Modeling and Generating Multivariate Time-Series Input Processes Using a Vector Autoregressive Technique

BAHAR BILLER
Carnegie Mellon University
and
BARRY L. NELSON
Northwestern University

We present a model for representing stationary multivariate time-series input processes with marginal distributions from the Johnson translation system and an autocorrelation structure specified through some finite lag. We then describe how to generate data accurately to drive computer simulations. The central idea is to transform a Gaussian vector autoregressive process into the desired multivariate time-series input process that we presume as having a VARTA (Vector-Autoregressive-To-Anything) distribution. We manipulate the autocorrelation structure of the Gaussian vector autoregressive process so that we achieve the desired autocorrelation structure for the simulation input process. We call this the *correlation-matching problem* and solve it by an algorithm that incorporates a numerical-search procedure and a numerical-integration technique. An illustrative example is included.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]—*time series analysis*; I.6.5 [**Simulation and Modeling**]: Model Development—*modeling methodologies*

General Terms: Experimentation, Languages

Additional Key Words and Phrases: Input modeling, multivariate time series, numerical integration, vector autoregressive process

1. INTRODUCTION

Representing the uncertainty in a simulated system by an input model is one of the challenging problems in the application of computer simulation. There

This research was partially supported by National Science Foundation Grant numbers DMI-9821011 and DMI-9900164, Sigma Xi Scientific Research Society grant number 142, and by Nortel Networks, Symix/Pritsker Division, and Autosimulations, Inc.
Authors' addresses: B. Biller, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213; email: billerb@andrew.cmu.edu; B. L. Nelson, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3119; email: nelsonb@iems.nwu.edu.

are an abundance of examples, from manufacturing to service applications, where input modeling is critical, including modeling the processing times of a workpiece across several workcenters, modeling the medical characteristics of organ-transplant donors and recipients [Pritsker et al. 1995], or modeling the arrival streams of packets in ATM telecommunications networks [Livny et al. 1993]. Building a large-scale discrete-event stochastic simulation model may require the development of a substantial number of, possibly multivariate, input models. Development of these models is facilitated by accurate and automated (or nearly automated) input modeling support. The ability of an input model to represent the underlying uncertainty is essential because even the most detailed logical model combined with a sound experimental design and thorough output analysis cannot compensate for inaccurate or irrelevant input models.

The interest among researchers and practitioners in modeling and generating input processes for stochastic simulation has led to commercial development of a number of input modeling packages, including ExpertFit (Averill M. Law and Associates, Inc.), the Arena Input Analyzer (Rockwell Software Inc.), Stat::Fit (Geer Mountain Software Corporation), and BestFit (Palisade Corporation). These products are most useful when data on the process of interest are available. The approach that they take is to exhaustively fit and evaluate the fit of the standard families of distributions (e.g., beta, Erlang, exponential, gamma, lognormal, normal, Poisson, triangular, uniform, or Weibull), and recommend the one with the best summary measures as the input model. The major drawback of the input models incorporated in these packages is that they emphasize independent and identically distributed (i.i.d.) processes with limited shapes that may not be flexible enough to represent some characteristics of the observed data or some known properties of the process that generates the data. However, dependent and multivariate time-series input processes with nonstandard marginal distributions occur naturally in the simulation of many service, communications, and manufacturing systems (e.g., Melamed et al. [1992] and Ware et al. [1998]). Input models that ignore dependence can lead to performance measures that are seriously in error and a significant distortion of the simulated system. This is illustrated in Livny et al. [1993], who examined the impact of autocorrelation on queueing systems.

In this article, we provide a model that represents dependencies in time sequence and with respect to other input processes in the simulation. Our goal is to match prespecified properties of the input process, rather than to fit the model to a sample of data. More specifically, we consider the case in which the first four moments of all of the marginal distributions, and the autocorrelation structure through some finite lag, are given, and we want to drive our simulation with vector time series that have these properties. The related problem of fitting our model to historical data is addressed in Biller and Nelson [2002, 2003a].

Our input-modeling framework is based on the ability to represent and generate continuous-valued random variates from a stationary k-variate time series $\{\mathbf{X}_t;\ t = 0, 1, 2, \ldots\}$, a model that includes univariate independent and identically distributed processes, univariate time-series processes, and finite-dimensional random vectors as special cases. Thus, our philosophy is

to develop a single, but very general, input model rather than a long list of more specialized models. Specifically, we let each component time series $\{X_{i,t}; i = 1, 2, \ldots, k;\ t = 0, 1, 2, \ldots\}$ have a marginal distribution from the Johnson translation system [Johnson 1949a] to achieve a wide variety of distributional shapes; and we reflect the desired dependence structure via Pearson product–moment correlations, $\rho_{\mathbf{X}}(i, j, h) \equiv \text{Corr}[X_{i,t}, X_{j,t-h}]$, for $h = 0, 1, 2, \ldots, p$. We achieve this using a transformation-oriented approach that invokes the theory behind the standardized Gaussian vector autoregressive process. Therefore, we refer to $\mathbf{X}_t$ as having a VARTA (Vector-Autoregressive-To-Anything) distribution. For $i = 1, 2, \ldots, k$, we take $\{Z_{i,t};\ t = 0, 1, 2, \ldots\}$ to be the ith component series of the k-variate Gaussian autoregressive base process of order p, where p is the maximum lag for which an input correlation is specified. Then, we obtain the ith time series via the transformation $X_{i,t} = F_{X_i}^{-1}[\Phi(Z_{i,t})]$, where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution and F_{X_i} is the Johnson-type cdf suggested for the ith component series of the input process. This transformation-oriented approach requires matching the desired autocorrelation structure of the input process by manipulating the autocorrelation structure of the Gaussian vector autoregressive base process. In order to make this method practically feasible, we propose a numerical scheme to solve correlation-matching problems accurately for VARTA processes.

The remainder of the article is organized as follows: In Section 2, we review the literature related to modeling and generating multivariate input processes for stochastic simulation. The comprehensive framework we employ, together with background information on vector autoregressive processes and the Johnson translation system, is presented in Section 3. The numerical-search and numerical-integration procedures are described in Section 4. Section 5 contains examples and Section 6 provides concluding remarks.

2. MODELING AND GENERATING MULTIVARIATE INPUT PROCESSES

A review of the literature on input modeling reveals a variety of models for representing and generating input processes for stochastic simulation. We restrict our attention to models that account for dependence in the input process, and refer the reader to Nelson and Yamnitsky [1998] and Law and Kelton [2000] for detailed surveys of the existing input-modeling tools.

When the problem of interest is to construct a stationary univariate time series with given marginal distribution and autocorrelation structure, there are two basic approaches: (i) Construct a time-series process exploiting properties specific to the marginal distribution of interest; or (ii) construct a series of autocorrelated uniform random variables, $\{U_t; t = 0, 1, 2, \ldots\}$, as a base process and transform it to the input process via $X_t = G_X^{-1}(U_t)$, where G_X is an arbitrary cumulative distribution function. The basic idea is to achieve the target autocorrelation structure of the input process X_t by adjusting the autocorrelation structure of the base process U_t.

The primary shortcoming of approach (i) is that it is not general: a different model is required for each marginal distribution of interest and the sample paths of these processes, while adhering to the desired marginal distribution

and autocorrelation structure, sometimes have unexpected features. An example is given by Lewis et al. [1989], who constructed time series with gamma marginals. In this paper, we take the latter approach (ii), which is more general and has been used previously by various researchers including Melamed [1991], Melamed et al. [1992], Willemain and Desautels [1993], Song et al. [1996], and Cario and Nelson [1996, 1998]. Of these, the most general model is given by Cario and Nelson, who redefined the base process as a Gaussian autoregressive process from which a series of autocorrelated uniform random variables is constructed via the probability-integral transformation. Further, their model controls the autocorrelations at lags of higher order than the others can handle. Our approach is very similar to the one in that study, but we define the base process by a *vector* autoregressive process that allows the modeling and generation of multivariate time-series processes.

The literature reveals a significant interest in the construction of random vectors with dependent components, which is a special case of our model. There are an abundance of models for representing and generating random vectors with marginal distributions from a common family. Excellent surveys can be found in Devroye [1986] and Johnson [1987]. However, when the component random variables have different marginal distributions from different families, there are few alternatives available. One approach is to transform multivariate normal vectors into vectors with arbitrary marginal distributions. The first reference to this idea appears to be Mardia [1970], who studied the bivariate case. Li and Hammond [1975] discussed the extension to random vectors of any finite dimension having continuous marginal distributions.

There are numerous other references that take a similar approach. Among these, we refer the interested reader to Chen [2001] and Cario et al. [2001], who generated random vectors with arbitrary marginal distributions and correlation matrix by the so-called NORTA (Normal-To-Anything) method, involving a componentwise transformation of a multivariate normal random vector. Cario et al. also discussed the extension of their idea to discrete and mixed marginal distributions. Their results can be considered as broadening the results of Cario and Nelson [1996] beyond a common marginal distribution. Recently, Lurie and Goldberg [1998] implemented a variant of the NORTA method for generating samples of predetermined size, while Clemen and Reilly [1999] described how to use the NORTA procedure to induce a desired rank correlation in the context of decision and risk analysis.

The transformation-oriented approach taken in this paper is related to methods that transform a random vector with uniformly distributed marginals into a vector with arbitrary marginal distributions; for example, Cook and Johnson [1981] and Ghosh and Henderson [2002]. However, it is quite different from techniques that construct joint distributions as mixtures of distributions with extreme correlations among their components [Hill and Reilly 1994]. While the mixture method is very effective for random vectors of low dimension (e.g., $k \leq 3$), the computational requirements quickly become expensive for higher dimensional random vectors.

The primary contribution of this article is to develop a comprehensive input-modeling framework that pulls together the theory behind univariate

time series and random vectors with dependent components and extends it to the multivariate time series, while also providing a numerical method to implement it.

3. THE MODEL

In this section, we present the VARTA framework together with the theory that supports it and the implementation problems that must be solved.

3.1 Background

Our premise is that searching among a list of input models for the "true, correct" model is neither a theoretically supportable nor practically useful paradigm upon which to base general-purpose input-modeling tools. Instead, we view input modeling as customizing a highly flexible model that can capture the important features of interest, while being easy to use, adjust, and understand. We achieve flexibility by incorporating vector autoregressive processes and the Johnson translation system into the model in order to characterize the process dependence and marginal distributions, respectively. We define the base process $\mathbf{Z}_t$ as a standard Gaussian vector autoregressive process whose autocorrelation structure is adjusted in order to achieve the desired autocorrelation structure of the input process $\mathbf{X}_t$. Then, we construct a series of autocorrelated uniform random variables, $\{U_{i,t};\ i = 1, 2, \ldots, k;\ t = 0, 1, 2, \ldots\}$, using the probability-integral transformation $U_{i,t} = \Phi(Z_{i,t})$. Finally, for $i = 1, 2, \ldots, k$, we apply the transformation $X_{i,t} = F_{X_i}^{-1}[U_{i,t}]$, which ensures that the ith component series, $\{X_{i,t};\ t = 0, 1, 2, \ldots\}$, has the desired Johnson-type marginal distribution F_{X_i}.

Below, we provide a brief review of the features of vector autoregressive processes and the Johnson translation system that we exploit; we then present the framework.

3.1.1 *The VAR$_k$(p) Model.* In a k-variate vector autoregressive process of order p (the VAR$_k(p)$ model) the presence of each variable is represented by a linear combination of a finite number of past observations of the variables plus a random error. This is written in matrix notation as[1]

$$\mathbf{Z}_t = \boldsymbol{\alpha}_1\mathbf{Z}_{t-1} + \boldsymbol{\alpha}_2\mathbf{Z}_{t-2} + \cdots + \boldsymbol{\alpha}_p\mathbf{Z}_{t-p} + \mathbf{u}_t, \quad t = 0, \pm1, \pm2, \ldots, \tag{1}$$

where $\mathbf{Z}_t = (Z_{1,t}, Z_{2,t}, \ldots, Z_{k,t})'$ is a $(k \times 1)$ random vector of the observations at time t and the $\boldsymbol{\alpha}_i$, $i = 1, 2, \ldots, p$, are fixed $(k \times k)$ autoregressive coefficient matrices. Finally, $\mathbf{u}_t = (u_{1,t}, u_{2,t}, \ldots, u_{k,t})'$ is a k-dimensional white noise vector representing the part of $\mathbf{Z}_t$ that is not linearly dependent on past observations; it has $(k \times k)$ covariance matrix Σ_u such that

$$\mathrm{E}[\mathbf{u}_t] = \mathbf{0}_{(k\times1)} \text{ and } \mathrm{E}[\mathbf{u}_t\mathbf{u}_{t-h}'] = \begin{cases} \Sigma_u & \text{if } h = 0, \\ \mathbf{0}_{(k\times k)} & \text{otherwise.} \end{cases}$$

The covariance matrix Σ_u is assumed to be positive definite.

[1]Although it is sometimes assumed that a process is started in a specified period, we find it more convenient to assume that it has been started in the infinite past.

Although the definition of the $\text{VAR}_k(p)$ model does not require the multivariate white noise vector, $\mathbf{u}_t$, to be Gaussian, our model makes this assumption. We also assume stability, meaning that the roots of the reverse characteristic polynomial, $|\mathbf{I}_{(k\times k)} - \boldsymbol{\alpha}_1 z - \boldsymbol{\alpha}_2 z^2 - \cdots - \boldsymbol{\alpha}_p z^p| = 0$, lie outside of the unit circle in the complex plane ($\mathbf{I}_{(k\times k)}$ is the $(k \times k)$ identity matrix). This further implies stationarity of the corresponding $\text{VAR}_k(p)$ process [Lütkepohl 1993, Proposition 2.1].

A first-order vector autoregressive process (the $\text{VAR}_k(1)$ model) can be expressed in terms of past and present white noise vectors as

$$\mathbf{Z}_t = \sum_{i=0}^{\infty} \boldsymbol{\alpha}_1^i \mathbf{u}_{t-i}, \qquad t = 0, \pm 1, \pm 2, \ldots. \tag{2}$$

[Lütkepohl 1993, page 10]. Since the assumption of stability makes the sequence $\{\boldsymbol{\alpha}_1^i; i = 0, 1, 2, \ldots\}$ absolutely summable [Lütkepohl 1993; Appendix A, Section A.9.1], the infinite sum (2) exists in mean square [Lütkepohl 1993; Appendix C, Proposition C.7]. Therefore, using the representation in (2), the first and second (time-invariant) moments of the $\text{VAR}_k(1)$ model are obtained as

$$\begin{aligned}
\mathrm{E}[\mathbf{Z}_t] &= \mathbf{0}_{(k\times 1)} \text{ for all } t, \\
\Sigma_Z(h) &= \mathrm{E}[(\mathbf{Z}_t - \mathrm{E}[\mathbf{Z}_t])(\mathbf{Z}_{t-h} - \mathrm{E}[\mathbf{Z}_{t-h}])'] \\
&= \lim_{n\to\infty} \sum_{i=0}^{n} \sum_{j=0}^{n} \boldsymbol{\alpha}_1^i \mathrm{E}[\mathbf{u}_{t-i}\mathbf{u}'_{t-h-j}] \left(\boldsymbol{\alpha}_1^j\right)' \\
&= \lim_{n\to\infty} \sum_{i=0}^{n} \boldsymbol{\alpha}_1^{i+h} \Sigma_u \left(\boldsymbol{\alpha}_1^i\right)' = \sum_{i=0}^{\infty} \boldsymbol{\alpha}_1^{i+h} \Sigma_u \left(\boldsymbol{\alpha}_1^i\right)',
\end{aligned}$$

because $\mathrm{E}[\mathbf{u}_t\mathbf{u}'_s] = \mathbf{0}$ for $t \neq s$ and $\mathrm{E}[\mathbf{u}_t\mathbf{u}'_t] = \Sigma_u$ for all t [Lütkepohl 1993, Appendix C.3, Proposition C.8]. We use the covariance matrices $\Sigma_Z(h)$, $h = 0, 1, \ldots, p$, to characterize the autocovariance structure of the base process as

$$\Sigma_{\mathbf{Z}} = \begin{pmatrix} \Sigma_Z(0) & \Sigma_Z(1) & \ldots & \Sigma_Z(p-2) & \Sigma_Z(p-1) \\ \Sigma'_Z(1) & \Sigma_Z(0) & \ldots & \Sigma_Z(p-3) & \Sigma_Z(p-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Sigma'_Z(p-1) & \Sigma'_Z(p-2) & \ldots & \Sigma'_Z(1) & \Sigma_Z(0) \end{pmatrix}_{(kp\times kp)}. \tag{3}$$

In this article, we assume that the autocovariance matrix, $\Sigma_{\mathbf{Z}}$, is positive definite.

We can extend the discussion above to $\text{VAR}_k(p)$ processes with $p > 1$ because any $\text{VAR}_k(p)$ process can be written in the first-order vector autoregressive form. More precisely, if $\mathbf{Z}_t$ is a $\text{VAR}_k(p)$ model defined as in (1), a corresponding kp-dimensional first-order vector autoregressive process

$$\bar{\mathbf{Z}}_t = \bar{\alpha}_1 \bar{\mathbf{Z}}_{t-1} + \bar{\mathbf{u}}_t \tag{4}$$

can be defined, where

$$\bar{\mathbf{Z}}_t = \begin{pmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \\ \mathbf{Z}_{t-2} \\ \vdots \\ \mathbf{Z}_{t-p+1} \end{pmatrix}_{(kp\times 1)} \quad \bar{\boldsymbol{\alpha}}_1 = \begin{pmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \dots & \boldsymbol{\alpha}_{p-1} & \boldsymbol{\alpha}_p \\ \mathbf{I}_{(k\times k)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(k\times k)} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{(k\times k)} & \mathbf{0} \end{pmatrix}_{(kp\times kp)} \quad \bar{\mathbf{u}}_t = \begin{pmatrix} \mathbf{u}_t \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}_{(kp\times 1)}.$$

This is known as "the state-space model" of the k-variate autoregressive process of order p [Lütkepohl 1993, page 418]. Following the foregoing discussion, the first and second moments of $\bar{\mathbf{Z}}_t$ are

$$\mathrm{E}[\bar{\mathbf{Z}}_t] = \mathbf{0}_{(kp\times 1)} \text{ for all } t \text{ and } \Sigma_{\bar{Z}}(h) = \sum_{i=0}^{\infty} \bar{\boldsymbol{\alpha}}_1^{i+h} \Sigma_{\bar{u}} \left(\bar{\boldsymbol{\alpha}}_1^{i}\right)', \tag{5}$$

where $\Sigma_{\bar{u}} = \mathrm{E}[\bar{\mathbf{u}}_t \bar{\mathbf{u}}_t']$ for all t. Using the $(k \times kp)$ matrix $\mathbf{J} = (\mathbf{I}_{(k\times k)}\ \mathbf{0} \cdots \mathbf{0})$, the process $\mathbf{Z}_t$ is obtained as $\mathbf{Z}_t = \mathbf{J}\bar{\mathbf{Z}}_t$. Since $\bar{\mathbf{Z}}_t$ is a well-defined stochastic process, the same is true for $\mathbf{Z}_t$. The mean $\mathrm{E}[\mathbf{Z}_t]$ is zero for all t and the (time-invariant) covariance matrices of the $\mathrm{VAR}_k(p)$ model are given by $\Sigma_Z(h) = \mathbf{J}\Sigma_{\bar{Z}}(h)\mathbf{J}'$.

We can describe the $\mathrm{VAR}_k(p)$ model using either its autocovariance structure, $\Sigma_Z(h)$ for $h = 0, 1, \dots, p$, or its parameters, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_p$ and Σ_u. In input-modeling problems, we directly adjust $\Sigma_Z(h)$, $h = 0, 1, \dots, p$, to achieve the desired autocorrelation structure of $\mathbf{X}_t$. To determine $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_p$ and Σ_u from $\Sigma_Z(h)$, $h = 0, 1, \dots, p$, we simply solve the multivariate Yule–Walker equations [Lütkepohl 1993, page 21] given by $\boldsymbol{\alpha} = \Sigma\Sigma_{\mathbf{Z}}^{-1}$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_p)_{(k\times kp)}$ and $\Sigma = (\Sigma_Z(1), \Sigma_Z(2), \dots, \Sigma_Z(p))_{(k\times kp)}$. Once $\boldsymbol{\alpha}$ is obtained, Σ_u can be determined from

$$\Sigma_u = \Sigma_Z(0) - \boldsymbol{\alpha}_1\Sigma_Z'(1) - \cdots - \boldsymbol{\alpha}_p\Sigma_Z'(p). \tag{6}$$

Our motivation for defining the base process, $\mathbf{Z}_t$, as a standard Gaussian vector autoregressive process is that it enables us to obtain the desired marginal distributions while incorporating the process dependence into the generated values implicitly. Further, it brings significant flexibility to the framework through its ability to characterize dependencies both in time sequence and with respect to other component series in the input process. We ensure that each component series of the input process $\{X_{i,t}; i = 1, 2, \dots, k;\ t = 0, 1, 2, \dots\}$ has the desired marginal distribution F_{X_i} by applying the transformation $X_{i,t} = F_{X_i}^{-1}[\Phi(Z_{i,t})]$. This works, provided each $Z_{i,t}$ is a standard normal random variable. The assumption of Gaussian white noise implies that $\mathbf{Z}_t$ is a Gaussian process[2] with mean $\mathbf{0}$. This further implies that the random vector $(Z_{i,t}, Z_{j,t-h})'$ has a bivariate normal distribution and, hence, $Z_{i,t}$ is a normal random variable

[2]This is considered a standard result in the time-series literature and stated without proof in several books, for example, Lütkepohl [1993, page 12]. However, the reader can find the corresponding proof together with the distributional properties of Gaussian vector autoregressive base processes in the online companion [Biller and Nelson 2003c].

(bivariate normality will be exploited when we solve the correlation-matching problem).

We force $Z_{i,t}$ to be standard normal by defining $\Sigma_Z(0)$ to be a correlation matrix and all entries in $\Sigma_Z(h), h = 1, 2, \ldots, p$ to be correlations. For this reason, we will use the terms "autocovariance" and "autocorrelation" interchangeably in the remainder of the article. We now state more formally the result that the random vector $(Z_{i,t}, Z_{j,t-h})'$ is bivariate normal; the proof, together with additional distributional properties, is in Biller and Nelson [2003c].

THEOREM 3.1. *Let $\mathbf{Z}_t$ denote a stable pth-order vector autoregressive process, $VAR_k(p)$, as defined in* (1) *with a positive definite autocorrelation matrix $\Sigma_{\mathbf{Z}}$ given by* (3). *The random variable $\tilde{Z} = (Z_{i,t}, Z_{j,t-h})'$, for $i, j = 1, 2, \ldots, k$ and $h = 0, 1, 2, \ldots$ (except $i = j$ when $h = 0$) has a nonsingular bivariate normal distribution with density function given by*

$$f(\tilde{z}; \Sigma_2) = \frac{1}{2\pi |\Sigma_2|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\tilde{z}'\Sigma_2^{-1}\tilde{z}\right), \ \tilde{z} \in \Re^2,$$

$$\Sigma_2 = \begin{pmatrix} 1 & \rho_{\mathbf{Z}}(i, j, h) \\ \rho_{\mathbf{Z}}(i, j, h) & 1 \end{pmatrix}_{(2\times 2)}.$$

PROOF. See Theorem 1 of Biller and Nelson [2003c]. □

Using the distributional properties provided in this section, we can achieve the desired autocorrelation structure of the input process by adjusting the autocorrelation structure of the Gaussian vector autoregressive base process as described in Section 3.2 below.

To generate a multivariate time series with given Johnson-type marginals and autocorrelation structure specified through lag p, we need to be able to generate realizations from a k-variate Gaussian vector autoregressive process of any required length, say T. We now explain how to do this using standard theory [Lütkepohl 1993, Appendix D.1]:

—First, we obtain the starting values, $\mathbf{z}_{-p+1}, \mathbf{z}_{-p+2}, \ldots, \mathbf{z}_0$, using the autocorrelation structure, $\Sigma_Z(h)$, $h = 0, 1, \ldots, p$, and the implied parameters, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p$ and Σ_u. We also obtain a series of Gaussian white noise vectors, $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_T$. Then, we generate the time series $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T$ recursively as $\mathbf{z}_t = \boldsymbol{\alpha}_1\mathbf{z}_{t-1} + \cdots + \boldsymbol{\alpha}_p\mathbf{z}_{t-p} + \mathbf{u}_t$ for $t = 1, 2, \ldots, T$.

—To generate $\mathbf{z}_{-p+1}, \mathbf{z}_{-p+2}, \ldots, \mathbf{z}_0$ as realizations of $\mathbf{Z}_{-p+1}, \mathbf{Z}_{-p+2}, \ldots, \mathbf{Z}_0$ whose joint distribution is given by a nonsingular kp-dimensional multivariate normal distribution (Biller and Nelson [2003c, Remark 1]), we choose a $(kp \times kp)$ matrix $\mathbf{Q}$ such that $\mathbf{Q}\mathbf{Q}' = \Sigma_{\mathbf{Z}}$. Then we obtain the starting-value vector as $(\mathbf{z}_0', \mathbf{z}_{-1}', \ldots, \mathbf{z}_{-p+1}')' = \mathbf{Q}\,(v_1, \ldots, v_{kp})'$, where the v_i's are independent standard normal random variates. In this way, we ensure that the process starts stationary.

—To obtain the series of independent Gaussian white noise vectors, $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_T$, we first choose k independent univariate standard normal variates $v_1, v_2, \ldots, v_k$, and then multiply by a $(k \times k)$ matrix $\mathbf{P}$ for which $\mathbf{P}\mathbf{P}' = \Sigma_u$; that is, $\mathbf{u}_t = \mathbf{P}\,(v_1, v_2, \ldots, v_k)'$. We repeat this process T times.

3.1.2 *The Johnson Translation System.* In the case of modeling data with an unknown distribution, an alternative to using a standard family of distributions is to use a more flexible system of distributions. We propose using the Johnson translation system [Johnson 1949a]. Our motivation for using this system is practical, rather than theoretical: In many applications, simulation output performance measures are insensitive to the specific input distribution chosen provided that enough moments of the distribution are correct, for example, Gross and Juttijudata [1997]. The Johnson system can match any feasible first four moments, while the standard input models incorporated in some existing software packages and simulation languages match only one or two moments. Thus, our goal is to represent key features of the process of interest, as opposed to finding the "true" distribution.

The Johnson translation system for a random variable X is defined by a cdf of the form

$$F_X(x) = \Phi\left\{\gamma + \delta f\left[\frac{x-\xi}{\lambda}\right]\right\}, \tag{7}$$

where γ and δ are shape parameters, ξ is a location parameter, λ is a scale parameter, and $f(\cdot)$ is one of the following transformations:

$$f(y) = \begin{cases} \log(y) & \text{for the } S_L \text{ (lognormal) family,} \\ \log(y + \sqrt{y^2+1}) & \text{for the } S_U \text{ (unbounded) family,} \\ \log\left(\frac{y}{1-y}\right) & \text{for the } S_B \text{ (bounded) family,} \\ y & \text{for the } S_N \text{ (normal) family.} \end{cases}$$

There is a unique family (choice of f) for each feasible combination of the skewness and the kurtosis that determine the parameters γ and δ. Any mean and (positive) variance can be attained by any one of the families by the manipulation of the parameters λ and ξ. Within each family, a distribution is completely specified by the values of the parameters $[\gamma, \delta, \lambda, \xi]$ and the range of X depends on the family of interest.

The Johnson translation system provides good representations for unimodal distributions and can represent certain bimodal shapes, but not three or more modes. In spite of this, the Johnson translation system enables us to achieve a wide variety of distributional shapes. A detailed illustration for the shapes of the Johnson-type probability density functions can be found in Johnson [1987].

3.2 The Model

In this section we describe a model for a stationary k-variate time-series input process $\{\mathbf{X}_t;\ t = 0, 1, 2, \ldots\}$ with the following properties:

(1) Each component time series $\{X_{i,t};\ t = 0, 1, 2, \ldots\}$ has a Johnson-type marginal distribution that can be defined by F_{X_i}. In other words, $X_{i,t} \sim F_{X_i}$ for $t = 0, 1, 2, \ldots$ and $i = 1, 2, \ldots, k$.

(2) The dependence structure is specified via Pearson product—moment correlations $\rho_{\mathbf{X}}(i, j, h) = \text{Corr}[X_{i,t}, X_{j,t-h}]$, for $h = 0, 1, \ldots, p$ and $i, j = 1, 2, \ldots, k$. Equivalently, the lag-h correlation matrices are defined by

$\Sigma_X(h) = \text{Corr}\left[\mathbf{X}_t, \mathbf{X}_{t-h}\right] = [\rho_{\mathbf{X}}(i, j, h)]_{(k\times k)}$, for $h = 0, 1, \ldots, p$, where $\rho_{\mathbf{X}}(i, i, 0) = 1$. Using the first $h = 0, 1, \ldots, p-1$ of these matrices, we define $\Sigma_{\mathbf{X}}$ analogously to $\Sigma_{\mathbf{Z}}$.

Accounting for dependence via Pearson product—moment correlation is a practical compromise we make in the model. Many other measures of dependence have been defined (e.g., Nelsen [1998]) and they are arguably more informative than the product—moment correlation for some distribution pairs. However, product—moment correlation is the only measure of dependence that is widely used and understood in engineering applications. We believe that making it possible for simulation users to incorporate dependence via product—moment correlation, while limited, is substantially better than ignoring dependence. Further, our model is flexible enough to incorporate dependence measures that remain unchanged under strictly increasing transformations of the random variables, such as Spearman's rank correlation and Kendall's τ, should those measures be desired.

We obtain the ith time series via the transformation $X_{i,t} = F_{X_i}^{-1}[\Phi(Z_{i,t})]$, which ensures that $X_{i,t}$ has distribution F_{X_i} by well-known properties of the inverse cumulative distribution function. Therefore, the central problem is to select the autocorrelation structure, $\Sigma_Z(h)$, $h = 0, 1, \ldots, p$, for the base process that gives the desired autocorrelation structure, $\Sigma_X(h)$, $h = 0, 1, \ldots, p$, for the input process.

We let $\rho_{\mathbf{Z}}(i, j, h)$ be the (i, j)th element of the lag-h correlation matrix, $\Sigma_Z(h)$, and let $\rho_{\mathbf{X}}(i, j, h)$ be the (i, j)th element of $\Sigma_X(h)$. The correlation matrix of the base process $\mathbf{Z}_t$ directly determines the correlation matrix of the input process $\mathbf{X}_t$, because

$$\rho_{\mathbf{X}}(i, j, h) = \text{Corr}[X_{i,t}, X_{j,t-h}] = \text{Corr}\left[F_{X_i}^{-1}[\Phi(Z_{i,t})], F_{X_j}^{-1}[\Phi(Z_{j,t-h})]\right]$$

for all $i, j = 1, 2, \ldots, k$ and $h = 0, 1, 2, \ldots, p$, excluding the case $i = j$ when $h = 0$. Further, only $\text{E}[X_{i,t}X_{j,t-h}]$ depends on Σ_Z, since

$$\text{Corr}[X_{i,t}, X_{j,t-h}] = \frac{\text{E}[X_{i,t}X_{j,t-h}] - \text{E}[X_{i,t}]\text{E}[X_{j,t-h}]}{\sqrt{\text{Var}[X_{i,t}]\text{Var}[X_{j,t-h}]}}$$

and $\text{E}[X_{i,t}]$, $\text{E}[X_{j,t-h}]$, $\text{Var}[X_{i,t}]$, $\text{Var}[X_{j,t-h}]$ are fixed by F_{X_i} and F_{X_j} (i.e., $\mu_i = \text{E}[X_{i,t}]$, $\mu_j = \text{E}[X_{j,t-h}]$, $\sigma_i^2 = \text{Var}[X_{i,t}]$ and $\sigma_j^2 = \text{Var}[X_{j,t-h}]$ are properties of F_{X_i} and F_{X_j}). Since $(Z_{i,t}, Z_{j,t-h})'$ has a nonsingular standard bivariate normal distribution with correlation $\rho_{\mathbf{Z}}(i, j, h)$ (Theorem 3.1), we have

$$\begin{aligned}\text{E}[X_{i,t}X_{j,t-h}] &= \text{E}\left[F_{X_i}^{-1}[\Phi(Z_{i,t})]F_{X_j}^{-1}[\Phi(Z_{j,t-h})]\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_i}^{-1}[\Phi(z_{i,t})]F_{X_j}^{-1}[\Phi(z_{j,t-h})]\vartheta_{\rho_{\mathbf{Z}}(i,j,h)}(z_{i,t}, z_{j,t-h})\,dz_{i,t}\,dz_{j,t-h},\end{aligned} \tag{8}$$

where $\vartheta_{\rho_{\mathbf{Z}}(i,j,h)}$ is the standard bivariate normal probability density function with correlation $\rho_{\mathbf{Z}}(i, j, h)$.

This development is valid for any marginal distributions F_{X_i} and F_{X_j} for which the expectation (8) exists. However, since $Z_{i,t}$ and $Z_{j,t-h}$ are standard normal random variables with a nonsingular bivariate distribution, the joint

distribution of $X_{i,t}$ and $X_{j,t-h}$ is well-defined and the expectation (8) always exists in the case of Johnson marginals. Further, the Johnson translation system is a particularly good choice because

$$\begin{aligned} X_{i,t} &= F_{X_i}^{-1}[\Phi(Z_{i,t})] = \xi_i + \lambda_i f_i^{-1}\left[\frac{Z_{i,t}-\gamma_i}{\delta_i}\right] \\ X_{j,t-h} &= F_{X_j}^{-1}[\Phi(Z_{j,t-h})] = \xi_j + \lambda_j f_j^{-1}\left[\frac{Z_{j,t-h}-\gamma_j}{\delta_j}\right], \end{aligned} \tag{9}$$

avoiding the need to evaluate $\Phi(\cdot)$. Notice that the Eq. (9) defines a bivariate Johnson distribution as in Johnson [1949b].

From (8) we see that the correlation between $X_{i,t}$ and $X_{j,t-h}$ is a function only of the correlation between $Z_{i,t}$ and $Z_{j,t-h}$, which appears in the expression for $\vartheta_{\rho_{\mathbf{Z}}(i,j,h)}$. We denote the implied correlation $\mathrm{Corr}[X_{i,t}, X_{j,t-h}]$ by the function $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)]$ defined as

$$\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_i}^{-1}[\Phi(z_{i,t})]F_{X_j}^{-1}[\Phi(z_{j,t-h})]\vartheta_{\rho_{\mathbf{Z}}(i,j,h)}(z_{i,t}, z_{j,t-h})\,dz_{i,t}\,dz_{j,t-h} - \mu_i\mu_j}{\sigma_i\sigma_j}.$$

Thus, the problem of determining $\Sigma_Z(h)$, $h = 0, 1, \ldots, p$, that gives the desired input correlation matrices $\Sigma_X(h)$, $h = 0, 1, \ldots, p$, reduces to $pk^2 + k(k-1)/2$ individual matching problems in which we try to find the value $\rho_{\mathbf{Z}}(i,j,h)$ that makes $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)] = \rho_{\mathbf{X}}(i,j,h)$. Unfortunately, it is not possible to find the $\rho_{\mathbf{Z}}(i,j,h)$ values analytically except in special cases [Li and Hammond 1975]. Instead, we establish some properties of the function $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)]$ that enable us to perform a numerical search to find the $\rho_{\mathbf{Z}}(i,j,h)$ values within a predetermined precision. We primarily extend the results in Cambanis and Marsy [1978], Cario and Nelson [1996], and Cario et al. [2001]—which apply to time-series input processes with identical marginal distributions and random vectors with arbitrary marginal distributions—to the multivariate time-series input processes with arbitrary marginal distributions. The proofs of all results can be found in the Appendix.

The first two properties concern the sign and the range of $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)]$ for $-1 \le \rho_{\mathbf{Z}}(i,j,h) \le 1$.

PROPOSITION 3.2. *For any distributions F_{X_i} and F_{X_j}, $c_{ijh}(0) = 0$ and $\rho_{\mathbf{Z}}(i,j,h) \ge 0$ (≤ 0) implies that $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)] \ge 0$ (≤ 0).*

It follows from the proof of Proposition 3.2 that taking $\rho_{\mathbf{Z}}(i,j,h) = 0$ results in a multivariate time series in which $X_{i,t}$ and $X_{j,t-h}$ are not only uncorrelated, but are also independent. The following property shows that the minimum and maximum possible input correlations are attainable.

PROPOSITION 3.3. *Let $\underline{\rho}_{ij}$ and $\bar{\rho}_{ij}$ be the minimum and maximum possible bivariate correlations, respectively, for random variables having marginal distributions F_{X_i} and F_{X_j}. Then, $c_{ijh}[-1] = \underline{\rho}_{ij}$ and $c_{ijh}[1] = \bar{\rho}_{ij}$.*

The next two results shed light on the shape of the function $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)]$.

THEOREM 3.4. *The function* $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)]$ *is nondecreasing for* $-1 \leq \rho_{\mathbf{Z}}(i,j,h) \leq 1$.

THEOREM 3.5. *If there exists* $\epsilon > 0$ *such that*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sup_{\rho_{\mathbf{Z}}(i,j,h)\in[-1,1]} \left\{ \left| F_{X_i}^{-1}[\Phi(z_{i,t})] F_{X_j}^{-1}[\Phi(z_{j,t-h})] \right|^{1+\epsilon} \vartheta_{\rho_{\mathbf{Z}}(i,j,h)}(z_{i,t}, z_{j,t-h}) \right\} dz_{i,t}\, dz_{j,t-h} < \infty,$$

then the function $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)]$ *is continuous for* $-1 \leq \rho_{\mathbf{Z}}(i,j,h) \leq 1$.

Since $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)]$ is a continuous, nondecreasing function under the mild conditions stated in Theorem 3.5, any reasonable search procedure can be used to find $\rho_{\mathbf{Z}}(i,j,h)$ such that $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)] \approx \rho_{\mathbf{X}}(i,j,h)$ (although perhaps not efficiently). Proposition 3.2 provides the initial bounds for such a procedure. Proposition 3.3 shows that the extremal values of $\rho_{\mathbf{X}}(i,j,h)$ are attainable under the model. Furthermore, from Proposition 3.3, Theorem 3.5, and the Intermediate Value Theorem, any possible bivariate correlation for F_{X_i} and F_{X_j} is attainable under the model. Theorem 3.4 provides the theoretical basis for adjusting the values of $\rho_{\mathbf{Z}}(i,j,h)$ and is the key to developing a computationally accurate numerical scheme, which we present in the following section.

Throughout the previous discussion, we assumed that there exists a joint distribution with marginal distributions F_{X_i}, for $i = 1, 2, \ldots, k$, and an autocorrelation structure characterized by $\Sigma_X(h)$, $h = 0, 1, \ldots, p$. However, not all combinations of F_{X_i}, $i = 1, 2, \ldots, k$, and $\Sigma_X(h)$, $h = 0, 1, \ldots, p$, are feasible. Clearly, for the autocorrelation structure to be feasible, we must have $\underline{\rho}_{ij} \leq \rho_{\mathbf{X}}(i,j,h) \leq \bar{\rho}_{ij}$ for each $i, j = 1, 2, \ldots, k$ and $h = 0, 1, \ldots, p$. In addition, $\Sigma_{\mathbf{X}}$ must be positive definite and this can be ensured by selecting a positive definite base autocorrelation matrix $\Sigma_{\mathbf{Z}}$ (J. R. Wilson, personal communication). Unfortunately, the converse of this result does not necessarily hold; that is, there exist sets of marginals with a feasible autocorrelation structure that are not representable by the VARTA transformation. Both Li and Hammond [1975] and Lurie and Goldberg [1998] gave examples where this appears to be the case and recently Ghosh and Henderson [2002] proved the existence of a joint distribution that is not representable as a transformation of a multivariate normal random vector. Although these studies primarily focus on the NORTA procedure, they can be extended to the VARTA case. However, Ghosh and Henderson's computational experience suggests that the failure of the NORTA method is rare. Further, inspection of the input correlation matrices for which the NORTA method does not work shows that the correlations lie either on the boundary or in close proximity to the set of achievable correlations specified by the marginals of the input process. We have observed that using the Johnson translation system tends to mitigate this problem because it provides a relatively comprehensive set of achievable correlations. If, after solving the bivariate correlation-matching problems, the base correlation matrix is not positive definite, then Ghosh and Henderson [2002] suggest the application of semidefinite programming on the base matrix, which is completed to be positive semidefinite. Motivated by this idea, we incorporate a modification step to the

data generation routine, which we present in more detail in our technical report [Biller and Nelson 2003b].

Our next result indicates that the input process $\mathbf{X}_t$ is stationary if the base $\text{VAR}_k(p)$ process $\mathbf{Z}_t$ is, and it follows immediately from the definition of strict stationarity.

PROPOSITION 3.6. *If $\mathbf{Z}_t$ is strictly stationary, then $\mathbf{X}_t$ is strictly stationary.*[3]

4. IMPLEMENTATION

In this section, we consider the problem of solving the correlation-matching problem for a fully specified VARTA process. Our objective is to find $\hat{\rho}_{\mathbf{Z}}(i, j, h)$ such that $c_{ijh}[\hat{\rho}_{\mathbf{Z}}(i, j, h)] \approx \rho_{\mathbf{X}}(i, j, h)$ for $i, j = 1, 2, \ldots, k$ and $h = 0, 1, \ldots, p$ (excluding the case $i = j$ when $h = 0$). The idea is to take some initial base correlations, transform them into the implied correlations for the specified pair of marginals (using a numerical integration technique), and then employ a search method until we find a base correlation that approximates the desired input correlation within a prespecified level of accuracy.

This problem was previously studied by Cario and Nelson [1998], Chen [2001], and Cario et al. [2001]. Since the only term in (8) that is a function of ρ is ϑ_ρ, Cario and Nelson suggest the use of a numerical integration procedure in which points (z_i, z_j) at which the integrand is evaluated do not depend on ρ and a grid of values are evaluated simultaneously by reweighting the $F_{X_i}^{-1}[\Phi(z_i)]F_{X_j}^{-1}[\Phi(z_j)]$ terms by different ϑ_ρ values. They refine the grid until one of the grid points $\hat{\rho}_{\mathbf{Z}}(i, j, h)$ satisfies $c_{ijh}[\hat{\rho}_{\mathbf{Z}}(i, j, h)] \approx \rho_{\mathbf{X}}(i, j, h)$, for $h = 0, 1, \ldots, p$. This approach makes particularly good sense in their case because all of their matching problems share a common marginal distribution, so many of the grid points will be useful. Chen and Cario et al. evaluate (8) using sampling techniques and apply stochastic root-finding algorithms to search for the correlation of interest within a predetermined precision. This approach is very general and makes good sense when the dimension of the problem is small and a diverse collection of marginal distributions might be considered.

Contrary to the situations presented in these papers, evaluating the function $F_{X_i}^{-1}[\Phi(z_i)]F_{X_j}^{-1}[\Phi(z_j)]$ is not computationally expensive for us because the Johnson translation system is based on transforming standard normal random variates. Thus, we avoid evaluating $\Phi(z_i)$ and $\Phi(z_j)$. However, we may face a very large number of matching problems, specifically $pk^2+k(k-1)/2$ such problems. Our approach is to take advantage of the superior accuracy of a numerical integration technique that supports a numerical-search procedure without suffering a substantial computational burden. We will address the efficiency of this technique in detail in our technical report [Biller and Nelson 2003b].

4.1 Numerical Integration Technique

This section briefly summarizes how we numerically evaluate $\text{E}[X_{i,t}\, X_{j,t-h}]$ given the marginals, F_{X_i} and F_{X_j}, and the associated correlation, $\rho_{\mathbf{Z}}(i, j, h)$.

[3]Note that for a Gaussian process, strict stationarity and weak stationarity are equivalent properties.

Since we characterize the input process using the Johnson translation system, evaluation of the composite function $F_X^{-1}[\Phi(z)]$ is significantly simplified because $F_X^{-1}[\Phi(z)] = \xi + \lambda f^{-1}[(z-\gamma)/\delta]$, where

$$f^{-1}(a) = \begin{cases} \exp(a) & \text{for the } S_L \text{ (lognormal) family,} \\ \dfrac{\exp(a) - \exp(-a)}{2} & \text{for the } S_U \text{ (unbounded) family,} \\ \dfrac{1}{1+\exp(-a)} & \text{for the } S_B \text{ (bounded) family,} \\ a & \text{for the } S_N \text{ (normal) family.} \end{cases}$$

Letting $i = 1$, $j = 2$, and $\rho_{\mathbf{Z}}(i, j, h) = \rho$ for convenience, the integral we need to evaluate can be written as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\xi_1 + \lambda_1 f_1^{-1}[(z_1-\gamma_1)/\delta_1]\right) \times \left(\xi_2 + \lambda_2 f_2^{-1}[(z_2-\gamma_2)/\delta_2]\right) \frac{\exp(-(z_1^2 - 2\rho z_1 z_2 + z_2^2)/2(1-\rho^2))}{2\pi\sqrt{1-\rho^2}}\, dz_1\, dz_2. \quad (10)$$

The expansion of the formula (10), based on the families to which f_1^{-1} and f_2^{-1} might belong, takes us to a number of different subformulas, but all with a similar form of

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w[z_1, z_2]\; g[z_1, z_2, \rho]\; dz_1\, dz_2,$$

where $w[z_1, z_2] = \exp(-(z_1^2 + z_2^2))$, but the definition of $g[\cdot]$ changes from one subproblem to another. Notice that the integral (8) exists only if $|\rho| < 1$, but we can solve the problem for $|\rho| = 1$ using the discussion in the proof of Proposition 3.3 (see the Appendix).

Our problem falls under the broad class of numerical integration problems for which there exists an extensive literature. Despite the wide-ranging and detailed discussion of its theoretical and practical aspects, computing a numerical approximation of a definite double integral with infinite support (called a cubature problem) reliably and efficiently is often a highly complex task. As far as we are aware, there are only two published softwares, "Ditamo" [Robinson and Doncker 1981] and "Cubpack" [Cools et al. 1997], which were specifically designed for solving cubature problems. While preparing the numerical integration routine for our software, we primarily benefited from the work accomplished in the latter reference.

As suggested by the numerical integration literature (e.g., Krommer and Ueberhuber [1994]), we use a global adaptive integration algorithm, based on transformations and subdivisions of regions, for an accurate and efficient solution of our cubature problem. The key to a good solution is the choice of an appropriate transformation from the infinite integration region of the original problem to a suitable finite region for the adaptive algorithm. Therefore, we transform the point (z_1, z_2) from the infinite region $[-\infty, \infty]^2$ to the finite region $[-1, 1]^2$ by using a doubly infinite hypercube transformation

$z_i = \psi_i(z_i^*) = \tan(\pi z_i^*/2)$ for $-1 < z_i^* < 1$ and $i = 1, 2$. Because $d\psi_i(z_i^*)/dz_i^* = (\pi/2)[1+\tan^2(\pi z_i^*/2)]$, the integral (10) is transformed into one of the following forms:

$$\int_{-1}^{1}\int_{-1}^{1} \frac{w[\tan(\pi z_1^*/2), \tan(\pi z_2^*/2)]\, g[\tan(\pi z_1^*/2), \tan(\pi z_2^*/2), \rho]}{4/\pi^2\left[1+\tan^2(\pi z_1^*/2)\right]^{-1}\left[1+\tan^2(\pi z_2^*/2)\right]^{-1}}\, dz_1^*\, dz_2^*, |\rho| < 1$$

$$\int_{-1}^{1}\int_{-1}^{1} \frac{\prod_{i=1}^{2}\left(\xi_i + \lambda_i f_i^{-1}[(\tan(\pi z_i^*/2) - \gamma_i)/\delta_i]\right)}{4\sqrt{2/\pi}\exp\left[\frac{1}{2}\tan^2(\pi z_1^*/2)\right]\left[1+\tan^2(\pi z_1^*/2)\right]^{-1}}\, dz_1^*\, dz_2^*, \rho = 1 \qquad (11)$$

$$\int_{-1}^{1}\int_{-1}^{1} \frac{\left(\xi_1 + \lambda_1 f_1^{-1}[(\tan(\pi z_1^*/2) - \gamma_1)/\delta_1]\right)\left(\xi_2 + \lambda_2 f_2^{-1}[(\tan(\pi z_1^*/2) - \gamma_2)/\delta_2]\right)}{4\sqrt{2/\pi}\exp\left[\frac{1}{2}\tan^2(\pi z_1^*/2)\right]\left[1+\tan^2(\pi z_1^*/2)\right]^{-1}}$$
$$\times\, dz_1^*\, dz_2^*, \rho = -1.$$

Although the $\rho = \pm 1$ cases could be expressed as a single integral, we express them as double integrals in order to take advantage of the accurate and reliable error estimation strategy developed specifically for cubature problems.

As a check on consistency and efficiency of the transformation $\psi(z^*) = \tan(\pi z^*/2)$, we compared its performance with other doubly infinite hypercube transformations including $\psi(z^*) = z^*/(1 - |z^*|)$, $\psi(z^*) = \text{sign}(z^*)(-\nu \ln(|z^*|))^{\frac{1}{2}}$, and $\psi(z^*) = \text{sign}(z^*)(-\nu \ln(1 - |z^*|))^{\frac{1}{2}}$ for some $\nu > 0$, as suggested by Genz [1992]. While $d\psi(z^*)/dz^*$ is generally singular at the points z^* for which $\psi(z^*) = \pm\infty$, and this entails singularities of the transformed integrand in the case of the doubly infinite hypercube transformations listed above, we do not need to deal with this problem when we use $\psi(z^*) = \tan(\pi z^*/2)$ for $-1 < z^* < 1$. Further, we empirically observed that the transformation $\psi(z^*) = \tan(\pi z^*/2)$ leads to relatively smooth shapes to be integrated, increasing the effectiveness of the global adaptive integration algorithm for solving the correlation-matching problem.

Since the integration regions in the formulas (11) correspond to squares defined over $[-1, 1]^2$, we can use a variety of cubature formulas developed for integration in a unit-square region and accommodate any rectangular region using the standard affine transformations (scaling and translation). Therefore, our numerical integration routine requires the central data structure to be a collection of rectangles. This allows us to take full advantage of polymorphism of C++ when we incorporate this routine in the software. Figure 1 provides a high-level view of how the algorithm works. In the figure, we use $\text{C}(\ell; B)$ and $\text{E}(\ell; B)$ to denote the cubature formula and the error estimation strategy, respectively, applied to the integrand ℓ over the region B. Further, $\text{I}(\ell; B)$ corresponds to the true value of the integral.

As the criterion for success, we define the maximum allowable error level as

$$\max(\epsilon_{\text{abs}}, \epsilon_{\text{rel}} \times \text{C}(|\ell|; B)),$$

where ϵ_{abs} corresponds to the requested absolute error and ϵ_{rel} is the requested relative error. This definition is a combination of a pure test for convergence with respect the absolute error ($\epsilon_{\text{rel}} = 0$ and $|\text{E}(\ell; B)| < \epsilon_{\text{abs}}$) and a pure test for convergence with respect to the relative error ($\epsilon_{\text{abs}} = 0$ and

calculate the total approximate integral value, $c := C(\ell; B)$, over region B;
calculate the total error estimate, $e := E(\ell; B)$, over region B;
insert (B, c, e) into the data structure;
while ($e >$ maximum allowable error level) & (size of the data structure $\neq 0$) **do**
{
  choose the element of the data structure (index s) with the largest error estimate, e_s;
  subdivide the chosen region B_s into subregions: $B_{s,r}$, $r = 1, 2, \ldots, R$ ($R = 2$ or $R = 4$);
  calculate approximate integral values in each subregion: $c_{s,r} = C(\ell; B_{s,r})$, $r = 1, 2, \ldots, R$;
  calculate error estimates in each subregion: $e_{s,r} = E(\ell; B_{s,r})$, $r = 1, 2, \ldots, R$;
  determine whether there is a decrease in the total error estimate:
  if $(e_s > e_{s,1} + \cdots + e_{s,R})$ then
  {
    insert $(B_{s,1}, c_{s,1}, e_{s,1}), \ldots, (B_{s,R}, c_{s,R}, e_{s,R})$ into the data structure;
    $c := c - c_s + \sum_{r=1}^{R} c_{s,r}$;
    $e := e - e_s + \sum_{r=1}^{R} e_{s,r}$;
    remove (B_s, c_s, e_s) from the data structure;
  }
  else
  {
    delete the newly created subregions, $B_{s,r}$, $r = 1, 2, \ldots, R$;
    remove (B_s, c_s, e_s) from the data structure and mark hopeless;
  }
}
end do
return the total approximate integral value, c, with its error estimate e;

Fig. 1. Meta algorithm for the numerical integration routine.

$|\mathrm{E}(\ell; B)|/\mathrm{C}(|\ell|; B) < \epsilon_{\mathrm{abs}}$). The constants ϵ_{abs} and ϵ_{rel} are defined in our software [see Section 5], in which we can also force one or the other of these criteria to be satisfied by specifying the error for the other to be zero. Notice that we define the maximum allowable error level using $\mathrm{C}(|\ell|; B)$ instead of $|\mathrm{C}(\ell; B)|$. This avoids heavy cancelation that might occur during the calculation of the approximate value $\mathrm{C}(\ell; B) \approx 0$, although the function values in the integration problems might not be small. For the full motivation behind this convergence test, we refer the reader to Krommer and Ueberhuber [1994]. The additional calculation of $\mathrm{C}(|\ell|; B)$ causes only a minor increase in the overall computational effort as no additional function evaluations are needed.

After we select the rectangular region with the largest error estimate, we dissect it into two or four smaller subregions, which are affinely similar to the original one, by lines running parallel to the sides [Cools 1994]. Adopting the "C2rule13" routine of the Cubpack software, we approximate the integral and the error associated with each subregion using a fully symmetric cubature formula of degree 13 with 37 points [Rabinowitz and Richter 1969; Stroud 1971] and a sequence of null rules with different degrees of accuracy. If the subdivision decreases the total error estimate, then the descendants (subregions) of the selected region are added to the collection of rectangular regions over which the function ℓ is integrated, the total approximate integral and error estimates are updated, and finally the selected rectangle is removed from the collection. Otherwise, the selected rectangle is considered to be hopeless, which means that the current error estimate for that region cannot be reduced further. When either the total error estimate falls below the maximum error level, or all regions are marked as hopeless, we stop the integration routine and report the result.

Due to the importance of the error estimation strategy in solving the correlation-matching problem accurately, we next give a brief description of null rules and the motivation for using them, and explain how we calculate an error estimate from null rules. Readers who are not interested in the specifics of the numerical integration technique may skip the remainder of this subsection.

Krommer and Ueberhuber [1994] define an n-point d-degree null rule as the sum $\mathrm{N}_d(\ell) = \sum_{i=1}^{n} u_i\ \ell(x_i)$ with at least one non-zero weight and the condition that $\sum_{i=1}^{n} u_i = 0$, where $x_i,\ i = 1, 2, \ldots, n$ and $u_i,\ i = 1, 2, \ldots, n$ correspond to the abscissas and the weights of the null rule, respectively, and $\ell(x_i)$ is the value the integrand ℓ takes at the abscissa x_i. The null rule $\mathrm{N}_d(\ell)$ is furthermore said to have degree d if it assigns zero to all polynomials of degree not more than d, but not all polynomials of degree $d + 1$. When two null rules of the same degree exist, say $\mathrm{N}_{d,1}(\ell)$ and $\mathrm{N}_{d,2}(\ell)$, the number $\mathrm{N}_d(\ell) = \sqrt{\mathrm{N}_{d,1}^2(\ell) + \mathrm{N}_{d,2}^2(\ell)}$ is computed and called a combined rule. We use the tuple $(\cdot,\cdot)$ to refer to such a combined null rule and $(\cdot)$ to refer to a single null rule.

For any given set of n distinct points, there is a manifold of null rules, but we restrict ourselves to the "equally strong" null rules whose weights have the same norm as the coefficients of the cubature formula. The advantage of using the equally strong null rules is that if we consider the error estimate coming from a sequence of null rules and the true error of the numerical integration as random variables, then they can be shown to have the same mean and standard deviation [Krommer and Ueberhuber 1994, page 171]. This fact is exploited to provide an error estimate.

Next, we explain the motivation for using null rules: A common error estimation procedure is based on using two polynomial integration formulas, $\mathrm{C}_{n_1}(\ell; B)$ and $\mathrm{C}_{n_2}(\ell; B)$, with different degrees of accuracy,[4] n_1 and n_2 such that $n_1 < n_2$, that is, $\mathrm{C}_{n_2}(\ell; B)$ is expected to give more accurate results than $\mathrm{C}_{n_1}(\ell; B)$. The integration formula with the higher degree is taken as the approximation of the

[4]The degree of accuracy of a cubature formula $\mathrm{C}_D(\ell; B)$ is D if $\mathrm{C}_D(\ell; B)$ is exact for all polynomials of degree $d \leq D$, but not exact for all polynomials of degree $d = D+1$. In our notation, the subscript on C indicates the degree.

true integral and $|C_{n_1}(\ell; B) - C_{n_2}(\ell; B)|$ is taken as the error estimate. Although reasonably good estimates can be obtained if the integrand ℓ is sufficiently smooth and the region B is small, this approach is in general problematic. Since error estimation depends on the underlying formulas, we can accidentally find values of $|C_{n_1}(\ell; B) - C_{n_2}(\ell; B)|$ that are too small when compared to $|C_{n_2}(\ell; B) - I_{n_2}(\ell; B)|$, resulting in a significant underestimation of the true error. At the same time, it is possible that as the degree of the polynomial approximating the true integral increases, the error terms do not decrease. Therefore, extensive experiments are often needed for each pair of integration formulas to ensure satisfactory reliability and accuracy of the estimates. Using sequences of null rules is an approach designed to overcome these difficulties with the following features: (i) The abscissas and weights of a null rule are independent of the integrand ℓ. (ii) Extensive function evaluations are avoided by using the same integrand evaluations used for approximating the integral. (iii) The procedure identifies the type of the asymptotic behavior that the sequences of null rules, $\{N_d(\ell),\ d = 0, \ldots, n-2\}$, follows and, accordingly, it calculates an error estimate for $|C(\ell; B) - I(\ell; B)|$.

The major difficulty in the application of the null rules is to decide how to extract an error estimate from the numbers produced by the null rules with different degrees of accuracy. The approach is to heuristically distinguish the behavior of the sequence $\{N_d(\ell),\ d = 0, \ldots, n-2\}$ among three possible types of behavior, which are nonasymptotic, weakly asymptotic, and strongly asymptotic. Following Cools et al. [1997], we use seven independent fully symmetric null rules of degrees (1), (3, 3), (5, 5), and (7, 7) to obtain $N_1(\ell)$, $N_3(\ell)$, $N_5(\ell)$, and $N_7(\ell)$, which are used to conduct a test for observable asymptotic behavior: The test for strong asymptotic behavior requires r to be less than a certain critical value, r_{crit}, where r is taken to be the maximum of the quantities $\sqrt{N_7(\ell)/N_5(\ell)}$, $\sqrt{N_5(\ell)/N_3(\ell)}$, and $\sqrt{N_3(\ell)/N_1(\ell)}$. This leads to the error estimate $|C(\ell; B) - I(\ell; B)| \approx K r_{\text{crit}}^{s-q+2} r^{q-s} N_s(\ell)$, where K is a safety factor, s is the highest value among the possible degrees attained by a null rule, and q is the degree of the corresponding cubature formula. If $r > 1$, then there is assumed to be no asymptotic behavior at all and the error estimate is $K N_s(\ell)$. The condition $r_{\text{crit}} \leq r \leq 1$ denotes the weak asymptotic behavior and we use the error estimate $K r^2 N_s(\ell)$. For the derivation of the formulas suggested for error estimates with different types of asymptotic behavior, we refer the reader to Berntsen and Espelid [1991] and Laurie [1994]. In order to attain optimal (or nearly optimal) computational efficiency, the free parameters, r_{crit} and K, need to be tuned on a battery of test integrals to get the best trade-off between reliability and efficiency. In our software, we make full use of the test results provided by Cools et al. [1997].

4.2 Numerical Search Procedure

The numerical integration scheme allows us to accurately determine the input correlation implied by any base correlation. To search for the base correlation that provides a match to the desired input correlation, we use the secant method (also called regula falsi), which is basically a modified version

of Newton's method. We use Υ to denote the function to which the search procedure is applied and define it as the difference between the function $c_{ijh}[\rho_{\mathbf{Z}}]$ evaluated at the unknown base correlation $\rho_{\mathbf{Z}}$ and the given input correlation $\rho_{\mathbf{X}}$, that is, $\Upsilon(\rho_{\mathbf{Z}}) = c_{ijh}[\rho_{\mathbf{Z}}] - \rho_{\mathbf{X}}$. Since the objective is to find $\hat{\rho}_{\mathbf{Z}}$ for which $c_{ijh}[\hat{\rho}_{\mathbf{Z}}] = \rho_{\mathbf{X}}$ holds, we reduce the matching problem to finding zeroes of the function Υ.

In the secant method, the first derivative of the function $\Upsilon(\rho_{\mathbf{Z},m})$ evaluated at point $\rho_{\mathbf{Z},m}$ of iteration m, $d\Upsilon(\rho_{\mathbf{Z},m})/d\rho_{\mathbf{Z},m}$, is approximated by the difference quotient, $[\Upsilon(\rho_{\mathbf{Z},m}) - \Upsilon(\rho_{\mathbf{Z},m-1})]/(\rho_{\mathbf{Z},m} - \rho_{\mathbf{Z},m-1})$ [Blum 1972]. The iterative procedure is given by

$$\rho_{\mathbf{Z},m+1} = \rho_{\mathbf{Z},m} - \Upsilon(\rho_{\mathbf{Z},m}) \left(\frac{\rho_{\mathbf{Z},m} - \rho_{\mathbf{Z},m-1}}{\Upsilon(\rho_{\mathbf{Z},m}) - \Upsilon(\rho_{\mathbf{Z},m-1})} \right) \tag{12}$$

and it is stopped when the values obtained in consecutive iterations ($\rho_{\mathbf{Z},m}$ and $\rho_{\mathbf{Z},m+1}$) are close enough, for instance $|\rho_{\mathbf{Z},m} - \rho_{\mathbf{Z},m+1}| < 10^{-8}$. Clearly, the procedure (12) amounts to approximating the curve $y_m = \Upsilon(\rho_{\mathbf{Z},m})$ by the secant (or chord) joining the points $(\rho_{\mathbf{Z},m}, \Upsilon(\rho_{\mathbf{Z},m}))$ and $(\rho_{\mathbf{Z},m-1}, \Upsilon(\rho_{\mathbf{Z},m-1}))$. Since the problem of interest is to find $\hat{\rho}_{\mathbf{Z}} = \Upsilon^{-1}(0)$, we can regard (12) as a linear interpolation formula for Υ^{-1}; that is, we wish to find the unknown value $\Upsilon^{-1}(0)$ by interpolating the known values $\Upsilon^{-1}(y_m)$ and $\Upsilon^{-1}(y_{m-1})$.

In the one-dimensional case, the secant method can be modified in a way that ensures convergence for any continuous function Υ [Blum 1972]: Following from Proposition 3.2, we choose $\rho_{\mathbf{Z},0} = 0$ and $\rho_{\mathbf{Z},1} = 1$, or $\rho_{\mathbf{Z},0} = 0$ and $\rho_{\mathbf{Z},1} = -1$, depending on whether $\rho_{\mathbf{X}} > 0$ or $\rho_{\mathbf{X}} < 0$, respectively. Therefore, the functions $\Upsilon(\rho_{\mathbf{Z},0})$ and $\Upsilon(\rho_{\mathbf{Z},1})$ have opposite signs. Then there exists a $\hat{\rho}_{\mathbf{Z}}$ between $\rho_{\mathbf{Z},0}$ and $\rho_{\mathbf{Z},1}$, which satisfies $c_{ijh}(\hat{\rho}_{\mathbf{Z}}) - \rho_{\mathbf{X}} = 0$. Next, we determine $\rho_{\mathbf{Z},2}$ by formula (12). Before proceeding with the next iteration, we determine which of the two points $\rho_{\mathbf{Z},0}$, $\rho_{\mathbf{Z},1}$ is such that the value of Υ has the opposite sign to $\Upsilon(\rho_{\mathbf{Z},2})$. We relabel that point as $\rho'_{\mathbf{Z},1}$ and proceed to find $\rho_{\mathbf{Z},3}$ using $\rho_{\mathbf{Z},2}$ and $\rho'_{\mathbf{Z},1}$. This ensures that $\hat{\rho}_{\mathbf{Z}}$ is enclosed in a sequence of intervals $[a_m, b_m]$ such that $a_m \leq a_{m+1} \leq b_{m+1} \leq b_m$ for all m and $b_m - a_m \to 0$ for some m. Since the corresponding function is strictly increasing (J. R. Wilson, personal communication) and quite smooth in the case of the Johnson translation system, the application of this method gives accurate and reliable results converging in a small amount of time, reducing the effort required to solve a large number of matching problems.

5. EXAMPLE

In this section, we present an example that gives an explicit illustration of the framework described in Sections 3 and 4. We select a problem that will be difficult for our technique: The true marginal distribution, which we know, is not Johnson and therefore must be approximated as Johnson by matching the first four moments. Further, for the true marginal (which is uniform), the correlation-matching problem can be solved exactly. However, for our Johnson approximation, we solve the correlation-matching problem using our numerical technique. This allows us to compare a perfectly specified VARTA representation (correct marginals, correct correlations) to our approximation (closest Johnson marginal, numerically matched correlations). However, in both cases,

we achieve the desired autocorrelation structure for the input process by manipulating the autocorrelation structure of the Gaussian vector autoregressive process as suggested in Section 3.

Suppose that we require a trivariate ($k = 3$) random variable with $(0, 1)$ uniform marginal distributions. The correlation matrices are specified at lags 0 and 1 (i.e., $p = 1$) as

$$\Sigma_X(0) = \begin{pmatrix} 1.00000 & 0.36459 & 0.40851 \\ 0.36459 & 1.00000 & 0.25707 \\ 0.40851 & 0.25707 & 1.00000 \end{pmatrix}$$

and

$$\Sigma_X(1) = \begin{pmatrix} 0.28741 & 0.23215 & 0.10367 \\ 0.12960 & 0.28062 & 0.28992 \\ 0.11742 & 0.25951 & 0.16939 \end{pmatrix},$$

respectively.

First, we need to select an autocorrelation structure for the underlying base process, $\text{VAR}_3(1)$, by solving the correlation-matching problem. This is equivalent to solving 12 individual matching problems, each of which can be solved in two different ways.

Case 1. Since the marginals are $(0, 1)$ uniform distributions, it is possible to find the unknown base correlation, $\rho_{\mathbf{Z}}$, by using the relationship

$$\rho_{\mathbf{Z}} = 2 \sin(\pi \rho_{\mathbf{X}}/6),$$

where $\rho_{\mathbf{X}}$ is the desired input correlation [Kruskal 1958].

Case 2. The individual matching problems are solved through the use of the numerical schema suggested in Section 4.

The $(0, 1)$ uniform distribution is approximated by a Johnson-bounded distribution ($\gamma_i = 0.000,\ \delta_i = 0.646,\ \lambda_i = 1.048,\ \xi_i = -0.024$ for $i = 1, 2, 3$), whose first four moments are identical to the first four moments of the uniform distribution, using the AS99 algorithm of Hill et al. [1976]. The probability density functions for the uniform and the approximating Johnson-type distribution are given in Figure 2. The uniform distribution is not a member of the Johnson system, as can be easily seen from the figure: The approximating Johnson bounded distribution has two modes, one antimode, and a range of $[-0.024, 1.024]$. More visually pleasing approximations are possible, but they do not match the moments of the uniform distribution exactly, which is our goal. However, we could solve the correlation matching problem for any approximating distribution that is chosen.

Having solved the correlation-matching problem in two different ways, we solve the multivariate Yule–Walker equations for the autoregressive coefficient matrices and the covariance matrices of the white noise. In each case, the vector autoregressive base process is stationary with a positive definite autocorrelation matrix. Finally, we generate realizations from the underlying vector autoregressive processes and transform the standard normal random variates $z_{i,t}$

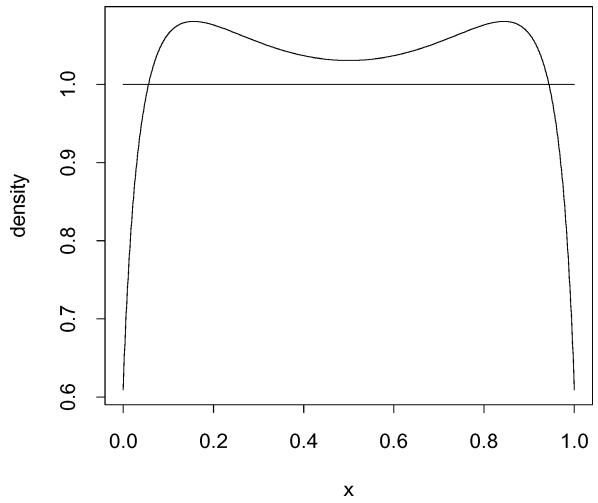

Fig. 2. Probability density functions for uniform and approximating Johnson bounded distributions.

Table I. Kolmogorov-Smirnov Test Statistics for each Component Series

$KS_{\mathbf{X}}$	Case 1	Case 2
$\boldsymbol{X}_1$	0.964	0.929
$\boldsymbol{X}_2$	1.709	1.875
$\boldsymbol{X}_3$	1.055	1.092

into $x_{i,t}$ using the transformations $\Phi(z_{i,t})$ and $\xi_i + \lambda_i(1 + \exp(-(z_{i,t} - \gamma_i)/\delta_i))^{-1}$ for Cases 1 and 2, respectively, for $i = 1, 2, 3$ and $t = 0, 1, \ldots, 10000$.

Next, we evaluate how well the desired marginals and autocorrelation structure of the input process are represented in 10000 generated data points. In Table I, we report the adjusted Kolmogorov-Smirnov ($KS_{\mathbf{X}}$) test statistics [Stephens 1974] indicating the maximum absolute differences between the cdfs of the empirical distribution and the $(0, 1)$ uniform marginal distribution for each component series. As noted by Moore [1982] and Gleser and Moore [1983] in the context of short-memory processes, the critical values and the corresponding nominal levels of significance of goodness-of-fit tests for independent and identically distributed data can be grossly incorrect when observations are dependent. Thus, we use the 5% critical value of 1.358 as a rough guide for judging the adequacy of the fit and provide the quantile—quantile $(Q - Q)$ plots comparing the ith quantile of the empirical distribution function, $X_{(i)}$, with the ith quantile of the uniform distribution function, $(i - 0.5)/10000$, and the Johnson bounded distribution function, $\xi + \lambda f^{-1}[((i - 0.5)/10000 - \gamma)/\delta]$ for $i = 1, 2 \ldots, 10000$, in Figures 3, 4, and 5. It is visually obvious that the generation schema reproduced the desired time series reasonably well. Notice that the second component series represents the desired marginal and autocorrelation structure as successfully as the first and third component series even though the test statistics for the second component series are larger than the ones of

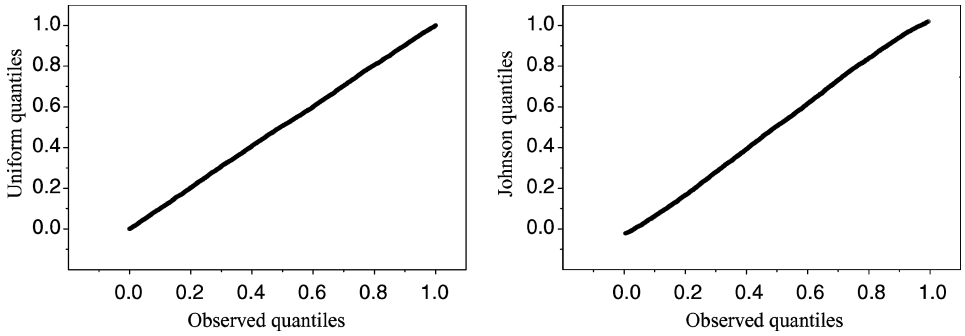

Fig. 3. (Left) Q-Q Plot Comparing the Empirical and Uniform Distribution Functions of the First Component Series (Right) Q-Q Plot Comparing the Empirical and Approximating Johnson Bounded Distribution Functions of the First Component Series

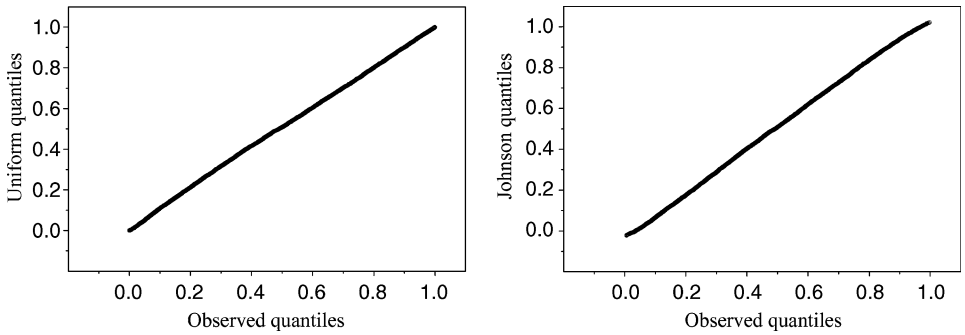

Fig. 4. (Left) Q-Q Plot Comparing the Empirical and Uniform Distribution Functions of the Second Component Series (Right) Q-Q Plot Comparing the Empirical and Approximating Johnson Bounded Distribution Functions of the Second Component Series

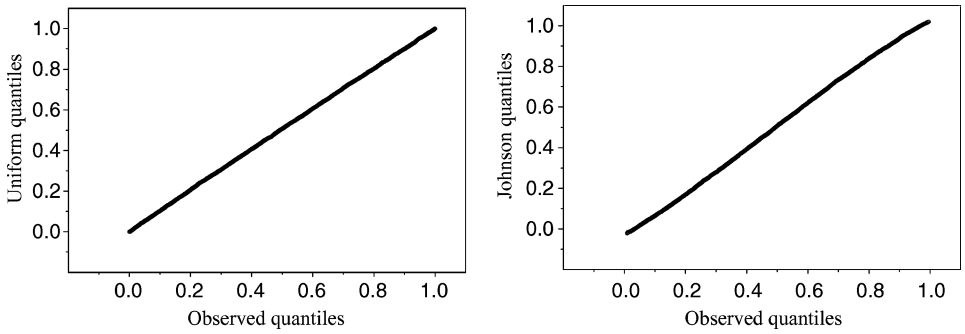

Fig. 5. (Left) Q-Q Plot Comparing the Empirical and Uniform Distribution Functions of the Third Component Series (Right) Q-Q Plot Comparing the Empirical and Approximating Johnson Bounded Distribution Functions of the Third Component Series

the first and third component series. Although the range of the corresponding Johnson-bounded distribution is $(-0.024, 1.024)$ as opposed to $(0, 1)$, we find the Johnson translation system is successful in representing the key features of the desired marginal distributions.

Table II. Absolute Difference (E_1) and Relative Percent Difference (E_2) between the Estimates and the True Parameters for the Input Autocorrelation Structure under Case 1 and Case 2

	E_1		E_2	
$\rho_{\mathbf{X}}(i, j, h)$	Case 1	Case 2	Case 1	Case 2
$\rho_{\mathbf{X}}(1, 2, 0)$	0.004	0.004	0.972	0.983
$\rho_{\mathbf{X}}(1, 3, 0)$	0.001	0.011	0.185	2.617
$\rho_{\mathbf{X}}(2, 3, 0)$	0.002	0.002	0.779	0.784
$\rho_{\mathbf{X}}(1, 1, 1)$	0.008	0.008	2.953	2.946
$\rho_{\mathbf{X}}(1, 2, 1)$	0.003	0.003	1.242	1.237
$\rho_{\mathbf{X}}(1, 3, 1)$	0.006	0.006	5.951	5.951
$\rho_{\mathbf{X}}(2, 1, 1)$	0.009	0.009	7.569	7.568
$\rho_{\mathbf{X}}(2, 2, 1)$	0.008	0.002	2.939	0.617
$\rho_{\mathbf{X}}(2, 3, 1)$	0.009	0.001	3.057	0.386
$\rho_{\mathbf{X}}(3, 1, 1)$	0.002	0.002	2.036	2.038
$\rho_{\mathbf{X}}(3, 2, 1)$	0.010	0.010	3.992	3.987
$\rho_{\mathbf{X}}(3, 3, 1)$	0.000	0.000	0.124	0.127

Finally, in Table II, we report the absolute difference (E_1) and the relative percent difference (E_2) for statistically significant digits between the estimated input autocorrelation structure and the desired input autocorrelation structure used to generate the data. For example, when $\rho_{\mathbf{X}}(2, 1, 1)$ is of interest, we observe an absolute difference of 0.009 and a relative difference of 7.568% between the estimated and true autocorrelation structures under Case 2. We find that Case 2—the VARTA approach—performs as well as Case 1 in incorporating the desired autocorrelation structure into the generated data.

We have developed a stand-alone, PC-based program that implements the VARTA framework with the suggested search and numerical-integration procedures for simulating input processes. The key computational components of the software are written in portable `C++` code and it is available at `<www.andrew.cmu.edu/billerb/>`.

6. CONCLUSION AND FUTURE RESEARCH

In this article, we provide a general-purpose tool for modeling and generating dependent and multivariate input processes. We reduce the setup time for generating each VARTA variate by solving the correlation-matching problem with a numerical method that exploits the features of the Johnson translation system. The evaluation of the composite function $F_X^{-1}[\Phi(\cdot)]$ could be slow and memory intensive in the case of the standard families of distributions, but not Johnson.

However, the framework requires the full characterization of the Johnson-type marginal distribution through parameters $[\gamma, \delta, \lambda, \xi]$ and function $f(\cdot)$ corresponding to the Johnson family of interest. Swain et al. [1988] fit Johnson-type marginals to independent and identically distributed univariate data, but dependent, multivariate data sets are of interest in this paper. Therefore, it would be quite useful to estimate the underlying VARTA model from a given historical data set. This requires the determination of the type of Johnson family and the parameters of the corresponding distribution in such a way that the

dependence structure in the multivariate input data is captured. These issues are the subject of Biller and Nelson [2002, 2003a].

APPENDIX

PROOF OF PROPOSITION 3.2. If $\rho_{\mathbf{Z}}(i, j, h) = 0$, then

$$\begin{aligned} \mathrm{E}[X_{i,t}X_{j,t-h}] &= \mathrm{E}\{F_{X_i}^{-1}[\Phi(Z_{i,t})]F_{X_j}^{-1}[\Phi(Z_{j,t-h})]\} \\ &= \mathrm{E}\{F_{X_i}^{-1}[\Phi(Z_{i,t})]\}\mathrm{E}\{F_{X_j}^{-1}[\Phi(Z_{j,t-h})]\} \\ &= \mathrm{E}[X_{i,t}]E[X_{j,t-h}], \end{aligned}$$

because $\rho_{\mathbf{Z}}(i, j, h) = 0$ implies that $Z_{i,t}$ and $Z_{j,t-h}$ are independent. If $\rho_{\mathbf{Z}}(i, j, h) \geq 0\ (\leq 0)$, then, from the association property [Tong 1990],

$$\mathrm{Cov}[g_1(Z_{i,t}, Z_{j,t-h}), g_2(Z_{i,t}, Z_{j,t-h})] \geq 0 (\leq 0)$$

holds for all nondecreasing functions g_1 and g_2 such that the covariance exists. Selection of $g_1(Z_{i,t}, Z_{j,t-h}) \equiv F_{X_i}^{-1}[\Phi(Z_{i,t})]$ and $g_2(Z_{i,t}, Z_{j,t-h}) \equiv F_{X_j}^{-1}[\Phi(Z_{j,t-h})]$ together with the association property implies the result because $F_{X_\nu}^{-1}[\Phi(\cdot)]$ for $\nu \in \{i, j\}$ is a nondecreasing function from the definition of a cumulative distribution function. □

PROOF OF PROPOSITION 3.3. A correlation of 1 is the maximum possible for bivariate normal random variables. Therefore, taking $\rho_{\mathbf{Z}}(i, j, h) = 1$ is equivalent (in distribution) to setting $Z_{i,t} \leftarrow \Phi^{-1}(U)$ and $Z_{j,t-h} \leftarrow \Phi^{-1}(U)$, where U is a $U(0, 1)$ random variable [Whitt 1976]. This definition of $Z_{i,t}$ and $Z_{j,t-h}$ implies that $X_{i,t} \leftarrow F_{X_i}^{-1}[U]$ and $X_{j,t-h} \leftarrow F_{X_j}^{-1}[U]$, from which it follows that $c_{ijh}(1) = \bar{\rho}_{ij}$ by the same reasoning. Similarly, taking $\rho_{\mathbf{Z}}(i, j, h) = -1$ is equivalent to setting $X_{i,t} \leftarrow F_{X_i}^{-1}[U]$ and $X_{j,t-h} \leftarrow F_{X_j}^{-1}[1 - U]$, from which it follows that $c_{ijh}(-1) = \underline{\rho}_{ij}$. □

LEMMA A.1. *Let $g(z_{i,t}, z_{j,t-h}) \equiv F_{X_i}^{-1}[\Phi[z_{i,t}]]F_{X_j}^{-1}[\Phi[z_{j,t-h}]]$ for given cumulative distribution functions F_{X_i} and F_{X_j}. Then the function g is superadditive.*

PROOF. The result follows immediately from Lemma 1 of Cario et al. [2001] with $z_1 = z_{i,t}$, $z_2 = z_{j,t-h}$, $X_1 = X_i$, and $X_2 = X_j$. □

PROOF OF THEOREM 3.4. It is sufficient to show that, if $\rho_{\mathbf{Z}}^* \geq \rho_{\mathbf{Z}}$ then $c_{ijh}[\rho_{\mathbf{Z}}^*] \geq c_{ijh}[\rho_{\mathbf{Z}}]$, where for brevity we let $\rho_{\mathbf{Z}} = \rho_{\mathbf{Z}}(i, j, h)$ and $\rho_{\mathbf{Z}}^* = \rho_{\mathbf{Z}}^*(i, j, h)$. Following the definition of the function c_{ijh}, this is equivalent to saying that, if $\rho_{\mathbf{Z}}^* \geq \rho_{\mathbf{Z}}$, then $\mathrm{E}_{\rho_{\mathbf{Z}}^*}[X_{i,t}X_{j,t-h}] \geq \mathrm{E}_{\rho_{\mathbf{Z}}}[X_{i,t}X_{j,t-h}]$.

Let $\Phi_{\rho_{\mathbf{Z}}}[z_{i,t}, z_{j,t-h}]$ be the joint cdf of $Z_{i,t}$ and $Z_{j,t-h}$, which is the standard bivariate normal distribution with correlation $\rho_{\mathbf{Z}}$. From Slepian's inequality [Tong 1990], it follows that

$$\Phi_{\rho_{\mathbf{Z}}^*}[z_{i,t}, z_{j,t-h}] \geq \Phi_{\rho_{\mathbf{Z}}}[z_{i,t}, z_{j,t-h}]$$

for all $z_{i,t}$ and $z_{j,t-h}$ if $\rho_{\mathbf{Z}}^* \geq \rho_{\mathbf{Z}}$.

Let $g(z_{i,t}, z_{j,t-h}) \equiv F_{X_i}^{-1}[\Phi[z_{i,t}]]F_{X_j}^{-1}[\Phi[z_{j,t-h}]]$. The result we need is a consequence of Corollary 2.1 of Tchen [1980]. Specializing Corollary 2.1 to the case

$n = 2$ and continuous joint distribution function $\Phi_{\rho_{\mathbf{Z}}}$, Tchen [1980] shows that

$$\begin{aligned}
&\mathrm{E}_{\rho^*_{\mathbf{Z}}}[X_{i,t}X_{j,t-h}] - \mathrm{E}_{\rho_{\mathbf{Z}}}[X_{i,t}X_{j,t-h}] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(z_{i,t}, z_{j,t-h})\, d\Phi_{\rho^*_{\mathbf{Z}}}(z_{i,t}, z_{j,t-h}) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(z_{i,t}, z_{j,t-h})\, d\Phi_{\rho_{\mathbf{Z}}}(z_{i,t}, z_{j,t-h}) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\Phi_{\rho^*_{\mathbf{Z}}}(z_{i,t}, z_{j,t-h}) - \Phi_{\rho_{\mathbf{Z}}}(z_{i,t}, z_{j,t-h})]\, dK(z_{i,t}, z_{j,t-h})
\end{aligned}$$

for some positive measure K, provided that $g(z_{i,t}, z_{j,t-h})$ is "2-positive" (which is implied by superadditivity, see Lemma A.1), and a bounding condition on $g(z_{i,t}, z_{j,t-h})$ holds (the condition is trivially satisfied here). But, as a consequence of Slepian's inequality,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\Phi_{\rho^*_{\mathbf{Z}}}(z_{i,t}, z_{j,t-h}) - \Phi_{\rho_{\mathbf{Z}}}(z_{i,t}, z_{j,t-h})]\, dK(z_{i,t}, z_{j,t-h}) \geq 0$$

establishing the result. □

PROOF OF THEOREM 3.5. Theorem 3.5 follows immediately from Lemma 2 of Cario et al. [2001] with $Z_1 \equiv Z_{i,t}$, $Z_2 \equiv Z_{j,t-h}$, $X_1 \equiv X_{i,t}$, $X_2 \equiv X_{j,t-h}$, and $\rho = \rho_{\mathbf{Z}}(i, j, h)$. □

ACKNOWLEDGMENTS

The authors thank the referees, and especially James R. Wilson, for providing numerous improvements to the article.

REFERENCES

BERNTSEN, J. AND ESPELID, T. O. 1991. Error estimation in automatic quadrature routines. *ACM Trans. Math. Soft. 17*, 233–252.

BILLER, B. AND NELSON, B. L. 2002. Parameter estimation for ARTA processes. In *Proceedings of the 2002 Winter Simulation Conference*, E. Yucesan, C. H. Chen, J. L. Snowdon, and J. M. Charnes, Eds. Institute of Electrical and Electronics Engineers, Piscataway, N.J., 255–262.

BILLER, B. AND NELSON, B. L. 2003a. Fitting time-series input processes for simulation. Tech. rep., Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA.

BILLER, B. AND NELSON, B. L. 2003b. On the performance of the ARTA fitting algorithm for stochastic simulation. Tech. Rep., Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, Pa.

BILLER, B. AND NELSON, B. L. 2003c. Online companion to "modeling and generating multivariate time-series input processes using a vector autoregressive technique. Tech. Rep., Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Ill.

BLUM, E. K. 1972. *Numerical Analysis and Computation Theory and Practice*. Addison-Wesley, Reading, Mass.

CAMBANIS, S. AND MASRY, E. 1978. On the reconstruction of the covariance of stationary Gaussian processes observed through zero-memory nonlinearities. *IEEE Trans. Inf. Theory 24*, 485–494.

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

CARIO, M. C. AND NELSON, B. L. 1998. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS J. Comput. 10*, 72–81.

CARIO, M. C., NELSON, B. L., ROBERTS, S. D., AND WILSON, J. R. 2001. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Tech. Rep., Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Ill.

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

CLEMEN, R. T. AND REILLY, T. 1999. Correlations and copulas for decision and risk analysis. *Manage. Sci. 45*, 208–224.

COOK, R. D. AND JOHNSON, M. E. 1981. A family of distributions for modeling non-elliptically symmetric multivariate data. *J. Roy. Stat. Soc. B 43*, 210–218.

COOLS, R. 1994. The subdivision strategy and reliability in adaptive integration revisited. Tech. Rep., Department of Computer Science, Katholieke University Leuven, Leuven, Belgium.

COOLS, R., LAURIE, D., AND PLUYM, L. 1997. Algorithm 764: `Cubpack++`: A `C++` package for automatic two-dimensional cubature. *ACM Trans. Math. Softw. 23*, 1–15.

DEVROYE, L. 1986. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

GENZ, A. 1992. Statistics applications of subregion adaptive multiple numerical integration. In *Numerical Integration—Recent Developments, Software, and Applications*. 267–280.

GHOSH, S. AND HENDERSON, S. G. 2002. Chessboard distributions and random vectors with specified marginals and covariance matrix. *Oper. Res. 50*, 820–834.

GLESER, L. J. AND MOORE, D. S. 1983. The effect of dependence on chi-squared test and empiric distribution tests of fit. *Ann. Stat. 11*, 1100–1108.

GROSS, D. AND JUTTIJUDATA, M. 1997. Sensitivity of output performance measures to input distributions in queueing simulation modeling. In *Proceedings of the 1997 Winter Simulation Conference*, D. H. Withers, B. L. Nelson, S. Andradóttir, and K. J. Healy, Eds. Institute of Electrical and Electronics Engineers, Piscataway, N.J., 296–302.

HILL, I. D., HILL, R., AND HOLDER, R. L. 1976. Fitting Johnson curves by moments. *Appl. Stat. 25*, 180–189.

HILL, R. R. AND REILLY, C. H. 1994. Composition for multivariate random vectors. In *Proceedings of the 1994 Winter Simulation Conference*, D. A. Sadowski, A. F. Seila, J. D. Tew, and S. Manivannan, Eds. Institute of Electrical and Electronics Engineers, Piscataway, N.J., 332–339.

JOHNSON, M. E. 1987. *Multivariate Statistical Simulation*. Wiley, New York.

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.

KROMMER, A. R. AND UEBERHUBER, C. W. 1994. *Numerical Integration on Advanced Computer Systems*. Springer-Verlag, New York.

KRUSKAL, W. 1958. Ordinal measures of association. *J. Ameri. Stat, Assoc. 53*, 814–861.

LAURIE, D. P. 1994. Null rules and orthogonal expansions. In *Proceedings of the International Conference on Special Functions, Approximation, Numerical Quadrature and Orthogonal Polynomials*, R. V. Zahar, Ed., Birkhävser, 359–370.

LAW, A. M. AND KELTON, W. D. 2000. *Simulation Modeling and Analysis*. McGraw-Hill, New York.

LEWIS, P. A. W., MCKENZIE, E., AND HUGUS, D. K. 1989. Gamma processes. *Commun. Stat. Stoch. Models 5*, 1–30.

LI, S. T. AND HAMMOND, J. L. 1975. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Trans. Syst., Man, and Cybernet. 5*, 557–561.

LIVNY, M., MELAMED, B., AND TSIOLIS, A. K. 1993. The impact of autocorrelation on queueing systems. *Manage. Sci. 39*, 322–339.

LURIE, P. M. AND GOLDBERG, M. S. 1998. An approximate method for sampling correlated random variables from partially-specified distributions. *Manage. Sci. 44*, 203–218.

LUTKEPOHL, H. 1993. *Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York.

MARDIA, K. V. 1970. A translation family of bivariate distributions and Frèchet's bounds. *Sankhya A 32*, 119–122.

MELAMED, B. 1991. TES: A class of methods for generating autocorrelated uniform variates. *ORSA J. Comput. 3*, 317–329.

MELAMED, B., HILL, J. R., AND GOLDSMAN, D. 1992. The TES methodology: Modeling empirical stationary time series. In *Proceedings of the 1992 Winter Simulation Conference*, R. C. Crain, J. R. Wilson, J. J. Swain, and D. Goldsman, Eds. Institute of Electrical and Electronics Engineers, Piscataway, N.J., 135–144.

MOORE, D. S. 1982. The effect of dependence on chi-squared tests of fit. *Ann. Stat. 4*, 357–369.
NELSEN, R. B. 1998. *An Introduction to Copulas*. Springer-Verlag, New York.
NELSON, B. L. AND YAMNITSKY, M. 1998. Input modeling tools for complex problems. In *Proceedings of the 1998 Winter Simulation Conference*, J. S. Carson, M. S. Manivannan, D. J. Medeiros, and E. F. Watson, Eds. Institute of Electrical and Electronics Engineers, Piscataway, N.J., 105–112.
PRITSKER, A. A. B., MARTIN, D. L., REUST, J. S., WAGNER, M. A. F., WILSON, J. R., KUHL, M. E., ROBERTS, J. P., DAILY, O. P., HARPER, A. M., EDWARDS, E. B., BENNETT, L., BURDICK, J. F., AND ALLEN, M. D. 1995. Organ transplantation policy evaluation. In *Proceedings of the 1995 Winter Simulation Conference*, W. R. Lilegdon, D. Goldsman, C. Alexopoulos, and K. Kang, Eds. Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1314–1323.
RABINOWITZ, P. AND RICHTER, N. 1969. Perfectly symmetric two-dimensional integration formulas with minimal number of points. *Math. Comput. 23*, 765–799.
ROBINSON, I. AND DONCKER, E. D. 1981. Algorithm 45: Automatic computation of improper integrals over a bounded or unbounded planar region. *Computing 27*, 253–284.
SONG, W. T., HSIAO, L., AND CHEN, Y. 1996. Generating pseudorandom time series with specified marginal distributions. *European Journal of Operational Research 93*, 1–12.
STEPHENS, M. A. 1974. EDF statistics for goodness of fit and some comparisons. *J. Amer. Stat. Assoc. 69*, 730–737.
STROUD, A. H. 1971. *Approximate Calculation of Multiple Integrals*. Prentice-Hall, Englewood Cliffs. N.J.
SWAIN, J. J., VENKATRAMAN, S., AND WILSON, J. R. 1988. Least-squares estimation of distribution functions in Johnson's translation system. *J. Stat. Comput. Simul. 29*, 271–297.
TCHEN, A. H. 1980. Inequalities for distributions with given marginals. *Ann. Probab. 8*, 814–827.
TONG, Y. L. 1990. *The Multivariate Normal Distribution*. Springer-Verlag, New York.
WARE, P. P., PAGE, T. W., AND NELSON, B. L. 1998. Automatic modeling of file system workloads using two-level arrival processes. *ACM Trans. Model. Comput. Simul. 8*, 305–330.
WHITT, W. 1976. Bivariate distributions with given marginals. *Ann. Stat. 4*, 1280–1289.
WILLEMAIN, T. R. AND DESAUTELS, P. A. 1993. A method to generate autocorrelated uniform random numbers. *J. Stat. Comput. Simul. 45*, 23–31.

Received July 2001; revised February 2002 and October 2002; accepted April 2003