INFORMS Journal on Computing
Vol. 20, No. 3, Summer 2008, pp. 485–498
ISSN 1091-9856 | EISSN 1526-5528 | 08 | 2003 | 0485

DOI 10.1287/ijoc.1070.0261

Evaluation of the ARTAFIT Method for Fitting Time-Series Input Processes for Simulation

Bahar Biller
Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, billerb@andrew.cmu.edu

Barry L. Nelson
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, nelsonb@northwestern.edu

Time-series input processes occur naturally in the stochastic simulation of many service, communications, and manufacturing systems, and there are a variety of time-series input models available to match a given collection of properties, typically a marginal distribution and an autocorrelation structure specified via the use of one or more time lags. The focus of this paper is the situation in which the collection of properties are not "given," but data are available from which a time-series input model is to be estimated. The input model we consider is the very flexible autoregressive-to-anything (ARTA) model of Cario and Nelson [Cario, M. C., B. L. Nelson. 1996. Autoregressive to anything: Time-series input processes for simulation. *Oper. Res. Lett.* **19** 51–58]. Recently, we developed a statistically valid algorithm (ARTAFIT) for fitting this model to stationary univariate time-series data using marginal distributions from the Johnson translation system. In this paper, we perform a comprehensive numerical study to assess the performance of our algorithm relative to the two most commonly used approaches: (a) fitting the marginal distribution but ignoring the autocorrelation structure, and (b) fitting separately the marginal distribution as in (a) and the autocorrelation structure using the sample autocorrelation function. We find that ARTAFIT, which fits the marginal distribution and the autocorrelation structure jointly, outperforms both (a) and (b), and we demonstrate the importance of taking dependencies into account while developing input models for stochastic simulation.

Key words: simulation; statistical analysis
History: Accepted by Marvin Nakayama, Area Editor for Simulation; received February 2006; revised February 2007, November 2007; accepted November 2007. Published online in *Articles in Advance* May 30, 2008.

1. Introduction

Modeling the uncertainty in the input of the system being studied is one of the challenging problems in the design of stochastic simulation experiments. Input modeling is often characterized as selecting appropriate univariate probability distributions to represent the primitive inputs of interest, and it would indeed be this simple if the relevant input processes could all be represented as sequences of independent random variables having identical marginal distributions. When such univariate models do apply, there exist a number of fitting methods with good statistical properties, e.g., maximum likelihood, least squares, and moment matching, for estimating the marginal distributions of the underlying input processes. We refer the interested reader to Vincent (1998) and Law and Kelton (2000) for comprehensive reviews of the input modeling tools available for independent and identically distributed input processes.

However, dependent time-series input processes occur naturally in the stochastic simulation of many service, communications, and manufacturing systems (see, e.g., Melamed et al. 1992, Ware et al. 1998). Ignoring dependence in these settings can lead to very poor estimates of performance measures, as illustrated by Livny et al. (1993), who examine the impact of autocorrelated interarrival times on the mean waiting time of a single-server queue.

Motivated by the severe consequences of ignoring dependence in real-life systems, input modeling for dependent processes has attracted the attention of a number of researchers. The focus has largely been on two types of input processes: random vectors and time series, with the latter being the subject of this paper. Specifically, we consider a sequence of random variables $\{X_t;\ t = 1, 2, \ldots, n\}$ of length n with discrete time index t whose full characterization is the n-dimensional distribution of all individual random variables X_t, $t = 1, 2, \ldots, n$, where n may be arbitrarily large. Of course, the amount of expert knowledge or data needed for the full characterization of this time series is typically prohibitive. The classical approach for getting around this problem is to match only certain key properties, such as the marginal distribution of the individual random elements and the autocorrelation structure of the sequence. This may not—and typically will not—uniquely or even correctly specify the joint distribution of interest, but the hope is that these key characteristics are sufficient to produce accurate simulation results. Thus, there has been significant interest in the simulation community on fitting time-series input processes to a given collection of properties, such as the marginal distribution moments and the autocorrelations up to a finite lag. We refer the reader to Song et al. (1996), Cario and Nelson (1998), Chen (2001), and Biller and Nelson (2003) for example studies.

However, it is often the case that the desired properties are not "given," but time-series data from the process of interest are available. The common practice is to ignore dependence, assuming an independent and identically distributed (i.i.d.) input process and fitting the parameters of the marginal distribution of this process using maximum likelihood or least squares methods. It is important to note that when the data are actually dependent, but this dependence is ignored in the likelihood function, the resulting estimators are not guaranteed to possess the statistical properties of the standard maximum likelihood estimators. An improvement over ignoring dependence is to match the autocorrelation structure to the sample autocorrelation function but fit the marginal distribution assuming an i.i.d. input process. In this paper, we demonstrate that a good alternative to both of these approaches is to use the ARTAFIT algorithm of Biller and Nelson (2005) for fitting an autoregressive-to-anything (ARTA) process of Cario and Nelson (1996), which is regarded as one of the most general models for univariate time series in the simulation input modeling literature. The ARTAFIT algorithm chooses parameters for the marginal distribution from the Johnson translation system (Johnson 1949a) and the autocorrelation structure jointly to obtain a better overall fit. We will investigate the performance of ARTAFIT in recovering the true model parameters for different sample sizes and a wide variety of distributional shapes and dependence characteristics when data actually come from an ARTA process. We will also provide empirical evidence that the ARTA process combined with the ARTAFIT algorithm can represent a wide variety of time-series input processes that are not ARTA. This large-scale empirical study is an important companion to the theoretical properties established in Biller and Nelson (2005) because it demonstrates that the model and the fitting method can be expected to perform well in practice.

The remainder of this paper is organized as follows. In §2, we review the time-series input models used in our empirical study and describe the estimation methods we use to fit these input models to data samples. We discuss the selection of the experimental factors and the performance metrics in §3, present our findings in §4, and conclude with future research directions in §5. For the interested reader, we provide background information on the Johnson translation system, on the gamma process of Lewis et al. (1989), and on the implementation of the ARTAFIT algorithm in the Online Supplement to this paper (available at http://joc.pubs.informs.org/ecompanion.html).

2. Fitting Time-Series Input Processes for Simulation

In §2.1, we describe two different time-series models for stochastic simulation: the ARTA process of Cario and Nelson (1996) and the gamma process of Lewis et al. (1989). The evaluation of the ARTAFIT procedure of Biller and Nelson (2005) is the major focus of this paper. While the ARTA process is the target process for this procedure, the gamma process is the one we choose to stress test the ARTA model and the ARTAFIT algorithm. Later in §2.2, we discuss the problem of fitting the ARTA input process to data generated by ARTA and gamma processes.

2.1. Time-Series Input Processes

2.1.1. ARTA Processes. An ARTA process $\{X_t;\ t=1,2,\ldots\}$ is a univariate time series with an arbitrary marginal distribution F and autocorrelations $\rho_X(h)$, $h=1,2,\ldots,p$ specified through finite lag p. In this paper, we call the parameter p the order of dependence and construct the ARTA process $\{X_t;\ t=1,2,\ldots\}$ via the transformation $U_t=\Phi(Z_t)$, $t=1,2,\ldots,$ where $\{U_t;\ t=1,2,\ldots\}$ corresponds to a time series with uniform marginals on $(0,1]$ and the base process $\{Z_t;\ t=1,2,\ldots\}$ is a stationary, standard Gaussian autoregressive process of order p with the representation

$$Z_t=\sum_{h=1}^{p}\alpha_h Z_{t-h}+Y_t,\quad t=p+1,p+2,\ldots.$$

The α_h, $h=1,2,\ldots,p$ are fixed autoregressive coefficients that uniquely determine the autocorrelation structure of the base process, $\rho_Z(h)$, $h=1,2,\ldots,p$, and Y_t, $t=p+1,p+2,\ldots$ are i.i.d. Gaussian random variables with mean zero and variance σ_Y^2. Choosing $\sigma_Y^2=1-\sum_{h=1}^{p}\alpha_h\rho_Z(h)$ ensures that each Z_t is marginally standard normal. The time series $\{X_t;\ t=1,2,\ldots\}$ is obtained via the transformation

$$X_t=F^{-1}(U_t)=F^{-1}[\Phi(Z_t)],$$

which ensures that X_t has distribution F by the well-known properties of the inverse cumulative distribution function. We note that the inverse cumulative distribution function method is an essential ingredient in the construction of the ARTA input process and it works for *any* marginal distribution, although F^{-1} might have to be evaluated by an approximate numerical method when it has no convenient closed-form expression.

The challenge in the construction of the ARTA process comes from the fact that the autocorrelations of the input process X_t are not the same as the autocorrelations of the base process Z_t unless X_t is marginally standard normal. Thus, the central problem is to find the base autocorrelation $\rho_Z(h)$ satisfying the following equality for the input autocorrelation $\rho_X(h)$:

$$\begin{aligned}\rho_X(h) &= \mathrm{Corr}[X_0, X_h] = \mathrm{Corr}\big[F^{-1}[\Phi(Z_0)], F^{-1}[\Phi(Z_h)]\big] \\ &= \sigma^{-2}\bigg\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F^{-1}[\Phi(z_0)]F^{-1}[\Phi(z_h)] \\ &\qquad\qquad \cdot \vartheta_{\rho_Z(h)}(z_0, z_h)\, dz_0\, dz_h - \mu^2\bigg\}, \qquad (1)\end{aligned}$$

where μ and σ^2 are the mean and the variance of the stationary time-series input process X_t and $\vartheta_{\rho_Z(h)}$ is the bivariate standard normal probability density function with correlation $\rho_Z(h)$. When the marginal distribution is uniform on $(0, 1]$, this correlation-matching problem reduces to $\rho_X(h) = 6/\pi \sin^{-1}[\rho_Z(h)/2]$ (Kruskal 1958). For other marginal distributions, solving this problem is not an easy task. Fortunately, the correlation-matching problem has functional properties that allow the development of efficient numerical search procedures to find $\rho_Z(h)$ within a predetermined precision. We refer the reader to Cario and Nelson (1996) for a thorough discussion of those properties and to Song et al. (1996), Cario and Nelson (1998), Chen (2001), and Biller and Nelson (2003) for the procedures that have been developed to date.

2.1.2. Gamma Processes. Over the past two decades there has been considerable research on modeling univariate time series with gamma marginal distributions in fields such as operations analysis, hydrology, and meteorology. As a result, there exist a variety of gamma processes with the ability to generate stationary first-order time series. For example studies, we refer the reader to Jacobs and Lewis (1977), Gaver and Lewis (1980), and Lawrence and Lewis (1980, 1981, 1985). In this paper, we choose to work with the process developed by Lewis et al. (1989) to represent a time series with gamma marginal distribution and positive autocorrelation specified at lag one. The objective is twofold: (i) to investigate how well the characteristics of the gamma marginal distribution are captured by the ARTAFIT algorithm, and (ii) to challenge the autoregressive dependence structure that is central to the construction of the ARTA process. Because the focus of this paper is the evaluation of the ARTAFIT algorithm, we provide the technical details on how to construct the gamma processes, which will serve the purpose of this paper, in the Online Supplement.

2.2. Fitting Time-Series Input Processes for Simulation

In this section, we take the data $\{x_t;\ t = 1, 2, \ldots, n\}$ generated by ARTA and gamma processes as input but assume that the underlying input process is unknown and try to fit one. We do this using three different algorithms, each of which assumes a different input model. The first one, which we call the MARGINAL algorithm, ignores dependence, assumes an i.i.d. process, and uses the least squares estimation method to determine the parameters of a Johnson distribution (Swain et al. 1988). In Biller and Nelson (2005), we show that the resulting parameter estimates have asymptotical properties such as strong consistency and normality as long as data are generated by an i.i.d. Johnson input process. Although independence is, in fact, the assumption most commercial input-modeling software packages make, we provide experimental results later in §4 demonstrating the failure of this assumption in providing a good input-process representation when dependence actually exists.

The second algorithm, which Cario and Nelson (1996) call ARTAFACTS, is composed of two steps. In the first step, it determines the parameters of a Johnson marginal distribution using the least squares estimation method under the assumption of independence, i.e., simply the MARGINAL algorithm. In the second step, the autocorrelations are estimated using the sample autocorrelation function defined as

$$\hat\rho_X(h) = \frac{\sum_{i=1}^{n-h}(x_i - \hat\mu)(x_{i+h} - \hat\mu)}{\sum_{i=1}^{n}(x_i - \hat\mu)^2},$$

where $\hat\mu = \sum_{i=1}^{n} x_i/n$ is the sample mean of the time series (Wei 1990). It then solves the correlation-matching problem (1) to find the base autocorrelations $\hat\rho_Z(h)$, $h = 1, 2, \ldots, p$ corresponding to the underlying input autocorrelations $\hat\rho_X(h)$, $h = 1, 2, \ldots, p$. We pick a value for the order of dependence p by minimizing the criterion $\ln|\tilde\sigma_Y(p)| + [\ln(n)/n]p$, where $\tilde\sigma_Y(p)$ denotes the (pseudo) maximum likelihood estimator of σ_Y for an autoregressive model of order p. This simple procedure, developed by Schwarz (1978), is known to find the true order with probability one as the sample size goes to infinity. After the identification of the order of dependence and the base autocorrelations, we solve for the autoregressive coefficients of the underlying ARTA process using Yule-Walker equations (Wei 1990).

The last algorithm, which we call ARTAFIT, was developed by Biller and Nelson (2005) for fitting ARTA processes with marginal distributions from the Johnson translation system to stationary univariate time-series data, i.e., estimating the vector of ARTA parameters denoted by $\boldsymbol\psi = (\boldsymbol\theta, \alpha_1, \alpha_2, \ldots, \alpha_p)'$, where

$\boldsymbol{\theta}$ is the vector of marginal distribution parameters (e.g., $\boldsymbol{\theta} = (\gamma, \delta, \lambda, \xi)$ for the Johnson marginal distribution of type f and $\boldsymbol{\theta} = (k, \beta)$ for a gamma marginal distribution) and p is the order of dependence to be estimated from stationary univariate time-series data. The key to the construction of the ARTAFIT algorithm is to focus on the distributional properties of the white-noise process $\{Y_t;\ t = p+1, p+2, \ldots\}$ of the base process $\{Z_t;\ t = 1, 2, \ldots\}$. In other words, the key is to find the ARTA parameters ensuring that random variables Y_t, $t = p+1, p+2, \ldots$ are i.i.d. and Gaussian. In Biller and Nelson (2005), we do this by using the standardized white-noise process

$$V_t(\boldsymbol{\psi}) = \frac{Y_t}{\sigma_Y} = \frac{Z_t - \sum_{h=1}^{p} \alpha_h Z_{t-h}}{\sqrt{1 - \sum_{h=1}^{p} \alpha_h \rho_Z(h)}}, \quad t = p+1, p+2, \ldots, n,$$

the probability integral transformation $Z_t = \Phi^{-1}[F_{\boldsymbol{\theta}}(X_t)]$, and searching for the parameters that make $\Phi[V_t(\boldsymbol{\psi})]$, $t = p+1, p+2, \ldots, n$ appear to be a sample of i.i.d. uniform random variables on $(0, 1]$ by minimizing the objective function

$$S(\boldsymbol{\psi}) = \frac{1}{(n-p)^2} \sum_{t=p+1}^{n} \frac{(n-p+1)^2(n-p+2)}{(t-p)(n+1-t)} \cdot \left(\Phi[V_{(t)}(\boldsymbol{\psi})] - \frac{t-p}{n-p+1} \right)^2$$

subject to the constraint $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is the feasible region for the parameters of the marginal distribution function and the autoregressive coefficients of the base process. We refer the reader to Biller and Nelson (2005) for the derivation of this objective function utilizing the theory behind the translated uniform order statistics.

In Biller and Nelson (2005), we show that the objective function $S(\boldsymbol{\psi})$ is a three-times continuously differentiable function for every $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ and that the objective function $S(\boldsymbol{\psi})$ is convex around any unconstrained local minimum. We also show that using a general-purpose optimization algorithm with local convergence properties ensures that we reach a local minimum solution when we start in its convex surrounding region. Although these results have been obtained under the assumption of Johnson marginal distributions, the distribution-free formulation of the ARTAFIT algorithm ensures its convergence also for gamma marginals. When the marginal distribution of interest comes from the Johnson translation system, we additionally prove the strong consistency of location and scale parameters of the Johnson marginal distribution and the strong consistency of the estimators of the parameters of the marginal distribution conditional on knowing the true autoregressive base process. Similar statistical properties extend to the case in which the ARTAFIT algorithm is implemented assuming a gamma marginal distribution. We further empirically observe that solving $S(\boldsymbol{\psi})$ for any fixed, feasible γ, δ, λ, and ξ provides robust estimates of $\alpha_1, \alpha_2, \ldots, \alpha_p$. Therefore, an effective way to minimize $S(\boldsymbol{\psi})$ subject to $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ is to decompose the optimization problem into the estimation of the marginal distribution parameters and the estimation of the base process parameters and work iteratively between improving the estimates for $(\gamma, \delta, \lambda, \xi)$ and $(\alpha_1, \alpha_2, \ldots, \alpha_p)$. We provide the details on the ARTAFIT algorithm based on such an iterative procedure in the Online Supplement.

3. Experimental Design

Our objective is to compare the performance of the ARTAFIT algorithm to the MARGINAL and ARTAFACTS algorithms in capturing the key characteristics of two different sets of data: (i) data that come from ARTA processes with Johnson marginals, and (ii) data generated by gamma processes. The purpose of experimenting with the first data set is to analyze the sensitivity of recovering the true ARTA processes with respect to the experimental factors of §3.1, while the objective of using the second data set is to see how well the ARTA model captures the characteristics of a process that we know is not ARTA. Next, we discuss the experimental factors and the performance metrics for the evaluation of the resulting fits.

3.1. Factor Selection

In this section, we provide a brief discussion on the selection of the sample size and the characteristics of the marginal distribution and the autocorrelation structure as the experimental factors.

3.1.1. Sample Size. Although we proved the statistical properties of the ARTAFIT estimators as the sample size n approaches infinity in Biller and Nelson (2005), we do not know much about the small-sample properties of these estimators. Therefore, we let the sample size n take the values of 30, 50, 100, 500, 1,000, 5,000, and 10,000 and investigate how the goodness of the resulting fits change as a function of the size of the data samples.

3.1.2. Marginal Distribution. When the data come from an ARTA process with Johnson marginals, we implement the ARTAFIT algorithm assuming Johnson marginals, and when the data come from a gamma process, we implement the algorithm using both Johnson and gamma marginals. The reason behind using the Johnson distribution is its ability to represent any feasible, finite first four moments, providing a great deal of flexibility that is sufficient for many practical problems. We refer the reader to the

Online Supplement for the shapes of the Johnson density functions for which we summarize our findings in §4.1. Although heavy-tailed distributions cannot be adequately represented by the Johnson distributions due to the finiteness of the moments, Johnson (1949a) describes how one could modify the Johnson cumulative distribution function to attain infinite moments. However, this is beyond the scope of this paper.

Similarly, the reason behind using the gamma distribution in our comparative study is its ability to model a wide variety of distributional shapes for random quantities with positive values. Although any gamma distribution is specified by the shape parameter $k > 0$ and the scale parameter $\beta > 0$, the latter is known to have no impact on the coefficient of skewness, the coefficient of kurtosis, and the coefficient of variation. Thus, we take $\beta = 1$ in the remainder of the paper and experiment with different values of the shape parameter k. We refer the reader to the Online Supplement for the shapes of the gamma density functions for which we summarize our findings in §4.2.

3.1.3. Autocorrelation Structure. We characterize the autocorrelation structure of an input process through the order of dependence p that we control and its autoregressive coefficients defined up to order p, i.e., α_h for $h = 1, 2, \ldots, p$. For any given value of p, we choose α_h, $h = 1, 2, \ldots, p$ in such a way that the underlying base process is stationary. In other words, the reverse characteristic polynomial has no roots in or on the complex unit circle; i.e., $1 - \sum_{h=1}^{p} \alpha_h z^p \neq 0$ for $|z| \leq 1$ (Wei 1990). This is an important condition to satisfy as it ensures the stationarity of the input process (Cario and Nelson 1996). We note that $p = 0$ corresponds to an independent process.

Next, we present three different experimental factors, each of which is associated with a different aspect of the autocorrelation structure: strength of dependence, pattern of dependence, and form of dependence.

Strength of Dependence. We use this factor to summarize in a single number the autocorrelation structure jointly specified by the order of dependence and the autoregressive coefficients. We denote the strength of dependence by η and define it as

$$\eta = 1 + 2 \lim_{n\to\infty} \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) |\rho_X(h)|. \qquad (2)$$

Clearly, η measures the rate of decay in the absolute magnitudes of the autocorrelations as seen on a correlogram. In the case of positive dependence, i.e., $\rho_X(h) \geq 0$ for $h \geq 1$, it can further be thought of as the number of dependent observations equivalent to one independent observation. It is worth noting that an autoregressive base process with higher order of dependence does not necessarily imply a time series with stronger dependence. For example, η is equal to 2.636 for $\rho_X(1) = 0.45$ when $p = 1$ and 2.211 for $\rho_X(1) = 0.35$ and $\rho_X(2) = 0.15$ when $p = 2$. On the other hand, η is 3.831 for $\rho_X(1) = 0.45$ and $\rho_X(2) = 0.35$ when $p = 2$. In our experiments, we choose the value of p together with α_h, $h = 1, 2, \ldots, p$, in such a way that larger values of p result in larger values for η.

Pattern of Dependence. Figure 1 provides the evolution of the autocorrelations of two first-order ($p = 1$) autoregressive processes, one with $\rho_X(1) = 0.50$ and the other with $\rho_X(1) = -0.50$, as the lag index takes values between 0 and 10. Because the absolute magnitudes of the autocorrelations decay at the same rate for both of these processes, the values for the strength of dependence η are the same. Clearly, we cannot identify whether the autocorrelations of the process with $\rho_X(1) = -0.50$ decay with an alternating pattern by looking at the strength of dependence. Thus, we consider the pattern of dependence as another experimental factor and investigate how well the 2^p such sign patterns can be successfully

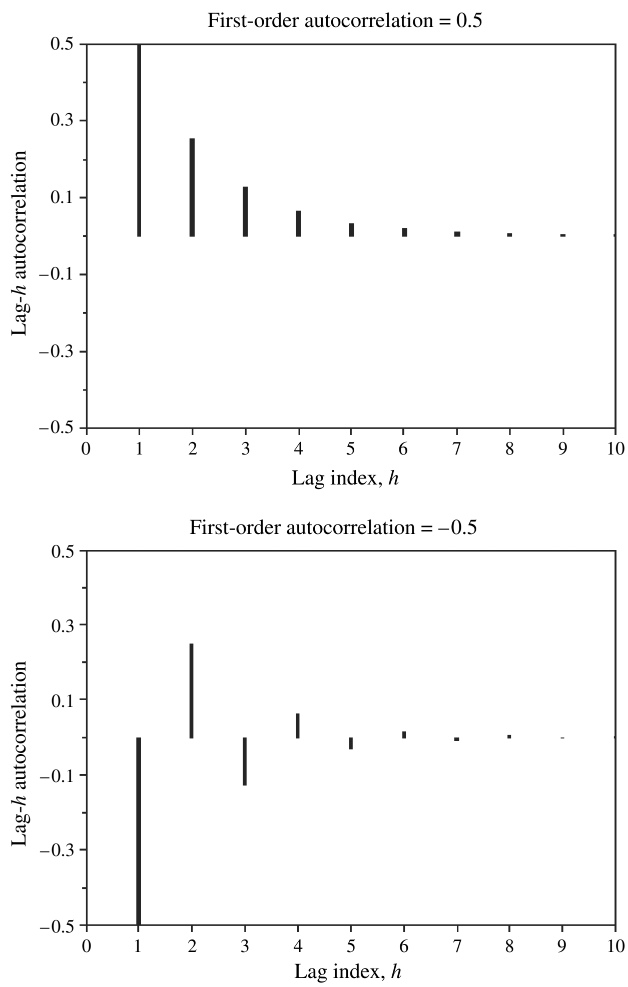

Figure 1 Autocorrelations of First-Order Autoregressive Processes with Different Patterns

identified by ARTAFACTS and ARTAFIT algorithms when data come from autoregressive processes with orders $p = 1, 2, 3$.

Form of Dependence. The objective of this paper is not only to investigate the performance of the ARTAFIT algorithm when data come from a true ARTA process, but also to assess its performance against data generated by a process we know that is not ARTA. Thus, we choose the form of dependence as another experimental factor and use the gamma process, which allows a mix of autoregressive and moving average forms of dependence, to investigate the impact of this factor on the goodness of the resulting fits.

3.2. Performance Metric Selection

In this section, we present the performance metrics we use to evaluate the fits resulting from the implementation of the MARGINAL, ARTAFACTS, and ARTAFIT algorithms. Our focus is on the evaluation of the fitted marginal distribution, autocorrelations, and joint distribution.

To calculate the goodness of the marginal distribution fit $\hat{F}$, we use the Kolmogorov-Smirnov and Anderson-Darling test statistics that we denote by K-S and A-D, respectively. Both of these tests compare the fitted cumulative distribution function $\hat{F}$ to the empirical cumulative distribution function F_n. While the K-S test statistic corresponds to the largest distance between $F_n(x)$ and $\hat{F}(x)$, i.e., $\sup_x\{|F_n(x) - \hat{F}(x)|\}$, the A-D test statistic is the weighted average of the squared differences $[F_n(x) - \hat{F}(x)]^2$, where the weights are the largest for $\hat{F}(x)$ close to zero and one. Thus, the A-D test is particularly good at detecting discrepancies in the tails. Because the critical values and the corresponding nominal levels of significance of these tests for i.i.d. data can be grossly incorrect when observations are dependent (Moore 1982, Gleser and Moore 1983), we use the 5% critical values for the K-S and A-D test statistics (i.e., 0.895 and 0.751) only as a rough guide for judging the adequacy of the marginal distribution fits.

We evaluate the goodness of the fit of the autocorrelation structure by comparing the strength of dependence of the true process (η) to the strength of dependence implied by the fitted input model ($\hat{\eta}$). In particular, we track the percent absolute difference between η and $\hat{\eta}$ (i.e., $|\eta - \hat{\eta}|/\eta * 100$). Obviously, the convergence of $\hat{\eta}$ to η does not guarantee the convergence of the estimated (individual) autocorrelations to the true autocorrelations, but comparing $\hat{\eta}$ to η summarizes the comparison between the true and fitted autocorrelation structures via the use of a single test statistic.

Because a dependent input model is jointly characterized by the marginal distribution function and the autocorrelation structure, neither a pure marginal fit nor a pure autocorrelation fit is sufficient for choosing a good representation. To support our observations with a test that measures the goodness of the joint distribution fit, we use the two-dimensional K-S test that is a generalization of the one-dimensional K-S test to bivariate distributions. Using this test, we measure how well the fitted ARTA distribution of the random variables X_t and X_{t-h}, which is jointly characterized by marginal cumulative distribution function $\hat{F}$ and lag-h autocorrelation $\hat{\rho}_X(h)$, represents the joint empirical distribution function. Under the assumption of a Johnson marginal distribution, the corresponding joint distribution is simply the bivariate Johnson distribution as in Johnson (1949b). To compute the two-dimensional K-S test statistic, we first calculate the fraction of data points falling in each of the four natural quadrants around each point (x_t, x_{t-h}), $t = h+1, h+2, \ldots, n$, and then calculate the probabilities of $(X_t > x_t, X_{t-h} > x_{t-h})$, $(X_t < x_t, X_{t-h} > x_{t-h})$, $(X_t < x_t, X_{t-h} < x_{t-h})$, and $(X_t > x_t, X_{t-h} < x_{t-h})$ for the same set of points using the fitted parameters $\hat{F}$ and $\hat{\rho}_X(h)$ for $h = 1, 2, \ldots, \hat{p}$. By ranging both over the data points and over the quadrants, we take the maximum difference of the corresponding probabilities as the two-dimensional K-S test statistic. We refer the reader to Press et al. (1992) for further details such as the computation of the significance levels for this test as well as the computer codes that can be readily integrated into any software for the calculation of the two-dimensional K-S test statistics.

4. Experimental Results

We present only the representative results here from the comprehensive empirical comparison/analysis performed. The plots of this section are drawn using the test statistics introduced in the previous section and averaged over a number of replications of the entire experiment determined as follows: each experiment starts with an initial run of 30 replications, and the number of replications is increased whenever necessary to ensure an absolute error of no more than 0.1 on both the one-dimensional K-S test statistic and the two-dimensional K-S test statistic. We note that the plots of this section are obtained by implementing the MARGINAL, ARTAFACTS, and ARTAFIT algorithms via the use of portable C++ codes and then connecting the test statistics computed for the experimental settings of interest; i.e., these plots are not obtained by fitting curves to the corresponding test statistics.

4.1. Experimenting with Data from True ARTA Processes

In this section, we experiment with data that come from ARTA processes with Johnson marginals. We present our findings in the following three subsections experimenting with data generated from a Johnson bounded distribution whose shape, location, and scale parameters are taken as $\gamma = 0$, $\delta = 2$, $\xi = 0$, and $\lambda = 1$ and whose probability density function is available in the Online Supplement. Our selection of a Johnson bounded distribution for the presentation of the representative results here is motivated by the challenge of finding good fits for the bounded family.

The findings reported in this section are obtained as a result of using MARGINAL, ARTAFACTS, and ARTAFIT algorithms to fit input models to data we generate from processes with strength of dependence ranging between $\eta = 1$ and $\eta = 38.850$ that correspond to an independent process (i.e., $p = 0$) and a first-order autocorrelated process with $\rho_Z = 0.95$, respectively.

Note that although the experimental factors and the performance metrics we choose to work with hold for any order of dependence, we present representative results for first-order autoregressive processes in this section. We choose to do that to allow the comparison of the goodness of the fits reported in this section to the fits obtained as a result of experimenting with first-order (non-ARTA) gamma processes. The experimental results obtained for higher orders of dependence are consistent with the ones reported in this paper.

4.1.1. Goodness of the Marginal Fits. We present the plots of the K-S test statistic versus the sample size for different levels of the strength of dependence (as denoted by η in the upper right corner) in Figure 2. The top plot is obtained from the implementation of the MARGINAL and ARTAFACTS algorithms, while the bottom plot is given by the ARTAFIT algorithm. We present the plots obtained similarly for the A-D test statistic in Figure 3.

We observe that the goodness of the marginal distribution fits deteriorates with increasing values of the strength of dependence. If we compare the test statistics observed on the y-axes to the 5% critical values for the K-S and A-D test statistics (only as a rough guide), we observe that the K-S test statistics obtained from MARGINAL, ARTAFACTS, and ARTAFIT algorithms are quite comparable as long as $\eta \leq 7.000$, but capturing the tail behavior as represented by the

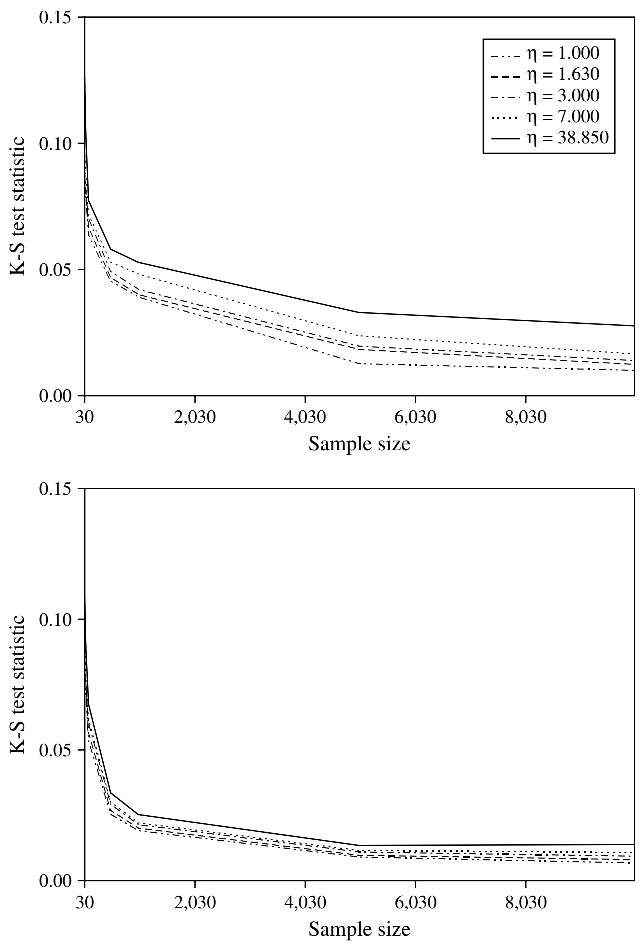

Figure 2 K-S Test Statistics Obtained from MARGINAL/ARTAFACTS (Top) and ARTAFIT (Bottom) Algorithms

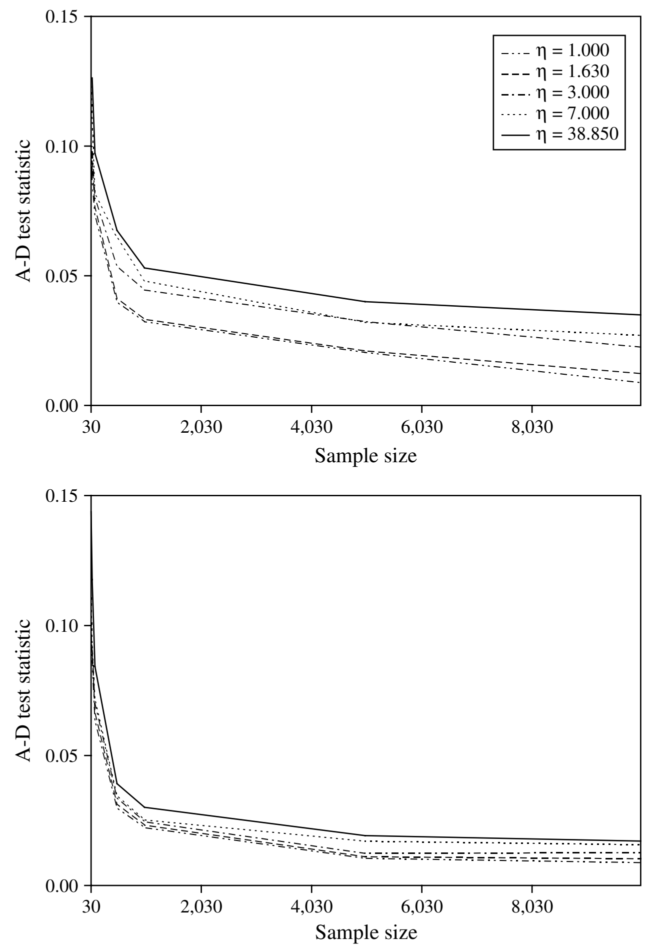

Figure 3 A-D Test Statistics Obtained from MARGINAL/ARTAFACTS (Top) and ARTAFIT (Bottom) Algorithms

A-D test statistics in Figure 3 poses a serious challenge, especially under the assumption of a strongly autocorrelated true process. Note that the strength of dependence shows its impact on the deteoriation of the K-S test statistics at $\eta = 38.850$ in Figure 2, while we identify a lower threshold (i.e., $\eta = 3.000$) beyond which the negative impact of the strength of dependence on the goodness of the MARGINAL and ARTAFACTS fits is obvious. In all of the experiments we have performed, we have observed the superior performance of the ARTAFIT algorithm showing itself much earlier in the A-D test statistics than in the K-S test statistics, especially when strong dependencies exist in the input process data. Thus, we conclude that the ARTAFIT algorithm performs better than both the MARGINAL and ARTAFACTS algorithms, particularly in capturing the tail behavior of the underlying input process.

4.1.2. Goodness of the Autocorrelation Fits. When we use the MARGINAL algorithm to estimate the parameters of the underlying input model, we have $\hat{\eta} = 1$ following directly from the assumption of independence; i.e., $\hat{\rho}_X(h) = 0$ for $h \geq 1$. Clearly, ignoring dependence leads to very poor estimates for η unless the process generating the data is independent. Therefore, we will restrict our attention to the autocorrelation fits obtained from the ARTAFACTS and ARTAFIT algorithms in the remainder of the section.

We present the percent absolute difference between the true η and the $\hat{\eta}$ estimated by the ARTAFACTS and ARTAFIT algorithms for different values of the

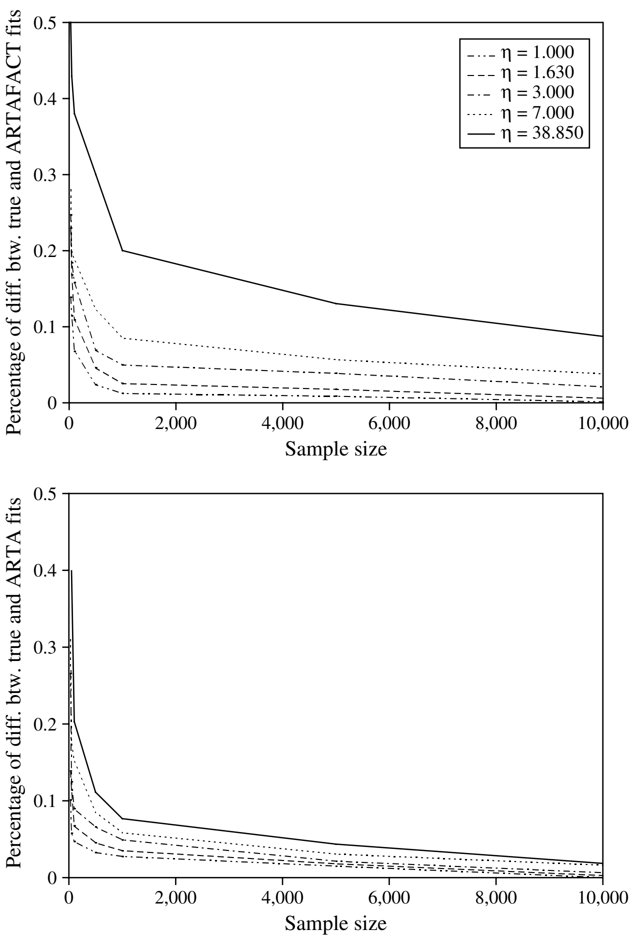

Figure 4 Percent Difference Between η and $\hat{\eta}$ Estimated by ARTAFACTS (Top) and ARTAFIT (Bottom) Algorithms

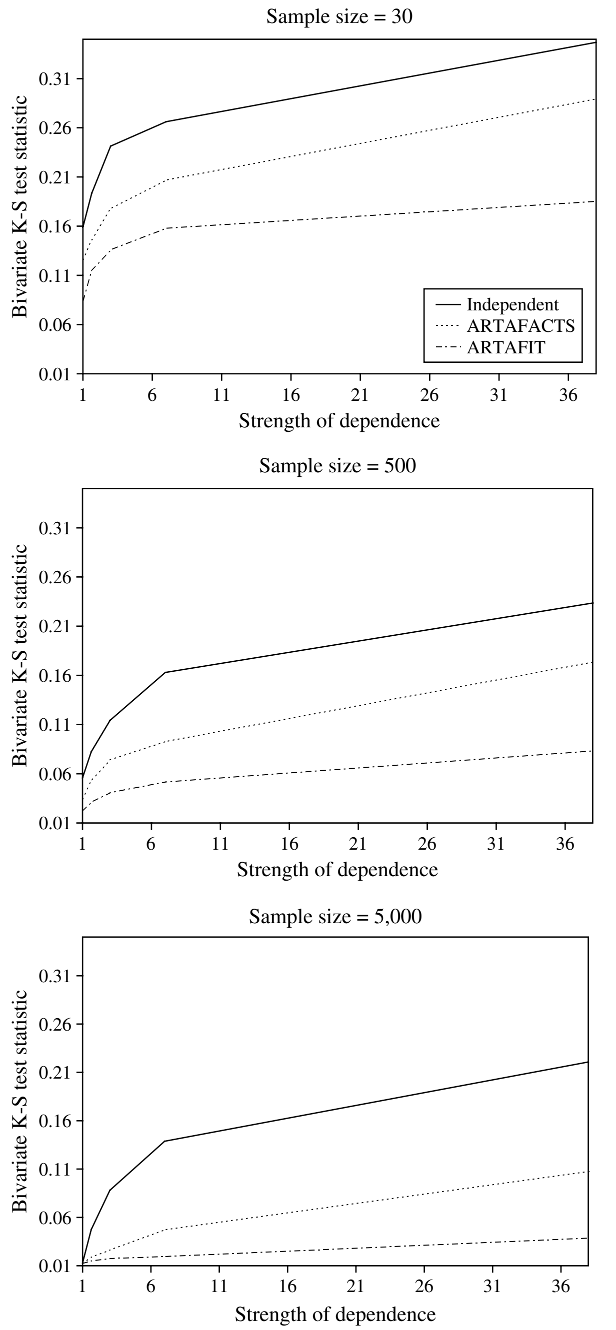

Figure 5 Two-Dimensional K-S Test Statistics Obtained from the Implementation of the MARGINAL, ARTAFACTS, and ARTAFIT Algorithms

sample size and the true strength of dependence in Figure 4. We observe that both the ARTAFACTS algorithm and the ARTAFIT algorithm perform quite well in predicting the true value for η when the strength of dependence takes low to medium values, but the performance of the ARTAFIT algorithm is slightly better. For increasing values of the strength of dependence, we observe that the ARTAFIT algorithm outperforms the ARTAFACTS algorithm in representing the autocorrelation structure of the underlying input processes.

4.1.3. Goodness of the Joint Distribution Fits. We present the comparison of the two-dimensional K-S test statistics (at lag one) obtained from the MARGINAL, ARTAFACTS, and ARTAFIT algorithms for different values of strength of dependence and sample size in Figure 5. Each of the three plots provided here assume a different sample size, 30, 500, and 5,000, that we consider as small, medium, and large sample sizes. We observe that the two-dimensional K-S test statistics of the three fits are closest to each other when the true process is independent, especially in the large-sample case. As the strength of dependence increases, the failure of the independence assumption in developing good input models is obvious. While the ARTAFACTS algorithm performs significantly better than the MARGINAL algorithm in capturing the underlying joint distribution, the ARTAFIT algorithm provides the best fits by far. As the sample size and the strength of dependence increase, we further observe that the test statistics obtained from the ARTAFIT algorithm increase at a rate that is significantly smaller than the rate observed for the fits of the MARGINAL and ARTAFACTS algorithms. Thus, the ARTAFIT algorithm is much better suited for solving the data-fitting problem when we have a large number of data points in our samples.

It is interesting how the fits of the ARTAFACTS algorithm compare to the fits of the MARGINAL and ARTAFIT algorithms for different sample sizes. The test statistics obtained from the ARTAFACTS algorithm are pretty close to the test statistics of the independent model in the small sample, while they improve and get closer to the test statistics of the ARTAFIT algorithm as the sample size gets larger. This can be explained by the strong consistency of the sample autocorrelation function (Wei 1990). However,

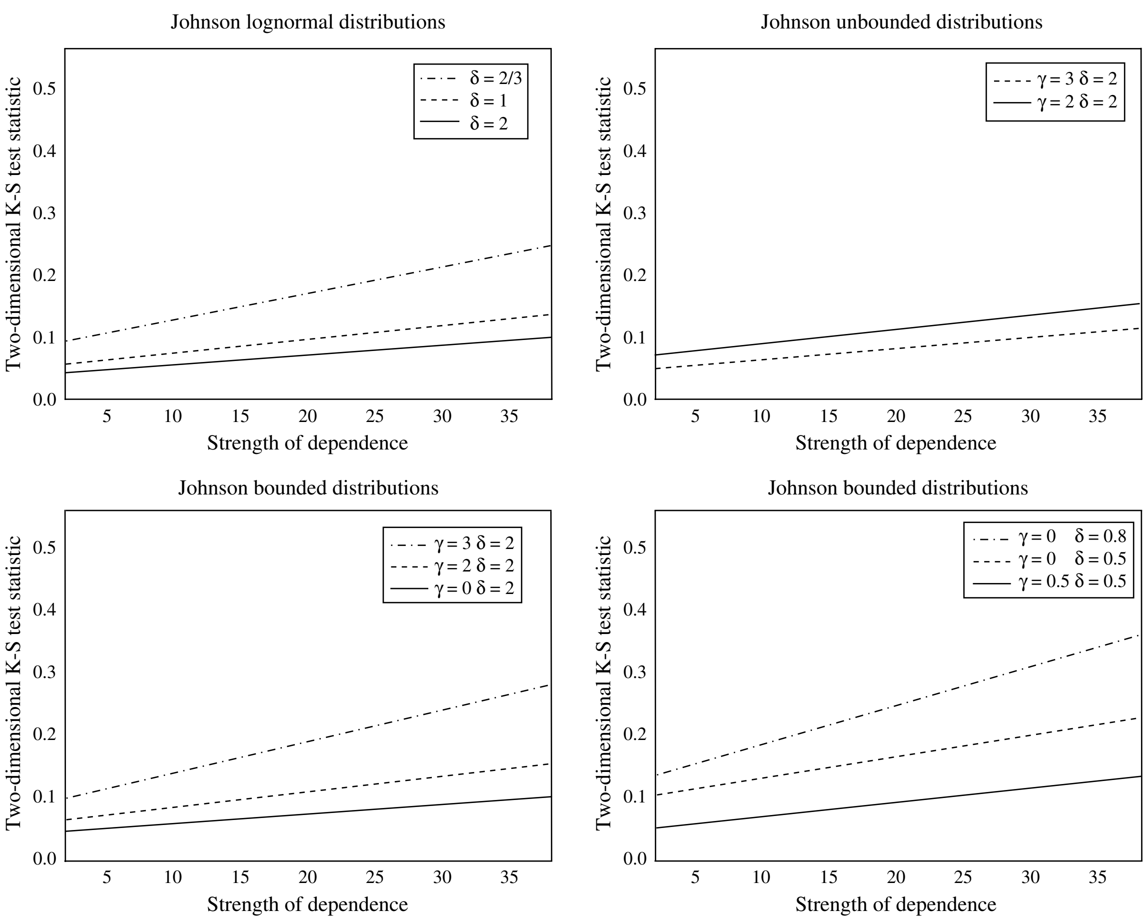

Figure 6 Two-Dimensional K-S Test Statistics Obtained from the ARTAFIT Algorithm

the impact of the strength of dependence dominates the impact of the sample size and the discrepancy between the two-dimensional K-S test statistics of the ARTAFACTS and ARTAFIT results increases as the autocorrelations of the experimental data samples get stronger.

In the remainder of the section, we investigate the impact of the marginal distribution characteristics on the goodness of the ARTA fits. We present our findings in Figure 6 with respect to different levels of the strength of dependence in samples of 5,000 data points for each one of the Johnson probability density functions presented in the Online Supplement. The subtitles and the legends in Figure 6 provide the descriptions of the Johnson marginal distributions to which the results belong. It is important to note that the results presented in the lower left corner of Figure 6 are obtained as a result of experimenting with unimodal Johnson bounded distributional shapes, while bimodal Johnson bounded marginal distributions are used to obtain the results reported in the lower right corner. We find that the joint distribution fits are quite sensitive to the skewness as well as the bimodality in the marginal distribution shapes. As the skewness decreases leading to a symmetric probability density function, we observe significant improvement in the goodness of the fits obtained for every Johnson family from which we generated experimental data.

Although we include the pattern of dependence as a factor in our experimental design, we do not provide any plots summarizing its impact because we have not found it to be significant on the goodness of the resulting fits. We have not encountered a case in which the ARTAFIT algorithm incorrectly identifies the pattern of dependence. Thus, we conclude that it is the strength, rather than the pattern, of dependence that has critical impact on the resulting fits.

4.2. Experimenting with Data from Gamma Processes

In this section, we experiment with data that come from first-order gamma processes. Our objective is to assess the performance of the ARTAFIT algorithm in capturing the key characteristics of the underlying model that we know is not ARTA.

We first assume that the type of the marginal distribution is known. Thus, we implement the ARTAFIT algorithm assuming a gamma distribution with

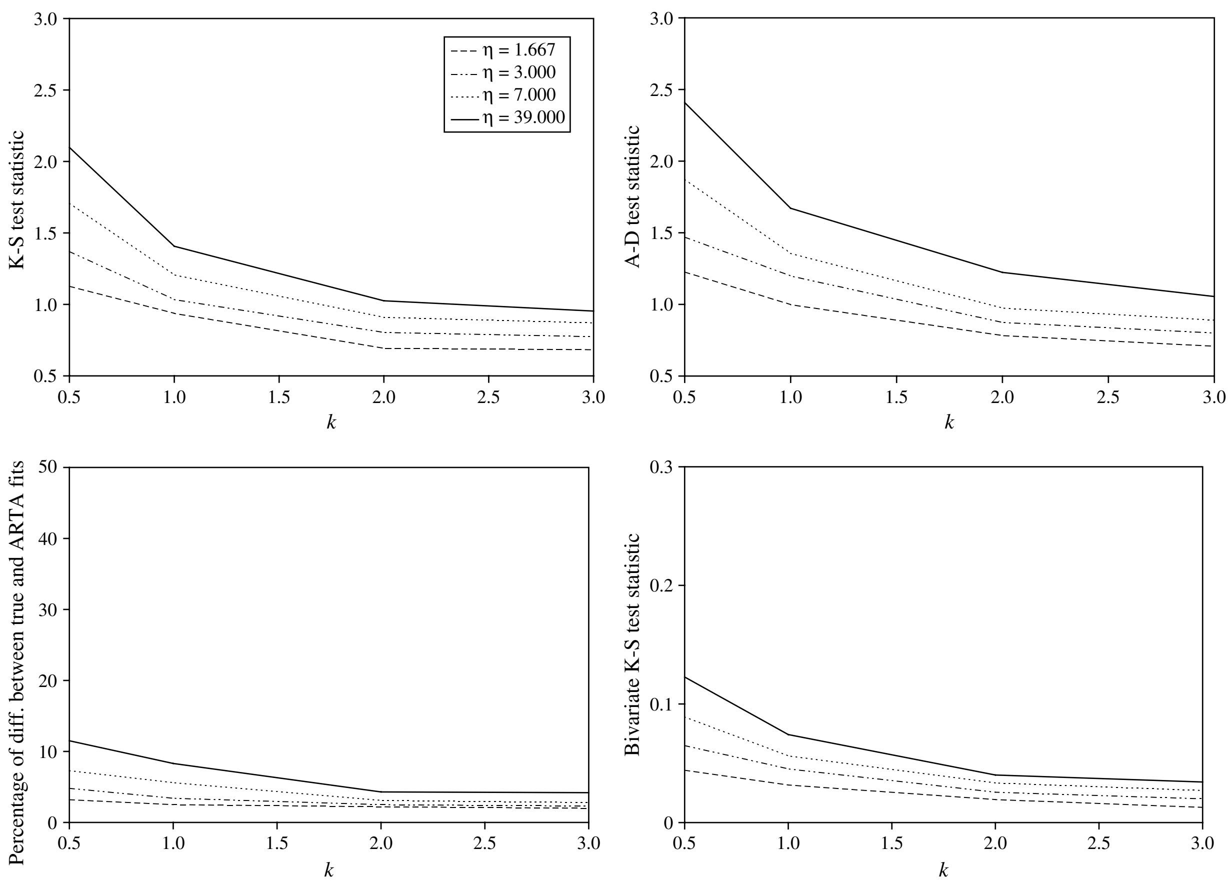

Figure 7 The Goodness of the ARTA Models with Gamma Marginals Fitted to Data from a Gamma Process

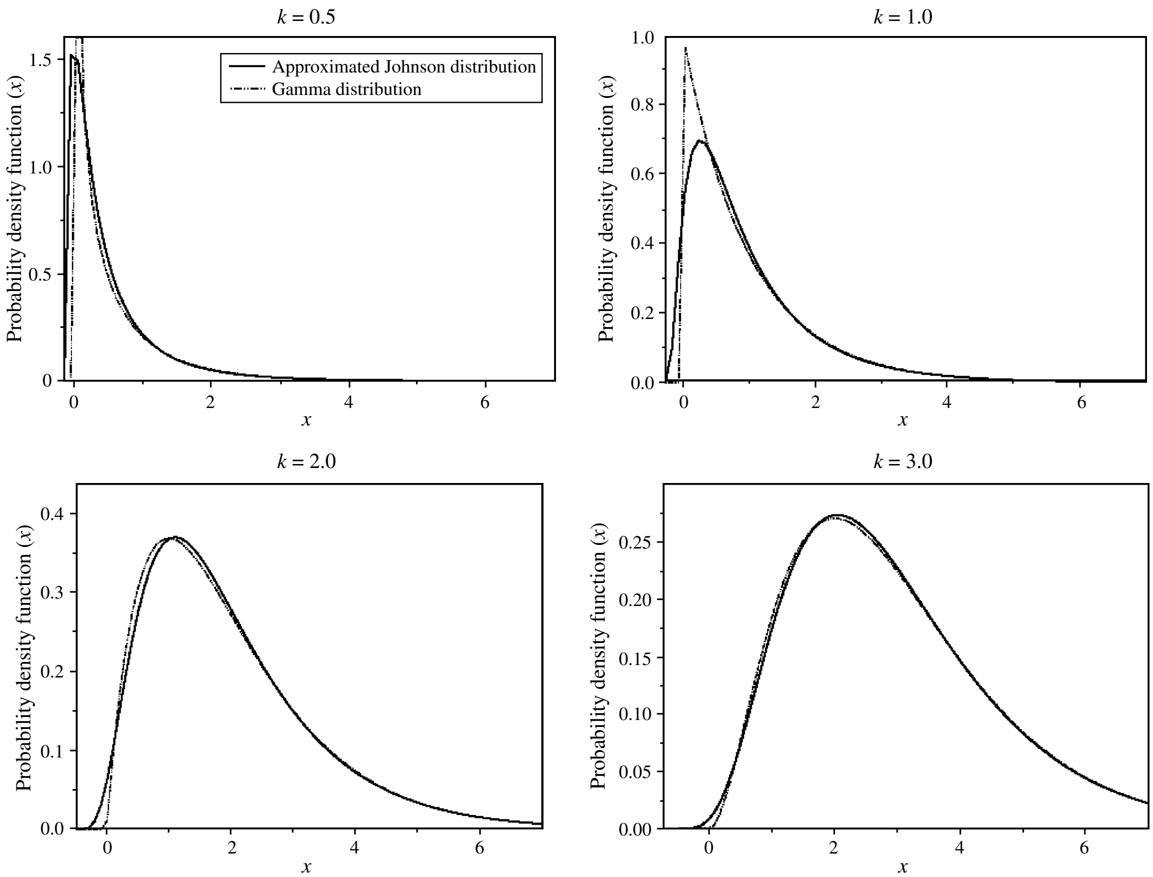

Figure 8 Comparison of the Gamma Density Functions with Approximating Johnson Distributions

unknown shape and scale parameters. We present representative plots of our findings for different values of the shape parameter (k) and the strength of dependence (η) in Figure 7. Taking the sample size as 5,000 and assuming that the underlying marginal distribution is gamma in all cases, we provide plots corresponding to the K-S test statistic, the A-D test statistic, the percent absolute difference between (true) η and (fitted) $\hat{\eta}$, and finally the two-dimensional K-S test statistic for the random variables that are lag-one apart. These results suggest that the ARTA model captures the underlying Markovian structure reasonably well. However, when the mode of the marginal distribution is near zero, which exists when $k \leq 1$, this poses a serious challenge for the ARTAFIT algorithm.

Next, we assume that we do not know the type of the marginal distribution. Therefore, we use the Johnson translation system to represent the underlying marginal distribution. In this case, we find that the fitted marginal distribution, which is typically chosen

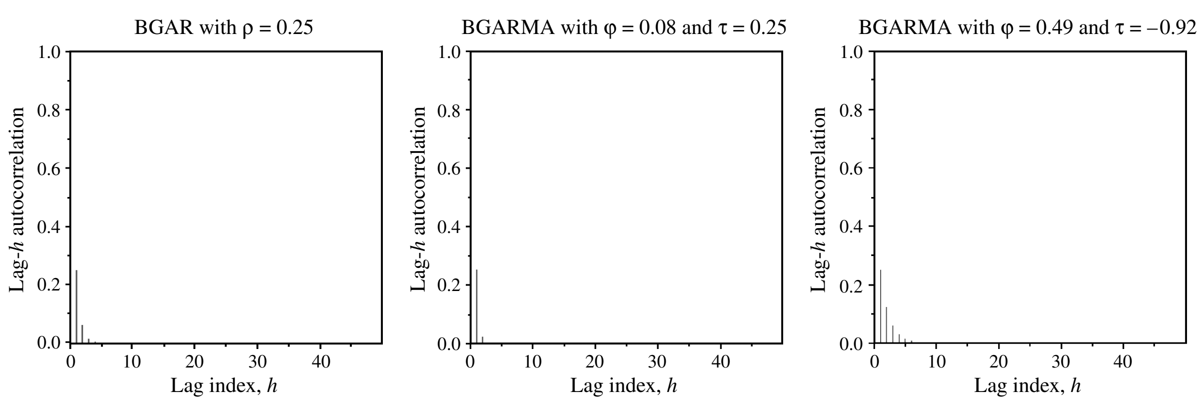

Figure 9 Autocorrelations of BGAR and BGARMA When $\rho_X(1) = 0.25$

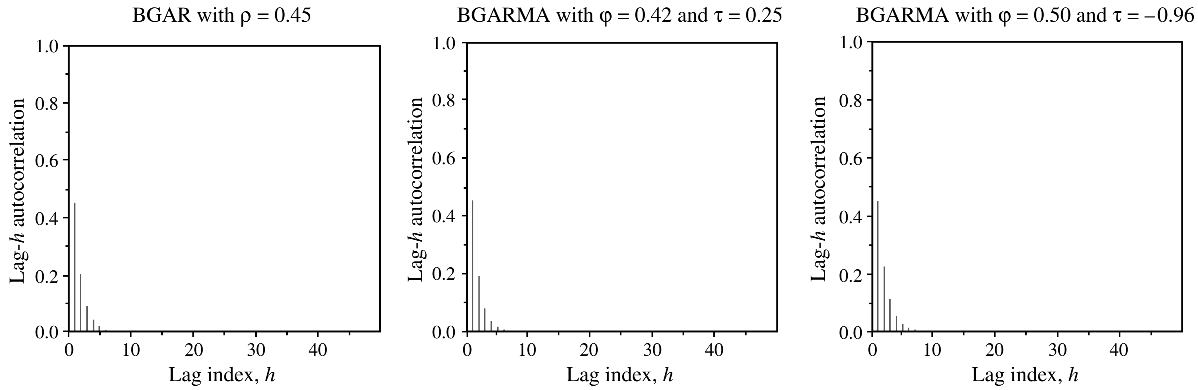

Figure 10 **Autocorrelations of BGAR and BGARMA When $\rho_X(1) = 0.45$**

from the bounded family, allows negative values. It is important to note that we obtain test statistics as good as the ones presented in Figure 7 only when we enforce $\xi = 0$, which ensures positive random variables, in the definition of the underlying Johnson cumulative distribution function. To provide further insight on this problem, we compare the gamma density function to its approximation with the same first four moments in the Johnson translation system for different values of k in Figure 8. We note that this approximation is obtained fitting a Johnson curve to the probability density function of the corresponding gamma random variable via the moment-matching algorithm of Hill et al. (1976). The first row corresponds to the probability density functions with $k = 0.5$ and $k = 1.0$, while the second row has the probability density functions for $k = 2.0$ and $k = 3.0$. The second-row approximations provide reasonably good representation for the gamma distribution, while the first-row approximations fail to capture the mode near zero. However, the lower bounds of all of the approximations go below zero. When we enforce $\xi = 0$ in the definition of the underlying Johnson cumulative distribution function or the constraint $\xi \geq 0$ in the definition of the feasible region for the parameters of the marginal distribution, which also ensures a positive-valued time series, the ARTAFIT algorithm provides significantly better approximations as the probability associated with $\Pr\{X < 0\}$ in the probability density function of the approximated Johnson distribution in Figure 8 is redistributed to ensure $\Pr\{X \geq 0\} = 1$. Thus, it is critical to incorporate any knowledge about the distributional parameters into the constraint set, which in turn is used for the construction of the feasible parameter region, or into the functional form of the Johnson cumulative distribution function when applicable.

Finally, we investigate how the form of dependence affects the goodness of the fits obtained from the ARTAFIT algorithm. To do that, we first generate data from three gamma processes with identical lag-one autocorrelations but different autocorrelation structures, each of which is a mix of moving-average and autoregressive components. Then, we implement the ARTAFIT algorithm under the assumption of a gamma marginal distribution and assess how well an ARTA process represents such a dependence structure. We refer the reader to Figures 9–11 for the illustration of these three gamma processes (BGARMA) as well as their comparison to the gamma processes with no moving-average components (BGAR), but with the same lag-one autocorrelations we chose as 0.25, 0.45, and 0.95, respectively.

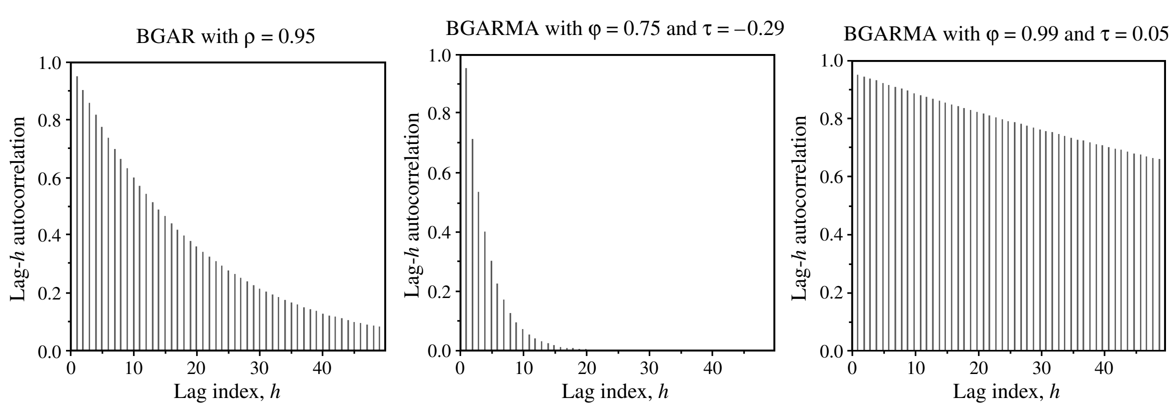

Figure 11 **Autocorrelations of BGAR and BGARMA When $\rho_X(1) = 0.95$**

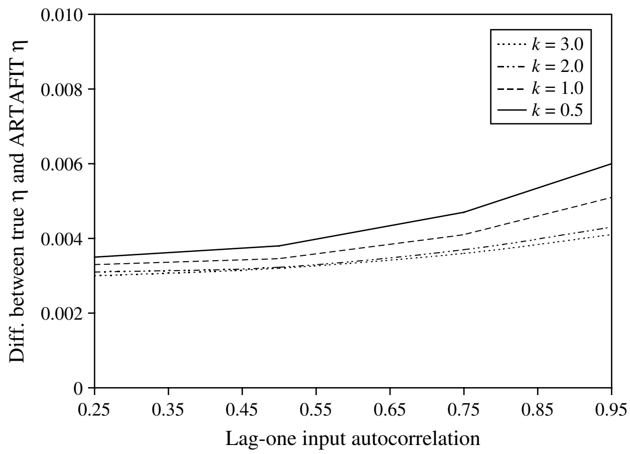

Figure 12 **The Goodness of the ARTA Model Fitted to the BGARMA Data**

Each plot in these figures corresponds to a correlogram, i.e., $\rho_X(h)$ versus h for $h \geq 0$. Clearly, the autocorrelation structure represented in the middle plots decays faster than the autoregressive autocorrelation structure, while this rate is slower in the plots placed in the right-hand side of Figures 9–11, even though they all share the same lag-one autocorrelations.

We present the percent difference in the two-dimensional K-S test statistics for random variables that are lag-one apart in Figure 12. We observe that the ARTAFIT algorithm performs very well in capturing the lag-one autocorrelation of the gamma process. The performance of the algorithm is slightly better for $k > 1$, but the percent difference between the true and fitted lag-one autocorrelations is less than 1% in all cases, which we find quite satisfactory. Next, we investigate how well the ARTAFIT algorithm captures the high-order autocorrelations; i.e., $\rho_X(h)$ for $h \geq 2$. We find that the performance of the algorithm is very much dependent on the strength of dependence. We refer the reader to Figure 13 for the absolute differences between the autocorrelations of the ARTA fit and the true BGARMA process whose lag-one autocorrelations are taken as 0.25, 0.45, and 0.95, and higher-order autocorrelations have been displayed in Figures 9–11. The subtitle in each plot of this figure presents the strength of dependence for the ARTA fit in parentheses, while the legend in each of these plots presents the strength of dependence of the true BGARMA processes with decay rates smaller and larger than the geometric rate of a pure autoregressive process. While our results suggest that the ARTA model is still a plausible model to use for the first two cases in which $\rho_X(1) = 0.25$ and $\rho_X(1) = 0.45$, the ARTA model might fall short in representing a nonautoregressive form of dependence, especially when $\rho_X(1) = 0.95$.

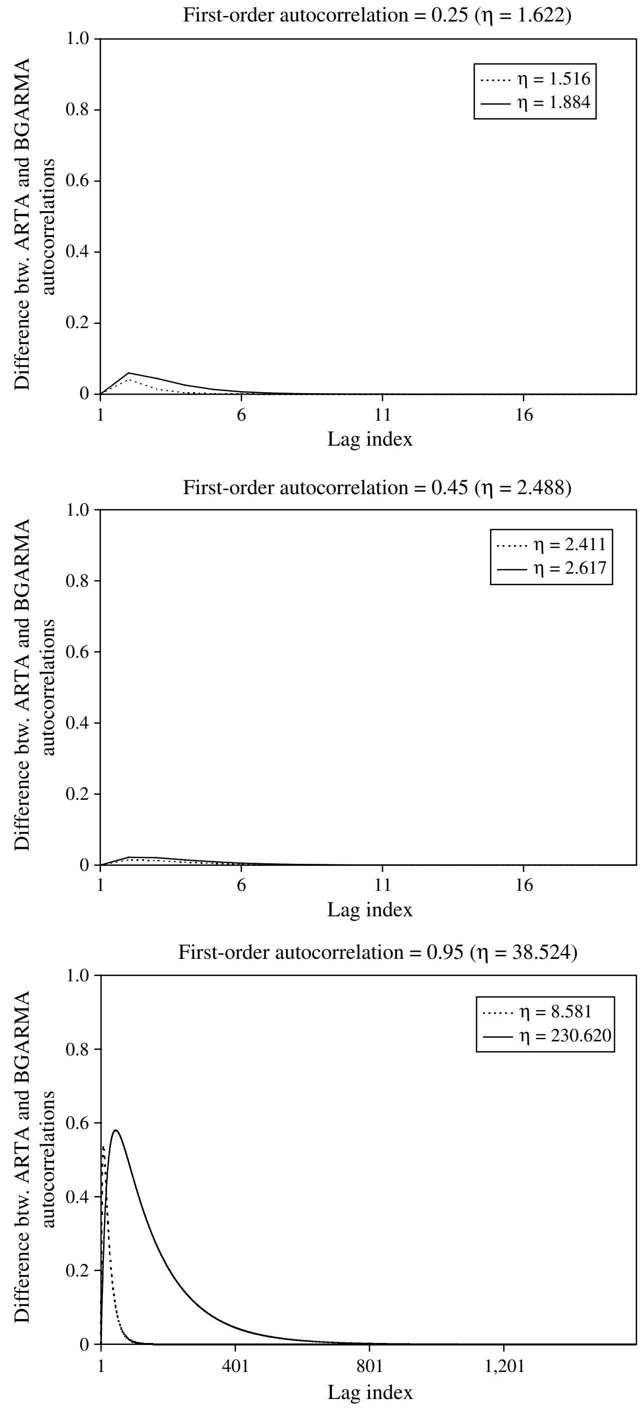

Figure 13 **Comparison of the Autocorrelation Structure of the True BGARMA Process to the Autocorrelation Structure of the Fitted ARTA Process**

5. Conclusion

In this paper, we report the results of a comprehensive numerical study comparing the goodness of fitting independent, ARTAFACTS, and ARTA input models to stationary time-series data with respect to sample sizes and characteristics of the marginal distributions and autocorrelation structures. Our experimental results demonstrate the importance of taking dependencies into account while estimating input processes for simulation applications as well as the use of an algorithm that jointly solves for the marginal distribution and autocorrelation parameters of the underlying input model.

When we generate data from a true ARTA process, we find quite satisfactory fits as a result of using the ARTAFIT algorithm. While we obtain robust fits for the autocorrelation structure, highly skewed bimodal distributional shapes pose serious challenges for the ARTAFIT algorithm. Fortunately, it is possible to improve the performance of the algorithm significantly when one incorporates prior information on the distributional parameters into the constraint set, which in turn is used for the construction of the feasible parameter region of the data-fitting problem.

We observe that the ARTAFIT algorithm outperforms the MARGINAL and ARTAFACTS algorithms in all cases. More specifically, both ARTAFIT and ARTAFACTS give good fits in small-sample cases even though we find that ARTAFIT still performs better than ARTAFACTS. However, the ARTAFIT algorithm begins to distinguish itself once more data become available. One might argue that the implementation of the ARTAFACTS algorithm is much simpler than the ARTAFIT algorithm, and thus, it should be the preferred method of fitting. Considering the importance of good input models on the accuracy of decisions supported by stochastic simulations and the availability of high computing power, we recommend the use of the ARTAFIT algorithm despite the setup required for its implementation. Additionally, what is considered a small sample or a large sample is often dependent on the problem of interest, and thus, we can rely on the ARTAFIT algorithm to provide good fits in all cases.

When we generate data from a gamma process, which we know is not ARTA, the ARTAFIT algorithm performs extremely well in capturing the lag-one joint distributional properties, while it might fall short in representing time series whose autocorrelations decay at nongeometric rates. We believe that this problem can be overcome by simply replacing the underlying autoregressive base process with a mixed autoregressive, moving-average process. This is a subject of future research.

Acknowledgments

The authors thank the referees for providing numerous improvements to this paper.

References

Biller, B., B. L. Nelson. 2003. Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Trans. Model. Comput. Simulation* **13** 1–27.

Biller, B., B. L. Nelson. 2005. Fitting time-series input processes for simulation. *Oper. Res.* **53**(3) 549–559.

Cario, M. C., B. L. Nelson. 1996. Autoregressive to anything: Time-series input processes for simulation. *Oper. Res. Lett.* **19** 51–58.

Cario, M. C., B. L. Nelson. 1998. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS J. Comput.* **10**(1) 72–81.

Chen, H. 2001. Initialization for NORTA: Generation of random vectors with specified marginals and correlations. *INFORMS J. Comput.* **13**(4) 312–331.

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

Gleser, L. J., D. S. Moore. 1983. The effect of dependence on chi-squared test and empiric distribution tests of fit. *Ann. Statist.* **11** 1100–1108.

Hill, I. D., R. Hill, R. L. Holder. 1976. Fitting Johnson curves by moments. *Appl. Statist.* **25** 180–189.

Jacobs, P. A., P. A. W. Lewis. 1977. A mixed autoregressive-moving average exponential sequence and point process (EARMA$(1,1)$). *Adv. Appl. Probab.* **9** 87–104.

Johnson, N. L. 1949a. Systems of frequency curves generated by methods of translation. *Biometrika* **36** 149–176.

Johnson, N. L. 1949b. Bivariate distributions based on simple translation systems. *Biometrika* **36** 297–304.

Kruskal, W. 1958. Ordinal measures of association. *J. Amer. Statist. Assoc.* **53** 814–861.

Law, A. M., W. D. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, New York.

Lawrence, A. J., P. A. W. Lewis. 1980. The exponential autoregressive-moving average process (EARMA(p,q)). *J. Roy. Statist. Soc.* **42** 150–161.

Lawrence, A. J., P. A. W. Lewis. 1981. A new autoregressive time series model in exponential variables (NEAR(1)). *Adv. Appl. Probab.* **13** 826–845.

Lawrence, A. J., P. A. W. Lewis. 1985. Modeling and residual analysis of nonlinear autoregressive time series in exponential variables (with discussion). *J. Roy. Statist. Soc.* **47** 165–202.

Lewis, P. A. W., E. McKenzie, D. K. Hugus. 1989. Gamma processes. *Comm. Statist.–Stochastic Models* **5** 1–30.

Livny, M., B. Melamed, A. K. Tsiolis. 1993. The impact of autocorrelation on queueing systems. *Management Sci.* **39**(3) 322–339.

Melamed, B., J. R. Hill, D. Goldsman. 1992. The TES methodology: Modeling empirical stationary time series. J. J. Swain, D. Goldsman, R. C. Crain, J. R. Wilson, eds. *Proc. 1992 Winter Simulation Conf.*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, 135–144.

Moore, D. S. 1982. The effect of dependence on chi-squared tests of fit. *Ann. Statist.* **4** 357–369.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery. 1992. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, Cambridge, UK.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.

Song, W. T., L. Hsiao, Y. Chen. 1996. Generating pseudorandom time series with specified marginal distributions. *Eur. J. Oper. Res.* **93** 1–12.

Swain, J. J., S. Venkatraman, J. R. Wilson. 1988. Least-squares estimation of distribution functions in Johnson's translation system. *J. Statist. Comput. Simulation* **29** 271–297.

Vincent, S. 1998. Input data analysis. J. Banks, ed. *The Handbook of Simulation*. John Wiley & Sons, New York, 55–91.

Ware, P. P., T. W. Page, B. L. Nelson. 1998. Automatic modeling of file system workloads using two-level arrival processes. *ACM Trans. Model. Comput. Simulation* **8** 305–330.

Wei, W. W. S. 1990. *Time Series Analysis: Univariate and Multivariate Methods*. Addison Wesley, New York.