Raftery (1982) generalizes their models to include multivariate time series with exponential marginals. Gaver and Lewis (1980) present univariate AR-type models with gamma marginals, and Jacobs and Lewis (1978, 1983) present ARMA-type models for univariate discrete-valued time series. Block, Langberg, and Stoffer (1988) present bivariate ARMA-type models with bivariate exponential and geometric marginals, and Langberg and Stoffer (1987) develop bivariate MA-type processes with exponential and geometric marginals. Models and statistical methodologies for the analysis of categorical time series have recently been developed by Fahrmeir and Kauffmann (1987), Kauffmann (1987), and Stoffer (1987).

In this paper we focus on three different problems. In Section 2 we present a model and corresponding statistical methodology for analyzing univariate time series with arbitrary continuous marginal distributions. The model is then used to analyze wind speed data. In Section 3 we present the bivariate exponential autoregressive (BEAR) model which can be used to analyze bivariate time series in which the process of interest has bivariate exponential marginals. We show that a BEAR process can have well known bivariate exponential marginal distribution such as the Marshall and Olkin (1967) bivariate exponential distribution. The theory of positive dependence is used to show that the BEAR model can consist of associated random variables. Finally, in Section 4 we present a multivariate AR-type model with Dirichlet marginal distributions. This model is useful for modeling and forecasting vector processes in which the distribution of the random vector at each point in time is a Dirichlet distribution. Estimation and prediction methods are given for the model and the techniques are illustrated on a data set from soil science.

2. Univariate Processes With Arbitrary Continuous Marginals. In this section we present models and corresponding statistical methodology that will allow an investigator to model and analyze time series with any continuous marginal distribution, whether or not the investigator is willing to specify the family of distributions. The method used here is essentially the translation method

(see Mardia, 1970) via the probability integral transform. That is, the processes are constructed by a monotone nondecreasing transform of a Gaussian linear process and the data are assumed to be obtained by a simple instantaneous nonlinear filter acting on a Gaussian process (see Hannan, 1970, section 2.7, for a discussion of nonlinear filters applied to Gaussian processes). Both a nonparametric approach (when the family of distributions is not specified) and a parametric approach (when the family of distributions is specified) are taken. The techniques of this section are then illustrated by analyzing a real data set.

Underlying both the nonparametric models and the parametric models is a mean zero Gaussian process $\{Y_n, n \in \mathcal{Z}\}$, such that $E\{Y_n^2\} = 1$. Thus, for each n, $Y_n \sim \Phi$ where Φ represents the Gaussian cdf. As an example of such a process, consider a linear process where $\{\epsilon_n, n \in \mathcal{Z}\}$ is a sequence of iid $N(0, \sigma^2)$ random variables, $\{\psi_0, \psi_1, \ldots\}$ is a sequence of parameters such that $\Sigma_{q=0}^{\infty} \mid \psi_q \mid < \infty$ and $\sigma^{-2} = \Sigma_{q=0}^{\infty} \psi_q^2$, and

$$Y_n = \Sigma_{q=0}^{\infty} \psi_q \epsilon_{n-q}, \quad n \in \mathcal{Z}. \tag{2}$$

First we present the nonparametric approach. Let $\{X_n, n \in \mathcal{Z}\}$ be a process of interest such that the cdf of X_n is H (that is, $X_n \sim H, n \in \mathcal{Z}$) where H is a continuous, unspecified cdf. For H, let H^{-1} be a right continuous inverse of H given by:

$$H^{-1}(p) = \begin{cases} \inf \{x : H(x) > p\} & 0 \leq p < 1 \\ \sup \{x : H(x) < 1\} & p = 1. \end{cases} \tag{3}$$

We define the process $\{X_n, n \in \mathcal{Z}\}$ as follows:

$$X_n = H^{-1}[\Phi(Y_n)], \quad n \in \mathcal{Z} \tag{4}$$

where $\{Y_n, n \in \mathcal{Z}\}$ is the Gaussian process. Noting that for all n, $\Phi(Y_n)$ has a uniform cdf on (0,1) it follows that $X_n \sim H, n \in \mathcal{Z}$.

Next we present the parametric approach. Let Ω be a subset of the kth Euclidean space and let $\{H_\theta : \boldsymbol{\theta} \in \Omega\}$ be a specified family of absolutely continuous cdf's which depends on the parameter $\boldsymbol{\theta}$. We define the process $\{X_n(\boldsymbol{\theta}); n \in \mathcal{Z}, \boldsymbol{\theta} \in \Omega\}$ as follows:

$$X_n(\boldsymbol{\theta}) = H_\theta^{-1}[\Phi(Y_n)], \quad n \in \mathcal{Z}. \tag{5}$$

By a preceding argument, it follows that $X_n(\boldsymbol{\theta}) \sim H_\theta; n \in \mathcal{Z}, \boldsymbol{\theta} \in \Omega$.

For an example of some parametric models, let $\Omega = (0, \infty) \times (-\infty, \infty)$, let Z be a random variable with cdf F, and let H_θ be the cdf of $\theta_1 Z + \theta_2, \boldsymbol{\theta} \in \Omega$. Then the class of processes $\{X_n(\boldsymbol{\theta}) = \theta_1 F^{-1}[\Phi(Y_n)] + \theta_2; n \in \mathcal{Z}, \boldsymbol{\theta} \in \Omega\}$ has H_θ marginals. So, for example, the class of processes could have Cauchy marginals: $F^{-1}(p) = \tan \pi p$, logistic marginals: $F^{-1}(p) = -\ell n(p^{-1} - 1)$, or shifted exponential marginals: $F^{-1}(p) = -\ell n(1 - p)$, to mention a few.

Estimation and prediction for each of the models is relatively simple if the underlying Gaussian process $\{Y_n, n \in \mathcal{Z}\}$ is an ARMA(p,q) process. Henceforce we assume that

$$(6) \qquad \Sigma_{j=0}^{p}\alpha_j Y_{n-j} = \Sigma_{k=0}^{q}\beta_k \epsilon_{n-k}, \quad n \in \mathcal{Z}, \quad \alpha_0 = \beta_0 = 1$$

where $\{\epsilon_n, n \in \mathcal{Z}\}$ is the process described in (2), and with appropriate conditions on $\{\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q\}$ for (6) to be a stationary, causal and invertible process (see for example, Brockwell and Davis, 1987). If $x_1, \ldots, x_N$ are N observations from a process given by (4) or (5), let $x_{(1)}, \ldots, x_{(N)}$ be the corresponding order statistics and let $H_N(t), t \in (-\infty, \infty)$, be the empirical cdf of the observations given by:

$$(7) \qquad H_N(t) = N^{-1}\Sigma_{n=1}^{N-1} I(t)_{[x_{(n)},\infty)} + N^{-1} I(t)_{(x_{(N)},\infty)}$$

where I_A denotes the indicator function of A. Note that in (7), $H_N(t) < 1$ for $t \leq x_{(N)}$; the necessity of this definition will be apparent in (8). The properties of the empirical cdf for dependent processes (that are mixing) are given in Gastwirth and Rubin (1975). A Kolmogorov-Smirnov statistic for (correlated) data sampled from an autoregressive process is given in Weiss (1978).

For the nonparametric model, estimation and prediction are accomplished by first transforming the data as follows:

$$(8) \qquad \hat{y}_n = \Phi^{-1}[H_N(x_n)], \quad n = 1, \ldots, N$$

where H_N is given in (7), and then by treating the $\hat{y}_n$ as a sample of length N from the Gaussian process (6). Parameter estimates and forecasts based on the $\hat{y}_n$ are obtained via standard methods and the forecasts of the original model are approximated by setting $x_{N+j}^N = H_N^{-1}[\Phi(\hat{y}_{N+j}^N)]$, $j = 1, 2, \ldots$, where $\hat{y}_{N+j}^N$ are the estimated j-step-ahead forecasts of the Y_n process based on the data $\hat{y}_1, \ldots, \hat{y}_N$, and H_N^{-1} is given in (3).

For the parametric model, parameter estimation can be obtained via maximum likelihood. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p), \boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)$, and for $y_1, \ldots, y_N$, N observations from the Gaussian ARMA(p, q) process given by (6), let $L_Y(\boldsymbol{\alpha}, \boldsymbol{\beta}, y_1, \ldots, y_N)$ be the likelihood function of $y_1, \ldots, y_N$. Recall that σ^2 is a function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ alone. The likelihood function of the data $x_1, \ldots, x_N$, is given by

$$(9) \quad L_X(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, x_1, \ldots, x_N) = L_Y(\boldsymbol{\alpha}, \boldsymbol{\beta}, \Phi^{-1}[H_\theta(x_1)], \ldots, \Phi^{-1}[H_\theta(x_N)])$$

$$\times \Pi_{n=1}^{N}\{\dot{\Phi}^{-1}[H_\theta(x_n)]\}\dot{H}_\theta(x_n)$$

where $\cdot$ denotes differentiation. We note that in (9) we may write

$$\dot{\Phi}^{-1}[H_\theta(x_n)] = 1/\dot{\Phi}\{\Phi^{-1}[H_\theta(x_n)]\}.$$

Let $\hat{\boldsymbol{\theta}}_N$ denote the MLE of $\boldsymbol{\theta}$, and let

$$\tilde{y}_n = \Phi^{-1}[H_{\hat{\theta}_N}(x_n)], \quad n = 1, 2, \ldots, N, \tag{10}$$

then the forecasts for the parametric process (5) may be approximated by setting $x^N_{N+j}(\hat{\boldsymbol{\theta}}_N) = H^{-1}_{\hat{\theta}_N}[\Phi(\tilde{y}^N_{N+j})], j = 1, 2, \ldots,$ where $\tilde{y}^N_{N+j}$ denotes the forecast obtained from the $\tilde{y}_n$, $n = 1, \ldots, N$, based on the MLE's of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and the model (6).

As an example of the kind of data that can be handled by the models, we fit both a parametric and a nonparametric model to wind speed (mph) measurements in Washington, D.C., May – September, 1977, 133 observations (which we denote $x_0, \ldots, x_{132}$) made daily at noon. Since wind speeds have been modeled using the Weibull distribution (see Lawrance and Lewis, 1985), our parametric model assumes that the data comes from a process with shifted Weibull marginals, that is,

$$H_\theta(x) = 1 - \exp\{-\mu(x - \xi)^\gamma\}, \quad x \geq \xi \tag{11}$$

where $\boldsymbol{\theta} = (\gamma, \mu, \xi)$ with $\gamma > 0$, $\mu > 0$, and $-\infty < \xi < \infty$. Also, the data indicated an AR(1) model for the $\{Y_n\}$ process given by (6), that is,

$$Y_n = \alpha Y_{n-1} + \epsilon_n, \quad |\alpha| < 1 \tag{12}$$

where $\{\epsilon_n\}$ is white Gaussian noise, $\epsilon_n \sim N(0, \sigma^2)$, with $\sigma^2 = 1 - \alpha^2$.

Figure 2.1 shows a plot of the data as well as the one-step-ahead forecasts using the parametric approach. The actual data is shown by a solid line in Figure 2.1, with the data points represented by circles; the extreme observations are 5.0 mph and 18.8 mph. Maximum likelihood estimation (cf. 9) under model assumptions (5), (11) and (12) yielded the following MLE's:

$$\hat{\gamma}_N = 1.732, \ \hat{\mu}_N = .069, \ \hat{\xi}_N = 4.920, \ \hat{\alpha}_N = .388.$$

The one-step-ahead forecasts (shown in Figure 2.1 by a dashed line) were calculated by first computing the transformed data (10), then forecasting $\tilde{y}_n$ as $\tilde{y}^{n-1}_n = .388\tilde{y}_{n-1}$, and then setting $x^{n-1}_n(\hat{\boldsymbol{\theta}}_N) = H^{-1}_{\hat{\theta}_N}[\Phi(\tilde{y}^{n-1}_n)]$, $n = 1, \ldots, 132$, as the one-step-ahead wind speed forecasts. The analysis of the $\tilde{y}_n$ values ($n = 0, 1, \ldots, 132$) using the parametric model verified the model assumption (12).

Figure 2.2 compares the empirical distribution function (solid line) of the wind speed data with $H_{\hat{\theta}_N}$ (dashed line), showing satisfactory results. It is clear from Figure 2.2 that the nonparametric approach for this data set leads to relatively the same results as the parametric approach. The only parameter to be estimated in the nonparametric case is α in (12). For this example we obtained a value of .386 as an estimate of α; this compares well with the corresponding estimate $\hat{\alpha}_N$ in the parametric case. As the one-step-ahead forecasts of the wind speed data do not differ visually from those obtained from the parametric model, we do not show these forecasts. The analysis of the $\hat{y}_n (n = 0, \ldots, 132)$ using the nonparametric approach (cf. 8) verified the model assumption (12).

Figure 2.1: Observed (solid line) and predicted (dashed line) wind speed data (mph) using a parametric model.

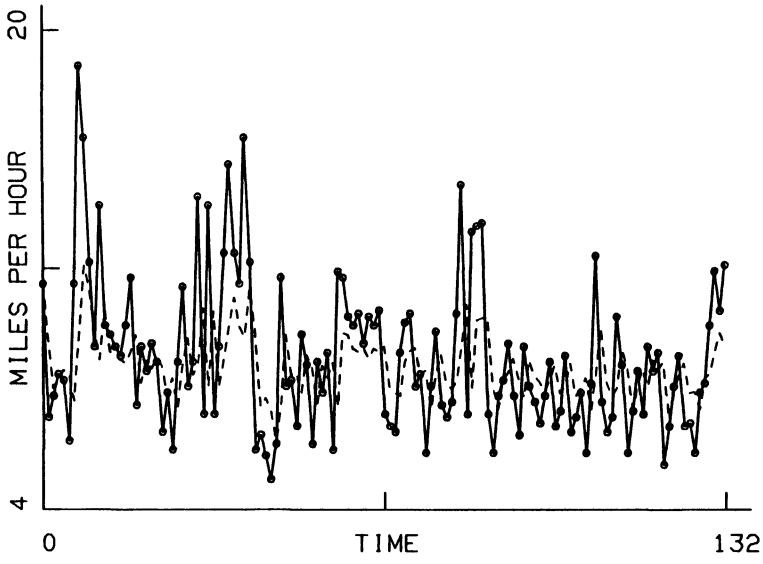
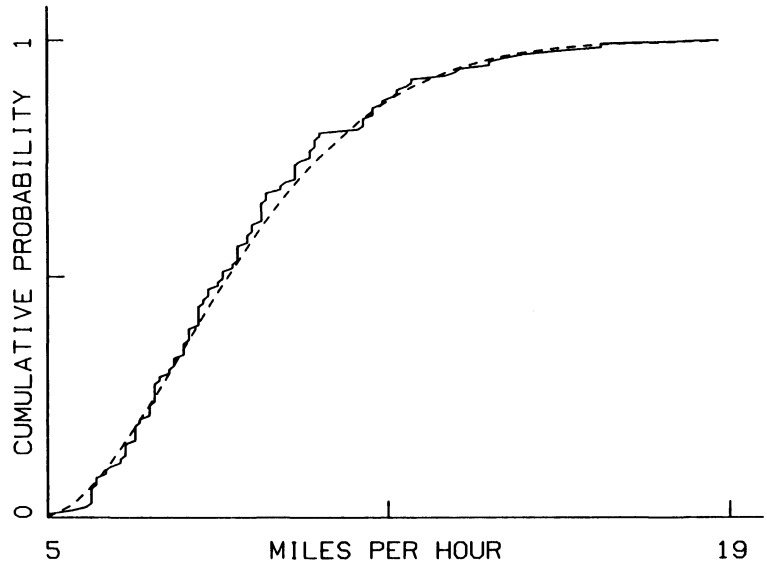


Figure 2.2: Empirical distribution (solid line) of the wind speed measurements and the fitted Weibull c.d.f. (dashed line).



3. Bivariate Exponential Autoregressive Processes. In this section we present the BEAR(1) model and discuss some of its properties. A detailed account of the BEAR(1) process as well as some related models, statistical methodology, bivariate dependence mechanisms, and a data example may be found in Block, Langberg, and Stoffer (1988). Throughout this section we will say that a random vector $\mathbf{E} = (E_1, E_2)$ has a bivariate exponential distribution if the marginal distributions of E_1 and E_2 are exponential. First we introduce the concepts of random mixing (Lemma 3.1) and random summation (Lemma 3.2). We shall write $U \stackrel{S}{=} V$ to mean that random variables (or vectors) U and V are stochastically equal.

LEMMA 3.1. *(Random Mixing). Let X and Z be independent random variables with exponential distributions where $X \stackrel{S}{=} Z$. Let I be a Bernoulli random variable independent of X and Z, and such that $\Pr\{I = 1\} = 1 - \pi$. Then the random variable given by*

$$Y \stackrel{S}{=} IZ + \pi X$$

has the same exponential distribution as X and Z.

LEMMA 3.2. *(Random Summation). Let X_j, $j = 1, 2, \ldots$, be iid exponential random variables with mean $\pi/\lambda, 0 < \pi < 1$. Let N have a geometric distribution $(N \geq 1)$ with mean π^{-1} and be independent of the X_j. Then the random variable given by*

$$Y \stackrel{S}{=} \Sigma_{j=1}^{N} X_j$$

has an exponential distribution with mean λ^{-1}.

The proofs of Lemmas 3.1 and 3.2 follow easily by computing characteristic functions. In the following theorem we connect the concepts of these two lemmas and show how these results may be used to extend the concepts of random mixing and random summation to bivariate exponential random vectors. The proof of this theorem may be found in Block, Langberg, and Stoffer (1988, Lemma 2.17).

THEOREM 3.1. *Let (X_{1j}, X_{2j}), $j = 1, 2, \ldots$, be iid random vectors with a fixed bivariate exponential distribution. Let (N_1, N_2) be independent of the (X_{1j}, X_{2j}) and have the bivariate geometric distribution given by*

$$\Pr\{N_1 > n_1, N_2 > n_2\} = \begin{cases} p_{11}^{n_1}[p_{01} + p_{11}]^{n_2 - n_1} & 0 \leq n_1 \leq n_2 \\ p_{11}^{n_2}[p_{10} + p_{11}]^{n_1 - n_2} & 0 \leq n_2 \leq n_1 \end{cases} \tag{13}$$

where $p_{ij} \geq 0$ $(i, j = 0, 1)$, $p_{00} + p_{01} + p_{10} + p_{11} = 1$, $p_{10} + p_{11} < 1$, and $p_{01} + p_{11} < 1$. Let (I_1, I_2) be a bivariate Bernoulli random vector where $\Pr\{I_1 = i, I_2 = j\} = p_{ij}$, $i, j = 0, 1$, such that $1 - \pi_1 = p_{10} + p_{11} < 1$, and $1 - \pi_2 = p_{01} + p_{11} < 1$. Then

$$\left(\Sigma_{j=1}^{N_1} X_{1j}, \Sigma_{j=1}^{N_2} X_{2j}\right) \stackrel{S}{=} (I_1 Z_1, I_2 Z_2) + (\pi_1 X_1, \pi_2 X_2)$$

where $(Z_1, Z_2) \stackrel{S}{=} \left(\Sigma_{j=1}^{N_1} X_{1j}, \Sigma_{j=1}^{N_2} X_{2j}\right)$, $(X_1, X_2) \stackrel{S}{=} (\pi_1^{-1} X_{1j}, \pi_2^{-1} X_{2j})$, *and* (I_1, I_2), (X_1, X_2), *and* (Z_1, Z_2) *are mutually independent.*

The bivariate geometric distribution given in (13) is discussed in detail in Block (1977). In order to provide insight into the connection between the concepts of random summation and random mixing and the BEAR(1) process, we present the following theorem.

THEOREM 3.2. *Let* (Y_1, Y_2) *and* (Z_1, Z_2) *have the same unknown bivariate exponential distribution. Let* (X_1, X_2) *have a known bivariate exponential distribution. Let* (I_1, I_2) *be as defined in Theorem 3.1, and let all the aforementioned random vectors be mutually independent. If*

$$(Y_1, Y_2) \stackrel{S}{=} (I_1 Z_1, I_2 Z_2) + (\pi_1 X_1, \pi_2 X_2)$$

then (Y_1, Y_2) *and* (Z_1, Z_2) *have the same bivariate exponential distribution as* (X_1, X_2).

PROOF. Let Ψ_Y, Ψ_X, and Ψ_Z be the characteristic functions of the random vectors (Y_1, Y_2), (X_1, X_2), and (Z_1, Z_2), respectively, and note that $\Psi_Y = \Psi_Z$. Then for real numbers u, v, we have the characteristic function equation:

$$\Psi_Y(u, v) = \Psi_X(\pi_1 u, \pi_2 v)[p_{00} + p_{10}\Psi_Z(u, 0) + p_{01}\Psi_Z(0, v) + p_{11}\Psi_Z(u, v)].$$

Letting $u = 0$, we may solve for $\Psi_Y(0, v)$ and $\Psi_Z(0, v)$. Similarly, letting $v = 0$, we may solve for $\Psi_Y(u, 0)$ and $\Psi_Z(u, 0)$. Inserting these results into the characteristic function equation and noting that Ψ_X is specified, the result then follows easily by solving for $\Psi_Y(u, v)$ and $\Psi_Z(u, v)$. ||

We are now ready to present the BEAR(1) model. First, we shall need some notation. Let $(I_1(n), I_2(n))$ and (N_1, N_2) be as in Theorem 3.1 and let $\{\mathbf{E}(n), n = \pm 1, \pm 2, \ldots\}$ be an iid sequence of bivariate exponential random vectors with mean vector $(\lambda_1^{-1}, \lambda_2^{-1})$ which is independent of (N_1, N_2) and $(I_1(n), I_2(n))$. We set

$$\mathbf{E}(0) = \left(\Sigma_{j=1}^{N_1} \pi_1 E_1(-j),\ \Sigma_{j=1}^{N_2} \pi_2 E_2(-j)\right)' \tag{14}$$

where we have written $\mathbf{E}(n) = (E_1(n), E_2(n))'$. Note that by Lemma 3.2, $\mathbf{E}(0)$ has a bivariate exponential distribution with mean vector $(\lambda_1^{-1}, \lambda_2^{-1})$, but the joint distribution of $\mathbf{E}(0)$ is not completely specified unless the distribution of $\mathbf{E}(n)$ is given. Define $A(n)$ to be the 2×2 diagonal random matrix $A(n) = \text{diag}\{I_1(n), I_2(n)\}$ and B to be the 2×2 diagonal matrix $B = \text{diag}\{\pi_1, \pi_2\}$.

The BEAR(1) process is defined as follows:

$$\mathbf{X}(n) = \begin{cases} \mathbf{E}(0) & n = 0 \\ A(n)\mathbf{X}(n-1) + B\mathbf{E}(n) & n = 1, 2, \ldots . \end{cases}$$

The following useful characterization of the BEAR(1) process is established in Block, Langberg, and Stoffer (1988). This characterization is primarily based on the preceding lemmas and theorems.

LEMMA 3.3. *Let* $\{\mathbf{X}(n), n = 0,1,2,\ldots\}$ *be a BEAR(1) process. Then for* $n = 0,1,2,\ldots,$

$$\mathbf{X}(n) \stackrel{S}{=} \mathbf{E}(0)$$

where $\mathbf{E}(0)$ *is given in (14).*

From Lemma 3.3 and results given in Block (1977), we may show that a BEAR(1) process can have well known bivariate exponential distributions. These include the Marshall-Olkin (1967), Downton (1970), Hawkes (1972), and Paulson (1973) bivariate exponential distributions. Details may be found in Block, Langberg, and Stoffer (1988, Remark 4.5). As an example we specify a particular $\mathbf{E}(0)$. Let E be an exponential random variable with mean θ, $0 < \theta < (\lambda_1 + \lambda_2)^{-1}$, $\pi_1 = \lambda_1\theta$, $\pi_2 = \lambda_2\theta$, and let $\mathbf{E}(1) = (\pi_1^{-1}E, \pi_2^{-1}E)'$. Let b_1, b_2, b_{12} be nonnegative real numbers such that $\lambda_1 = b_1 + b_{12}$, $\lambda_2 = b_2 + b_{12}$, and let $p_{00} = \theta b_{12}$, $p_{10} = \theta b_2$, $p_{01} = \theta b_1$, and $p_{11} = 1 - \theta(b_1 + b_2 + b_{12})$, then the resulting $\mathbf{X}(n)$ has a Marshall-Olkin (1967) bivariate exponential distribution.

In view of Lemma 3.3, we also have the following result on the positive dependence of the random variables of the BEAR(1) process. If $\mathbf{Z} = (Z_1, \ldots, Z_q)$, $q = 1,2,\ldots,$ is a random vector we say that the random variables $Z_1, \ldots, Z_q$ are *associated* if $\operatorname{cov}\{f(Z), g(Z)\} \geq 0$ for all f and g monotonically nondecreasing in each argument, such that the expectations exist.

LEMMA 3.4. *Let* $\{\mathbf{X}(n)\}$ *be a BEAR(1) process and suppose* $E_1(1)$ *and* $E_2(1)$ *are associated, then* $\{X_i(n_j); i = 1,2, j = 1,2,\ldots,k\}$, $k > 0$ *integer, are associated.*

The proof of Lemma 3.4 follows by using the result of Lemma 3.3 in conjunction with a theorem on the association of a sequence of random variables that is conditionally increasing in sequence, see Barlow and Proschan (1981, Theorem 4.7).

4. Multivariate Dirichlet Processes. In this section we are concerned with modeling $(k+1)$-dimensional series, say $\mathbf{P}(n) = (P_1(n), \ldots, P_{k+1}(n))$; $n = 0,1,2,\ldots,$ where at each time point n, $\mathbf{P}(n)$ has a Dirichlet distribution with parameter vector $(\alpha_1, \ldots, \alpha_{k+1})$. That is, we are interested in modeling and forecasting multivariate time series in which the data are proportions and are constrained so that $\Sigma_{j=1}^{k+1} P_j(n) = 1$ at each point in time.

Before presenting the model we establish some results. Throughout this section let $\alpha_1, \ldots, \alpha_{k+1} > 0$, let Y_j be independent gamma$(1, \alpha_j)$ random variables, and define $Z_j = Y_j / \Sigma_{\ell=1}^{j} Y_\ell$, $j = 1, \ldots, k+1$. Note that $Z_1 \equiv 1$ and that Z_j, $j = 2, \ldots, k+1$ are independent beta$(\alpha_j, \Sigma_{\ell=1}^{j-1}\alpha_\ell)$ random variables. Finally let

$$U_j = \begin{cases} Z_j \Pi_{\ell=j+1}^{k+1}(1 - Z_\ell) & j = 1, \ldots, k \\ Z_{k+1} & j = k+1 \end{cases}$$

and note that $U_j = Y_j / \Sigma_{\ell=1}^{k+1} Y_\ell$, $j = 1, \ldots, k+1$. It then follows that

$$(U_1, \ldots, U_{k+1}) \sim \text{ Dirichlet } (\alpha_1, \ldots, \alpha_{k+1}).$$

We now define a univariate autoregressive-type process with beta marginals. This process will then be used to build the multivariate Dirichlet process. To ease the notation, we define $A_q = \Sigma_{\ell=1}^{q} \alpha_\ell$.

LEMMA 4.1. *(Beta Processes). For each j, $j = 2, \ldots, k+1$, let $\{B_{nj}, n = 1, 2, \ldots\}$ be a sequence of iid Bernoulli random variables such that $\Pr\{B_{nj} = 1\} = \alpha_j A_j^{-1}$, let $\{Q_{nj}, n = 1, 2, \ldots\}$ be a sequence of iid beta$(1, A_j)$ random variables, and let Z_{0j} be a beta(α_j, A_{j-1}) random variable. Assume that $\{B_{nj}\}$, $\{Q_{nj}\}$, and $\{Z_{0j}\}$ are mutually independent, $n = 1, 2, \ldots; j = 2, \ldots, k+1$. Then the processes $\{Z_{nj}, n = 1, 2, \ldots; j = 2, \ldots, k+1\}$ defined by*

$$Z_{nj} = Q_{nj} B_{nj} + (1 - Q_{nj}) Z_{n-1,j} \tag{15}$$

have beta(α_j, A_{j-1}) marginals, that is,

$$Z_{nj} \overset{S}{=} Z_{0j}, \quad n = 1, 2, \ldots; \ j = 2, \ldots, k+1. \tag{16}$$

PROOF. Let Y_0 be a gamma(1,1) random variable which is independent of $Y_1, \ldots, Y_{k+1}$. Then we have the following results:

(i) $Q_{nj} \overset{S}{=} Y_0 (\Sigma_{\ell=0}^{j} Y_\ell)^{-1}$, and

(ii) $Z_{nj} \overset{S}{=} Y_j (\Sigma_{\ell=1}^{j} Y_\ell)^{-1}$ for $n = 1, 2, \ldots; \ j = 2, \ldots, k+1$,

and by Basu's Theorem the random variables

(iii) $Y_0 (\Sigma_{\ell=0}^{j} Y_\ell)^{-1}$ and $Y_j (\Sigma_{\ell=0}^{j} Y_\ell)^{-1}$ are independent, $j = 2, \ldots, k+1$.

We now establish (16) by an induction argument on n. By (i), (ii), (iii), and (15) we have that

$$\begin{aligned} Z_{1j} &\overset{S}{=} Y_0 (\Sigma_{\ell=0}^{j} Y_\ell)^{-1} B_{1j} \\ &\quad + [1 - Y_0 (\Sigma_{\ell=0}^{j} Y_\ell)^{-1}] Y_j (\Sigma_{\ell=1}^{j} Y_\ell)^{-1} \\ &= [Y_0 B_{1j} + Y_j] (\Sigma_{\ell=0}^{j} Y_\ell)^{-1}, \quad j = 2, \ldots, k+1. \end{aligned} \tag{17}$$

Let f_j, g_j be the density function of Z_{1j}, Z_{0j}, respectively, $j = 2, \ldots, k+1$, and let $y \in (0,1)$. Then by (17), for $j = 2, \ldots, k+1$,

$$\begin{aligned} f_j(y) &= \alpha_j A_j^{-1} \alpha_j^{-1} A_j y g_j(y) + A_{j-1} A_j^{-1} A_j A_{j-1}^{-1} (1-y) g_j(y) \\ &= g_j(y). \end{aligned}$$

Hence (16) holds for $n = 1$. The induction proof now proceeds by assuming that (16) holds for a given n, $n > 1$, then using a similar argument, one can show that (16) holds for $n+1$. Hence (16) holds for all n. ||

We are now ready to define the stationary autoregressive-type sequence of Dirichlet random vectors, $\{\mathbf{P}(n), n = 0, 1, 2, \ldots\}$. Let $Z_{n1} \equiv 1, n = 0, 1, 2, \ldots$, and let $\{Z_{nj}, n = 0, 1, 2, \ldots; j = 2, \ldots k+1\}$ be as defined as (15). Let

$$P_j(n) = \begin{cases} Z_{nj} \Pi_{\ell=j+1}^{k+1} (1 - Z_{n\ell}) & j = 1, \ldots, k \\ Z_{n,k+1} & j = k+1 \end{cases} \tag{18}$$

and let

$$\mathbf{P}(n) = (P_1(n), \ldots, P_{k+1}(n)), \quad n = 0, 1, 2, \ldots. \tag{19}$$

It follows by Lemma 4.1 and the preceding results on the joint distribution of $(U_1, \ldots, U_{k+1})$ that for each n, $\mathbf{P}(n)$ has a Dirichlet distribution with parameter vector $(\alpha_1, \ldots, \alpha_{k+1}), n = 0, 1, 2, \ldots$.

Estimation of the parameter vector $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_{k+1})$ of the Dirichlet process (19) may be carried out by maximum likelihood. Let $\mathbf{P}(1), \ldots, \mathbf{P}(N)$, be a realization of length N of the process (19). Maximum likelihood estimation is accomplished as follows:

(1) Transform the data to the beta sequences by setting

$$Z_{nj} = P_j(n)[\Sigma_{\ell=1}^{j} P_\ell]^{-1}, \quad j = 2, \ldots, k+1; n = 1, \ldots, N.$$

(2) From the Z_{nj}, calculate the Q_{nj} and B_{nj} defined in Lemma 4.1:

(a) If $Z_{nj} > Z_{n-1,j}$ then $B_{nj} = 1$ and

$$Q_{nj} = (Z_{nj} - Z_{n-1,j})/(1 - Z_{n-1,j}),$$

(b) If $Z_{nj} \leq Z_{n-1,j}$ then $B_{nj} = 0$ and

$$Q_{nj} = (Z_{nj} - Z_{n-1,j})/(-Z_{n-1,j}),$$

for $j = 2, \ldots, k+1$, and $n = 1, \ldots, N$. These calculations are direct consequences of (15).

(3) Maximize the log-likelihood function of the observations $\{Q_{nj}, B_{nj}; n = 1, \ldots, N, j = 2, \ldots, k+1\}$, say $L(\theta)$, given by

$$\begin{aligned} L(\theta) &= \Sigma_{\ell=2}^{k+1}\{A_\ell \ell n \Delta_\ell + S_\ell \ell n(A_\ell A_{\ell-1}^{-1} - 1)\} \\ &+ N\Sigma_{\ell=1}^{k} \ell n A_\ell - \Sigma_{\ell=2}^{k+1} \ell n \Delta_\ell, \end{aligned}$$

where $\Delta_\ell = \Pi_{n=1}^{N}(1 - Q_{n\ell})$, and $S_\ell = \Sigma_{n=1}^{N} B_{n\ell}$, for $\ell = 2, \ldots, k+1$.

For the cases $k = 1, 2$ an explicit solution to $\partial L(\theta)/\partial A_\ell = 0$, $\ell = 1, \ldots, k+1$, exists so that for $k = 1$ (that is, the process of interest is the univariate beta process given in (15) with $j = 2$)

$$\hat{\alpha}_1 = (N - S_2)/(-\ell n \Delta_2)$$

$$\hat{\alpha}_2 = S_2/(-\ell n \Delta_2)$$

and for $k = 2$,

$$\hat{\alpha}_1 = (2N - S_3)(N - S_2)/[-N\ell n(\Delta_2 \Delta_3)]$$

$$\hat{\alpha}_2 = (2N - S_3)S_2/[-N\ell n(\Delta_2 \Delta_3)]$$

$$\hat{\alpha}_3 = S_3/(-\ell n(\Delta_3)$$

are the MLE's. For arbitrary k, the MLE's can be obtained by a numerical method such as a Newton-Raphson or scoring procedure.

As an example of the kind of data that can be handled by the model, we consider spatial data presented in Mechergui (1984). As part of a study of the water table in the vicinity of drainage tiles, Mechergui analyzes the content of 72 equally (and linearly) spaced auger holes each dug to a depth of 2.4 meters. The sampling rate was 10 auger holes per 52 meters and the proportions of silt, clay, and sand are obtained for each sample with the constraint that the percentages of silt, clay, and sand in each auger hole sample sum to 100%.

We fit a Dirichlet model (19) to the proportions of silt, clay, and sand, $P_1(n)$, $P_2(n)$, and $P_3(n)$, respectively, based on the 72 observation vectors $\mathbf{P}(n) = (P_1(n), P_2(n), P_3(n))$, $n = 0, 1, 2, \ldots, 71$, and then used the estimated model to obtain one-step-ahead forecasts for the time points $n = 1, \ldots, 71$. Figure 4.1 shows the silt series, $P_1(n)$, as well as the one-step-ahead forecasts based on the estimated model; Figure 4.2 shows similar plots for the clay series, $P_2(n)$.

Figure 4.1: Observed (solid line) and predicted (dashed line) silt content series.

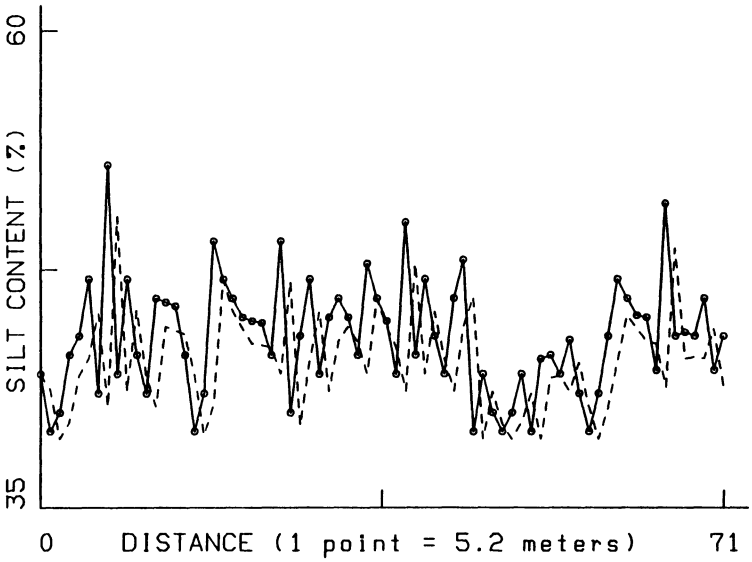
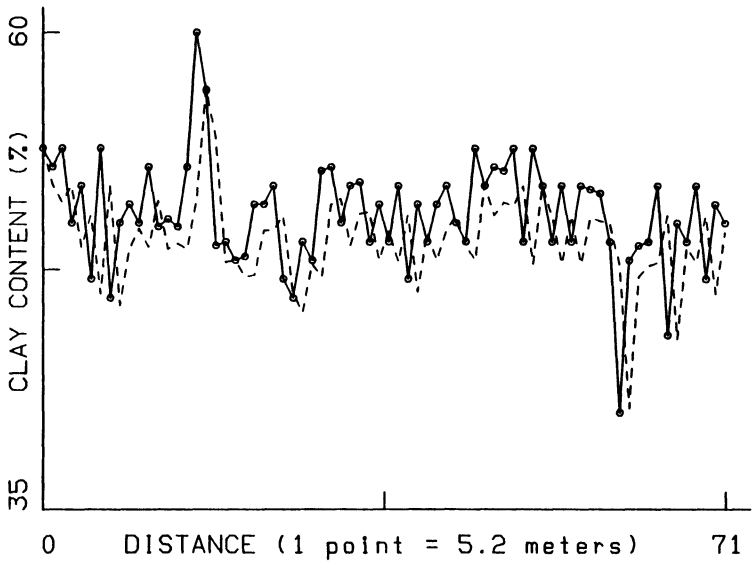


Figure 4.2: Observed (solid line) and predicted (dashed line) clay content series.



Maximum likelihood estimation was carried out as previously discussed. In this case we found the MLE's to be:

$$\hat{\alpha}_1 = 2.573, \quad \hat{\alpha}_2 = 2.963, \quad \hat{\alpha}_3 = 2.137.$$

Forecasting was accomplished by first forecasting Z_{n2} and Z_{n3} obtained in estimation step (1), and then using these forecasts to predict $\mathbf{P}(n)$ via the transformation (18). Then one-step-ahead forecasts of Z_{nj}, $j = 2,3$ based on (15) is given as follows:

$$\hat{Z}_{nj} = (1 + \Sigma_{\ell=1}^{j}\hat{\alpha}_\ell)^{-1}[\hat{\alpha}_j(\Sigma_{\ell=1}^{j}\hat{\alpha}_\ell)^{-1} + (\Sigma_{\ell=1}^{j}\hat{\alpha}_\ell)Z_{n-1,j}], \quad j = 2,3$$

for $n = 1,\ldots,71$. Recall that $Z_{n1} \equiv 1$ for all n, and hence we put $\hat{Z}_{n1} \equiv 1$ for all n. From (18) we obtain the one-step-ahead forecasts for $P_j(n)$, $j = 1,2,3$:

$$\hat{P}_j(n) = \begin{cases} \hat{Z}_{nj}\Pi_{\ell=j+1}^{3}(1 - \hat{Z}_{n\ell}) & j = 1,2 \\ \hat{Z}_{n3}, & j = 3. \end{cases} \tag{20}$$

The root mean square prediction errors for the silt, clay, and sand series based on the preceding prediction equations are .040, .036, and .037 respectively.

REFERENCES

BARLOW, R.E. and PROSCHAN, F. (1981). *Statistical Theory of Reliability and Life Testing: Probability Models.* To Begin With, Silver Spring, MD.

BLOCK, H.W. (1977). A family of bivariate life distributions. In *The Theory and Applications of Reliability, Vol. 1*, C.P. Tsokos and I. Shimi, Eds. Academic Press, New York.

BLOCK, H.W., LANGBERG, N.A., and STOFFER, D.S. (1988). Bivariate exponential and geometric autoregressive and autoregressive-moving average models. *Advances in Applied Probability* **20** 798–821.

BROCKWELL, P.J. and DAVIS, R.A. (1987). *Time Series: Theory and Methods*, Springer-Verlag, New York.

DOWNTON, F. (1970). Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society, Ser. B.* **32** 408–417.

FAHMEIR, L. and KAUFMANN, H. (1987). Regression models for nonstationary categorical time series. *Journal of Time Series Analysis* **8** 147–160.

GASTWIRTH, J.L. and RUBIN, H. (1975). The asymptotic distribution theory of the empiric c.d.f. for mixing processes. *Annals of Statistics* **3** 809–824.

GAVER, D.P. and LEWIS, P.A.W. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability* **12** 727–745.

HANNAN, E.J. (1970). *Multiple Time Series.* Wiley and Sons, New York.

HAWKES, A.G. (1972). A bivariate exponential distribution with applications in reliability. *Journal of the Royal Statistical Society, Ser. B.* **34** 129–131.

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. (1978). Discrete time series generated by mixtures I: Correlational and runs properties. *Journal of the Royal Statistical Society, Ser. B* **40** 94–105.

JACOBS, P.A. and LEWIS, P.A.W. (1983). Stationary discrete autoregressive moving average time series generated by mixtures. *Journal of Time Series Analysis* **4** 18–36.

KAUFMANN, H. (1987). Regression models for nonstationary categorical time series: Asymptotic estimation theory. *Annals of Statistics* **15** 79–98.

LANGBERG, N.A. and STOFFER, D.S. (1987). Moving average models with bivariate exponential and geometric distributions. *Journal of Applied Probability* **24** 48–61.

LAWRANCE, A.J. and LEWIS, P.A.W. (1980). The exponential autoregressive-moving average process EARMA (p, q). *Journal of the Royal Statistical Society, Ser. B* **42** 150–161.

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

LAWRANCE, A.J. and LEWIS, P.A.W. (1985). Modeling and residual analysis of nonlinear autoregressive time series in exponential variables (with Discussion). *Journal of the Royal Statistical Society, Ser. 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, P.R. Krishnaiah, ed., North Holland, 151–166.

MARDIA, K.V. (1970). *Families of Bivariate Distributions.* Hafner Publishing Co., Darien, Conn.

MARSHALL, A.W. and OLKIN, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association* **62** 30–44.

MECHERGUI, M. (1984). Stochastic modeling of the water table in the vicinity of drainage tiles. Unpublished dissertation in Soil Science, Graduate Division, University of California, Davis.

PAULSON, A.S. (1973). A characterization of the exponential distribution and a bivariate exponential distribution. *Sankhyā, Ser. A* **35** 69–78.

RAFTERY, A.E. (1982). Generalized non-normal time series. In *Time Series Analysis: Theory and Practice I*, O.D. Anderson, ed., North Holland, Amsterdam, 621–640.

STOFFER, D.S. (1987). Walsh-Fourier analysis of discrete-valued time series. *Journal of Time Series Analysis* **8** 449–467.

WEISS, M.S. (1978). Modification of the Kolmogorov-Smirnov statistic for use with correlated data. *Journal of the American Statistical Association* **73** 872–875.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PA 15260

DEPARTMENT OF STATISTICS
HAIFA UNIVERSITY
MOUNT CARMEL, HAIFA, ISRAEL

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PA 15260