Modeling and Generating Multivariate Time Series with Arbitrary Marginals Using a Vector Autoregressive Technique

Bahar Deler
Barry L. Nelson
Dept. of IE & MS, Northwestern University

September 2000

Abstract

We present a model for representing stationary multivariate time series with arbitrary marginal distributions and autocorrelation structures and describe how to generate data quickly and 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 correlation structure of the Gaussian vector autoregressive process so that we achieve the desired correlation structure for the simulation input process. For the purpose of computational efficiency, we provide a numerical method, which incorporates a numerical-search procedure and a numerical-integration technique, for solving this correlation-matching problem.

Keywords: Computer simulation, vector autoregressive process, vector time series, multivariate input modeling, numerical integration.

1 Introduction

Representing the uncertainty or randomness in a simulated system by an input model is one of the challenging problems in the practical application of computer simulation. There are an abundance of examples, from manufacturing to service applications, where input modeling is critical; e.g., modeling the time to failure for a machining process, the demand per unit time for inventory of a product, or the times between arrivals of calls to a call center. Further, building a large-scale discrete-event stochastic simulation model may require the development of a large 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 and 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 Processor (Rockwell Software Inc.), Stat::Fit (Geer Mountain Software Corporation), and BestFit (Palisade Corporation). However, the input models incorporated in these packages will sometimes fall short because they emphasize good representations for the marginal distribution of independent and identically distributed (i.i.d.) processes. However, dependent, multivariate time-series input processes occur naturally in the simulation of many service, communications, and manufacturing systems (e.g., Melamed, Hill, and Goldsman 1992, Ware, Page, and Nelson 1998), so there may be dependencies in time sequence or with respect to other input processes in the simulation. Ignoring dependence can lead to performance measures that are seriously in error and a significant distortion of the simulated system. When it was noted that the limited shapes represented by the standard families of distributions (e.g., beta, Erlang, exponential, gamma, lognormal, normal, Poisson, triangular, uniform, or Weibull) were not flexible enough to represent some of the characteristics of the marginal input processes, these input-modeling packages were improved by expanding the list of families of distributions. If the same philosophy is applied to modeling dependence, then the list of candidate families of distributions quickly explodes.

In this paper, we provide an input modeling framework for continuous distributions which addresses some of the limitations of the current input models. The framework is based on the ability to represent and generate random variates from a stationary k-variate vector time series $\{\mathbf{X}_t;\ t = 0, 1, 2, \ldots\}$, a model that includes univariate i.i.d. 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 modeling framework rather than a long list of more specialized models.

We let each component time series $\{X_{i,t};\ i = 1, 2, \ldots, k;\ t = 0, 1, 2, \ldots\}$ have a Johnson marginal distribution to achieve a wide variety of distributional shapes, while accurately reflecting the desired dependence structure via 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 use a transformation-oriented approach that invokes the theory behind the standardized Gaussian vector autoregressive processes. Therefore, we refer to $\mathbf{X}_t$ as having a VARTA (Vector-Autoregressive-To-Anything) distribution. The i^{th} time series is obtained via the transformation $X_{i,t} = F_{X_i}^{-1}[\Phi(Z_{i,t})]$, where F_{X_i} is the Johnson-type cumulative distribution function suggested for the i^{th} component series of the input process and $\{Z_{i,t};\ i = 1, 2, \ldots, k;\ t = 0, 1, 2, \ldots\}$ is the i^{th} component series of the k-variate Gaussian autoregressive process of order p with the representation $\mathbf{Z}_t = \sum_{h=1}^{p} \boldsymbol{\alpha}_h \mathbf{Z}_{t+h} + \mathbf{u}_t$ (see Section 3.1.1). This transformation-oriented approach requires matching the desired correlation structure of the input process by manipulating the correlation structure of the Gaussian vector autoregressive process. In order to make this method practically feasible, we propose an efficient numerical scheme to solve the correlation-matching problem for generating VARTA processes.

The remainder of this paper 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 the background information on vector autoregressive models and the Johnson family of distributions, is presented in Section 3. We describe the numerical search procedure supported by a numerical integration technique in the same section. Finally, an example is provided in Section 4 and concluding remarks are given in Section 5.

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

emphasize the dependence structure of the data, and we refer the reader to Law and Kelton (2000) and Nelson and Yamnitsky (1998) for detailed surveys of 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; and (ii) construct a univariate series of autocorrelated uniform random variables, $\{U_t;\ t = 0, 1, 2, \ldots\}$, as the base process and transform it to the univariate 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 the former approach 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, McKenzie, and Hugus (1989), who constructed time series with gamma marginals. In this paper, we take the latter approach, which is more general and has been used previously by various researchers including Melamed (1991), Melamed, Hill, and Goldsman (1992), Willemain and Desautels (1993), Song, Hsiao, and Chen (1996), and Cario and Nelson (1996). Among all of these, the most general model is given by Cario and Nelson, who redefined the base process as a Gaussian autoregressive model from which a series of autocorrelated uniform random variables are constructed via the probability-integral transformation. Further, their model controls the correlations between 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 model 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 dependent components and 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 taken to model such processes is to transform multivariate normal vectors into vectors with arbitrary marginal distributions, which is related to the methods that transform a random vector with uniformly distributed marginals (Cook and

Johnson 1981 and Ghosh and Henderson 2000). The first reference to the idea of transforming multivariate normal vectors 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 hint at the same idea. Among these, we refer the interested reader to Cario, Nelson, Roberts, and Wilson (2001) and Chen (2000), who generate random vectors with arbitrary marginal distributions and correlation matrix by a so-called NORTA (Normal-To-Anything) method, involving a component-wise transformation of a multivariate normal random vector. Cario, Nelson, Roberts, and Wilson also discuss the extension of their idea to discrete and mixed marginal distributions. Their results can be considered as an extension of 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.

Our transformation-oriented approach is quite different from techniques that randomly mix distributions with extremal correlations to obtain intermediate correlations. See Hill and Reilly 1994 for an example of the mixing technique.

The primary contribution of this paper 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. In other words, univariate time-series processes and finite-dimensional random vectors are special cases of our model.

3 The Model, Theory, and Implementation

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

3.1 Background

Our premise in the development of the VARTA framework 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 present in

data, while being easy to use, adjust, and understand. We achieve this flexibility by incorporating the vector autoregressive processes and the Johnson family of distributions into our model in order to characterize the process dependence and marginal distributions, respectively. We define the base process as a standard Gaussian vector autoregressive process $\mathbf{Z}_t$ whose correlation structure is adjusted in order to achieve the target correlation 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, we apply the transformation $X_{i,t} = F_{X_i}^{-1}[U_{i,t}]$, which ensures that each component series, $\{X_{i,t};\ i = 1, 2, \ldots, k;\ t = 0, 1, 2, \ldots\}$, has the desired Johnson-type marginal distribution, F_{X_i}.

Below, we provide brief background information on vector autoregressive processes and the Johnson family of distributions; we then present the framework.

3.1.1 The VAR_k(p) Model

In a k-variate vector autoregressive model of order p (the $\mathrm{VAR}_k(p)$ model) the presence of each variable is represented by a linear combination of a finite number of past observations of all the variables plus a random error. This is written in matrix notation as

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

where $\mathbf{Z}_t = (Z_{1,t}, Z_{2,t}, \cdots, Z_{k,t})'$ is a $(k \times 1)$ random vector of the observations recorded at time t and the $\boldsymbol{\alpha}_i$ are fixed $(k \times k)$ autoregressive coefficient matrices for $i = 1, \ldots, p$. Although we assume zero mean $\mathrm{E}[\mathbf{Z}_t]$, the definition (1) can be modified to allow the possibility of a non-zero mean by adding a constant term to its right-hand side. 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, with $(k \times k)$ covariance matrix $\boldsymbol{\Sigma}_u$ so that

$$\mathrm{E}[\mathbf{u}_t] = \mathbf{0}_{(k \times 1)} \ \text{ and } \ \mathrm{E}[\mathbf{u}_t \mathbf{u}_{t+h}'] = \begin{cases} \boldsymbol{\Sigma}_u & \text{if} \quad h = 0, \\ \mathbf{0}_{(k \times k)} & \text{otherwise.} \end{cases}$$

Although the definition of the $\mathrm{VAR}_k(p)$ model does not require the multivariate white noise to be Gaussian, our model makes this assumption. We also assume stability, which further implies

the stationarity of a $\text{VAR}_k(p)$ process started in the infinite past; that is, the $\boldsymbol{\alpha}_i$'s do not depend on time and they satisfy the condition 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).

Under the stability assumption, a first-order vector autoregressive process, $\mathbf{Z}_t$, can also be expressed in terms of past and present white noise vectors as

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

Using the representation (2), the first and second moments of the $\mathbf{Z}_t$ process are seen to be

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

because $\mathrm{E}[\mathbf{u}_t\mathbf{u}_s'] = 0$ for $t \neq s$ and $\mathrm{E}[\mathbf{u}_t\mathbf{u}_t'] = \boldsymbol{\Sigma}_u, \ \forall\, t$ (Lutkepohl 1993; Appendix C.3, Proposition C.8). We use the covariance matrices $\boldsymbol{\Sigma}_Z(h)$, $h = 0, \ldots, p$, in order to characterize the covariance structure of the corresponding process.

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 model

$$\bar{\mathbf{Z}}_t = \bar{\boldsymbol{\alpha}}_1 \bar{\mathbf{Z}}_{t-1} + \bar{\mathbf{u}}_t \tag{3}$$

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 also known as "the state-space model" of the k-variate autoregressive process of order p. Following the foregoing discussion, the first and second moments of $\bar{\mathbf{Z}}_t$ are written as

$$\mathrm{E}[\bar{\mathbf{Z}}_t] = \mathbf{0}_{(kp\times 1)} \text{ and } \boldsymbol{\Sigma}_{\bar{Z}_t}(h) = \sum_{i=0}^{\infty} \bar{\boldsymbol{\alpha}}_1^i \boldsymbol{\Sigma}_{\bar{u}} (\bar{\boldsymbol{\alpha}}_1^{i+h})', \tag{4}$$

where $\boldsymbol{\Sigma}_{\bar{u}} = \mathrm{E}[\bar{\mathbf{u}}_t \bar{\mathbf{u}}_t']$. 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 covariances of the $\mathrm{VAR}_k(p)$ model are given by $\boldsymbol{\Sigma}_{Z_t}(h) = \mathbf{J}\boldsymbol{\Sigma}_{\bar{Z}_t}(h)\mathbf{J}'$, and are also time-invariant (Lutkepohl 1993).

We can describe the $\mathrm{VAR}_k(p)$ model using either its covariance structure, $\boldsymbol{\Sigma}_Z(h)$ for $h = 0, \dots, p$, or its system parameters, $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p$ and $\boldsymbol{\Sigma}_u$. In input modeling problems we directly adjust $\boldsymbol{\Sigma}_Z(h)$ to achieve the desired correlation structure of $\mathbf{X}_t$. To determine $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p$ and $\boldsymbol{\Sigma}_u$ from $\boldsymbol{\Sigma}_Z(h)$, $h = 0, \dots, p$, we simply solve the multivariate Yule-Walker equations given by $\boldsymbol{\alpha} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_p)_{(k\times kp)}$, $\boldsymbol{\Sigma} = \Big(\boldsymbol{\Sigma}_Z(1), \boldsymbol{\Sigma}_Z(2), \cdots, \boldsymbol{\Sigma}_Z(p)\Big)_{(k\times kp)}$, and

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

Once the $\boldsymbol{\alpha}$ is obtained, $\boldsymbol{\Sigma}_u$ can be determined from

$$\boldsymbol{\Sigma}_u = \boldsymbol{\Sigma}_Z(0) - \boldsymbol{\alpha}_1\boldsymbol{\Sigma}_Z'(1) - \cdots - \boldsymbol{\alpha}_p\boldsymbol{\Sigma}_Z'(p). \tag{5}$$

We ensure that each component series of the input process $\{X_{i,t};\ i = 1, 2, \ldots, k;\ t = 0, 1, 2, \ldots\}$ 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 with mean $\mathbf{0}$ (Appendix A; Lemma 3.13). 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. We force $Z_{i,t}$ to be standard normal by defining $\mathbf{\Sigma}_Z(0)$ to be a correlation matrix, and all entries in $\mathbf{\Sigma}_Z(h), h = 1, 2, \ldots, p$ to be correlations. For this reason we will use the terms covariance structure and correlation structure of $\mathbf{Z}_t$ interchangably in the remainder of this paper.

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 the additional distributional properties, is in Appendix A.

Theorem 3.1 *Let $\mathbf{Z}_t$ denote the stationary p^{th}-order vector autoregressive process, $VAR_k(p)$, as defined in (1). The random variable $\mathbf{Z} = (Z_{i,t}, Z_{j,t+h})'$, $i \neq j$, has a bivariate normal distribution with density function*

$$f(\mathbf{z}; \mathbf{\Sigma}_2) = \frac{1}{2\pi|\mathbf{\Sigma}_2|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{z}'\mathbf{\Sigma}_2^{-1}\mathbf{z}},\ \ \mathbf{z} \in \Re^2,\ \textit{where}\ \ \mathbf{\Sigma}_2 = \begin{pmatrix} 1 & \rho_{\mathbf{Z}}(i,j,h) \\ \rho_{\mathbf{Z}}(i,j,h) & 1 \end{pmatrix}_{(2\times 2)}.$$

Further, the corresponding distribution is non-singular for $|\rho_{\mathbf{Z}}(i, j, h)| < 1$.

Proof. See Appendix A.

Using the distributional properties provided in this section, we can achieve the target autocorrelation structure of the input process $\mathbf{X}_t$ by adjusting the autocorrelation structure of the Gaussian vector autoregressive process $\mathbf{Z}_t$ as described in Section 3.2 below.

To generate a multivariate time series with given Johnson-type marginals and correlation structure, we need to be able to simulate a $\text{VAR}_k(p)$ series of any required length, say T. We now review explain how to do this using standard theory (Lutkepohl 1993):

- First, obtain the starting values, $\mathbf{z}_{-p+1}, \ldots, \mathbf{z}_0$, and a series of Gaussian white noise vectors, $\mathbf{u}_1, \ldots, \mathbf{u}_T$, using the covariance structure given by $\mathbf{\Sigma}_Z(h)$ for $h = 0, \ldots, p$ and the implied system parameters $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p$ and $\mathbf{\Sigma}_u$. Then generate 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 the Gaussian disturbances, choose k independent univariate standard normal variates $v_1, \ldots, v_k$, multiply by a $(k \times k)$ matrix $\mathbf{P}$ for which $\mathbf{PP'} = \mathbf{\Sigma}_u$, and repeat this process T times.

- To generate the starting values $\mathbf{z}_{-p+1}, \ldots, \mathbf{z}_0$ whose joint distribution is given by a non-singular p-variate normal distribution (Appendix A; Lemma 3.13), we choose a $(kp \times kp)$ matrix $\mathbf{Q}$ such that $\mathbf{QQ'} = \mathbf{\Sigma_Z}$. Then we obtain p initial starting vectors as $(\mathbf{z}'_0, \cdots, \mathbf{z}'_{-p+1}) = \mathbf{Q}\,(v_1, \cdots, v_{kp})'$, where the v_i's are independent normal variates with mean zero and unit variance. In this way, we ensure the same correlation structure for the initial values and the rest of the time series in such a way that the process starts stationary.

3.1.2 Johnson Family of Distributions

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 of distributions (Johnson 1949) for input modeling problems in which data are plentiful and nearly automated input modeling is required. Our motivation for using the Johnson 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. The Johnson system can match any feasible first four moments, while the standard input models incorporated in some existing software packages and simulation languages typically match only one or two moments. Thus, our goal is to match or represent key features of the data at hand, as opposed to finding the "true" distribution that was the source of the data.

The Johnson translation system for a random variable X, whose range depends on the family of interest, is defined by a cumulative distribution function (cdf) of the form

$$F_X(x) = \Phi\{\gamma + \delta f[(x - \xi)/\lambda]\}, \tag{6}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution, 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:

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

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

3.2 The Model

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

(1) Each component time series $\{X_{i,t};\ t = 0, 1, 2, \ldots\}$ has a Johnson 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 $i = 1, \ldots, k$.

(2) The dependence structure is specified via 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 $\mathbf{\Sigma}_X(h) = \text{Corr}(\mathbf{X}_t, \mathbf{X}_{t+h}) = [\rho_{\mathbf{X}}(i, j, h)]_{(k \times k)}$ for $h = 0, 1, \ldots, p$ where $\rho_{\mathbf{X}}(i, i, 0) = 1$. Using the first $h = 0, 1, \ldots, p$ of these matrices, we define $\mathbf{\Sigma}_{\mathbf{X}}$ analogously to $\mathbf{\Sigma}_{\mathbf{Z}}$.

Accounting for dependence via Pearson product-moment correlation is a practical compromise we make in our 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 i^{th} 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 correlation structure, $\mathbf{\Sigma}_Z(h)$, $h = 0, 1, \ldots, p$, for the base process that gives the desired correlation structure, $\mathbf{\Sigma}_X(h)$, $h = 0, 1, \ldots, p$, for the input process.

For $i \neq j$ and $h = 0, 1, 2, \ldots, p$, we let $\rho_{\mathbf{Z}}(i, j, h)$ be the $(i, j)^{\text{th}}$ element of the lag-h correlation matrix, $\mathbf{\Sigma}_Z(h)$, and let $\rho_{\mathbf{X}}(i, j, h)$ be the $(i, j)^{\text{th}}$ element of $\mathbf{\Sigma}_X(h)$. The correlation matrix of $\mathbf{Z}$ directly determines the correlation matrix of $\mathbf{X}$, because

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

for all $i \neq j$ and $h = 0, 1, 2, \ldots, p$. Notice that only $\text{E}[X_{i,t}X_{j,t+h}]$ depends on $\mathbf{Sigma}_Z$, 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_i and F_j). Since $(Z_{i,t}, Z_{j,t+h})'$ has a standard bivariate normal distribution with correlation $\rho_{\mathbf{Z}}(i, j, h)$ (Theorem 3.1), we have

$$\begin{aligned}\text{E}[X_{i,t}X_{j,t+h}] &= \text{E}\Big[F_{X_i}^{-1}[\Phi(Z_{i,t})]F_{X_j}^{-1}[\Phi(Z_{j,t+h})]\Big] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_i}^{-1}[\Phi(z_{i,t})]F_{X_j}^{-1}[\Phi(z_{j,t+h})]\vartheta_{\rho_{\mathbf{Z}}(i,j,h)}(z_{i,t}, z_{j,t+h})dz_{i,t}dz_{j,t+h}, \qquad (7)\end{aligned}$$

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_i and F_j for which the expectation (7) exists. However, since $Z_{i,t}$ and $Z_{j,t+h}$ are standard normal random variables, the Johnson translation system is a particularly good choice because

$$X_{i,t} = F_{X_i}^{-1}\left[\Phi(Z_{i,t})\right] = \xi_i + \lambda_i f_i^{-1}\left[(z_{i,t} - \gamma_i)/\delta_i\right]$$

$$X_{j,t+h} = F_{X_j}^{-1}\left[\Phi(Z_{j,t+h})\right] = \xi_j + \lambda_j f_j^{-1}\left[(z_{j,t+h} - \gamma_j)/\delta_j\right]$$

avoiding the need to evaluate $\Phi(Z)$.

From (7) 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 this function by $c_{ijh}[\rho_{\mathbf{Z}}(i,j,h)]$ and define it as

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

Thus, the problem of determining $\mathbf{\Sigma}_Z(h)$, $h = 0, 1, \ldots, p$, for $\mathbf{Z}$ that gives the desired correlation matrices $\mathbf{\Sigma}_X(h)$, $h = 0, 1, \ldots, p$, for $\mathbf{X}$ 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 an efficient numerical search to find the $\rho_{\mathbf{Z}}(i,j,h)$ values within a predetermined precision. We primarily extend the results in Cambaris and Marsy (1978), Cario and Nelson (1996), and Cario, Nelson, Roberts, and Wilson (2001)—which apply to time-series input processes with identical marginal distributions and random vectors with arbitrary marginal distributions—to the vector time-series input processes with arbitrary marginal distributions. The proofs of all results may be found in Appendix B.

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 distribution F_{X_i} and F_{X_j}, $c_{ijh}(0) = 0$ and $\rho_{\mathbf{Z}}(i,j,h) \ge 0$ (≤ 0) implies that $c_{ijh}\left[\rho_{\mathbf{Z}}(i,j,h)\right] =\ge 0$ (≤ 0).*

It follows from the proof of Proposition 3.2 that taking $\rho_{\mathbf{Z}}(i,j,h) = 0$ results in a vector 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 correlations are attainable.

Proposition 3.3 *Let $\underline{\rho}_{ij}$ and $\bar{\rho}_{ij}$ be the minimum and maximum feasible 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 $c_{ijh}\left[\rho_{\mathbf{Z}}(i,j,h)\right]$.

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

Since $c_{ijh}\left[\rho_{\mathbf{Z}}(i,j,h)\right]$ 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 $c_{ijh}\left[\rho_{\mathbf{Z}}(i,j,h)\right] \approx \rho_{\mathbf{X}}(i,j,h)$. 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 our model. Furthermore, from Proposition 3.3, Theorem 3.5, and the Intermediate Value Theorem, any feasible bivariate correlation for F_{X_i} and F_{X_j} is attainable under our 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 our computationally accurate and efficient numerical scheme, which we present in the following section.

Throughout the previous discussion we assumed that there exists a joint distribution with marginal distributions $F_{X_i},\ i = 1,2,\ldots,k$ and correlation structure characterized by $\mathbf{\Sigma}_X(h)$ for $h = 0,1,\ldots,p$. However, not all combinations of $F_{X_i},\ i = 1,\ldots,k$, and $\mathbf{\Sigma}_X(h)$, $h = 0,1,\ldots,p$, are feasible. Clearly, for the correlation structure to be feasible, we must have $\underline{\rho}_{ij} \leq \rho_{\mathbf{X}}(i,j,h) \leq \bar{\rho}_{ij}$ for each $i \neq j$ and $h = 0,1,\ldots,p$. In addition, $\mathbf{\Sigma_X}$ must be positive definite and this can be ensured by selecting a positive definite base correlation matrix $\mathbf{\Sigma_Z}$. We present this result in the following proposition.

Proposition 3.6 *If* $\mathbf{\Sigma_Z}$ *is nonnegative definite, then so is* $\mathbf{\Sigma_X}$ *implied by the VARTA transformation.*

Unfortunately, the converse of the above proposition does not necessarily hold; that is, there exists sets of marginals with feasible correlation structure that are not representable by the VARTA transformation. Both Li and Hammond (1975) and Lurie and Goldberg (1998) give examples where this appears to be the case and recently Ghosh and Henderson (2000) prove the existence of a joint

distribution that is not representable as a transformation of multivariate normal random vector. Although these studies primarily focus on the NORTA procedure, their discussion can be readily 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. Fortunately, using the Johnson family of distributions tends to mitigate this problem because they provide a relatively comprehensive set of achievable correlations. Given that the base correlation matrix is not positive definite, Ghosh and Henderson (2000) suggest the application of semi-definite programming on the nonpositive definite base matrix, which is completed to be positive definite. Using this idea, we incorporate a modification step to our data generation routine, for which we present a more detailed discussion in our forthcoming paper.

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.

Proposition 3.7 *If $\mathbf{Z}_t$ is strictly stationary, then $\mathbf{X}_t$ is strictly stationary*[1].

4 Implementation

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

This problem was previously studied by Cario and Nelson (1998), Chen (2000), and Cario, Nelson, Roberts, and Wilson (2001). Since the only term in (7) 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 then simultaneously evaluating an initial grid

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

of values 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 $c_{ijh}\left[\hat{\rho}_{\mathbf{Z}}(i,j,h)\right] \approx \rho_{\mathbf{X}}(i,j,h)$, for $h = 0, 1, \cdots, p$. This approach makes particularly good sense in their case because all of their correlation matching problems share a common marginal distribution, so many of the grid points will be useful. Chen, and Cario, Nelson, Roberts, and Wilson evaluate (7) 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 $F_{X_i}^{-1}[\Phi(z_i)]F_{X_j}^{-1}[\Phi(z_j)]$ is not computationally expensive for us because the Johnson 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 correlation matching problems, specifically $pk^2 + k(k-1)/2$ such problems. Therefore, our approach is to provide a computationally efficient method based on the implementation of a numerical search procedure supported by a numerical integration technique, which we discuss in detail in the succeeding section. We thus take advantage of the superior accuracy of a numerical integration technique, without suffering substantial computation burden.

4.1 Numerical Integration Technique

This section briefly summarizes how we numerically evaluate $\mathrm{E}[X_{i,t}\, X_{j,t+h}]$ given the marginals, F_{X_i} and F_{X_j}, and the associated correlation, $\rho_{\mathbf{Z}}(i,j,h)$. Since we characterize the input process using the Johnson 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

$$f^{-1}(a) = \begin{cases} e^a & \text{for the } \mathrm{S}_L \text{ (lognormal) family,} \\ \frac{e^a - e^{-a}}{2} & \text{for the } \mathrm{S}_U \text{ (unbounded) family,} \\ \frac{1}{1+e^{-a}} & \text{for the } \mathrm{S}_B \text{ (bounded) family,} \\ a & \text{for the } \mathrm{S}_N \text{ (normal) family.} \end{cases}$$

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

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(\xi_1 + \lambda_1 f_1^{-1}[(z_1-\gamma_1)/\delta_1]\Big)\Big(\xi_2 + \lambda_2 f_2^{-1}[(z_2-\gamma_2)/\delta_2]\Big)\frac{e^{-(z_1^2-2\rho z_1 z_2+z_2^2)/2(1-\rho^2)}}{2\pi\sqrt{1-\rho^2}}dz_1dz_2 \qquad (8)$$

The expansion of the formula (8), 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

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(z_1, z_2)g(z_1, z_2, \rho)dz_1dz_2$$

where $w(z_1, z_2) = e^{-(z_1^2+z_2^2)}$, but the definition of $g(\cdot)$ changes from one subproblem to another. Note that the integral (8) exists only if $|\rho| < 1$, but we can solve the problem for $|\rho| = 1$ using the representation that is presented in the proof of Proposition 3.3 (see Appendix B).

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 de Doncher 1981) and "Cubpack" (Cools, Laurie, and Pluym 1997), which were specifically designed for solving cubature problems. While preparing the numerical integration routine for our software, we primarily benefit from the work accomplished in the latter reference.

As suggested by the numerical integration literature (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 $\psi(z_i^*) = \tan(\pi z_i^*/2)$, $-1 < z_i^* < 1$, $i = 1, 2$. Because $d\psi(z_i^*)/dz_i^* =$

$\pi/2(1+\tan^2(\pi z_i^*/2))$, the integral (8) is transformed into one of the following forms:

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

$$\int_{-1}^{1}\int_{-1}^{1} \frac{\Big(\xi_1+\lambda_1 f_1^{-1}[(\tan(\pi z_1^*/2)-\gamma_1)/\delta_1]\Big)\Big(\xi_2+\lambda_2 f_2^{-1}[(\tan(\pi z_1^*/2)-\gamma_2)/\delta_2]\Big)}{\sqrt{\pi}^{-1}2\sqrt{2}e^{\tan^2(\pi z_1^*/2)/2}\Big(1+\tan^2(\pi z_1^*/2)\Big)^{-1}} dz_1^* dz_2^*,\ \rho=1 \quad (9)$$

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

Although the $\rho=\pm1$ cases could be expressed as a single integral, we express them as double integrals to take advantage of the accurate and reliable error estimation strategy that we developed specifically for the evaluation of (8).

As a check on consistency and efficiency of the transformation, $\psi(z_i^*)=\tan(\pi z_i^*/2)$, we compared its performance with other doubly infinite hypercube transformations including $\psi(z_i^*)=z_i^*/(1-|z_i^*|)$, $d\psi(z_i^*)/dz_i^*=1/(1-|z_i^*|)^2$, as suggested by Genz (1992). While $d\psi(z_i^*)/dz_i^*$ is generally singular at the points z_i^* for which $\psi(z_i^*)=\pm\infty$ and this entails singularities of the transformed integrand, we do not need to deal with this problem in the case of the transformation $\psi(z_i^*)=\tan(\pi z_i^*/2)$, $-1<z_i^*<1$. Further, the tan transformations leads to relatively smooth shapes to be integrated.

Since the integration regions of the formulae (9) are the squares defined over $[-1,1]^2$, we can use a variety of cubature formulas developed for unit squares and accommodate any rectangular regions using the standard affine transformations (scaling and translation). Therefore, our numerical integration routine requires the central data structure to be a collection of rectangles. Further, this allows us to take full advantage of polymorphism of `C++` when we incorporate this routine in our software. Figure 1 provides a high-level view of how our algorithm works. In the figure, we use ℓ to denote the integrand of interest on which the cubature formula $\mathrm{C}(|\cdot|;B)$, together with the error estimation strategy $\mathrm{E}(|\cdot|;B)$, are applied over the region B. Further, $\mathrm{I}(\ell;B)$ corresponds to the desired result.

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

$$\max\Big(\epsilon_{abs}, \epsilon_{rel} \times \mathrm{C}(|\cdot|;B)\Big)$$

where ϵ_{abs} is the requested absolute error and ϵ_{rel} is the requested relative error. This definition uses the relative L_1-test for convergence that Krommer and Ueberhuber (1994) define as $\epsilon_{abs} = 0$ and $|\mathrm{E}(\cdot;B)|/\mathrm{C}(|\cdot|;B) < \epsilon_{rel}$. By using $\mathrm{C}(|\cdot|;B)$ instead of $\mathrm{C}(\cdot;B)$, we avoid heavy cancellation that might occur during the calculation of the approximate value $\mathrm{C}(\cdot;B) \approx 0$, although the function values in our integration problems are not generally small. For a full motivation behind this convergence test, we refer the reader to Krommer and Ueberhuber (1994). The additional calculation of $\mathrm{C}(|\cdot|;B)$ causes only a minor increase in the overall computational effort as no additional function evaluations are needed.

After we select the rectangle 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 improves the total error estimate, the selected rectangle is removed from the collection, its descendants (one or more) are added to it, and the total approximate integral and error estimates are updated. Otherwise, the selected rectangle is considered to be hopeless, which means that the current error estimate for that region cannot be reduced further. When it is certain that any decrease in the error of approximation is not possible, we stop the integrator and report failure.

Due to the importance of an error estimation strategy in determining the performance of a numerical integration routine, we briefly review the concept of null rules and how we use them. 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 the abscissas and weights of the null rule are notated with x_i and u_i, respectively. A null rule $\mathrm{N}_d(\ell)$ is furthermore said to have degree d if it maps to zero all polynomials of degree not more than d, but not all polynomials of degree $d+1$. When two null rules $\mathrm{N}_{d,1}(\ell)$ and $\mathrm{N}_{d,2}(\ell)$ of the same degree exist, the

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: $B_{n,r}$, $r = 1, \ldots, R$ ($R = 2$ or $R = 4$);
  calculate error estimates in each subregion: $e_{n,r} = E(\ell; B_{n,r})$, $r = 1, 2, \ldots, R$;
  determine whether there is an improvement in the total error estimate:
  if ($e_s < e_{n,1} + \ldots + e_{n,R}$) then
    delete the newly created subregions, $B_{n,r}$, $r = 1, 2, \ldots, R$;
  else
  {
    calculate approximate integral values in each subregion: $c_{n,r} = C(\ell; B_{n,r})$, $r = 1, 2, \ldots, R$;
    insert $(B_{n,1}, c_{n,1}, e_{n,1}), \ldots, (B_{n,R}, c_{n,R}, e_{n,R})$ into the data structure;
    $c := c - c_s + \sum_{r=1}^{R} c_{n,r}$;
    $e := e - e_s + \sum_{r=1}^{R} e_{n,r}$;
  }
  remove (B_s, c_s, e_s) from the data structure;
}
end do
return the total approximate integral value, c, with its error estimate e;

Figure 1: Meta algorithm for the numerical integration routine.

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 that enforce the weights, u_i, to have the same norm as the coefficients of the cubature rule, η_i; i.e., $\sum_{i=1}^n u_i^2 = \sum_{i=1}^n \eta_i^2$. We can explain our motivation behind using them as follows: If the integrand ℓ produces random numbers of mean zero, then the expected value of $\mathrm{N}_d(\ell)$ is zero and its variance does not depend on d. Further, for equally strong null rules $\mathrm{N}_d(\ell)$, the true error $\mathrm{C}(\ell;B) - \mathrm{I}(\ell;B)$ and $\mathrm{N}_d(\ell)$ have the same mean and standard deviation (Krommer and Ueberhuber 1994; page 171). In this way, we hope to ensure a satisfactory reliability and accuracy for error estimation and also avoid extensive experiments by utilizing the same integrand evaluations in the null rule that we need for approximating the integral.

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 rule approximations. The approach is to heuristically distinguish the behavior of the sequence $\{\mathrm{N}_d(\ell),\ d = 0,\ldots,n-2\}$ among three possible types of behavior, which are non-asymptotic, weakly asymptotic, and strongly asymptotic. Following Cools, Laurie, and Pluym (1997), we use seven independent fully symmetric null rules of degrees $(7,7)$, $(5,5)$, $(3,3)$, and (1) to obtain $\mathrm{N}_1(\ell)$, $\mathrm{N}_3(\ell)$, $\mathrm{N}_5(\ell)$, and $\mathrm{N}_7(\ell)$, which are used to conduct a test for observable asymptotic behavior. In the numerical integration literature, it is a well-known result that as the integration interval, τ, in which the integrand ℓ is sufficiently smooth, tends to 0, the error of a null rule of degree d is roughly proportional to br^d for certain unknown constants r $(= O(\tau^4))$ and b. Thus, $\mathrm{N}_{d+2}(\ell)/\mathrm{N}_d(\ell) \approx r^2$ for $d = 1,3,5$ (Berntsen and Espelid 1991, Laurie 1994, Cools, Laurie, and Pluym 1997). This relation, termed "strong asymptotic behavior", leads to an optimistic error estimate based both on the knowledge of the null rules and the basic rule's degree of precision: $|\mathrm{C}(\ell) - \mathrm{I}(\ell)| \approx ar^{q-s}\mathrm{N}_s(\ell)$, where $a > 1$ is a safety factor, s is the highest value among the possible degrees attained by a null rule, q is the degree of the corresponding cubature formula, and r is taken to be the maximum of the quantities $\sqrt{\mathrm{N}_7(\ell)/\mathrm{N}_5(\ell)}$, $\sqrt{\mathrm{N}_5(\ell)/\mathrm{N}_3(\ell)}$, and $\sqrt{\mathrm{N}_3(\ell)/\mathrm{N}_1(\ell)}$. The test for strong asymptotic behavior requires that r is less than a certain critical value, r_{crit}. If $r > 1$, there is assumed to be no asymptotic behavior at all and the error estimate is $K\mathrm{N}_s(\ell)$, where K is another safety factor. The test $r_{crit} \le r \le 1$ denotes the weak test on asymptotic behavior and when this test is passed, we use the error estimate $Kr^2\mathrm{N}_s(\ell)$ where

the safety factors of the nonasymptotic behavior and strong asymptotic behavior are related by $a = Kr_{crit}^{s-q+2}$. In order to attain optimal (or nearly optimal) computational efficiency, the free parameters, r_{crit} and K, are 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, Laurie, and Pluym (1997).

A more detailed presentation of the implementation is the subject of our forthcoming paper.

4.2 Numerical Search Procedure

The numerical integration scheme allow 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 (or alternatively called regula falsi), which is basically the modified version of Newton's method. We use φ 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, $\rho_{\mathbf{X}}$, i.e., $\varphi(\rho_{\mathbf{Z}}) = c_{ijh}[\rho_{\mathbf{Z}}] - \rho_{\mathbf{X}}$. 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 correlation matching problem to finding zeroes of the function φ.

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

$$\rho_{\mathbf{Z},m+1} = \rho_{\mathbf{Z},m} - \varphi(\rho_{\mathbf{Z},m})\Big(\frac{\rho_{\mathbf{Z},m} - \rho_{\mathbf{Z},m-1}}{\varphi(\rho_{\mathbf{Z},m}) - \varphi(\rho_{\mathbf{Z},m-1})}\Big) \tag{10}$$

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 (10) amounts to approximating the curve $y_m = \varphi(\rho_{\mathbf{Z},m})$ by the secant (or chord) joining the points $(\rho_{\mathbf{Z},m}, \varphi(\rho_{\mathbf{Z},m}))$ and $(\rho_{\mathbf{Z},m-1}, \varphi(\rho_{\mathbf{Z},m-1}))$. Since the problem of interest is to find $\hat{\rho}_{\mathbf{Z}} = \varphi^{-1}(0)$, we can regard (10) as a linear interpolation formula for φ^{-1}, i.e., we wish to find the unknown value $\varphi^{-1}(0)$ by interpolating the known values $\varphi^{-1}(y_m)$ and $\varphi^{-1}(y_{m-1})$. Further, the definition of function φ modifies the method in a way that ensures convergence for any continuous function. The secant method requires that we choose two starting points, $\rho_{\mathbf{Z},0}$ and $\rho_{\mathbf{Z},1}$, sufficiently close to $\hat{\rho}_{\mathbf{Z}}$. 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 $\varphi(\rho_{\mathbf{Z},0})$ and $\varphi(\rho_{\mathbf{Z},1})$ have always opposite signs. Then there exists a $\hat{\rho}_{\mathbf{Z}}$ between $\rho_{\mathbf{Z},0}$ and $\rho_{\mathbf{Z},1}$, which satisfies $c_{ijh}(\hat{\rho}_{\mathbf{Z}}) - \rho_{\mathbf{X}} = 0$.

Since the corresponding function is strictly increasing (Wilson 2001) and quite smooth in the case of Johnson system, the application of the Secant method as a search procedure gives accurate and reliable results in a small amount of time, reducing the amount of effort required to solve large number of correlation matching problems.

5 Example

In this section we present an example that gives an explicit illustration of the multivariate time series generation with given marginals and correlation structure. Suppose that we require a trivariate ($k = 3$) random variable with Johnson-type marginals, which are lognormal ($\gamma_1 = -1.92720, \delta_1 = 2.46941, \lambda_1 = 1, \xi_1 = -2.36889$), unbounded ($\gamma_2 = 0, \delta_2 = 1.87474, \lambda_2 = 1.87474, \xi_2 = 0$), and bounded ($\gamma_3 = 0.47801, \delta_3 = 0.37459, \lambda_3 = 4.12314, \xi_3 = -0.85883$). The probability density function for each Johnson-type distribution is given in Figure 2. Further, the correlation matrices are specified at lags 0 and 1 (i.e., $p = 1$) as

$$\mathbf{\Sigma}_X(0) = \begin{pmatrix} 1.00000 & 0.05529 & 0.24462 \\ 0.05529 & 1.00000 & 0.66543 \\ 0.24462 & 0.66543 & 1.00000 \end{pmatrix} \text{ and } \mathbf{\Sigma}_X(1) = \begin{pmatrix} -0.93655 & -0.33202 & -0.06112 \\ -0.33202 & 0.78667 & 0.30587 \\ -0.66543 & 0.36718 & 0.06646 \end{pmatrix}.$$

Solving a total of 12 correlation matching problems including six different Johnson marginal pairs ($\mathrm{S}_L\mathrm{S}_L$, $\mathrm{S}_L\mathrm{S}_U$, $\mathrm{S}_L\mathrm{S}_B$, $\mathrm{S}_U\mathrm{S}_U$, $\mathrm{S}_U\mathrm{S}_B$, $\mathrm{S}_B\mathrm{S}_B$), we determine the underlying correlation structure as

$$\mathbf{\Sigma}_Z(0) = \begin{pmatrix} 1.00000 & 0.05000 & 0.40000 \\ 0.05000 & 1.00000 & 0.60000 \\ 0.40000 & 0.60000 & 1.00000 \end{pmatrix} \text{ and } \mathbf{\Sigma}_Z(1) = \begin{pmatrix} -0.70000 & -0.30000 & -0.10000 \\ -0.30000 & 0.80000 & 0.50000 \\ -0.60000 & 0.60000 & 0.05000 \end{pmatrix}.$$

Next, we solve the multivariate Yule-Walker equations for the system parameters of the under-

lying base process, $\mathrm{VAR}_3(1)$:

$$\alpha_1 = \begin{pmatrix} -0.93320 & -0.65204 & 0.66451 \\ -0.44865 & 0.64806 & 0.29063 \\ -0.57527 & 0.71984 & -0.15179 \end{pmatrix} \text{ and } \mathbf{\Sigma}_u = \begin{pmatrix} 0.21759 & -0.04058 & 0.19808 \\ -0.04058 & 0.20165 & -0.07256 \\ 0.19808 & -0.07256 & 0.23052 \end{pmatrix}.$$

The application of the corresponding tests reveals that $\mathbf{Z}_t$ is stationary with a positive definite correlation matrix $\mathbf{\Sigma_Z}$. Then, we simulate the underlying vector autoregressive process and transform the standard normal random variates $z_{i,t}$ into $x_{i,t}$ using the transformation $\xi_i + \lambda_i f_i^{-1}[(z_{i,t} - \gamma_i)/\delta_i]$ for $i = 1, \ldots, k$ and $t = 0, 1, \ldots$. The scatterplots of $(X_{i,t}, X_{j,t+h})$ for $i = 1, 2, 3$ and $h = 0, 1$ are given in Figure 3, and time-series plots in Figure 4

We have developed a stand-alone, PC-based program that implements the VARTA framework with the suggested search procedure and numerical integration techniques for simulating input processes. The key computational components of the software are written in portable C code in such a way that we can make them available individually for incorporation into commercial products. The detailed presentation of the software is the subject of our forthcoming paper.

6 Conclusion and Future Research

In this paper, 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, which exploits the features of Johnson system, specifically the evaluation of the composite function $F_X^{-1}[\Phi(.)]$ would be slow and memory intensive in the case of standard family of distributions. However, our framework requires the full characterization of the Johnson system through parameters $[\gamma, \delta, \lambda, \xi]$ and function $f(\cdot)$ corresponding to the Johnson family of interest. As an area for future research, it would be quite useful to be able to characterize the underlying system to which a given historical data set belongs. This requires the determination of the type of Johnson family to use together with the parameters of the corresponding distribution in a way that the dependence structure in the multivariate input data is captured. These issues are the subject of our ongoing research.

References

[1] Melamed, B., J. R. Hill, and D. Goldsman (1992). The TES methodology: Modeling empirical stationary time series. In *Proceedings of the 1992 Winter Simulation Conference*. Edited by J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, pp. 135–144. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

[2] Ware, P. P., T. W. Page, and B. L. Nelson (1998). Automatic modeling of file system workloads using two-level arrival processes. **ACM Transactions on Modeling and Computer Simulation**, **Vol. 8**, pp. 305–330.

[3] Law, A. M. and W. D. Kelton (2000). *Simulation Modeling and Analysis*. Third edition, McGraw-Hill, Inc., New York.

[4] Nelson, B. L. and M. Yamnitsky (1998). Input modeling tools for complex problems. In *Proceedings of the 1998 Winter Simulation Conference*. Edited by D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

[5] Lewis, P. A. W., E. McKenzie, and D. K. Hugus (1989). Gamma processes. **Communications in Statistics-Stochastic Models**, **5**, pp. 1–30.

[6] Melamed, B. (1991). TES: A class of methods for generating autocorrelated uniform variates. **ORSA Journal on Computing**, **3**, pp. 317–329.

[7] Willemain, T. R. and P. A. Desautels (1993). A method to generate autocorrelated uniform random numbers. **Journal of Statistical Computation and Simulation**, **45**, pp. 23–31.

[8] Song, W. T., L. Hsiao, and Y. Chen (1996). Generating pseudorandom time series with specified marginal distributions. **European Journal of Operational Research**, **93**, pp. 1–12.

[9] Cario, M. C. and B. L. Nelson (1996). Autoregressive to anything: Time series input processes for simulation. **Operations Research Letters**, **19**, pp. 51–58.

[10] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

[11] Johnson, M. E. (1987). it Multivariate Statistical Simulation. New York: John-Wiley.

[12] Hill, R. R. and C. H. Reilly (1994). Composition for multivariate random vectors. In *Proceedings of the 1994 Winter Simulation Conference*. Edited by J.D. Tew, S. Manivannan, D. A. Sadowsky, and A. F. Seila, pp. 332–339. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

[13] Cook, R. D. and M. E. Johnson (1981). A family of distributions for modeling non-elliptically symmetric multivariate data. **Journal of the Royal Statistical Society B**, **43**, pp. 210–218.

[14] Ghosh, S. and S. G. Henderson (2000). Chessboard distributions and random vectors with specified marginals and covariance matrix. Working paper, Department of Industrial and Operations Engineering , University of Michigan, Ann Arbor.

[15] Mardia, K. V. (1970). A translation family of bivariate distributions and Frèchet's bounds. **Sankhya A**, **32**, pp. 119–122.

[16] Li, S. T. and J. L. Hammond (1975). Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. **IEEE Transactions on Systems, Man, and Cybernetics**, **5**, pp. 557–561.

[17] Cario, M. C., B. L. Nelson, S. D. Roberts, and J. R. Wilson (2001). Numerical methods for fitting and simulating autoregressive-to-anything processes. **INFORMS Journal on Computing**, **10**, pp. 72–81.

[18] Chen, H. (2000). Initialization for NORTA: Generation of random vectors with specified marginals and correlations. Working Paper, Department of Industrial Engineering, Chung Yuan Christian University, Taiwan.

[19] Lurie, P. M. and M. S. Goldberg (1998). An approximate method for sampling correlated random variables from partially-specified distributions. **Management Science**, **44**, pp. 203–218.

[20] Clemen, R. T. and T. Reilly (1999). Correlations and copulas for decision and risk analysis. **Management Science**, **45**, pp. 208–224.

[21] Lutkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York.

[22] Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. **Biometrika**, **36**, pp. 297–304.

[23] Nelsen, R. B. (1998). *An Introduction to Copulas*. Lecture Notes in Statistics, Springer-Verlag Publishing Company, New York.

[24] Cambanis, S. and E. Marsy (1978). On the reconstruction of the covariance of stationary Gaussian processes observed through zero-memory nonlinearities. **IEEE Transactions on Information Theory**, **24**, pp. 485–494.

[25] Robinson, I. and E. De Doncker (1981). Algorithm 45: Automatic computation of improper integrals over a bounded or unbounded planar region. **Computing**, **27**, pp. 253-284.

[26] Cools, R., D. Laurie, and L. Pluym (1997). Algorithm 764: Cubpack++: A C++ package for automatic two-dimensional cubature. **ACM Transactions on Mathematical Software**, **23**, pp. 1-15.

[27] Krommer, A. R. and C. W. Ueberhuber (1994). *Numerical Integration on Advanced Computer Systems.* Lecture Notes in Computer Science, Springer-Verlag, New York.

[28] Genz, A. (1992). Statistics applications of subregion adaptive multiple numerical integration. In *Numerical Integration - Recent Developments, Software, and Applications*. Edited by T. O. Espelid and A. Genz. Kluwer Academic Publishers, Dordrecht, pp. 267–280.

[29] Cools, R. (1994). The subdivision strategy and reliability in adaptive integration revisited. Rep. TW 213, Dept. of Computer Science, Katholieke Univ. Leuven, Leuven, Belgium.

[30] Rabinowitz, P. and N. Richter (1969). Perfectly symmetric two-dimensional integration formulas with minimal number of points. **Mathematical Computations**, **23**, pp. 765–799.

[31] Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals*. Prentice-Hall, Englewood Cliffs, New Jersey.

[32] Berntsen, J. and T. O. Espelid (1991). Error estimation in automatic quadrature routines. **ACM Transactions on Mathematical Software**, **17**, pp. 233–252.

[33] Laurie, D. P. (1994). Null rules and orthogonal expansions. In *Approximation and Computation: A Festschrift in Honor of Walter Gautschi*. Edited by R. V. M. Zahar. International Series of Numerical Mathematics, vol. 119. Birkhäuser, pp. 359–370.

[34] Blum, E. K. (1972). *Numerical Analysis and Computation Theory and Practice*. Addison-Wesley Publishing Company, Massachusetts.

[35] Wilson, J. R. (2001). Personal Communication.

[36] Lehmann, E. L. (1998). *Elements of Large Sample Theory*. Springer Texts in Statistics. Springer Verlag, New York.

[37] Rencher, A. C. (1998). *Multivariate Statistical Inference and Applications*. John Wiley and Sons, New York.

[38] Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley and Sons, New York.

[39] Tong, Y. L. (1990). *The Multivariate Normal Distribution*. Springer Series in Statistics, Springer-Verlag, New York.

[40] Whitt, W. (1976). Bivariate distributions with given marginals. **The Annals of Statistics**, **4**, pp. 1280-1289.

Appendix A: On the Distribution of VARTA

Our approach is to show the result for the first-order autoregressive process and then generalize it to the p^{th}-order autoregressive process by using the fact that a $\text{VAR}_{kp}(1)$ model can provide a state-space representation of the $\text{VAR}_k(\text{p})$ model. We refer the interested reader to Lutkepohl (1993).

Lemma 6.1 *Let* $\mathbf{Z}_t$ *denote a stationary first-order vector autoregressive process,* $VAR_k(1)$*, where* $\mathbf{Z}_t = (Z_{1,t}, Z_{2,t}, \cdots, Z_{k,t})'$ *is a* $(k \times 1)$ *random vector and* $\boldsymbol{\alpha}_1$ *is a fixed* $(k \times k)$ *autoregressive coefficient matrix. Let* $\mathbf{u}_t = (u_{1,t}, u_{2,t}, \cdots, u_{k,t})'$ *be* k*-dimensional Gaussian white noise; that is* $\mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u)$ *for all* t*, and* $\mathbf{u}_t$ *and* $\mathbf{u}_s$ *are independent for* $s \neq t$*. Then* $\mathbf{Z}_t$ *is a Gaussian process.*

Proof. The $\text{VAR}_k(1)$ is given by

$$\mathbf{Z}_t = \boldsymbol{\alpha}_1 \mathbf{Z}_{t-1} + \mathbf{u}_t. \tag{11}$$

Thus starting at $t = 1$, we get

$$\begin{aligned}
\mathbf{Z}_1 &= \boldsymbol{\alpha}_1 \mathbf{Z}_0 + \mathbf{u}_1 \\
\mathbf{Z}_2 &= \boldsymbol{\alpha}_1 \mathbf{Z}_1 + \mathbf{u}_2 = \boldsymbol{\alpha}_1(\boldsymbol{\alpha}_1 \mathbf{Z}_0 + \mathbf{u}_1) + \mathbf{u}_2 = \boldsymbol{\alpha}_1^2 \mathbf{Z}_0 + \boldsymbol{\alpha}_1 \mathbf{u}_1 + \mathbf{u}_2 \\
&= \boldsymbol{\alpha}_1^2 \mathbf{Z}_0 + \boldsymbol{\alpha}_1 \mathbf{u}_1 + \mathbf{u}_2 \\
&\vdots \\
\mathbf{Z}_t &= \boldsymbol{\alpha}_1^t \mathbf{Z}_0 + \sum_{i=0}^{t-1} \boldsymbol{\alpha}_1^i \mathbf{u}_{t-i}.
\end{aligned} \tag{12}$$

Hence, the vectors $\mathbf{Z}_1, \ldots, \mathbf{Z}_t$ are uniquely determined by $\mathbf{Z}_0, \mathbf{u}_1, \ldots, \mathbf{u}_t$. Also, the joint distribution of $\mathbf{Z}_1, \ldots, \mathbf{Z}_t$ is determined by the joint distribution of $\mathbf{Z}_0, \mathbf{u}_1, \ldots, \mathbf{u}_t$.

Since the stationary vector autoregressive models are of interest in this study, it is convenient to assume that the process has been started in the infinite past. From (12), we have

$$\begin{aligned}
\mathbf{Z}_t &= \boldsymbol{\alpha}_1 \mathbf{Z}_{t-1} + \mathbf{u}_t \\
&= \boldsymbol{\alpha}_1^{j+1} \mathbf{Z}_{t-j-1} + \sum_{i=0}^{j} \boldsymbol{\alpha}_1^i \mathbf{u}_{t-i}, \quad \text{for any } \; j \geq 1.
\end{aligned}$$

Since the process (11) is stable, its reverse characteristic polynomial has no roots in or on the complex unit circle. By Rule 7 of Lutkepohl (1993), Appendix A.6, this is equivalent to the condition that all eigenvalues of $\boldsymbol{\alpha}_1$ have modulus less than 1. This makes the sequence $\{\boldsymbol{\alpha}_1^i, \;\; i = 0, 1, 2, \ldots\}$ absolutely summable (Lutkepohl 1993; Appendix A, Section A.9.1). Hence, the infinite sum $\sum_{i=1}^{\infty} \boldsymbol{\alpha}_1^i \mathbf{u}_{t-i}$ exists in mean square (Lutkepohl 1993; Appendix C, Proposition C.7). Furthermore, since $\boldsymbol{\alpha}_1^{j+1}$ converges to zero rapidly as $j \to \infty$, the term $\boldsymbol{\alpha}_1^{j+1} \mathbf{Z}_{t-j-1}$ vanishes in the limit (Slutsky's theorem, Lehmann 1998). Hence, given that all eigenvalues of $\boldsymbol{\alpha}_1$ have modulus less than 1, stating that $\mathbf{Z}_t$ is the $\text{VAR}_k(1)$ process (11) is equivalent in probability to $\mathbf{Z}_t = \sum_{i=0}^{\infty} \boldsymbol{\alpha}_1^i \mathbf{u}_{t-i}, \;\; t = 0, 1, 2, \ldots$. Since we assume that $\mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u)$ for all t and $\mathbf{u}_t$ and $\mathbf{u}_s$ are independent for $s \neq t$, the random variable $\sum_{i=0}^{\infty} \alpha_1^i \mathbf{u}_{t-i}$ has distribution $N(\mathbf{0}, \varsigma_{\mathbf{Z}})$, where $\varsigma_{\mathbf{Z}}$ is given by $\sum_{i=0}^{\infty} \alpha_1^i \boldsymbol{\Sigma}_u (\alpha_1^i)'$. Lutkepohl (1993) assures that the infinite sum $\sum_{i=0}^{\infty} \alpha_1^i \boldsymbol{\Sigma}_u (\alpha_1^i)'$ exists via Proposition C.8 on the moments

of infinite sums of random vectors. Thus, the distribution of $\mathbf{Z}_t$ is given by $N(\mathbf{0}, \varsigma_{\mathbf{Z}})$, from which follows that $\mathbf{Z}_t$ is a Gaussian process.

Lemma 6.2 *Let $\mathbf{Z}_t$ and $\mathbf{Z}_{t+h}$ be from a stationary first-order vector autoregressive process, as defined in Lemma 6.1. Then the random variable $\mathbf{Z} = (\mathbf{Z}_t, \mathbf{Z}_{t+h})'$ has a $2k$-dimensional normal distribution with density function*

$$f(\mathbf{z}; \boldsymbol{\Sigma}_{2k}) = \frac{1}{2\pi|\boldsymbol{\Sigma}_{2k}|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{z}'\boldsymbol{\Sigma}_{2k}^{-1}\mathbf{z}}, \;\; \mathbf{z} \in \Re^{2k}, \boldsymbol{\Sigma}_{2k} = \begin{pmatrix} \boldsymbol{\Sigma}_Z(0) & \boldsymbol{\Sigma}_Z(h) \\ \boldsymbol{\Sigma}_Z(h)' & \boldsymbol{\Sigma}_Z(0) \end{pmatrix} \tag{13}$$

Further, it is a nonsingular distribution if $|\boldsymbol{\Sigma}_Z(0)| > 0$ and $\left|\boldsymbol{\Sigma}_Z(0) - \boldsymbol{\Sigma}_Z(h)\boldsymbol{\Sigma}_Z^{-1}(0)\boldsymbol{\Sigma}_Z(h)'\right| > 0$.

Proof. From Lemma 6.1, $\mathbf{Z}_t \sim N(\mathbf{0}, \varsigma_{\mathbf{Z}})$ and $\mathbf{Z}_{t+h} \sim N(\mathbf{0}, \varsigma_{\mathbf{Z}})$. Next, we will find the conditional distribution of $\mathbf{Z}_{t+h}$ given $\mathbf{Z}_t$: If $\mathbf{Z}_t = \mathbf{z}_t$, then it follows from (12) that $\mathbf{Z}_{t+h} = \boldsymbol{\alpha}_1^h \mathbf{z}_t + \sum_{i=0}^{h-1} \boldsymbol{\alpha}_1^i \mathbf{u}_{t+h-i}$. From the discussion in Lemma 6.1, $\boldsymbol{\alpha}_1^i \mathbf{u}_{t+h-i} \sim N(\mathbf{0}, \boldsymbol{\alpha}_1^i \boldsymbol{\Sigma}_u (\boldsymbol{\alpha}_1^i)')$ and $\sum_{i=0}^{h-1} \boldsymbol{\alpha}_1^i \mathbf{u}_{t+h-i} \sim N(\mathbf{0}, \sum_{i=0}^{h-1} \boldsymbol{\alpha}_1^i \boldsymbol{\Sigma}_u (\boldsymbol{\alpha}_1^i)')$, with the necessary condition that $\boldsymbol{\alpha}_1$ is a $(k \times k)$ real matrix such that $|\boldsymbol{\alpha}_1| \neq 0$. Then, $\mathbf{Z}_{t+h}$ conditioned on $\mathbf{Z}_t = \mathbf{z}_t$ has a $N(\boldsymbol{\alpha}_1^h \mathbf{z}_t, \varsigma_{\mathbf{Z},h})$ distribution, where $\varsigma_{\mathbf{Z},h} = \sum_{i=0}^{h-1} \boldsymbol{\alpha}_1^i \boldsymbol{\Sigma}_u (\boldsymbol{\alpha}_1^i)'$. Thus, the probability density function of $\mathbf{Z}_t$ and the conditional distribution of $\mathbf{Z}_{t+h}$ can be written as

$$f(\mathbf{z}_t) = \frac{1}{(2\pi)^{\frac{k}{2}}|\varsigma_{\mathbf{Z}}|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{z}_t'\varsigma_{\mathbf{Z}}^{-1}\mathbf{z}_t}, \;\; \mathbf{z}_t \in \Re^k$$

and

$$f(\mathbf{z}_{t+h}|\mathbf{z}_t) = \frac{e^{-\frac{1}{2}(\mathbf{z}_{t+h} - \boldsymbol{\alpha}_1^h \mathbf{z}_t)'\varsigma_{\mathbf{Z},h}^{-1}(\mathbf{z}_{t+h} - \boldsymbol{\alpha}_1^h \mathbf{z}_t)}}{(2\pi)^{\frac{k}{2}}|\varsigma_{\mathbf{Z},h}|^{\frac{1}{2}}}, \; \mathbf{z}_{t+h}, \mathbf{z}_t \in \Re^k,$$

respectively. Following Theorem 6 of Rohatgi (1976),

$$\begin{aligned} f(\mathbf{z}_{t+h}, \mathbf{z}_t) &= f(\mathbf{z}_t) f(\mathbf{z}_{t+h}|\mathbf{z}_t) \\ &= \frac{e^{-\frac{1}{2}\mathbf{z}_t'\varsigma_{\mathbf{Z}}^{-1}\mathbf{z}_t}}{(2\pi)^{\frac{k}{2}}|\varsigma_{\mathbf{Z}}|^{\frac{1}{2}}} \frac{e^{-\frac{1}{2}(\mathbf{z}_{t+h} - \boldsymbol{\alpha}_1^h \mathbf{z}_t)'\varsigma_{\mathbf{Z},h}^{-1}(\mathbf{z}_{t+h} - \boldsymbol{\alpha}_1^h \mathbf{z}_t)}}{(2\pi)^{\frac{k}{2}}|\varsigma_{\mathbf{Z},h}|^{\frac{1}{2}}} \end{aligned} \tag{14}$$

The algebraic simplification of (14) results in the density function given by (13). Hence, the random variable $(\mathbf{Z}_t, \mathbf{Z}_{t+h})'$ has a $2k$-dimensional bivariate normal distribution.

Since nonsingularity is attained only if $\mathbf{\Sigma}_{2k}$ is positive definite, the conditions on nonsingularity can be written as $|\mathbf{\Sigma}_Z(0)| > 0$ and $|\mathbf{\Sigma}_Z(0) - \mathbf{\Sigma}_Z(h)\mathbf{\Sigma}_Z^{-1}(0)\mathbf{\Sigma}_Z(h)'| > 0$. ∎

Corollary 6.3 *Let $\mathbf{Z}_t$ denote a stationary first-order vector autoregressive process, $VAR_k(1)$, as defined in Lemma 6.1. The random variable $\mathbf{Z} = (Z_{i,t}, Z_{j,t+h})'$ has a bivariate normal distribution with density function*

$$f(\mathbf{z}; \mathbf{\Sigma}_2) = \frac{1}{2\pi|\mathbf{\Sigma}_2|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{z}'\mathbf{\Sigma}_2^{-1}\mathbf{z}}, \ \ \mathbf{z} \in \Re^2, \tag{15}$$

where

$$\mathbf{\Sigma}_2 = \begin{pmatrix} 1 & \rho_{\mathbf{Z}}(i,j,h) \\ \rho_{\mathbf{Z}}(i,j,h) & 1 \end{pmatrix}_{(2\times 2)}$$

. It is a nonsingular distribution if $|\rho_{\mathbf{Z}}(i,j,h)| < 1$.

Proof. From Lemma 6.2, the random variable $\mathbf{Z}_{2k} = (\mathbf{Z}_t, \mathbf{Z}_{t+h})'$ has a non-singular multivariate normal distribution with a density function given by (13). Further,

$$\mathbf{\Sigma}_Z(0) = \begin{pmatrix} 1 & \rho_{\mathbf{Z}}(1,2,0) & \dots & \rho_{\mathbf{Z}}(1,k,0) \\ \rho_{\mathbf{Z}}(1,2,0) & 1 & \dots & \rho_{\mathbf{Z}}(2,k,0) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\mathbf{Z}}(1,k,0) & \rho_{\mathbf{Z}}(2,k,0) & \dots & 1 \end{pmatrix}_{(k\times k)} \quad \text{and} \ \ |\mathbf{\Sigma}_Z(0)| > 0,$$

while

$$\mathbf{\Sigma}_Z(h) = \begin{pmatrix} \rho_{\mathbf{Z}}(1,1,h) & \rho_{\mathbf{Z}}(1,2,h) & \dots & \rho_{\mathbf{Z}}(1,k,h) \\ \rho_{\mathbf{Z}}(1,2,h) & \rho_{\mathbf{Z}}(2,2,h) & \dots & \rho_{\mathbf{Z}}(2,k,h) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\mathbf{Z}}(1,k,h) & \rho_{\mathbf{Z}}(2,k,h) & \dots & \rho_{\mathbf{Z}}(k,k,h) \end{pmatrix}_{(k\times k)} \quad \text{and} \ \ |\mathbf{\Sigma}_Z(0) - \mathbf{\Sigma}_Z(h)\mathbf{\Sigma}_Z^{-1}(0)\mathbf{\Sigma}_Z(h)'| > 0.$$

Note that the random variable $(\mathbf{Z}_t, \mathbf{Z}_{t+h})'$ can be equivalently represented as $(Z_{1,t}, Z_{2,t}, \cdots, Z_{k-1,t}, Z_{k,t}, Z_{1,t+h}, Z_{2,t+h}, \cdots, Z_{k-1,t+h}, Z_{k,t+h})'$. Hence, the random variable $(Z_{i,t}, Z_{j,t+h})'$ has a bivariate normal distribution with correlation $\rho_{\mathbf{Z}}(i,j,h)$ (Theorem 2.2.C, Rencher 1998). ∎

Lemma 6.4 *Let $\mathbf{Z}_t$ denote a stationary p^{th}-order vector autoregressive process, $VAR_k(p)$ as defined in (1). If $\mathbf{u}_t = (u_{1,t}, u_{2,t}, \cdots, u_{k,t})'$ is a k-dimensional Gaussian white noise, that is $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_u)$ for all t, and $\mathbf{u}_t$ and $\mathbf{u}_s$ are independent for $s \neq t$, then $\mathbf{Z}_t$ is a Gaussian process.*

Proof. Since $\mathbf{Z}_t$ is a $\text{VAR}_k(p)$ defined as in (1), its state-space representation as a $\text{VAR}_{kp}(1)$ is given by (3). Note that $\mathbf{\Sigma_U}$ in (4) is a positive semi-definite matrix, but its k^{th} principle minor, that characterizes the part of the state-space model we are interested in, is positive definite. From Lemma 6.1, it follows that $\bar{\mathbf{Z}}_t$ has a kp-dimensional (singular) multivariate normal distribution. Thus, $\mathbf{Z}_t$, which is obtained by $\bar{\mathbf{J}}\bar{\mathbf{Z}}_t$ using the $(k \times kp)$ matrix $\bar{J} = (\mathbf{I}_{(k\times k)}\ \mathbf{0} \cdots \mathbf{0})$, has a k-dimensional (nonsingular) multivariate normal distribution function. Hence, $\mathbf{Z}_t$ is a Gaussian process. ∎

Lemma 6.5 *Let $\mathbf{Z}_t$ and $\mathbf{Z}_{t+h}$ be from a stationary p^{th}-order vector autoregressive process as defined in Lemma 6.4. Then, the random variable $\mathbf{Z} = (\mathbf{Z}_t, \mathbf{Z}_{t+h})'$ has a $2k$-dimensional multivariate normal distribution with a density function given by (13).*

Proof. Since $\mathbf{Z}_t$ and $\mathbf{Z}_{t+h}$ are $\text{VAR}_k(p)$ models defined as in (1), their corresponding state-space models are given by (3). Note that, for both of the models, $\mathbf{\Sigma_U}$ in (4) is a positive semi-definite matrix with a positive-definite k^{th} principle minor, which characterizes the part of the state-space models we are interested in. From Lemma 6.2, the distribution of $(\bar{\mathbf{Z}}_t, \bar{\mathbf{Z}}_{t+h})'$ is represented by a $2kp$-dimensional (singular) multivariate normal distribution. Note that $\mathbf{Z}_t$ and $\mathbf{Z}_{t+h}$ are obtained as $\bar{\mathbf{J}}\bar{\mathbf{Z}}_t$ and $\bar{\mathbf{J}}\bar{\mathbf{Z}}_{t+h}$, respectively, using the $(2k \times 2k)$ matrix $\bar{J} = (\mathbf{I}_{(2k\times 2k)}\ \mathbf{0} \cdots \mathbf{0})$. Hence, the random variable $(\mathbf{Z}_t, \mathbf{Z}_{t+h})'$ has a $2k$-dimensional non-singular bivariate normal distribution that is nonsingular since the corresponding variance-covariance matrix is positive definite. ∎

Proof of Proposition 3.1. From Lemma 6.5, the random variable $\mathbf{Z}_{2k} = (\mathbf{Z}_t, \mathbf{Z}_{t+h})'$ has a $2k$-dimensional multivariate normal distribution with a density function given by (13). Note that the random variable $(\mathbf{Z}_t, \mathbf{Z}_{t+h})'$ can be equivalently represented as

$$(Z_{1,t}, Z_{2,t}, \cdots, Z_{k-1,t}, Z_{k,t}, Z_{1,t+h}, Z_{2,t+h}, \cdots, Z_{k-1,t+h}, Z_{k,t+h})'.$$

Hence, the random variable $(Z_{i,t}, Z_{j,t+h})'$ has a non-singular bivariate normal distribution with correlation $|\rho_{\mathbf{Z}}(i, j, h)| < 1$ (Theorem 2.2.C, Rencher 1998). ∎

Appendix B: On the VARTA Properties

Proof of Proposition 3.2. If $\rho_{\mathbf{Z}}(i,j,h) = 0$, then

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

because $\rho_{\mathbf{Z}}(i,j,h) = 0$ implies that $Z_{i,t}$ and $Z_{j,t+h}$ are independent. If $\rho_{\mathbf{Z}}(i,j,h) \geq 0$ (≤ 0), then, from the association property (Tong 1990), $\mathrm{Cov}[g_1(Z_{i,t}, Z_{j,t+h}), g_2(Z_{i,t}, Z_{j,t+h})] \geq 0$ (≤ 0) holds for all non-decreasing 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 $F_{X_{(\cdot)}}^{-1}[\Phi(\cdot)]$ is a non-decreasing 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 6.6 *Let* $g(Z_{i,t}, Z_{j,t+h}) \equiv F_{X_i}^{-1}\Big[\Phi[Z_{i,t}]\Big]F_{X_j}^{-1}\Big[\Phi[Z_{j,t+h}]\Big]$ *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, Nelson, Roberts, and Wilson (2000) 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

$$\text{if } \rho_{\mathbf{Z}}^*(i,j,h) \geq \rho_{\mathbf{Z}}(i,j,h), \text{ then } \mathrm{c}_{ijh}(\rho_{\mathbf{Z}}^*(i,j,h)) \geq \mathrm{c}_{ijh}(\rho_{\mathbf{Z}}(i,j,h)). \tag{16}$$

Following the definition of function c_{ijh}, we can write (16) equivalently as

$$\text{if } \rho^*_{\mathbf{Z}}(i,j,h) \geq \rho_{\mathbf{Z}}(i,j,h), \text{ then } \mathrm{E}_{\rho^*_{\mathbf{Z}}(i,j,h)}[X_{i,t}X_{j,t+h}] \geq \mathrm{E}_{\rho_{\mathbf{Z}}(i,j,h)}[X_{i,t}X_{j,t+h}]. \tag{17}$$

We let $\Phi_\rho[Z_{i,t}, Z_{j,t+h}]$ be the joint cumulative distribution function of random variables $Z_{i,t}$ and $Z_{j,t+h}$, which is actually the standard bivariate normal distribution with correlation $\rho_{\mathbf{Z}}(i,j,h)$. In other words, $(Z_{i,t}, Z_{j,t+h})' \sim N(\mathbf{0}_2, \mathbf{\Sigma}_2)$. From Slepian's inequality (Tong 1990), it follows that

$$\Phi_{\rho^*_{\mathbf{Z}}(i,j,h)}[Z_{i,t}, Z_{j,t+h}] \geq \Phi_{\rho_{\mathbf{Z}}(i,j,h)}[Z_{i,t}, Z_{j,t+h}]$$

for all $Z_{i,t}$ and $Z_{j,t+h}$ if $\rho^*_{\mathbf{Z}}(i,j,h) \geq \rho_{\mathbf{Z}}(i,j,h)$. By definition, $\Phi_{\rho^*_{\mathbf{Z}}(i,j,h)}[Z_{i,t}, Z_{j,t+h}]$ is more concordant than $\Phi_{\rho_{\mathbf{Z}}(i,j,h)}[Z_{i,t}, Z_{j,t+h}]$ (Tchen 1980). Thus, if E_ρ denotes the expected value under distribution Φ_ρ, then

$$E_{\rho^*_Z(i,j,h)}[g(Z_{i,t}, Z_{j,t+h})] \geq E_{\rho_{\mathbf{Z}}(i,j,h)}[g(Z_{i,t}, Z_{j,t+h})]$$

for all superadditive functions g. From Lemma 6.6, $F^{-1}_{X_i}\Big[\Phi[Z_{i,t}]\Big]F^{-1}_{X_j}\Big[\Phi[Z_{j,t+h}]\Big]$ is superadditive. Thus, we let $g(Z_{i,t}, Z_{j,t+h}) \equiv F^{-1}_{X_i}\Big[\Phi[Z_{i,t}]\Big]F^{-1}_{X_j}\Big[\Phi[Z_{j,t+h}]\Big]$ and the equality (17) follows. Hence, the proof is complete. ∎

Proof of Theorem 3.5 Theorem 3.5 follows immediately from Lemma 2 of Cario, Nelson, Roberts, and Wilson (2000) 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 $\rho = \rho_{\mathbf{Z}}(i,j,h)$. ∎

Proof of Proposition 3.6

Proof. If we take an arbitrary collection of random variables implied by the VARTA transformation, i.e., X_{i,t_i} for some $i \in S$ and $t_i \geq 0$, then we can express the associated cumulative distribution function as follows:

$$\Pr\left\{\bigcap_{i\in S} X_{i,t_i} \leq x_i\right\} = \Pr\left\{\bigcap_{i\in S} Z_{i,t_i} \leq \Phi^{-1}\left[F_{X_i}(x_i)\right]\right\}.$$

This is a well-defined joint cumulative distribution function provided that $\mathbf{\Sigma}_{\mathbf{Z}}$ is nonnegative definite. Therefore, $\mathbf{\Sigma}_{\mathbf{X}}$ must also be nonnegative definite. ∎

Proof of Proposition 3.7

Proof. By definition, a process is called strictly s tationary if for each h the joint distribution of $\mathbf{Z}_t, \mathbf{Z}_{t+1}, \cdots, \mathbf{Z}_{t+h}$ is identical to the joint distribution of $\mathbf{Z}_s, \mathbf{Z}_{s+1}, \cdots, \mathbf{Z}_{s+h}$; that is, the distribution does not depend on the particular time point t. Thus, the proof follows immediately from the definition of strict stationarity. ∎

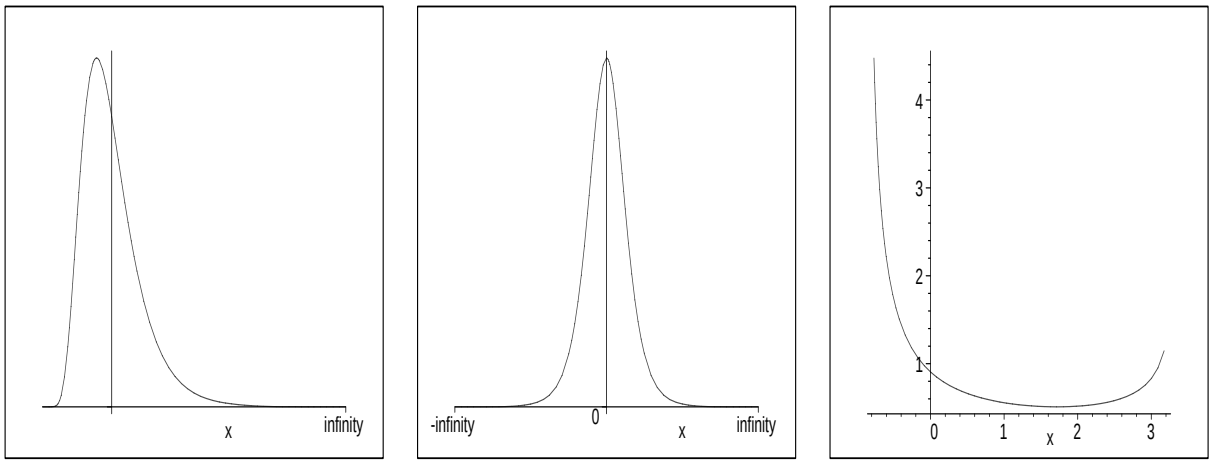

Figure 2: Probability density functions for the corresponding lognormal, unbounded, and bounded distributions, respectively.

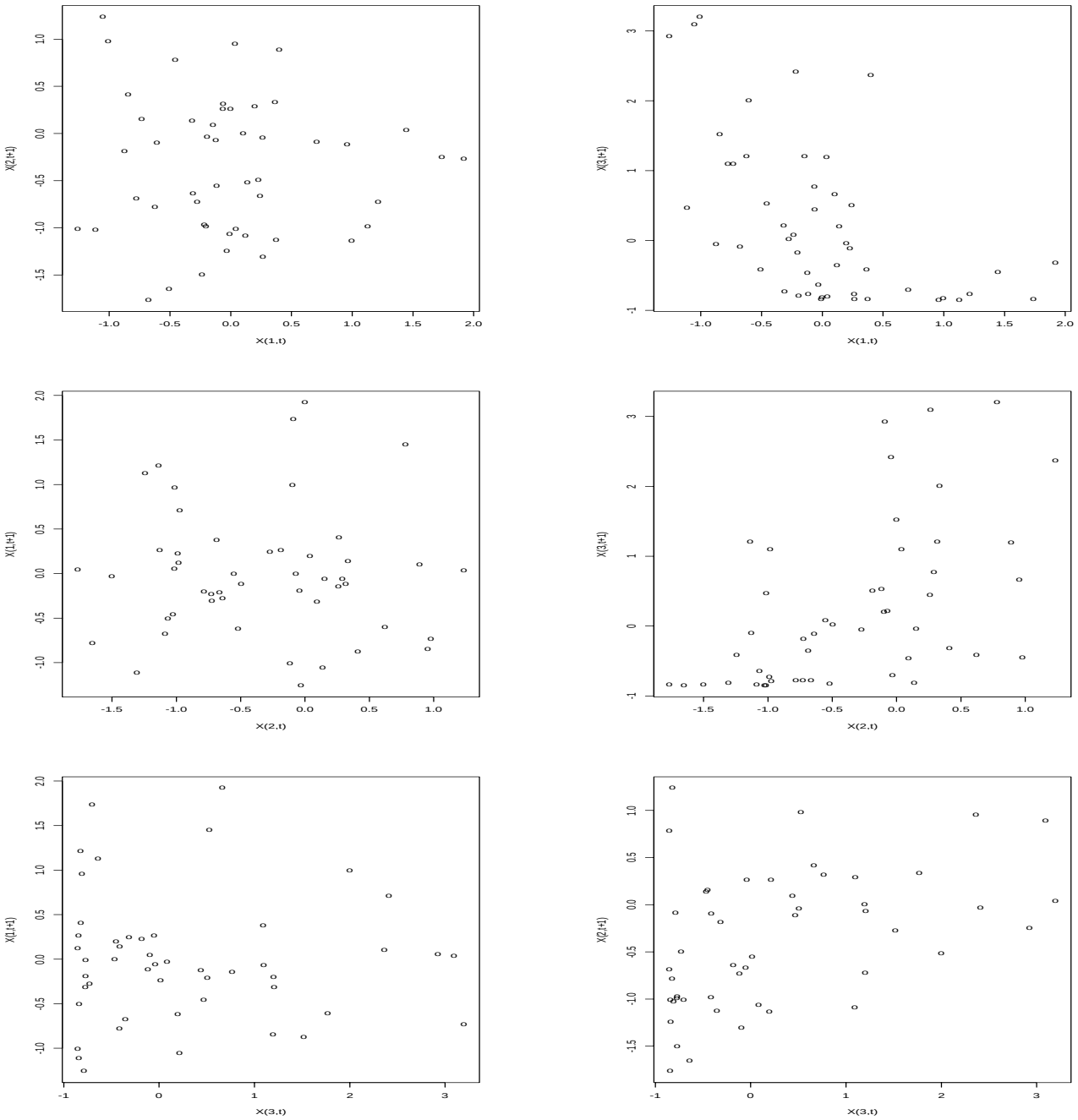

Figure 3: Scatter plots of $(X_{i,t}, X_{j,t+h})$ for $i = 1, 2, 3$ and $h = 0, 1$.

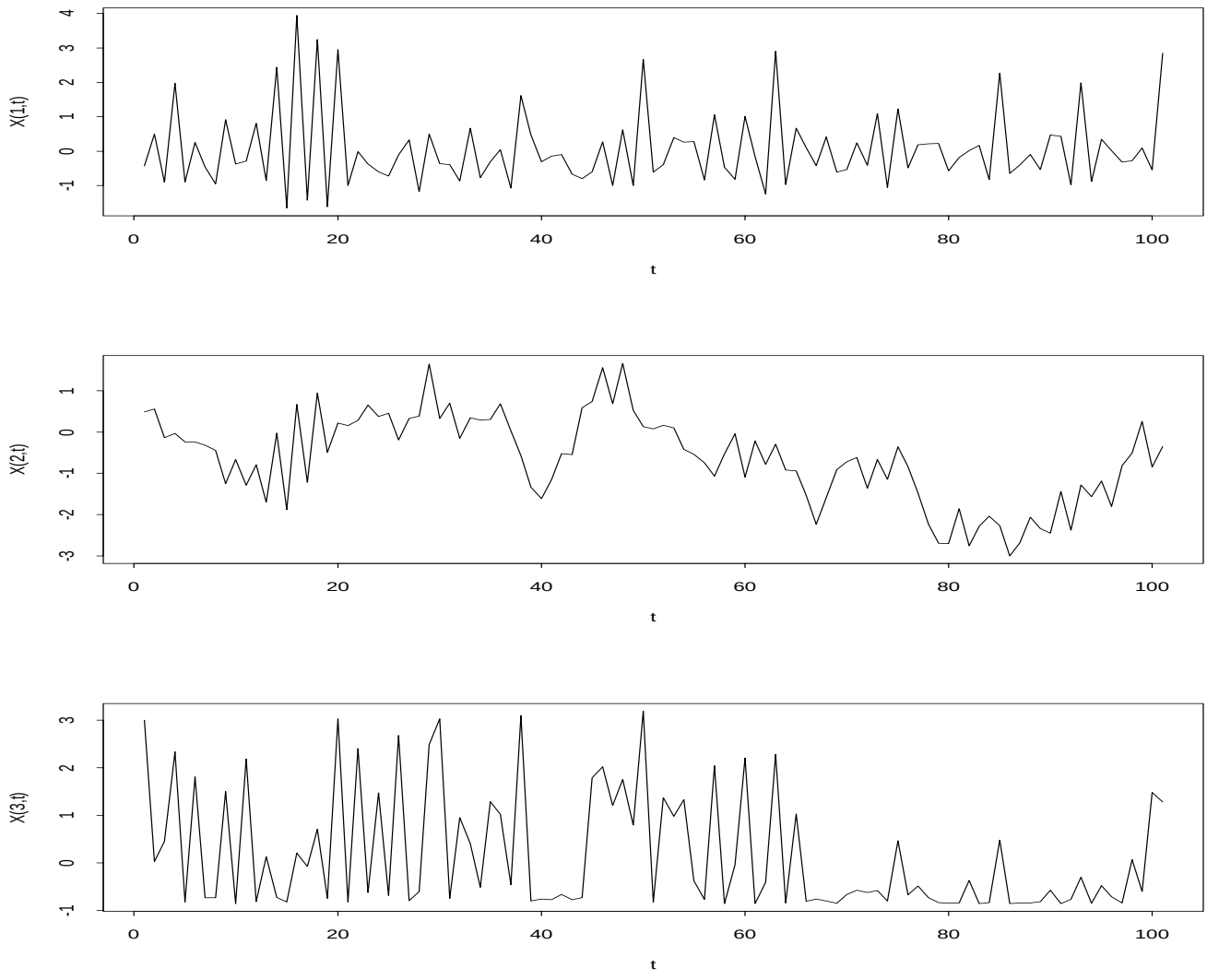

Figure 4: Generated component time series, $x_{1,t}$, $x_{2,t}$, and $x_{3,t}$ for $t = 0, 1, \ldots, 100$.