

QUANTIFYING INPUT UNCERTAINTY IN AN ASSEMBLE-TO-ORDER SYSTEM SIMULATION WITH CORRELATED INPUT VARIABLES OF MIXED TYPES

Alp Akcay

Department of Industrial Engineering
Bilkent University
06800, Ankara, TURKEY

Bahar Biller

Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA, 15213, USA

ABSTRACT

We consider an assemble-to-order production system where the product demands and the time since the last customer arrival are not independent. The simulation of this system requires a multivariate input model that generates random input vectors with correlated discrete and continuous components. In this paper, we capture the dependence between input variables in an undirected graphical model and decouple the statistical estimation of the univariate input distributions and the underlying dependence measure into separate problems. The estimation errors due to finiteness of the real-world data introduce the so-called input uncertainty in the simulation output. We propose a method that accounts for input uncertainty by sampling the univariate empirical distribution functions via bootstrapping and by maintaining a posterior distribution of the precision matrix that corresponds to the dependence structure of the graphical model. The method improves the coverages of the confidence intervals for the expected profit the per period.

1 INTRODUCTION

When we use simulation to estimate the performance of a stochastic system, it often includes multivariate input models (e.g., product demands and exchange rates in a global supply chain, processing times of a workpiece across several work centers) that are estimated using finite (and sometimes limited) amount of real-world data. The finite length of the historical data introduces estimation errors in the input models, implying additional uncertainty about the simulation outputs.

Quantification of input uncertainty in simulation output analysis is an active area of research, and a number of techniques have been proposed to address this problem. We refer reader to Barton (2012) for a detailed survey of the major approaches and a historical development of this literature. Ankenman and Nelson (2012), Song and Nelson (2013), Barton, Nelson, and Xie (2014) are the examples of some recent studies. The common assumption in the literature is that the input variables are statistically independent of each other. For many important real-world processes this is not the case, and it is the subject of this paper.

Our motivation comes from an assemble-to-order production (ATO) system simulation. In particular, we consider an ATO production system where the demands for the end products and the time since the last customer arrival are correlated. A simulation study of this system requires a multivariate input model that generates random input vectors with an appropriate dependence structure. More importantly, the input variables should be of mixed types in the sense that the univariate (marginal) distributions of the end product demands are discrete and the marginal distribution of the time since the last customer arrival is continuous.

In addition to accounting for dependence between input variables, we also consider the estimation of univariate input distributions with their empirical distribution functions, which are built directly from the past observations of the input variables. In an assemble to order system, for example, discrete demand arrives for each product and a parametric form may not be readily available to represent the demand-size distribution. In such cases, it is common to use the empirical distribution function in lieu of the unknown

demand-size distribution. An empirical distribution function is also practical to use for a decision maker as it avoids making any restrictive assumption on the parametric form of the demand distribution.

In a multivariate setting with correlated inputs, Biller and Corlu (2011) adopt the Bayesian model average approach of Chick (2001) and propose a method based on the copula-vine representation of the input model to efficiently sample the correlation matrix and the parameters of the known input distributions. However, the copula-vine representation of Biller and Corlu (2011) fails to work when at least one of the marginal distributions is discrete. Our paper fills this gap by allowing both continuous and discrete distributions. In the statistics literature, Pitt, Chan, and Kohn (2006) consider a multivariate model with both continuous and discrete distributions, however, they assume that each marginal distribution variable belongs to specified parametric families, and the use of empirical distributions is not considered. The semi-parametric copula estimation, introduced by Genest, Ghoudi, and Rivest (1995), estimates the marginal distributions and copula parameters separately by using the empirical distribution functions for the former. However, Hoff (2007) argues that the semiparametric estimators of the copula parameters, while well-behaved for continuous data, can fail for discrete data, making them inappropriate for a framework that captures input uncertainty under input variables of mixed types.

Considering this deficiency of using empirical distribution functions in a multivariate setting, Hoff (2007) has modeled the dependence structure by considering a multivariate normal distribution for a set of underlying latent variables. More recently, Dobra and Lenkoski (2011) extend this approach to a Gaussian copula graphical model that represents the conditional independencies between input variables. In this study, we build on this stream of work to obtain a posterior distribution for the correlation matrix of a normal copula function – which we use to model the joint distribution function of the input variables – and our objective is to use this posterior distribution to capture the input uncertainty in simulation output analysis. More specifically, we propose a procedure that integrates 1) the direct bootstrapping method (e.g., Barton and Schruben (1993), Barton and Schruben (2001), Barton, Nelson, and Xie (2010) to capture the uncertainty around the empirical distribution function associated with each input variable, and 2) the posterior sampling of the correlation matrix in a Gaussian copula graphical model to propagate the input uncertainty to the simulation output. The use of a graphical model is promising in multivariate input modeling as it gives additional flexibility to the simulation analyst to specify the conditional independencies in graph structure by considering the physics of the system being modeled, and hence, leads to a reduction in the number of parameters to be estimated from the historical data.

We organize the remainder of the paper as follows. In Section 2, we present the details of an undirected graphical model, which we consider as a promising approach for simulation input modeling. In Section 3, we describe the ATO system, illustrate how we build and estimate its multivariate input model based on an undirected graph, and discuss how we capture the input uncertainty. We present the results of numerical experiments in Section 4 and conclude the paper in Section 5.

2 GRAPHICAL REPRESENTATION OF THE DEPENDENCE STRUCTURE

The objective of this section is to model the joint distribution of a random vector $\mathbf{X} = (X_1, X_2, \dots, X_p)$. We let $V = \{1, 2, \dots, p\}$ for notational convenience. In our input modeling framework, we assume that there exists an ordering for the possible values of each input variable X_v , $v \in V$. This assumption is satisfied not only when X_v is continuous but also when X_v is binary, categorical with ordered categories, or count. We denote by F_v the univariate cumulative distribution function (cdf) of X_v , and $F_v^{-1}(u)$ the pseudo-inverse of F_v defined as $\inf\{x : F_v(x) \geq u\}$ for $u \in (0, 1)$.

We model the dependence structure in the joint distribution of \mathbf{X} in terms of a set of latent variables $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)$; see Hoff (2007) and Dobra and Lenkoski (2011) for the early studies following this approach. In particular, given a p by p precision matrix K , we let

$$\mathbf{Z} \sim N_p(0, K^{-1}),$$

where N_p denotes the multivariate normal distribution of dimension p . In this representation, K^{-1} is the covariance matrix. It is straightforward to see that the joint distribution of the scaled latent variables

$$\tilde{Z}_v := Z_v / \sqrt{(K^{-1})_{v,v}}, \quad v = 1, 2, \dots, p,$$

is then multivariate normal $\tilde{\mathbf{Z}} \sim N_p(0, \psi(K))$, where $\psi(K)$ is a correlation matrix with entries

$$\psi_{v_1, v_2}(K) = \frac{(K^{-1})_{v_1, v_2}}{\sqrt{(K^{-1})_{v_1, v_1}(K^{-1})_{v_2, v_2}}}. \quad (1)$$

Since $X_v = F_v^{-1}(\Phi(\tilde{Z}_v))$ for $v \in V$, the joint distribution F of \mathbf{X} is a function of the correlation matrix $\psi(K)$ and the univariate cumulative distribution functions F_v of X_v :

$$\begin{aligned} \mathbb{P}(X_1 \leq x_1, \dots, X_p \leq x_p) &:= F(x_1, \dots, x_p | \psi(K), F_1, \dots, F_p) \\ &= C(F_1(x_1), \dots, F_p(x_p) | \psi(K)), \end{aligned}$$

where

$$C(u_1, \dots, u_p | \psi) = \Phi_p(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p) | \psi) : [0, 1]^p \rightarrow [0, 1] \quad (2)$$

is the *Gaussian copula* with correlation matrix ψ (Nelsen 2006). In this representation, Φ is the cdf of the standard normal distribution, and $\Phi_p(\cdot | \psi)$ is the cdf of multivariate normal distribution with zero mean and correlation matrix ψ . When F_v is strictly monotonically increasing for all $v \in V$ (i.e., when all the margins are continuous), C is known to be unique. However, when one or more marginal distribution is discrete, this is no longer the case; see Genest and Nešlehová (2007). Nevertheless, the copula model in Equation (2) remains a well-defined joint distribution function for F (Smith and Khaled 2012).

The pairwise conditional independence relationships among the *latent* variables $\{Z_1, Z_2, \dots, Z_p\}$ can be encoded in a graph $G = (V, E)$, where each vertex $v \in V$ corresponds with a random variable Z_v and $E \subset V \times V$ are undirected edges (Whittaker 1990). In particular, the absence of an edge between Z_{v_1} and Z_{v_2} corresponds with the conditional independence of these two random variables given the remaining random variables under the joint distribution of \mathbf{Z} and is denoted by $Z_{v_1} \perp\!\!\!\perp Z_{v_2} | Z_{V \setminus \{v_1, v_2\}}$. This is called the pairwise Markov property and holds if and only if $K_{v_1, v_2} = 0$ in the precision matrix (Wasserman 2004). More generally, the edges of G correspond with the off-diagonal nonzero elements of K , that is, $E = \{(v_1, v_2) : K_{v_1, v_2} \neq 0, v_1 \neq v_2\}$. Furthermore, the global Markov property says that $\mathcal{A} \perp\!\!\!\perp \mathcal{B} | \mathcal{C}$, where \mathcal{A}, \mathcal{B} , and \mathcal{C} are distinct subsets of V and \mathcal{C} separates \mathcal{A} and \mathcal{B} (i.e., if every path from a variable in \mathcal{A} to a variable in \mathcal{B} intersects a variable in \mathcal{C}). It is well known that, if the distributions have positive continuous densities, the set of distributions associated with a graph that satisfy a pairwise Markov property is equivalent to the set of distributions that satisfy the global Markov property for that graph.

The construction of undirected graphs to encode the conditional independence relationships in the latent space provides a great flexibility to the simulation analyst by allowing any univariate continuous distribution and any discrete distribution with natural ordering in multivariate input modeling. We illustrate it in our assemble-to-order production system in Section 3.1. To the best of our knowledge, Dobra and Lenkoski (2011) is the first to construct a graphical model in the latent space – rather than the space of the original random variables – and it is referred as a *Gaussian copula graphical model*.

3 AN ASSEMBLE-TO-ORDER (ATO) PRODUCTION SYSTEM SIMULATION

We start with a description of the ATO production system in 3.1, where we also discuss the value of modeling the dependence structure with a graphical model to incorporate prior knowledge on conditional independence relationships. We discuss the issues of random-variate generation in Section 3.2 and the details of the Bayesian estimation framework for dependence parameters in Section 3.3. We discuss how we quantify the input uncertainty in Section 3.4.

3.1 System Description

Assemble-to-order systems have been widely studied in the supply chain management literature (Glasserman and Wang 1998, Cheng, Ettl, Lin, and Yao 2002, Iravani, Luangkesorn, and Simchi-Levi 2003). We will consider a modified version of the ATO system studied by Hong and Nelson (2006). This ATO system has the following general features: Components are made to stock to supply variable demands for finished products, and multiple finished products are assembled from the components. The system operates under a continuous-review base-stock policy under which each demand (of random size) triggers a replenishment order for that component. Components are produced one at a time on dedicated facilities, and production times are stochastic.

The system has six components and three finished products. Customers arrive at the system as a single Poisson process with rate λ . Each arriving customer declares how many units of each product will be purchased, denoted by D_i , $i = 1, 2, 3$. Each product requires a set of key components and a set of nonkey components. Each component sold brings a profit, p_j , $j = 1, 2, \dots, 6$, and each component in inventory has a holding cost h_j , $j = 1, 2, \dots, 6$, per unit time. If any of the key components is out of stock, the customer leaves. If all key components are in stock, the customer buys the product assembled from all the key components and the available nonkey components. If the available key items are not enough to meet all the demands placed by the customer, we assume that the demand is partially satisfied starting from the products that bring the most profit. The production time for each component is normally distributed with mean μ_j and variance σ_j^2 , truncated at zero. We tabulate all the parameters in Tables 1 and 2. The objective of the simulation study is to estimate the expected steady-state profit per period for a given value of the base-stock level for each component. We note that – even when the true values of the production-time parameters and the parameters of the demand arrival process are known – the performance of a policy under prespecified component base-stock levels can only be evaluated with simulation. We consider two types

Table 1: Parameters related to components.

Component	p_i	h_i	μ_i	σ_i
I	1	2	0.15	0.0225
II	2	2	0.40	0.0600
III	3	2	0.25	0.3750
IV	4	2	0.15	0.0225
V	5	2	0.25	0.0375
VI	6	2	0.08	0.0120

Table 2: Parameters related to products.

Product	Key component				Nonkey	
	I	II	III	IV	V	VI
1	1	0	0	1	0	1
2	1	1	0	0	1	0
3	0	0	1	1	1	0

of dependence in our input modeling framework: 1) The discrete demand random variables are correlated. One may argue that correlated end-product demands may arise when there are complementary or substitute product characteristics or when the demand for different product types are driven by similar underlying causes, e.g., specific marketing campaigns or special events. We assume there is no temporal dependence in the demand process. 2) We consider the correlations between the time since the last customer arrival, denoted by T , and the demand random variables. The dependence between an interarrival time and demand size has been studied in inventory management literature; see Willemain, Smart, Shockor, and DeSautels (1994), and Altay, Litteral, and Rudisill (2012).

In Figure 1, we illustrate the construction of the Gaussian copula graphical model to represent the dependence structure in the demand arrival process. The graph on the left is the *full graph* in which all the edges are present and none of the off-diagonal elements of the 5 by 5 precision matrix K is constrained to zero. Suppose that, given the knowledge of the realization of interarrival time T , knowing the demand size D_1 provides no information on the size of demand D_3 , and vice versa. In this case, we say that D_1 and

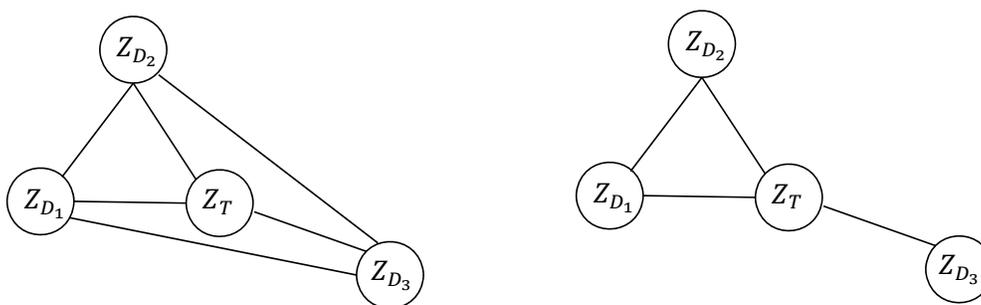


Figure 1: Illustration of the conditional independence structure in an ATO problem by an undirected graph: Full graph (left), $Z_{D_1} \perp\!\!\!\perp Z_{D_3} | Z_T$ and $Z_{D_2} \perp\!\!\!\perp Z_{D_3} | Z_T$ (right).

D_3 are independent conditional on T . We incorporate this information in the graph by omitting the edge between the vertices D_1 and D_3 , and set K_{D_1, D_3} equal to zero in the precision matrix. Similarly, the graph on the right illustrates that the demand random variables D_2 and D_3 are independent conditional on T .

The information on conditional independence can be inferred from the knowledge of the system physics. For example, the demand for products 1 and 3 can be only sensitive to the duration since last customer arrival without any apparent characteristics of complementary or substitute products that drive positive or negative demand correlation. Thus, once the duration since the last customer arrival is known, the demand for these two products can be regarded as independent. This is represented with a missing edge between latent-variable vertices Z_{D_1} and Z_{D_3} (which in turn implies conditional independence between D_1 and D_3) as illustrated in Figure 1 (right). On the other hand, suppose that products 1 and 2 are either complementary or substitute of each other and the duration since last customer arrival affects the demand for each product. Figure 1 (right) illustrates this situation with an edge between latent-variable vertices Z_{D_1} and Z_{D_2} .

In addition to the knowledge of system physics in constructing a graphical model, there are also methods to statistically test the existence of conditional independence between input random variables; see Borgelt, Steinbrecher, and Kruse (2009). In general, learning the graph structure from historical data is an active research area in statistics; we refer the reader to Liu, Han, Yuan, Lafferty, and Wasserman (2012) and the references therein. In this paper, we assume that the graph structure is known. However, we consider that the precision matrix is unknown and estimate it from historical data in a Bayesian fashion.

3.2 Sampling Random Variates

In this section, we briefly describe how the random instances of a multivariate input model (X_1, X_2, \dots, X_p) can be sampled to drive the simulation model. The Gaussian copula graphical model allows us to sample the latent variables from a multivariate normal distribution $N_p(0, \psi)$, where ψ is the correlation matrix in (1). The standard approach to the sampling of a multivariate normal random vector is based on the Cholesky decomposition of the correlation matrix ψ ; i.e., $\psi = AA'$ (Scheuer and Stoller 1962). More specifically, p independent standard normal random variates $\mathbf{v} = (v_1, v_2, \dots, v_p)'$ are first generated and then the scaled latent variables are set to $\tilde{\mathbf{Z}} = A\mathbf{v}$. Finally, the input random variables are obtained via the transformation $X_v = F_v^{-1}(\Phi(\tilde{Z}_v))$ for $v \in V$.

3.3 Statistical Estimation of the Marginal Distributions and the Dependence Structure

Since the true multivariate input model is unknown to the simulation analyst, it is common to estimate it from real-world data. In this study, we avoid making any assumption on the parametric form of any univariate input distribution F_i for $i \in V$. In particular, we consider the practical situation that the simulation analyst uses the empirical univariate cdf $\hat{F}_v(y) = (1/n) \sum_{t=1}^n \mathbf{1}(x_{v,t} \leq y)$ in lieu of the unknown cdf $F_v(\cdot)$ of the input random variable X_v . In this representation, n is the length of the historical data, $\mathbf{1}(\mathcal{E})$ is the

indicator of event \mathcal{E} , and $x_{t,v}$ is the realization of the input random variable X_v at time t for $v \in V$ and $t \in \{1, 2, \dots, n\}$.

In this way, we reduce the problem of statistical estimation of the multivariate input model to the estimation of the correlation matrix ψ for Gaussian copula in (2). This two-step method is known as *semiparametric* estimation and first introduced by Genest, Ghoudi, and Rivest (1995), who show that the resulting copula-parameter estimators are consistent and asymptotically normal when the univariate input cdfs are continuous. However, it is known that the semiparametric estimators of the copula parameters are problematic for discrete data because the transformation $\Phi^{-1}(\hat{F}_v(\cdot))$ does not change the data distribution – it only changes the sample space (Hoff 2007).

Let \mathbf{z}_t denote $(z_{t,1}, z_{t,2}, \dots, z_{t,p})'$, where $Z_{t,v}$ is the latent variable associated with the v th input random variable at time t . If the latent variables $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)'$ were observable, the simulation analyst could use them directly to make an inference about the correlation matrix ψ . However, the latent variables \mathbf{z} are not actually observed. Hoff (2007) suggests that the observed variables $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$, where $\mathbf{x}_t = (x_{t,1}, x_{t,2}, \dots, x_{t,p})'$, provide some information about the latent variables, even when the univariate cdfs F_v , $v \in V$, are unknown. In particular, the inequality $x_{t,v} < x_{t',v}$ holds for the historical data if and only if the inequality $z_{t,v} < z_{t',v}$ holds for the latent variables. This is simply because the pseudo-inverse $F_v^{-1}(\cdot)$ of the cdf of X_v , for $v \in V$, and the standard normal cdf $\Phi(\cdot)$ are nondecreasing functions. More generally, observing the historical data \mathbf{x} tells us that the latent variables \mathbf{z} must lie in the following set:

$$\mathcal{D} := \{ \mathbf{z} \in \mathbb{R}^{n \times p} : \max\{z_{t',v} : x_{t',v} < x_{t,v}\} < z_{t,v} < \min\{z_{t',v} : x_{t',v} < x_{t,v}\} \}.$$

In our search of an estimate for the correlation matrix ψ , we consider the occurrence of this event as our data. For notational convenience, we let \mathcal{D} denote the fixed space in $\mathbb{R}^{n \times p}$ associated with the latent variables \mathbf{z} given the historical data \mathbf{x} . In Sections 3.3.1 and 3.3.2, we present the likelihood function of the event $\mathbf{Z} \in \mathcal{D}$ for the full graphical model and a graphical model with encoded conditional independencies, respectively.

3.3.1 Full Graphical Model

We first present the likelihood of the observed data \mathbf{X} , which is given by

$$\begin{aligned} \mathbb{P}(\mathbf{X}|K, F_1, \dots, F_p) &= \mathbb{P}(\mathbf{Z} \in \mathcal{D}, \mathbf{X}|K, F_1, \dots, F_p) \\ &= \mathbb{P}(\mathbf{Z} \in \mathcal{D}|K, F_1, \dots, F_p) \mathbb{P}(\mathbf{X}|\mathbf{Z} \in \mathcal{D}, K, F_1, \dots, F_p) \\ &= \mathbb{P}(\mathbf{Z} \in \mathcal{D}|K) \mathbb{P}(\mathbf{X}|\mathbf{Z} \in \mathcal{D}, K, F_1, \dots, F_p). \end{aligned}$$

In this representation, the first equality follows because the event $\mathbf{Z} \in \mathcal{D}$ occurs whenever the historical data \mathbf{X} is observed. The last equality is based on the fact that the probability of the event $\mathbf{Z} \in \mathcal{D}$ does not depend on the univariate cdfs F_v , $v \in V$, which are treated as nuisance parameters. More specifically, we estimate the precision matrix K using only $\mathbb{P}(\mathbf{Z} \in \mathcal{D}|K)$, the part of the observed data likelihood that depends on the parameter of interest K but not on the nuisance parameters F_v , $v \in V$. There are two alternatives for the estimation of K . It can be done either by maximizing $\mathbb{P}(\mathbf{Z} \in \mathcal{D}|K)$ as a function of K (i.e., maximum likelihood estimation). However, the calculation of the likelihood function is problematic, and we will not follow it here. Instead, we will use the Bayesian approach to obtain the posterior distribution of K , which is given by

$$\mathbb{P}(K|\mathbf{Z} \in \mathcal{D}) \propto \mathbb{P}(K) \mathbb{P}(\mathbf{Z} \in \mathcal{D}|K).$$

Since the used likelihood function is based on the marginal probability of an event that is a superset of observing the ranks, Hoff (2007) refers it as the *extended rank likelihood*. We next present a Markov chain Monte Carlo sampler with stationary distribution $\mathbb{P}(K|\mathbf{Z} \in \mathcal{D})$ that is originally suggested by Hoff (2007) in terms of the covariance matrix K^{-1} . In particular, we construct a Gibbs sampler in terms of the precision matrix K , which is much easier to implement, noting that the distribution of Z_v conditional on the latent

variables $V \setminus \{v\}$ is normal with mean $-K_{v,v}^{-1}K_{v,-v}\mathbf{z}_{t,-v}$ and with variance $K_{v,v}^{-1}$. In this representation, $K_{v,-v}$ is the v th row of the precision matrix K with v th column removed. Similarly, $\mathbf{z}_{t,-v}$ is the latent variable vector $(z_{t,1}, z_{t,2}, \dots, z_{t,p})'$ with v th entry removed. Given the current state K^s of the chain, its next state K^{s+1} is generated by sequentially performing the following two steps:

Algorithm 1 Bayesian inference in full Gaussian copula graphical model

- 1: **Step 1:** Resample the latent data \mathbf{Z} :
 - 2: **for** $v = 1$ to p **do**
 - 3: **for** each $x \in \text{unique}\{x_{1,v}, \dots, x_{n,v}\}$ **do**
 - 4: **for** each t such that $x_{t,v} = x$ **do**
 - 5: $\mu_v \leftarrow -K_{v,-v}^s \mathbf{z}_{t,-v} / K_{v,v}^s$ and $\sigma_v^2 \leftarrow 1 / K_{v,v}^s$
 - 6: Calculate the bounds z_{lb} and z_{ub} in accordance with the set \mathcal{D} :
 - 7: $z_{lb} \leftarrow \max\{z_{t,v} : x_{t,v} < x\}$ and $z_{ub} \leftarrow \min\{z_{t,v} : x < x_{t,v}\}$
 - 8: Sample $u_{t,v}$ uniformly from the interval $\left[\Phi\left(\frac{z_{lb}-\mu_v}{\sigma_v}\right), \Phi\left(\frac{z_{ub}-\mu_v}{\sigma_v}\right)\right]$
 - 9: $z_{t,v} \leftarrow \mu_v + \Phi^{-1}(u_{t,v})\sigma_v$
 - 10: **Step 2:** Resample the precision matrix using the updated latent data:
 - 11: Sample K^{s+1} from a Wishart($\delta + n, K_0 + \mathbf{Z}'\mathbf{Z}$) distribution
-

In the sampling iteration provided in Algorithm 1, δ and K_0 are the prior parameters for the precision matrix K . We set $\delta = 3$ and $K_0 = I_p$, the p -dimensional identity matrix. The selection of this prior implies that the components of latent variables \mathbf{Z} are considered independent before observing any historical data and that the weight of the prior is equivalent to one realization of the multivariate data vector (Dobra and Lenkoski 2011).

In the next section, we provide a similar Markov chain Monte Carlo iteration to obtain the stationary distribution of $\mathbb{P}(K|\mathbf{Z} \in \mathcal{D})$ when the Gaussian copula graphical model encodes a given set of conditional independencies. To put it another way, we consider a precision matrix with some of the off-diagonal elements, that correspond to the missing edges in the graphical model, are constrained to be zero.

3.3.2 Graphical Model with Encoded Conditional Independencies

When the dependence structure is not a full graph, the Markov chain Monte Carlo iteration presented in Algorithm 1 requires two modifications. The first change is a simplification in Step 1 since the distribution of Z_v conditional on the latent variables $V \setminus \{v\}$ now depends only on the adjacent nodes of v , defined as $\text{adj}_G(v) = \{j \in V : (v, j) \in E\}$. More specifically, under the given graph structure G , the distribution of Z_v conditional on $\mathbf{Z}_{-v} = \mathbf{z}_{-v}$ is normal with mean $-\sum_{j \in \text{adj}_G(v)} K_{v,j} z_{t,j} / K_{v,v}$ and with variance $K_{v,v}^{-1}$. Consequently, we change the way we update the conditional mean μ_v in the fifth line of Algorithm 1 as in $\mu_v \leftarrow -\sum_{j \in \text{adj}_G(v)} K_{v,j} z_{t,j} / K_{v,v}$.

It is known that, given the graph structure G , the precision matrix K is constrained to the cone P_G of symmetric positive definite matrices with elements K_{v_1, v_2} equal to zero for all $(v_1, v_2) \notin E$ (Dobra and Lenkoski 2011). However, a sample of precision-matrix drawn from the Wishart distribution is not necessarily in P_G . Thus, the second change in Algorithm 1 is in Step 2, considering that the G -Wishart distribution is a conjugate prior for K when K is constrained by the graph G ; see Atay-Kayis and Massam (2005) and Wang and Li (2012) for details on G -Wishart distributions. To draw samples from a G -Wishart distribution, we present the block Gibbs sampler of Wang and Li (2012) in Algorithm 2 due to its convenience in implementation; see Wang and Li (2012) for alternative sampling methods.

The first task in sampling the precision matrix K from the G -Wishart distribution is identifying all the maximal cliques in the graph G ; i.e., note that a clique is a set of variables in G that are all adjacent to each other, and a set of variables is a maximal clique if it is a clique and if it is not possible to include

another variable and still be a clique. The notation K_{C_1, C_2} represents the matrix formed by the rows of K indexed by the variables in C_1 and the columns of K indexed by the variables in C_2 .

Algorithm 2 A block Gibbs sampler from G -Wishart(b, D) distribution

- 1: $\{C_j; j = 1 \dots, K\}$ is the set of maximum cliques in G ; $K^s \in P_G$ is the current value of precision matrix.
 - 2: **for** $j = 1$ to K **do**
 - 3: **Step 1:** Sample A from $\text{Wishart}(b, D_{C_j})$
 - 4: **Step 2:** $K_{C_j, C_j}^s \leftarrow A + K_{C_j, V \setminus C_j}^s (K_{V \setminus C_j, V \setminus C_j}^s)^{-1} K_{V \setminus C_j, C_j}^s$
 - 5: $K^{s+1} \leftarrow K^s$
-

To conclude, for a Gaussian copula graphical model with conditional independencies, the posterior distribution of the precision matrix K can be obtained after the second change in Algorithm 1 – the replacement of its eleventh line with Algorithm 2 to sample from G -Wishart($\delta + n, K_0 + \underline{\mathbf{Z}}'\underline{\mathbf{Z}}$) distribution.

3.4 Quantifying the Input Uncertainty

As we obtain the posterior distribution of the precision matrix (hence the correlation matrix) in Section 3.3 with a Markov chain Monte Carlo method, the posterior samples of the precision matrix are already on hand (after selection of the appropriate chain length, the length of scan intervals, and burn-in period). The key idea of our proposed method is to take advantage of these precision-matrix samples as we account for the uncertainty around the correlation coefficients in our multivariate input modeling framework. Furthermore, the separate handling of the statistical inference for marginal distributions and for the correlation coefficients allows us to propagate the uncertainty in the correlation parameters to the simulation output uncertainty by building on the well-known methods to capture input uncertainty.

An immediate way of accounting for the uncertainty around each univariate distribution is the direct bootstrapping method in a frequentist framework (Barton and Schruben 1993, Barton and Schruben 2001). As an alternative, a Bayesian model averaging strategy (Chick 2001) could be used to characterize the distribution of the expected simulation output by running simulations at each of many repeated samplings from a posterior distribution under a parametric input model. See Barton, Nelson, and Xie (2010) for a comparison of the common techniques to build confidence intervals that account for input uncertainty. In this study, we consider that the simulation analyst does not have any knowledge about the marginal distribution of each input variable except the available historical data, and the direct bootstrapping method is simply adopted to capture the uncertainty around the empirical distribution function of each input variable. The development of a Bayesian framework that integrates the use of the posterior distributions of marginal distributions and the posterior distribution of the precision matrix is the subject of the ongoing research.

We next present a simple procedure that incorporates the uncertainty in the correlation matrix without any extra computational effort to the direct bootstrapping method. In particular, a simulation replication at any particular set of empirical distribution functions is performed using a random correlation matrix sampled from its posterior distribution (Algorithm 3). The numerical experiments in Section 4 show that the resulting confidence intervals capture the uncertainty in the correlation matrix considerably.

In general, the accuracy of the resulting confidence intervals depend on having a large number of B , across which the total simulation budget is spread. Consequently, the mean simulation output may not be estimated precisely. Barton, Nelson, and Xie (2014) introduces a metamodeling approach that leverages all of the data from a designed experiment to provide improved predictions of the simulation output at each bootstrap point. The proposed procedure in Algorithm 3 can be conveniently adopted in such a metamodel where design points are sampled from the empirical moments of the univariate distributions and each simulation replication is performed with a new sample from the posterior distribution of the correlation matrix at a design point.

Algorithm 3 Direct-Bootstrapping Enhanced with Bayesian Sampling of Correlation Parameters

-
- 1: Set the number of bootstrap replications B , the significance level α , and the computational budget n_c .
 - 2: **for** $b = 1$ to B **do**
 - 3: Generate bootstrap samples \mathbf{x}^b and obtain the empirical distributions $\hat{F}_v^b, v \in V$.
 - 4: **for** $j = 1$ to n_c/B **do**
 - 5: Sample a correlation-matrix ψ^j from its posterior distribution.
 - 6: Compute the output Y_j of the simulation run driven by the inputs $\hat{F}_v^b, v \in V$, and ψ^j .
 - 7: Return the sample average \bar{Y}_b of $Y_j, j = 1, \dots, n_c/B$.
 - 8: Report the confidence interval $[\bar{Y}_{(\lceil B(\alpha/2) \rceil)}, \bar{Y}_{(\lceil B(1-\alpha/2) \rceil)}]$.
-

4 NUMERICAL RESULTS AND INSIGHTS

In this section, we continue our example on the ATO system, and focus on the input uncertainty associated with the demand-arrival process, a mix of continuous and discrete random variables. That is, without loss of generality, we assume that the simulation analyst knows the true distributions of the component production times.

The parameters of the demand-arrival process – which are unknown by the simulation analyst – are the cdf of the customer interarrival time (F_T), the cdf of demand size for Products 1, 2, and 3 (F_{D_1}, F_{D_2} , and F_{D_3}), and the precision matrix K that captures the relationships between four input models. We assume that the rate λ of the exponentially distributed interarrival time is 2, and that the demand for each product takes integer values between 1 and 10 with equal probabilities. The multivariate input vector consists of the demand-arrival input variables $V \in \{T, D_1, D_2, D_3\}$; i.e., the component production times are excluded for ease in presentation. We assume that the dependence structure is represented with a full graph, and specify a correlation matrix ψ with $\psi_{v_1, v_2} = \rho$ for all $v_1, v_2 \in V$ with $v_1 \neq v_2$.

It is worth noting that the simulation analyst does not specify any precision or correlation matrix in real world. These parameters – which represent the extent of dependence in the multivariate model – are learned from the available historical data. More specifically, we assume that there is not any information available about the demand-arrival process except a set of historical data $\mathbf{x}_t = (\tau_t, d_{1,t}, d_{2,t}, d_{3,t})$ for $t = 1, \dots, n$. Having this historical data on hand, the simulation analyst first uses the techniques in Section 3.3 to obtain a posterior distribution of the precision matrix, and then uses the empirical cdf of each input variable to generate random variates as described in Section 3.2. We simulate the ATO system for 70 time units, and remove the output of the first 20 time units to account for the initialization bias. Furthermore, the posterior distribution of the correlation matrix is obtained after 25,000 iteration of Algorithm 1; we remove the first 5,000 iterations and save the correlation matrices at every 10 iterations to eliminate the autocorrelation in the constructed Markov chain.

In Tables 3 and 4, we evaluate three procedures by estimating the confidence intervals for the expected per-period profit when the base-stock level of each component is set to ten units: 1) Conditional analysis which treat the empirical cdfs and the semiparametric maximum likelihood estimate of the correlation matrix as the true but unknown inputs; 2) the Bootstrap resampling method which captures the uncertainty around the empirical cdfs but not the correlation matrix; 3) the proposed procedure (Bootstrap + Bayesian) that accounts for the uncertainty around the correlation matrix as well.

It is important to note that a Bayesian method actually looks for a credible confidence interval that is intended to provide a posterior assessment of the analyst's remaining uncertainty about the mean simulation response rather than covering its fixed true value. However, we still use the coverage as a common measure of performance, and provide it at the first row of each column. We also report the average and standard deviation values of the confidence interval widths obtained from 500 macro replications in the second and third rows, respectively.

Not surprisingly, the conditional confidence interval provides very poor coverage as we only have 100 past observations of the input data vector. When the input uncertainty in both the empirical distribution

Table 3: $n = 100$, $\rho = 0.6$, $n_c = 4000$.

B	$1 - \alpha$	Conditional	Bootstrap	Bootstrap + Bayesian
100	0.95	0.11	0.82	0.96
		1.22	7.95	11.69
		0.07	2.01	3.18
200	0.95	0.13	0.85	0.94
		1.24	9.20	11.70
		0.11	3.10	4.94
400	0.95	0.17	0.95	0.99
		1.73	10.11	14.78
		0.13	2.50	3.50

Table 4: $n = 100$, $B = 200$, $n_c = 4000$.

ρ	$1 - \alpha$	Conditional	Bootstrap	Bootstrap + Bayesian
0.3	0.95	0.05	0.81	0.97
		1.10	7.01	11.20
		0.09	1.98	3.99
0.6	0.95	0.13	0.85	0.94
		1.24	9.20	11.70
		0.11	3.10	4.94
0.9	0.95	0.09	0.83	0.96
		1.05	9.17	13.47
		0.11	2.91	4.47

functions and the correlation matrix is captured, we observe that nominal coverage is attained in almost all the cases. However, the confidence intervals are likely to be conservative when the number of simulation replications is not sufficiently large per bootstrap of the empirical distribution function. For example, when B is equal to 400 in Table 3, each bootstrap step receives only 10 replications. Barton, Nelson, and Xie (2010) shows that this potential overcoverage problem can be effectively addressed by introducing a simulation metamodel that evaluates the simulation replications at carefully selected design points. The construction of a metamodel-assisted framework for capturing both empirical distribution and correlation matrix uncertainties is the subject of ongoing research.

5 CONCLUSION

In this paper, we illustrate the use of an undirected graphical model for simulation input modeling. In particular, the graphical representation of the dependence structure with a Gaussian normal copula allows us to consider any multivariate input model with a mix of continuous and discrete distributions. We obtain the posterior distribution of the correlation matrix associated with the normal copula and use the empirical distribution function of each input random variable. The separate handling of the statistical inference for marginal distributions and for the correlation coefficients allows us to combine different methods to propagate the uncertainty to the output uncertainty. Specifically, we capture the uncertainty around the empirical distribution functions with direct bootstrapping and capture the uncertainty around the correlation matrix by its posterior distribution. Our numerical analysis on an assemble-to-order system highlights the importance of the latter to achieve confidence intervals with better coverage.

REFERENCES

- Altay, N., L. Litteral, and F. Rudisill. 2012. "Effects of correlation on intermittent demand forecasting and stock control". *International Journal of Production Economics* 135 (1): 275–283.
- Ankenman, B., and B. L. Nelson. 2012. "A quick assessment of input uncertainty". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 241–250. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Atay-Kayis, A., and H. Massam. 2005. "A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models". *Biometrika* 92 (2): 317–335.
- Barton, R. R. 2012. "Tutorial: Input uncertainty in output analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 67–78. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Barton, R. R., B. L. Nelson, and W. Xie. 2010, December. "A framework for input uncertainty analysis". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Huan, and E. Yücesan, 1189–1198. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Barton, R. R., B. L. Nelson, and W. Xie. 2014. "Quantifying input uncertainty via simulation confidence intervals". *INFORMS Journal on Computing* 26 (1): 74–87.
- Barton, R. R., and L. W. Schruben. 1993. "Uniform and bootstrap resampling of empirical distributions". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. Evans, M. Mollaghasemi, E. Russell, and W. Biles, 503–508. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Barton, R. R., and L. W. Schruben. 2001. "Resampling methods for input modeling". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. Peters, J. Smith, D. Medeiros, and M. Rohrer, 372–378. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Biller, B., and C. Corlu. 2011. "Accounting for parameter uncertainty in large-scale stochastic simulations with correlated inputs". *Operations Research* 59 (3): 661–673.
- Borgelt, C., M. Steinbrecher, and R. R. Kruse. 2009. *Graphical models: representations for learning, reasoning and data mining*. John Wiley & Sons.
- Cheng, F., M. Ettl, G. Lin, and D. D. Yao. 2002. "Inventory-service optimization in configure-to-order systems". *Manufacturing & Service Operations Management* 4 (2): 114–132.
- Chick, S. E. 2001. "Input distribution selection for simulation experiments: accounting for input uncertainty". *Operations Research* 49 (5): 744–758.
- Dobra, A., and A. Lenkoski. 2011. "Copula Gaussian graphical models and their application to modeling functional disability data". *The Annals of Applied Statistics* 5 (2A): 969–993.
- Genest, C., K. Ghoudi, and L. Rivest. 1995. "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions". *Biometrika* 82 (3): 543–552.
- Genest, C., and J. Nešlehová. 2007. "A primer on copulas for count data". *Astin Bulletin* 37 (2): 475.
- Glasserman, P., and Y. Wang. 1998. "Leadtime-inventory trade-offs in assemble-to-order systems". *Operations Research* 46 (6): 858–871.
- Hoff, P. 2007. "Extending the rank likelihood for semiparametric copula estimation". *The Annals of Applied Statistics* 1 (1): 265–283.
- Hong, L. J., and B. L. Nelson. 2006. "Discrete optimization via simulation using COMPASS". *Operations Research* 54 (1): 115–129.
- Iravani, S., K. Luangkesorn, and D. Simchi-Levi. 2003. "On assemble-to-order systems with flexible customers". *IIE Transactions* 35 (5): 389–403.
- Liu, H., F. Han, M. Yuan, J. Lafferty, and L. Wasserman. 2012, 08. "High-dimensional semiparametric Gaussian copula graphical models". *The Annals of Statistics* 40 (4): 2293–2326.
- Nelsen, R. 2006. *An Introduction to Copulas*. 2nd edition, Springer, New York.

- Pitt, M., D. Chan, and R. Kohn. 2006. "Efficient Bayesian inference for Gaussian copula regression models". *Biometrika* 93 (3): 537–554.
- Scheuer, E. M., and D. S. Stoller. 1962. "On the generation of normal random vectors". *Technometrics* 4 (2): 278–281.
- Smith, M., and M. Khaled. 2012. "Estimation of copula models with discrete margins via Bayesian data augmentation". *Journal of the American Statistical Association* 107 (497): 290–303.
- Song, E., and B. Nelson. 2013. "A quicker assessment of input uncertainty". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 474–485. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Wang, H., and S. Li. 2012. "Efficient Gaussian graphical model determination under G-Wishart prior distributions". *Electronic Journal of Statistics* 6:168–198.
- Wasserman, L. 2004. *All of statistics: A concise course in statistical inference*. Springer.
- Whittaker, J. 1990. "Graphical models in applied multivariate analysis". *Chichester New York et al: John Wiley & Sons*.
- Willemain, T., C. Smart, J. Shockor, and P. DeSautels. 1994. "Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method". *International Journal of Forecasting* 10 (4): 529 – 538.

AUTHOR BIOGRAPHIES

ALP AKÇAY is an Assistant Professor of Industrial Engineering at Bilkent University. He received his Ph.D. in Operations Management and Manufacturing from Tepper School of Business at Carnegie Mellon University. His research interests include the design and analysis of stochastic system simulations and data-driven decision making under uncertainty with applications in operations management. His email address is alp.akçay@bilkent.edu.tr.

BAHAR BILLER is an Associate Professor of Operations Management and Manufacturing at Carnegie Mellon University. Her primary research interest lies in the area of computer simulation experiments for stochastic systems and more specifically, in multivariate input modeling for dependent input processes with applications to operations management and financial time-series modeling. Her email address is billerb@andrew.cmu.edu.