A Factor-Based Bayesian Framework for Risk Analysis in Stochastic Simulations

WEI XIE, Rensselaer Polytechnic Institute
CHENG LI, National University of Singapore
PU ZHANG, Rensselaer Polytechnic Institute

Simulation is commonly used to study the random behaviors of large-scale stochastic systems with correlated inputs. Since the input correlation is often induced by latent common factors in many situations, to facilitate system diagnostics and risk management, we introduce a factor-based Bayesian framework that can improve both computational and statistical efficiency, and also provide insights for system risk analysis. Specifically, we develop a flexible Gaussian copula based multivariate input model that can capture important properties in the real-world data. A nonparametric Bayesian approach is used to model marginal distributions and it can capture the properties, including multi-modality and skewness. We explore the factor structure of the underlying generative processes for the dependence. Both input and simulation estimation uncertainty are characterized by the posterior distributions. In addition, we interpret the latent factors and estimate their effects on the system performance, which could be used to support diagnostics and decision making for large-scale stochastic systems. Our approach is supported by both asymptotic theory and empirical study.

CCS Concepts: •**Computing methodologies** →*Modeling and simulation;* •**Applied computing** →*Operations research;*

Additional Key Words and Phrases: multivariate input model, nonparametric model, Gaussian copula, factor model, simulation, risk analysis

ACM Reference format:
WEI XIE, Cheng Li, and Pu Zhang. 2017. A Factor-Based Bayesian Framework for Risk Analysis in Stochastic Simulations. *ACM Trans. Model. Comput. Simul.* 9, 4, Article 39 (October 2017), 31 pages.
DOI: 0000001.0000001

1 INTRODUCTION

In the current interconnected world, the decision makers often face stochastic systems in large scale [46]. Simulation has become an important tool that is routinely used for designing and improving systems in a wide variety of fields, including manufacturing, supply chains and financial investments. As the systems become more complex and suffer from various sources of uncertainty, the decision makers are interested in correctly assessing system random behaviors, which could be quantified by a vector of quantiles or functions of quantiles, e.g., conditional value-at-risk (CVaR). In this paper, a new simulation methodology is introduced to correctly assess the risk performance for large-scale stochastic systems with correlated inputs.

The choice of *input models*, defined as the driving stochastic processes in the simulation experiments, has a direct impact on system performance estimation, and further affects decision making,

DOI: 0000001.0000001

such as ordering decisions in supply chains and investment decisions in portfolio management. Thus, to correctly assess the system random behaviors, input models should faithfully capture the important properties in the underlying physical processes, including heterogeneity, multi-modality, skewness and dependence. These properties are observed in the real data from automotive, electronics, biopharmaceutical and financial industries [1, 9, 43, 51]. The studies [2, 9, 44] and our empirical study also indicate that they could impact the system performance, especially risk behaviors.

Underlying unknown input models are often estimated from finite real-world data with error, called *input uncertainty*. There exist both input and simulation estimation errors. Ignoring either source of error could lead to unfounded confidence in the simulation assessment of system performance. Further, it is necessary to efficiently use the real-world data to reduce the overall estimation uncertainty of system performance.

Based on methodologies developed for quantifying the input model estimation uncertainty, existing approaches on input uncertainty can be divided into frequentist and Bayesian approaches; see the review in [6, 39, 50]. The frequentist approaches typically study the sampling distributions of point estimators of underlying input models. Since it could be hard to get the exact sampling distributions in many situations, the asymptotic approximation, including the normal approximation and the bootstrap, is often used to quantify the input model estimation uncertainty; see for example [5, 7, 8, 17, 18]. The asymptotic approximation is valid when the amount of real-world data is large. Compared to frequentist methods, Bayesian approaches derive the posterior distributions quantifying the input uncertainty and they do not need a large-sample asymptotic approximation for their validation. It is also straightforward for Bayesian approaches to incorporate the prior information about the underlying input models. See [54] for more detailed discussion on the comparison of frequentist and Bayesian approaches. In this paper, we focus on developing a new Bayesian approach.

Considering the amount of information required to construct a joint distribution, especially as the dimension of input models increases, *we assume that multivariate input models are characterized by marginal distributions and the dependence.* This assumption is commonly used in the simulation literature [12]. Thus, to correctly assess system performance, we need to faithfully capture the important features in the marginal distributions and the dependence. In this paper, we consider continuous input random variables.

Since the real-world data for each component often represent the variability coming from various latent sources of uncertainty, it could induce important properties, such as multi-modality, skewness and tails. For example, in a production system, a single raw material is shared by various production lines. The different stochastic status of production lines could induce the heterogeneity and multi-modality in the raw material demand data. The skewness and tails could be caused by the contamination in the production lines which leads to throwing away batches of products. These properties can have a great impact on system performance, e.g., the service levels in the inventory systems [2]. Thus, failing to capture them could lead to poor estimates for system performance.

Various approaches have been proposed in the literature to capture the important properties in the marginals (or univariate inputs); see the review and the tutorials on input uncertainty [6, 39, 50]. Many among them limit their choices of input models to parametric families with unknown parameter values, and characterize the input uncertainty by the posterior distributions of input parameters; see for example [13, 47, 54]. Among the parametric approaches, Johnson Transformation System (JTS) is relatively flexible. It can match any feasible combination of first four moments and capture a wide variety of unimodal and bimodal distributional shapes [13, 14]. However, parametric approaches tend to be limited, and the selected family may not be flexible enough to capture the rich properties in the real-world data. Model selection error does not disappear as the amount of real-world data increases.

Thus, Bayesian Model Averaging (BMA) was introduced to account for both families and parameters value uncertainty [19]. It was further extended to separate the input uncertainty and the simulation estimation error in [56, 57]. The family uncertainty is quantified by the posterior probabilities of a few *pre-specified* candidate parametric distributions. However, BMA relies on the assumption that all data are generated from a *single* true candidate distribution [15]. Without strong prior information about the distribution families of underlying physical processes, it could be hard to select appropriate candidate parametric distributions.

Considering that the real-world data represent the variability caused by various latent sources of uncertainty, we proposed a nonparametric Bayesian approach based on the Dirichlet Processes Mixture (DPM) to quantify the input uncertainty [53]. DPM approach was originally introduced in the statistics community; see for example [52], [25] and [30]. Since DPM models the underlying generative process, this approach provides a natural choice to characterize the input data coming from various latent sources of uncertainty. Thus, it can capture the important properties in the marginal distributions. The posteriors of flexible distributions can automatically account for the uncertainty of both model selection and parameter values. The nonparametric approach is asymptotically consistent under very general conditions, and it also demonstrates good and robust finite-sample performance [53].

In the simulation literature, various approaches were proposed to model input with dependence; see the review in [12]. Cario and Nelson introduced NORmal-To-Anything (NORTA) which can represent and generate random vectors with flexible marginal distributions and a correlation matrix [16]. However, since NORTA is based on moment-matching, it fails to represent an arbitrary feasible combination of marginals and a correlation matrix [55]. This issue becomes more severe as the dimension of input models increases [31]. Biller and Corlu proposed the use of Gaussian copula (GC) to model input joint distributions [10], which can avoid the NORTA infeasible issue. They further reviewed the copula-based multivariate input models that can capture the tail dependence in [11].

For large-scale stochastic systems with potential high-dimensional input models, we could have a limited amount of real-world data [9]. For example, in high-tech manufacturing industries, there could exist a large variety of products and a limited amount of real-world demand data because products tend to have the short life-cycle [35]. Since there are many correlation parameters to be estimated, statistical tests could suggest ignoring the *whole* dependence [9], which is important for the system risk performance assessment.

Correlated inputs could be induced by latent common factors in many situations. For example, in a project planning network, the activity durations for different tasks could be correlated because they are affected by the same nuisance factors, e.g., weather conditions. In inventory management for maintenance, the breakdowns of different components of complex systems, e.g., a jet engine and semiconductor production lines, could be dependent because they are impacted by the same underlying factors, e.g., the operating temperature. In portfolio investment, the return rates of different stocks could be dependent because they are impacted by common factors, e.g. the economic indicators of a certain industry.

In this paper, we propose a factor-based Bayesian framework so that we can correctly estimate the risk performance of stochastic systems with potential high-dimensional correlated inputs. Motivated by [45], we first develop a Gaussian copula based input model that can faithfully capture the important properties in real-world data, improve both computational and statistical efficiency, and facilitate the risk analysis for large-scale stochastic systems. We use the nonparametric Bayesian approach developed in [53] to model marginal distributions and explore the factor structure of the underlying generative processes for the dependence. Then, since our input models can not be specified by a fixed number of parameters, direct simulation is used to propagate the input uncertainty to outputs. Our framework delivers a posterior distribution and a credible interval (CrI) quantifying the overall uncertainty of system risk performance estimate. We can provide insights of underlying factors and estimate their effects on the system risk performance. *Therefore, our approach can correctly and efficiently assess the system risk behaviors, provide insights about the correlation, and further facilitate decision making to improve the system performance.*

In sum, the main contributions of this paper are as follows.

- Since the input correlation could be induced by latent common factors in many situations, we explore the factor structure in the correlation. Compared to the Gaussian copula model without exploring the factor structure in [10], our approach could improve both computational and statistical efficiency.
- We develop a factor-based Bayesian framework quantifying the overall uncertainty of the system risk performance estimates. We prove the asymptotic estimation consistency for the input models and the system performance.
- Our framework can provide insights of underlying factors. We further propose a procedure to estimate their effects, which could support diagnostics and decision making for large-scale stochastic systems.
- Since the marginal distributions tend to have relatively dominant impact on the system performance, we propose a flexible multivariate input model with nonparametric marginal distributions and a factor structure for the correlation. It can capture the important properties in real-world data, including heterogeneity, multi-modality, skewness and dependence.
- Even though there exists some simulation literature on assessing system risk performance, such as [34] and [36], the existing studies on the input uncertainty tend to focus on the system mean performance. In this paper, we study the impact of input uncertainty on the system risk performance characterized by a vector of percentiles.

In the next section, we provide a formal description of the problem of interest. In Section 3, we introduce a flexible multivariate input model and propose a factor-based Bayesian framework accounting for both input and simulation estimation uncertainty. We further interpret the latent common factors and estimate their effects on the system risk behaviors in Section 4. An empirical study on portfolio management is used to study the finite-sample performance of our input model and factor-based Bayesian framework in Section 5, and we conclude this paper in Section 6.

2 PROBLEM STATEMENT AND PROPOSED APPROACH

Given input models, denoted by F, the simulation outputs can be written as $\mathbf{Y}(F) = \{(Y_{r1}(F), Y_{r2}(F), \ldots, Y_{rL}(F)), r = 1, 2, \ldots, R\}$, where R is the number of replications and L is the run length. For example, in the inventory control, F is the distribution of product demands and $Y_{r\ell}$ is the overall cost occurring in the ℓth ordering time period. Notice that the simulation outputs depend on the choice of input models F that could be composed of mutually independent univariate and multivariate joint distributions. For notation simplification, suppose that there is only one multivariate joint distribution in F with the dimension, denoted by d.

We assume that the input joint distribution F is characterized by marginal distributions, denoted by $\{F_1, F_2, \ldots, F_d\}$, and a correlation matrix, denoted by $\mathbf{C}$. For an arbitrary feasible combination of marginals and a correlation matrix, there exists a Gaussian copula representation

$$F(x_1, x_2, \ldots, x_d) = \Phi_d\left(\Phi^{-1}\left[F_1(x_1)\right], \Phi^{-1}\left[F_2(x_2)\right], \ldots, \Phi^{-1}\left[F_d(x_d)\right]; \mathbf{C}\right)$$

where $\Phi_d(\cdot)$ and $\Phi(\cdot)$ denote the d-dimensional multivariate and univariate standard normal distributions. Gaussian copula can be interpreted as a transformation,

$$\mathbf{X} \overset{U_j=F_j(X_j)}{\longrightarrow} \mathbf{U} \overset{Z_j=\Phi^{-1}(U_j)}{\longrightarrow} \mathbf{Z}$$

for $j = 1, 2, \ldots, d$, where $\mathbf{U}$ follows a multivariate uniform distribution and $\mathbf{Z}$ follows a multivariate normal distribution, $\mathbf{Z} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{C})$; see [38, 49]. The unknown true input joint distribution, denoted by F^c, has the corresponding Gaussian copula representation specified by $(F_1^c, F_2^c, \ldots, F_d^c, \mathbf{C}^c)$.

In this paper, we propose a flexible multivariate input model to capture the important properties in the marginal distributions $F_1^c, F_2^c, \ldots, F_d^c$ and the correlation matrix $\mathbf{C}^c$. We model the marginals by a nonparametric Bayesian approach based on DPM [53]. Compared with parametric and BMA approaches, DPM does not require any strong prior information on the distribution families of $F_1^c, F_2^c, \ldots, F_d^c$ and it can capture the important properties, including multi-modality and skewness. In addition, considering that the correlated inputs could be induced by latent common factors in many situations, we explore the factor structure in the correlation $\mathbf{C}^c$. It not only improves the estimation efficiency of input model, but also provides insights of the dependence, especially for large-scale stochastic systems.

The underlying unknown input model F^c is estimated by m real-world data, denoted by a $(m \times d)$ matrix $\mathcal{X}_m^{(0)} \equiv \left(\mathbf{X}_1^{(0)}, \mathbf{X}_2^{(0)}, \ldots, \mathbf{X}_m^{(0)}\right)^\top$, with $\mathbf{X}_i^{(0)} \overset{i.i.d.}{\sim} F^c$ for $i = 1, 2, \ldots, m$. The input uncertainty is quantified by the posterior distribution $p(F|\mathcal{X}_m^{(0)})$. Since the input joint distribution could be specified by $(F_1, F_2, \ldots, F_d, \mathbf{C})$, the posterior distribution quantifying the input estimation uncertainty could be written as $p\left(F_1, F_2, \ldots, F_d, \mathbf{C}|\mathcal{X}_m^{(0)}\right)$. Since it is analytically and computationally intractable to do Bayesian inference on the marginals $F_1, F_2, \ldots, F_d$ and the correlation matrix $\mathbf{C}$ simultaneously, a two-stage estimation is typically used to do inference on the marginals and dependence separately [38, 49]. In the first stage, we estimate each marginal distribution F_j^c from the real-world data of the jth component, denoted by $\mathcal{X}_{jm}^{(0)} \equiv \{X_{j1}^{(0)}, X_{j2}^{(0)}, \ldots, X_{jm}^{(0)}\}$, for $j = 1, 2, \ldots, d$ with the estimation uncertainty quantified by the posterior distribution

$$p(F_j|\mathcal{X}_{jm}^{(0)}) \propto p(F_j) \cdot p(\mathcal{X}_{jm}^{(0)}|F_j)$$

where $p(F_j)$ denotes the prior and $p(\mathcal{X}_{jm}^{(0)}|F_j)$ represents the likelihood of data $\mathcal{X}_{jm}^{(0)}$. In the second stage, the marginals can be taken as "nuisance parameters" and the inference on the correction $\mathbf{C}$ is based on the summary statistics of $\mathbf{C}$, denoted by $D(\mathcal{X}_m^{(0)})$, independent of the marginals; see Section 8.3 in [48] and [33, 45]. We characterize the estimation uncertainty of the correlation matrix by the posterior distribution $p(\mathbf{C}|D(\mathcal{X}_m^{(0)}))$. Thus, the posteriors of the marginals and the correlation matrix can quantify the input uncertainty. Since the inference on dependence is only based on the summary statistics instead of the full real-world data, this two-stage approach simplifies estimation at the cost of only using partial information in the data. However, for continuous random vectors, this inference is asymptotically efficient [33].

We generate B posterior samples of input model to quantify the input uncertainty, denoted by $\{\tilde{F}^{(1)}, \tilde{F}^{(2)}, \ldots, \tilde{F}^{(B)}\}$, where $\tilde{F}^{(b)} \equiv \left(\tilde{F}_1^{(b)}, \tilde{F}_2^{(b)}, \ldots, \tilde{F}_d^{(b)}, \tilde{\mathbf{C}}^{(b)}\right)$ for $b = 1, 2, \ldots, B$. In this paper, we use $\tilde{\cdot}$ to denote the posterior sample. The sampling procedure for marginal distributions described in [53] can generate B samples for the marginal distributions, $\tilde{F}_j^{(b)} \sim p(F_j|\mathcal{X}_{jm}^{(0)})$, for $b = 1, 2, \ldots, B$ and $j = 1, 2, \ldots, d$. Then, we generate B samples of correlation matrix, $\tilde{\mathbf{C}}^{(b)} \sim p(\mathbf{C}|D(\mathcal{X}_m^{(0)}))$, for $b = 1, 2, \ldots, B$ by following the procedure in Section 3.1.

Given an input model F, a vector of quantiles of simulation outputs, denoted by $\mathbf{q}(F) = (q_1(F), q_2(F), \ldots, q_Y(F))$ with corresponding probabilities $0 < p_1 < p_2 < \ldots < p_Y < 1$, is used to characterize system random behaviors, where γ denotes a fixed positive integer, $q_\ell(F) \equiv \sup\{q \in \Re : G_{Y(F)}(q) \leq p_\ell\}$ for $\ell = 1, 2, \ldots, \gamma$ and $G_{Y(F)}$ is the cumulative distribution of the simulation output $Y(F)$.

The number of *active* input parameters, defined as parameters specifying the posterior sample of input model, $\tilde{F}^{(b)}$ for $b = 1, 2, \ldots, B$, depends on the estimated number of clusters for the marginal distributions and the estimated number of factors for input correlation. Since the number of parameters changes at different posterior samples of input model, it is challenging to construct a metamodel for system response. Thus, the direct simulation is used to propagate the input uncertainty to outputs. At each sample $\tilde{F}^{(b)}$, let $\mathbf{Y}^{(b)} \equiv \mathbf{Y}(\tilde{F}^{(b)})$ denote the simulation outputs. The simulation estimation uncertainty is characterized by the posterior distribution $p(\mathbf{q}(\tilde{F}^{(b)})|\mathbf{Y}^{(b)}, \tilde{F}^{(b)})$. Let $\tilde{\mathbf{q}}^{(b)} = \tilde{\mathbf{q}}(\tilde{F}^{(b)})$ denote a random draw from this posterior.

Therefore, in this paper, without strong prior information on the input model and the system response, a factor-based Bayesian framework is proposed to estimate the true quantiles $\mathbf{q}(F^c) = (q_1(F^c), q_2(F^c), \ldots, q_\gamma(F^c))$. It delivers a posterior and a CrI accounting for both input and simulation estimation uncertainty. Specifically, the input uncertainty is quantified by $\tilde{F} \sim p(F|\mathcal{X}_m^{(0)})$ and at any $\tilde{F}$, the simulation estimation uncertainty is quantified by $\tilde{\mathbf{q}}(\tilde{F}) \sim p(\mathbf{q}(\tilde{F})|\mathbf{Y}(\tilde{F}), \tilde{F})$. The overall estimation uncertainty for the quantiles is quantified by the posterior of the compound random vector $\tilde{\mathbf{q}}(\tilde{F})$. Our factor-based framework can efficiently use the real-world data $\mathcal{X}_m^{(0)}$ and reduce the estimation uncertainty of quantile responses. In addition, we can provide insights of the underlying common factors explaining the correlation and estimate their effects on the system risk performance, which could facilitate diagnostics and decision making for large-scale stochastic systems.

3 A BAYESIAN FRAMEWORK FOR RISK ANALYSIS

In this section, we propose a factor-based Bayesian framework to quantify the overall estimation uncertainty of system risk performance. In Section 3.1, we develop a flexible multivariate input model with nonparametric marginals and factor structure for the correlation. The input uncertainty is characterized by the posterior distribution of input model. Then, at each posterior sample of input model, we explore detailed simulation outputs and a nonparametric Bayesian approach is used to do inference on the system quantile response in Section 3.2. After that, a hierarchical sampling procedure is developed to deliver a posterior distribution and a CrI for system risk performance accounting for both input and simulation estimation uncertainty in Section 3.3. We show that as the amount of real-world data and the computational budget go to infinity, the system risk performance estimates converge to the true values.

3.1 Input Modeling and Input Uncertainty Quantification

We propose a flexible input model that can capture the important properties in the marginals and dependence respectively. Without strong prior information on the distribution families of marginals, a nonparametric Bayesian approach in Section 3.1.1 can capture the key properties and quantify the estimation uncertainty for marginals. Since the correlation could be induced by latent common factors in many cases, a factor model is used to explain the dependence in Section 3.1.2. Then, sampling approaches are proposed to do inference on the correlation matrix and the factor model in Sections 3.1.3 and 3.1.4. We prove the asymptotic consistency of the number of factors and the input joint distribution in Sections 3.1.5 and 3.1.6.

3.1.1 Nonparametric Marginal Distributions. Nonparametric DPM was introduced to quantify the input uncertainty for unit variate input models in [53]. Here, we briefly review it. Since the input data for each component could come from various latent sources of uncertainty, the density of the jth marginal distribution is modeled with DPM

$$f_j(x) = \sum_{\ell=1}^{+\infty} \pi_{j\ell} h_j(x|\boldsymbol{\phi}_{j\ell})$$

for $j = 1, 2, \ldots, d$, where $\pi_{j\ell}$ denotes the mixing weights and the kernel density function $h_j(\cdot|\boldsymbol{\phi}_{j\ell})$ is chosen based on the support of the distribution F_j^c. The mixing distribution of parameters $\left\{(\pi_{j\ell}, \boldsymbol{\phi}_{j\ell})_{\ell=1}^{+\infty}\right\}$, which is $\sum_{\ell=1}^{+\infty} \pi_{j\ell}\delta_{\boldsymbol{\phi}_{j\ell}}$ (δ_a is the Dirac function at a), is drawn from the Dirichlet process $\text{DP}(\alpha_j, G_{0j})$, where α_j is the dispersion parameter and G_{0j} is the base measure. Given the real-world data of the jth component $\mathcal{X}_{jm}^{(0)}$, the number of *active* clusters is finite and bounded from above by m.

DPM for the jth marginal distribution is specified by three key components: the kernel density $h_j(\cdot)$, the dispersion parameter α_j, and the base distribution G_{0j}. Considering the support of F_j^c commonly used in the simulation, we developed DPM models with three types of kernel densities, including Gaussian, Gamma and Beta, which account for the real-world data that are supported on the whole real line $\Re$, the half real line $\Re^+$, and a finite interval $[a_1, a_2]$ with $-\infty < a_1 < a_2 < \infty$ in [53]. Notice that the scaled version of DPM with Beta kernel could be applicable to continuous distributions with a finite support interval. Since DPM with a larger value of α_j tends to generate samples of the input density $f_j(\cdot)$ with more distinct active clusters, we infer the appropriate value of α_j from the real-world data. In addition, we choose G_{0j} to be as noninformative about the mixing parameters $\boldsymbol{\phi}_{j\cdot}$ as possible, and meanwhile take into account the ease of implementation. See the detailed description on the nonparametric univariate distribution in our study [53].

The posterior distribution $p(F_j|\mathcal{X}_{jm}^{(0)})$ is derived to characterize the estimation uncertainty for the marginal F_j with $j = 1, 2, \ldots, d$. Following the sampling procedure developed in [53], we can draw samples $\tilde{F}_j^{(b)} \sim p(F_j|\mathcal{X}_{jm}^{(0)})$ with $b = 1, 2, \ldots, B$ quantifying the marginal estimation uncertainty.

3.1.2 Factor Model for Correlated Inputs . Since the correlation between different components of $\mathbf{X}_i$ could be induced by latent common factors in many situations, we explore the factor model explaining the dependence in the latent random vector $\mathbf{Z}_i$

$$\mathbf{Z}_i = \underline{\Lambda}\boldsymbol{\eta}_i + \underline{\epsilon}_i$$

where $\underline{\Lambda}$ is a $(d \times k)$ loading matrix with k denoting the number of common factors, a $(k \times 1)$ vector $\boldsymbol{\eta}_i \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$ represents common factors with $\mathbf{I}_k$ denoting a $(k \times k)$ identity matrix, and a $(d \times 1)$ vector $\underline{\epsilon}_i \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$ represents Gaussian noises having the $(d \times d)$ diagonal covariance matrix $\boldsymbol{\Sigma} = \mathbf{C} - \underline{\Lambda\Lambda}^\top$. The common factors $\boldsymbol{\eta}_i$ and the noises ϵ_i are independent. The correlation matrix $\mathbf{C}$ has off-diagonal terms

$$\mathrm{C}_{jj'} \quad = \quad \sum_{h=1}^{k} \underline{\lambda}_{jh}\underline{\lambda}_{j'h} \text{ for } j \neq j' \text{ with } j, j' = 1, 2, \ldots, d. \tag{1}$$

To simplify the inference procedure, we consider the scaled random vector of $\mathbf{Z}_i$, denoted by $\mathbf{Q}_i$, following the factor model

$$\mathbf{Q}_i = \Lambda\boldsymbol{\eta}_i + \epsilon_i \tag{2}$$

with the noise $(d \times 1)$ vector $\epsilon_i \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$. Thus, the covariance matrix of $\underline{\mathbf{Q}_i}$ is $\Omega = \Lambda\Lambda^\top + \mathbf{I}_d$. Denote the data for scaled latent random vector by $\mathcal{Q}_m \equiv \left(\mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_m\right)^\top$. After doing the inference with the factor model in Equation (2), we can get

$$Z_{ij} = Q_{ij} \Big/ \sqrt{1 + \sum_{h=1}^{k} \lambda_{jh}^2}. \tag{3}$$

$\underline{\lambda}_{jh} = \lambda_{jh} / \sqrt{1 + \sum_{h=1}^{k} \lambda_{jh}^2}$ and $\underline{\epsilon}_{ij} = \epsilon_{ij} / \sqrt{1 + \sum_{h=1}^{k} \lambda_{jh}^2}$. Let $\mathcal{Z}_m \equiv (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_m)^\top$. *By Equation (1), the loading matrix* Λ *or* $\underline{\Lambda}$ *can characterize the correlation matrix* $\mathbf{C}$.

Here, we discuss the *intuitions* for modeling the input correlation with the factor model in Equation (2). First, in many situations, the input correlation could be induced by underlying common factors. For example, in the portfolio management example studied in Sections 5.2–5.3, the correlations among return rates of different stocks could be induced by various industry indexes. Thus, the factor model can be used to model the generative processes for the input correlation. Second, for many cases, the underlying factors represent aggregated effects, such as the industry indexes in the stock management example. Even though each component, e.g., the individual stock return rate, typically does not follow normal distribution, the normality assumption on the underlying factors, e.g., industry indexes, could hold by applying the central limit theorem.

Exploring the factor model on the input correlation leads to some *benefits*. First, the factor model can improve both computational and statistical efficiency. For the real-world data with limited sample size and relatively high dimension, the Gaussian copula factor model is more parsimoniously parametrized. Thus, it can reduce the computational time for the inference and improves the estimation accuracy. Second, since the factor model in (1) could model the generative processes for the input correlation, it can provide insights of underlying factors, which could facilitate the risk analysis for large-scale stochastic systems.

3.1.3 Bayesian Inference on the Correlation Matrix $\mathbf{C}$ *and the Loading Matrix* Λ. In this section, suppose that the number of common factors k is known. Given m real-world data, $\mathcal{X}_m^{(0)} = (\mathbf{X}_1^{(0)}, \mathbf{X}_2^{(0)}, \ldots, \mathbf{X}_m^{(0)})^\top$, we make inference on the correlation matrix $\mathbf{C}$ and the loading matrix Λ. If the marginals $F_1^c, F_2^c, \ldots, F_d^c$ are known, we can have the corresponding data on latent variables $\mathbf{Z}$ by applying the transformation $Z_{ij}^{(0)} = \Phi^{-1}\left[F_j^c(X_{ij}^{(0)})\right]$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, d$. For continuous marginals, there exists a one-to-one mapping between $\mathbf{X}_i^{(0)}$ and $\mathbf{Z}_i^{(0)}$. Since $\mathbf{Z}_i^{(0)} \overset{i.i.d.}{\sim} \mathcal{N}_d(\mathbf{0}, \mathbf{C})$ for $i = 1, 2, \ldots, m$, it is easy to derive the posterior distribution $p(\mathbf{C}|\mathbf{Z}_1^{(0)}, \mathbf{Z}_2^{(0)}, \ldots, \mathbf{Z}_m^{(0)})$; see [26].

However, the marginal distributions $F_1^c, F_2^c, \ldots, F_d^c$ are unknown, and the only information for the transformation $\Phi^{-1}\left[F_j^c(\cdot)\right]$ is an increasing function. Based on the marginal likelihood described in [48], the extended rank likelihood was proposed in [33] to generate a set of the $(m \times d)$ data matrix $\mathcal{Z}_m$ that is consistent with real-world data in terms of the relative order,

$$D(\mathcal{X}_m^{(0)}) \equiv \{\mathcal{Z}_m : X_{ij}^{(0)} < X_{i'j}^{(0)} \Rightarrow Z_{ij} < Z_{i'j}\}.$$

Thus, $\mathcal{Z}_m$ is independent on the marginals and only depends on the correlation $\mathbf{C}$.

Given the data $\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})$, the posterior $p(\mathbf{C}|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}))$ could be used to quantify the uncertainty of $\mathbf{C}$. Since the marginal distributions are unknown, $\mathcal{Z}_m$ consistent with $\mathcal{X}_m^{(0)}$ is not unique. To account for the impact from unknown marginals on the dependence estimation, we further generate samples of $\mathcal{Z}_m$ from $D(\mathcal{X}_m^{(0)})$. Therefore, the uncertainty of correlation matrix could be characterized by $p(\mathbf{C}|\mathcal{X}_m^{(0)}) \equiv \mathrm{E}\left[p(\mathbf{C}|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}))\right]$.

Since $\mathbf{C}$ is invariant to the orthogonal rotation of Λ, it is well known that the factor model has identification problems [1, 45]. Here we apply a commonly used set of *constraints on the loading* *matrix* Λ or $\underline{\Lambda}$: lower triangular matrix with all diagonal entries positive. Denote the restrict space by Θ_Λ. Given k, these constraints uniquely identify the factor loadings.

For the Bayesian inference on the loading matrix, a Gaussian prior satisfying the constraints is used,

$$\lambda_{jh}|\psi_{jh} \begin{cases} \sim \mathcal{N}(0, \psi_{jh}) & \text{if } j > h \\ \sim \text{TN}(0, \psi_{jh}, 0) & \text{if } j = h \\ = 0 & \text{if } j < h \end{cases}$$

for $j = 1, 2, \ldots, d$ and $h = 1, 2, \ldots, k$, where ψ_{jh} controls the prior variation of the loading of the hth factor on the jth component and $\text{TN}(0, \psi_{jh}, 0)$ denotes the normal with mean 0, variance ψ_{jh} and truncated to be positive. The prior for ψ_{jh} is Inverse-Gamma $(\alpha_0/2, \beta_0/2)$, where α_0 and β_0 are the hyper-parameters. Let $\boldsymbol{\Psi}$ denote a $(d \times k)$ matrix having elements ψ_{jh}. Let $\mathbf{H} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_m)^\top$ and $\mathbf{H}_h$ represents the hth column of $\mathbf{H}$ with $h = 1, 2, \ldots, k$. Let $\boldsymbol{Q}_{\cdot j}$ denote the jth column of $\boldsymbol{Q}_m$ with $j = 1, 2, \ldots, d$.

A Gibbs sampler is developed to generate posterior samples of parameters $(\Lambda, \mathbf{H}, \boldsymbol{\Psi})$ and latent variables $\mathcal{Z}_m$. The conditional posteriors of Λ, $\boldsymbol{\Psi}$, $\mathbf{H}$ are given in Equations (4)–(6)

$$\lambda_{jh}|\psi_{jh}, \boldsymbol{Q}_{\cdot j}, \mathbf{H} \begin{cases} \sim \mathcal{N}\left(v_{jh} \sum_{i=1}^{m} a_{ijh}\eta_{ih}, v_{jh}\right) & \text{if } j > h \\ \sim \text{TN}\left(v_{jh} \sum_{i=1}^{m} a_{ijh}\eta_{ih}, v_{jh}, 0\right) & \text{if } j = h \\ = 0 & \text{if } j < h \end{cases} \tag{4}$$

$$\psi_{jh}|\lambda_{jh} \sim \text{Inverse-Gamma}\left(\frac{\alpha_0 + 1}{2}, \frac{\beta_0 + \lambda_{jh}^2}{2}\right) \tag{5}$$

$$\boldsymbol{\eta}_i|\mathbf{Q}_i, \Lambda \sim \mathcal{N}_k\left((\Lambda^T\Lambda + \mathbf{I}_k)^{-1}\Lambda^T\mathbf{Q}_i, (\Lambda^T\Lambda + \mathbf{I}_k)^{-1}\right) \tag{6}$$

with $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, d$ and $h = 1, 2, \ldots, k$, where $v_{jh} = \left(\sum_{i=1}^m \eta_{ih}^2 + \psi_{jh}^{-1}\right)^{-1}$ and $a_{ijh} = Q_{ij} - \sum_{h' \neq h} \lambda_{jh'}\eta_{ih'}$. The detailed derivation for Equations (4)–(6) is provided in the online appendix. By the extended rank likelihood [33, 45], the conditional distribution of Z_{ij} is

$$Z_{ij}|\Lambda, \boldsymbol{\eta}_i \sim \text{TN}\left(\sum_{h=1}^{k} \underline{\lambda}_{jh}\eta_{hi}, \frac{1}{1 + \sum_{h=1}^{k} \lambda_{jh}^2}, Z_{ij}^{\ell}, Z_{ij}^{u}\right) \tag{7}$$

where $\text{TN}(u, \sigma^2, a, b)$ denotes the normal with mean u, variance σ^2 and truncated to (a, b), $Z_{ij}^{\ell} = \max\{Z_{i'j} : X_{i'j}^{(0)} < X_{ij}^{(0)}\}$ and $Z_{ij}^{u} = \min\{Z_{i'j} : X_{i'j}^{(0)} > X_{ij}^{(0)}\}$. Given Λ, we can get $\boldsymbol{Q}_m$ from $\mathcal{Z}_m$ by applying Equation (3).

Therefore, the Gibbs sampling can deliver posterior samples of the correlation and loading matrices, $\tilde{\Lambda} \sim p(\Lambda|\mathcal{X}_m^{(0)})$ and $\tilde{\mathrm{C}} \sim p(\mathrm{C}|\mathcal{X}_m^{(0)})$. The main steps in each iteration are in Algorithm 1.

Notice that the Gibbs sampler considers the uncertainty introduced from unknown marginals.

ALGORITHM 1: Gibbs Sampler to Generate Posterior Samples for Correlation and Loading Matrices

```
Input: The real-world data $\mathcal{X}_m^{(0)}$
Output: Posterior samples of $\tilde{\Lambda}$ and $\tilde{\mathbf{C}}$
1 Initialize the latent variables $\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})$, initialize the common factors $\mathbf{H}$, the loading matrix $\Lambda$ and the
    hyper-parameters $\mathbf{\Psi}$ by sampling from the priors.
2 Generate a sample $\lambda_{jh}$ for $j = 1, \ldots, d$ and $h = 1, \ldots, k$ by using Equation (4)
3 Generate a sample $\psi_{jh}$ for $j = 1, \ldots, d$ and $h = 1, \ldots, k$ by using Equation (5)
4 Generate a sample $\boldsymbol{\eta}_i$ for $i = 1, \ldots, m$ by using Equation (6)
5 Generate latent variables $Z_{ij}$ by using Equation (7), and then obtain $Q_{ij}$ by applying Equation (3) for
    $i = 1, \ldots, m$ and $j = 1, \ldots, d$
6 Repeat Steps 2-5 to generate posterior samples of loading matrix $\tilde{\Lambda}$ and correlation matrix $\tilde{\mathbf{C}}$ by applying
    Equation (1).
```

3.1.4 *Bayesian Inference on the Number of Common Factors k.*

Given a fixed number of common factors k, we develop a sampling procedure to do Bayesian inference on the loading and correlation matrices in Section 3.1.3. However, k is typically unknown, and factor models with different k could lead to the same correlation matrix; see [27] for this identification issue. Since the value of k controls the complexity of input model, we want to find the factor model with the smallest number of common factors explaining the input correlation, which could reduce the input uncertainty and also facilitate the interpretation of the correlated input.

In this section, we discuss a procedure to find the factor model with the smallest number of common factors. Specifically, a Bayesian sampling approach, the Reverse Jump Monte Carlo Markov Chain (RJMCMC), is used for the model selection; see [42]. It searches through models with different number of factors and selects the *simplest* one explaining the correlation. Denote the smallest number of factors that can explain the true correlation matrix $\mathbf{C}^c$ by k^0, and suppose that it is much smaller than d. We set the range of possible numbers of factors as $\mathcal{K} \equiv \{1, 2, \ldots, k_{\max}\}$, and ensure that $k_{\max}$ is large enough with $k_{\max} \geq k^0$. By Section 2.2 of [42], the largest possible number of latent factors $k_{\max}$ is bounded from above by the relation $d(d+1)/2 - d(k_{\max}+1) + k_{\max}(k_{\max}-1)/2 \geq 0$. In our empirical study, a uniform prior is imposed on k, $p(k) = 1/k_{\max}$ for any $k \in \mathcal{K}$.

Suppose that the current model is $\mathcal{M}_k$ with k common factors and loading matrix Λ_k. A new candidate model $\mathcal{M}_{k'}$ is generated according to the transition probabilities $J(k \longrightarrow k')$. In our study, we use the following transition probabilities:

- If $k = 1$, $J(k \longrightarrow k') = 1$ if $k' = k + 1$, and $= 0$ otherwise;
- If $k = k_{\max}$, $J(k \longrightarrow k') = 1$ if $k' = k - 1$, and $= 0$ otherwise;
- If $1 < k < k_{\max}$, $J(k \longrightarrow k') = 0.5$ if $k' = k + 1$ or $k' = k - 1$, and $= 0$ otherwise.

It ensures that the number of factors between each move can only increase or decrease by one within the range $\mathcal{K}$. Then, a Metropolis-Hasting approach is used to determine whether to accept the movement.

Based on the study [42], the Bayesian sampling procedure with RJMCMC for the inference on the number of common factors, the loading and correlation matrices is presented in Algorithm 2. In Step 1, we have preliminary MCMC analysis for each $k \in \mathcal{K}$ by following the sampling procedure described in Section 3.1.3, which allows us to estimate the posterior moments of loading matrix for any fixed k. In Steps 2 to 6, we find the proposal distribution and then apply the Metropolis-Hasting approach to do model selection. In our empirical study, we let the proposal distribution $g_{k'}(\Lambda_{k'})$ to be $\mathcal{N}(\mathbf{b}_{k'}, a\mathbf{B}_{k'})$, where $\mathbf{b}_{k'}$ and $\mathbf{B}_{k'}$ denote the approximate posterior mean and covariance matrix estimated from the preliminary MCMC analysis, and a denotes the scaling parameter. Similar to

Equation (4), we generate samples of $\Lambda_{k'}$ satisfying the lower triangular with positive diagonal entries. In Step 7, we update the number of common factor k, and then generate samples of loading matrix Λ_k by following the procedure in Section 3.1.3. By repeating Steps 2 to 7, we can get posterior samples of (k, Λ_k), and further obtain samples of C by applying Equation (1). Therefore, with posterior samples of marginal distributions obtained in Sections 3.1.1 and the correlation matrix obtained in this section, $\{\tilde{F}_1^{(b)}, \tilde{F}_2^{(b)}, \ldots, \tilde{F}_d^{(b)}, \tilde{\mathrm{C}}^{(b)}\}_{b=1}^{B}$, we can quantify the input uncertainty.

ALGORITHM 2: A Gibbs Sampler with RJMCMC for Posterior Inference of k, Λ_k and C

```
Input: The real-world data $\mathcal{X}_m^{(0)}$
Output: Posterior samples of $k, \Lambda_k, \tilde{\mathrm{C}}$
1 Preliminary MCMC analysis: For each $k \in \mathcal{K}$, run the Gibbs sampling procedure described in Section 3.1.3.
    Initialize a starting value of $k \in \mathcal{K}$ for the model search.
2 repeat
3     Draw a candidate number of factors $k'$ from $J(k \to k')$.
4     Draw a loading matrix $\tilde{\Lambda}_{k'}$ from the proposal distribution $g_{k'}(\Lambda_{k'})$.
5     Compute the accept ratio
          $\beta = \min\left\{1, \frac{p(\mathcal{Z}_m|k', \tilde{\Lambda}_{k'})p(\tilde{\Lambda}_{k'}|k')p(k')}{p(\mathcal{Z}_m|k, \Lambda_k)p(\Lambda_k|k)p(k)} \frac{g_k(\Lambda_k)J(k' \to k)}{g_{k'}(\tilde{\Lambda}_{k'})J(k \to k')}\right\}.$     (8)
      With probability $\beta$, accept the move to the $k'$-factor model, and set $k = k'$. Otherwise, keep $k$
      unchanged.
6 until convergence;
7 Within the updated model: Generate a new posterior sample of $\Lambda_k$ by applying the sampling procedure
    described in Section 3.1.3. Then, update the mean and variance parameters of $g_k(\Lambda_k)$ by applying the
    approach proposed in [32].
8 For $b = 1, \ldots, B$, redo step 3-5 and step 7 to draw a posterior sample of $(k, \Lambda_k)$. Calculate $\tilde{\mathrm{C}}^{(b)}$ by applying
    Equation (1).
```

3.1.5 Posterior Consistency of the Number of Common Factors. In this section, we show that the number of common factors can be consistently estimated by the sampling procedure in Section 3.1.4 and it converges to the *smallest* value k^0. *We assume that the true correlation matrix* C^c *has a sparse representation of* k^0 *factors:* $\mathrm{C}^c = \underline{\Lambda}^c \underline{\Lambda}^{c\top} + \Sigma^c$, where $\underline{\Lambda}^c$ is a $(d \times k^0)$ lower triangular loading matrix with positive diagonal entries and Σ^c is a $(d \times d)$ diagonal matrix with all diagonal entries positive.

For any $k \in \mathcal{K}$, we define a model space $\mathcal{M}_k$ with k factors as

$$\mathcal{M}_k = \left\{\mathrm{C} : \mathrm{C} = \underline{\Lambda\Lambda}^\top + \Sigma \qquad \text{such that } \underline{\Lambda} \in \Theta_\Lambda \text{ and } \Sigma \text{ has all diagonal entries positive}\right\}.$$

From Section 2.2 of [42], even under the restrictions that Λ and $\underline{\Lambda}$ are lower triangular with positive diagonal entries, the factor model still has the identification issue: $\mathcal{M}_1 \subset \mathcal{M}_2 \subset \cdots \subset \mathcal{M}_{k_{\max}}$, and $\mathrm{C}^c \in \mathcal{M}_k$ for all $k^0 \leq k \leq k_{\max}$. Thus, the factor model belongs to the so-called *singular models*, where Gaussian approximation of the posterior distribution fails to hold asymptotically [21, 22].

However, the posterior distribution on the space $\mathcal{M}_k$ still allows a different asymptotic approximation in the form of (2.7) in [22]. Let $p\big(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\big) = \mathrm{E}_{\mathrm{C}}\big[p\big(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathrm{C}, \mathcal{M}_k\big)\big]$ and $\widehat{\mathrm{C}}_k = \mathrm{argmax}_{\mathrm{C} \in \mathcal{M}_k} p\big(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathrm{C}, \mathcal{M}_k\big)$ be the marginal likelihood of $\mathcal{Z}_m$ and the maximum likelihood estimator of C on the model $\mathcal{M}_k$. Similar to the relation (2.7) and the assumptions (A1)–(A3) in [22], we make the following assumptions on the factor model.

(A1) For any $k \in \mathcal{K}$, $p\big(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\big)$ has the asymptotic expansion as $m \longrightarrow \infty$,

$$\log p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right) = \log p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \widehat{\mathrm{C}}_k, \mathcal{M}_k\right) - t_k \log m + s_k \log\log m + O_p(1), \qquad (9)$$

where $\{t_k\}$ and $\{s_k\}$ with $1 \leq k \leq k_{\max}$ are two sequences of positive constants which depend on the true number of factors k^0, and $\{t_k\}$ is strictly increasing in k.

(A2) For any integers k_1, k_2 that satisfies $k_1, k_2 \in \{k^0, k^0+1, \ldots, k_{\max}\}$, the likelihood ratio $p\big(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \widehat{\mathrm{C}}_{k_1}, \mathcal{M}_{k_1}\big)/p\big(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \widehat{\mathrm{C}}_{k_2}, \mathcal{M}_{k_2}\big)$ is bounded from above by constant in probability as $m \longrightarrow \infty$.

(A3) For any integer $k < k^0$ (if $k^0 > 1$), there exist positive constants δ_k such that $p\big(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \widehat{\mathrm{C}}_k, \mathcal{M}_k\big)/p\big(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \widehat{\mathrm{C}}_{k^0}, \mathcal{M}_{k^0}\big) < \exp(-\delta_k m)$ in probability as $m \longrightarrow \infty$.

(A4) For any integers $k_1, k_2 \in \mathcal{K}$, the prior on models satisfies that $p(\mathcal{M}_{k_1})/p(\mathcal{M}_{k_2})$ is bounded from above by constant.

In the above assumptions, t_k can be considered as the learning coefficient which is dependent on the data generating distribution, while s_k is related to the multiplicity of t_k; see [22, 23] for more details about t_k and s_k. Assumption (A1) is the asymptotic expansion (2.7) in [22]. The expansion is satisfied by the factor model with the sequence $\{t_k\}$ strictly increasing in k. This implies that for $k^0 \leq k_1 < k_2 \leq k_{\max}$, the second term on the right-hand side of Equation (9) penalizes the larger model $\mathcal{M}_{k_2}$ more than the smaller model $\mathcal{M}_{k_1}$. As a result, this together with Assumption (A2) implies that among all the models $\mathcal{M}_k$ with $k \geq k^0$, the posterior favors the most parsimonious model, which is the true model with k^0 nonzero common factors. Assumption (A3) is used to rule out those models with too few factors. Assumption (A4) is a mild assumption on the prior of models, which is satisfied trivially by the uniform prior.

Under Assumptions (A1)–(A4), Theorem 3.1 shows that as the amount of real-world data goes to infinity $m \longrightarrow \infty$, the posterior $p(\mathcal{M}_k \mid \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}))$ converges to the k^0-factor model. Thus, under the constraints specified by Θ_Λ, the posterior sample of loading matrix $\underline{\Lambda}$ converges to $\underline{\Lambda}^c$.

Theorem 3.1. *Suppose that Assumptions (A1)–(A4) hold. Then the posterior of* k *consistently estimates the true number of factors* k^0*, i.e. the posterior satisfies* $p\big(\mathcal{M}_{k^0} \mid \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})\big) \longrightarrow 1$ *in probability as* $m \longrightarrow \infty$.

The sampling procedure in Section 3.1.4 penalizes over- and under-parameterized factor model. Suppose that we choose the proposal distribution $g_k(\Lambda_k)$ to be the posterior distribution $p(\Lambda_k | k, \mathcal{Z}_m)$ and the uniform prior is used for the number of common factors $p(k)$. We can show that the part in the accept ratio in Equation (8), $p(\mathcal{Z}_m | k', \tilde{\Lambda}_{k'}) p(\tilde{\Lambda}_{k'} | k') p(k') g_k(\Lambda_k) / [p(\mathcal{Z}_m | k, \Lambda_k) p(\Lambda_k | k) p(k) g_{k'}(\tilde{\Lambda}_{k'})]$, equals to the marginal likelihood ratio $p(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_{k'})/p(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k)$. By applying Assumptions (A1)–(A2), for $k^0 < k \leq k_{\max}$, we have

$$p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right) / p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_{k^0}\right) \leq \exp[-(t_k - t_{k^0}) \log m/2]. \qquad (10)$$

By applying Assumption (A3), for $1 \leq k < k^0$, we have

$$p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right) / p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_{k^0}\right) \leq \exp(-\delta_0 m/2). \qquad (11)$$

See the proof of Theorem 3.1 in the appendix for the derivation of Equations (10)–(11). Thus, Step (1) of the sampling procedure in Section 3.1.4 tends to accept the moves to the k^0-factor model and the acceptance ratio goes to 1 as $m \longrightarrow \infty$. The empirical study in Section 5.1 demonstrates that the sampling procedure has good finite-sample performance on the model selection.

3.1.6 Asymptotic Convergence of Input Model Estimation. In this section, we show the asymptotic convergence of the estimated input model. Since the input joint distribution is characterized by the marginal distributions and a correlation matrix, Theorem 3.2 shows that under Conditions (B1) and (B2), as the amount of real-world data m goes to infinity, the estimated input joint distribution specified by $(\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_d, \tilde{\mathbf{C}})$ uniformly converges to the true input distribution specified by $(F_1^c, F_2^c, \ldots, F_d^c, \mathbf{C}^c)$. The proof for Theorem 3.2 is provided in the online Appendix.

(B1) The prior of correlation matrix, denoted by $p(\mathbf{C})$, has positive mass on any open neighborhood of $\mathbf{C}^c$ defined based on the Frobenius norm.

(B2) The true correlation matrix $\mathbf{C}^c$ has a factor decomposition in k^0 factors $(1 \leq k^0 \leq d)$.

(B3) The DPM posterior is consistent at each F_j with $j = 1, 2, \ldots, d$: if $\tilde{F}_j$ is drawn from the posterior $p(F_j|\mathcal{X}_{jm}^{(0)})$, then $\|\tilde{F}_j - F_j^c\|_\infty$ converges to zero in probability as $m \to \infty$ for all $j = 1, 2, \ldots, d$, where $\|\tilde{F}_j - F_j^c\|_\infty = \sup_{x \in \Re} \left|\tilde{F}_j(x) - F_j^c(x)\right|$.

Theorem 3.2. *Suppose that Conditions (B1), (B2) and (B3) hold, and $\tilde{F}_j$ is a posterior sample of marginal distribution from $p(F_j|\mathcal{X}_{jm}^{(0)})$ obtained by DPM for $j = 1, 2, \ldots, d$. Let $\tilde{\mathbf{C}}$ be a sample of correlation matrix drawn from $p(\mathbf{C}|\mathcal{X}_m^{(0)})$. The posterior sample $\tilde{\mathbf{F}}$ has the Gaussian copula representation*

$$\tilde{F}(x_1, x_2, \ldots, x_d) = \Phi_d\left(\Phi^{-1}\left[\tilde{F}_1(x_1)\right], \Phi^{-1}\left[\tilde{F}_2(x_2)\right], \ldots, \Phi^{-1}\left[\tilde{F}_d(x_d)\right]; \tilde{\mathbf{C}}\right).$$

Then almost surely under the true input model F^c, the posterior sample $\tilde{F}(x_1, x_2, \ldots, x_d)$ uniformly converges to $F^c(x_1, x_2, \ldots, x_d)$, i.e. $\|\tilde{F} - F^c\|_\infty \xrightarrow{p} 0$ as $m \to \infty$, where $\|\tilde{F} - F^c\|_\infty = \sup_{(x_1, x_2, \ldots, x_d) \in \Re^d} \left|\tilde{F}(x_1, x_2, \ldots, x_d) - F^c(x_1, x_2, \ldots, x_d)\right|$.

Assumption (B3) requires that the DPM posteriors of all marginal distributions to be consistently in the supremum norm of distribution functions, i.e. the Kolmogorov-Smirnov norm. Since the Kolmogorov-Smirnov norm is weaker than the total variation norm, (B3) readily holds for all DPM posteriors that are *strongly consistent.* The strong consistency of DPM has been well studied for the DPM of normals; see for example, [28, 29], etc.

3.2 Bayesian Inference for the Quantile Response

In Section 3.1, we provide a Bayesian approach to generate posterior samples of the input model, $\tilde{F}^{(b)}$ with $b = 1, 2, \ldots, B$, quantifying the input uncertainty. At each sample $\tilde{F}^{(b)}$, we run simulations to estimate the system performance characterized by a vector of quantiles. In this paper, we focus on the steady state behaviors. Since the distribution of simulation output $Y(\tilde{F}^{(b)})$ is unknown and it also depends on the input model $\tilde{F}^{(b)}$, the nonparametric Bayesian approach proposed in [24, 41] is used to quantify the simulation estimation uncertainty.

For a generic input model F which is a posterior sample of input model in our framework, the simulation outputs are $\mathbf{Y}(F) \equiv \{(Y_{r1}(F), Y_{r2}(F), \ldots, Y_{rL}(F)), r = 1, 2, \ldots, R\}$ with the number of replications R and the runlength L. We are interested in simultaneously estimating the quantiles of $Y(F)$ listed in the vector $\mathbf{q}(F) \equiv \left(q_1(F), q_2(F), \ldots, q_\gamma(F)\right)$ with probabilities $0 = p_0 < p_1 < p_2 < \ldots < p_\gamma < p_{\gamma+1} = 1$. The values of $q_0(F)$ and $q_{\gamma+1}(F)$ depend on the limits of the support for the simulation output distribution $G_{Y(F)}$. Since quantiles $\mathbf{q} = \left(q_1, q_2, \ldots, q_\gamma\right)$ divide all the outputs in $\mathbf{Y}(F)$ into $\gamma + 1$ groups, an approximate likelihood, denoted by $s(\cdot|\mathbf{q})$, follows a multinomial distribution [24, 37, 41]

$$s\left(\mathbf{Y}(F)|\,\mathbf{q}\right) = \binom{RL}{\mu_1, \mu_2, \ldots, \mu_{\gamma+1}} \prod_{\ell=1}^{\gamma+1} (p_\ell - p_{\ell-1})^{\mu_\ell} \tag{12}$$

where $\mu_\ell = \sum 1_{(q_{\ell-1}, q_\ell]}(\mathbf{Y}(F))$ denotes the number of entries in $\mathbf{Y}(F)$ located in the range $(q_{\ell-1}, q_\ell]$ for $\ell = 1, 2, \ldots, \gamma + 1$. This approximate likelihood is used in the posterior inference, by $p(\mathbf{q}|\mathbf{Y}(F)) \propto p(\mathbf{q})s(\mathbf{Y}(F)|\mathbf{q})$. The prior $p(\mathbf{q})$, such as a truncated normal distribution used in our empirical study, can guarantee the restriction $q_1 \leq q_2 \leq \ldots \leq q_\gamma$. Since the prior is not conjugate, MCMC with a Metropolis-Hasting step is used for posterior inference; see the procedure in [24].

At any fixed input model F, Theorem 3.3 shows the asymptotic consistency of the approximate posterior from (12) for each quantile $q_\ell(F)$ with $\ell = 1, 2, \ldots, \gamma$; see the proof of Proposition 1 in [40]. The convergence of the ratio $p(q|\mathbf{Y}(F))/p(q_\ell(F)|\mathbf{Y}(F))$ to zero for any $q \neq q_\ell(F)$ indicates that the approximate posterior from (12) asymptotically concentrates around the true system response quantiles $q_\ell(F)$, as the number of simulation outputs RL increases. Then, by applying the continuous mapping theorem, we have the consistency for the vector of quantiles $\mathbf{q}(F)$ and other risk measures that are continuous functions of quantiles.

(C1) For every posterior sample of input model F from the procedure in Section 3.1, the distribution of the simulation output $G_{Y(F)}$ is continuous at the quantiles $q_\ell(F)$ for all $\ell = 1, 2, \ldots, \gamma$.

Theorem 3.3. *(Lancaster and Jun [40] Proposition 1) Suppose that Condition (C1) holds. Then for any given input model F, if* $q \neq q_\ell(F)$ *for all* $\ell = 1, 2, \ldots, \gamma$ *and* $0 < G_{Y(F)}(q) < 1$*, the ratio* $p(q|\mathbf{Y}(F))/p(q_\ell(F)|\mathbf{Y}(F))$ *converges in probability to zero as* $RL \to \infty$*,.*

3.3 Procedure to Construct CrIs for the Risk Analysis

Built on our previous study [54], in this section, we provide the procedure in Algorithm 3 to construct the percentile CrIs for the quantile responses accounting for both input and simulation estimation uncertainty. It includes main steps as follows. Given a finite amount of real-world data, we generate posterior samples of input model, $\tilde{F}^{(b)}$ with $b = 1, 2, \ldots, B$, quantifying the input uncertainty in Step 1. Each sample $\tilde{F}^{(b)}$ is specified with marginals $(\tilde{F}_1^{(b)}, \tilde{F}_2^{(b)}, \ldots, \tilde{F}_d^{(b)})$ and the correlation matrix $\tilde{\mathbf{C}}^{(b)}$. Then, at each $\tilde{F}^{(b)}$, we run simulations with the runlength L and replications R, get the outputs $\mathbf{Y}^{(b)}$, and draw a sample $\tilde{q}_{\ell,b}$ from the posterior $p(q_\ell(\tilde{F}^{(b)})|\mathbf{Y}^{(b)}, \tilde{F}^{(b)})$ for $\ell = 1, 2, \ldots, \gamma$ quantifying the simulation estimation uncertainty in Steps 3 and 4. After that, we construct the percentile CIs for quantiles $(q_1, q_2, \ldots, q_\gamma)$ in Step 6. The similar procedure can be applied to other performance measures that are continuous functions of quantiles.

Theorem 3.4 shows the asymptotic consistency of the CrIs in (13). Under Conditions (B1), (B2) and (B3), Theorem 3.2 shows that the posterior distribution $p(F|\mathcal{X}_m^{(0)})$ converges to F^c as the amount of real-world data $m \to \infty$. According to Theorem 3.3, at any given posterior sample of input model $\tilde{F}^{(b)}$ for $b = 1, 2, \ldots, B$, the posterior distribution $p(q_\ell|\mathbf{Y}(\tilde{F}^{(b)}))$ converges to the true quantile $q_\ell(\tilde{F}^{(b)})$ as the simulation budget $RL \to \infty$ for $\ell = 1, 2, \ldots, \gamma$. Condition (D1) below assumes the continuity of quantile curve $q_\ell(F)$ around the true input model F^c with respect to the distance defined by the infinity norm, $\|F - F^c\|_\infty$. By applying the triangular inequality, we can show that the CrI in (13) shrinks to $q_\ell(F^c)$ as $RL \to \infty$ and $m \to \infty$. This convergence result is conditional on both the posterior samples $\tilde{F}^{(1)}, \tilde{F}^{(2)}, \ldots, \tilde{F}^{(B)}$ and the simulation outputs $\mathbf{Y}(\tilde{F}^{(1)}), \mathbf{Y}(\tilde{F}^{(2)}), \ldots, \mathbf{Y}(\tilde{F}^{(B)})$. The proof for Theorem 3.4 is in the online Appendix. Notice that this asymptotic consistency is fully driven by $RL, m \to \infty$ but does not require $B \to \infty$. For a vector of quantiles and risk measures that are continuous functions of quantiles, the consistency directly follows by applying the continuous mapping theorem. The empirical study in Section 5 demonstrates that our approach has good finite-sample performance.

(D1) For any $\epsilon > 0$, there exists $\delta > 0$ such that for any input model F, $\|F - F^c\|_\infty < \delta$ implies $|q_\ell(F) - q_\ell(F^c)| < \epsilon$ for $\ell = 1, 2, \ldots, \gamma$.

ALGORITHM 3: Construct CrIs for System Risk Performance

```
Input: The real-world data $\mathcal{X}_m^{(0)}$
Output: Percentile credible intervals for the quantile responses, CrI($q_\ell$) with $\ell = 1, 2, \ldots, \gamma$
Specify the priors for marginals and the correlation matrix. Apply Gibbs samplers described in Section 3.1 to
  generate $B$ posterior samples of input model $(\tilde{F}_1^{(b)}, \tilde{F}_2^{(b)}, \ldots, \tilde{F}_d^{(b)}, \tilde{\mathbf{C}}^{(b)})$ with $b = 1, 2, \ldots, B$.
for $b = 1, 2, \ldots, B$ do
    Generate $\mathbf{Z} \overset{i.i.d.}{\sim} \mathcal{N}_d(\mathbf{0}, \tilde{\mathbf{C}}^{(b)})$ and do transformation to obtain the input data $\mathbf{X}$ with
      $X_j = (\tilde{F}_j^{(b)})^{-1}\left[\Phi\left(Z_j\right)\right]$ for $j = 1, 2, \ldots, d$. Use these data to drive simulations and get the outputs
      $\mathbf{Y}^{(b)} = \mathbf{Y}(\tilde{F}^{(b)})$.
    Apply the inference in Section 3.2 and generate a posterior sample $\tilde{q}_{\ell,b} \sim p(q_\ell(\tilde{F}^{(b)}) | \mathbf{Y}^{(b)}, \tilde{F}^{(b)})$ for
      $\ell = 1, 2, \ldots, \gamma$.
end
Record the $(1-\alpha)100\%$ percentile CrIs quantifying the overall uncertainty for quantile $q_\ell$,
  $$\mathrm{CrI}(q_\ell) = [\tilde{q}_{\ell,(\lceil B\alpha/2 \rceil)}, \tilde{q}_{\ell,(\lceil B(1-\alpha/2) \rceil)}] \quad (13)$$
  where $\tilde{q}_{\ell,(b)}$ is the $b$th order statistic with $\tilde{q}_{\ell,(1)} \le \tilde{q}_{\ell,(2)} \le \cdots \le \tilde{q}_{\ell,(B)}$ for $\ell = 1, 2, \ldots, \gamma$.
```

Theorem 3.4. *Suppose that Conditions (B1), (B2), (B3), (C1) and (D1) hold, then conditional on* $\tilde{F}^{(1)}, \tilde{F}^{(2)}, \ldots, \tilde{F}^{(B)}$ *and* $\mathbf{Y}(\tilde{F}^{(1)}), \mathbf{Y}(\tilde{F}^{(2)}), \ldots, \mathbf{Y}(\tilde{F}^{(B)})$, *the CrI in Equation (13) asymptotically converges in probability to the true quantile response* $q_\ell(F^c)$ *for each* $\ell = 1, 2, \ldots, \gamma$, *as the simulation budget* $RL \longrightarrow \infty$ *and the amount of real-world data* $m \longrightarrow \infty$.

Following the similar procedure in [53], we can estimate the contributions from both input and simulation uncertainty. If the simulation uncertainty dominates the input uncertainty, additional simulation runs could further reduce the simulation estimation uncertainty. As we collect more simulation outputs, we can update our belief on quantiles by applying the Bayes' rule and stop the simulation when the contribution from the simulation estimation uncertainty reaches the desired level.

4 FACTOR STRUCTURE FOR DECISION MAKING

In this section, we further explore the factor structure of the input dependence to guide decision making for the risk analysis. First, we interpret the common factors to get insights of underlying generative processes for correlated inputs. Then, we propose a procedure to estimate their effects on the system performance, which could guide decision making to improve the system random behaviors. For example, the inventory management for a jet engine maintenance could involve ordering decisions for hundreds of components. The simultaneous breakdowns of different components can be induced by various lurking factors, e.g., the operating temperature and the contamination. Thus, based on the loading matrix, we could identify the active latent factors, and further find the dominant factors that have obvious impacts on the system performance so that one can develop strategies to improve the system performance.

To identify the latent factors, we first need to specify the number of common factors k. According to the literature on data analytics with factor model [1, 42], we take the mode of posterior $p(k|\mathcal{X}_m^{(0)})$, denoted by $\hat{k}$, as the true value. Theorem 3.1 shows that $\hat{k}$ asymptotically converges to k^0. The empirical study in Section 5 also demonstrates that the sampling procedure in Section 3.1.4 has a good finite-sample performance on determining k and the interpretation is robust to the model selection.

Then, given the $\hat{k}$-factor model, the posterior mean of $p(\underline{\Lambda}|\mathcal{X}_m^{(0)}, k = \hat{k})$, denoted by $\hat{\underline{\Lambda}}$, is used as a summary of the factor loadings for interpretation. The factor loadings represent the impact of the common factors on each component. From Equation (1), larger absolute values of $\hat{\underline{\lambda}}_{jh}$ and $\hat{\underline{\lambda}}_{j'h}$ imply stronger correlation between the jth and j'th components induced by the hth common factor, where $h = 1, 2, \ldots, \hat{k}$. In addition, $\hat{\underline{\lambda}}_{jh}^2 \in [0, 1]$ gives the proportion of variance of Z_j explained by the hth factor. *Thus, components with large loading $\hat{\underline{\lambda}}_{jh}^2$ can be used to identify the hth factor.*

After identifying $\hat{k}$ latent common factors, we propose a procedure to estimate the effect of each factor and also the effect of the whole correlation (all $\hat{k}$ factors) on the system quantile, say q. We first generate B_0 posterior samples of input model quantifying the input uncertainty in Step 1. Given finite real-world data and simulation budget, let Q_F, Q_0 and Q_{-h} represent random variables characterizing the quantile estimation uncertainty when the input model includes the whole correlation, no common factor (no input correlation) and all but the hth factor. By following the procedure in Section 3.3, we generate their posterior samples, denoted by $\tilde{q}_{F,b}, \tilde{q}_{0,b}$ and $\tilde{q}_{-h,b}$ for $b = 1, 2, \ldots, B_0$, in Steps 3, 4 and 6–7 correspondingly. Then, $\Delta Q_0 = Q_0 - Q_F$ characterizes the impact from the whole correlation and $\Delta Q_{-h} = Q_{-h} - Q_F$ characterizes the impact from the hth factor. We obtain their posterior samples, $\Delta\tilde{q}_{0,b} = \tilde{q}_{0,b} - \tilde{q}_{F,b}$ and $\Delta\tilde{q}_{-h,b} = \tilde{q}_{-h,b} - \tilde{q}_{F,b}$ for $b = 1, 2, \ldots, B_0$ and $h = 1, 2, \ldots, \hat{k}$ in Step 8. We further construct the percentile CrIs quantifying the estimation uncertainty of the effects from the whole correlation and the hth factor in Step 11. To reduce the estimation uncertainty for these effects, we run simulations in Steps 3, 4 and 7 with common random numbers. Similar procedure can be applied to a vector of quantiles and other performance measures that are functions of quantiles.

To study the effect of the hth factor, we set the hth column of loading matrix to be zero in Step 6, which requires that the order of underlying factors is fixed at posterior samples of loading matrix. However, when there exists any diagonal entry of loading matrix $\underline{\Lambda}$ close to 0, the identification constraint of positive diagonal entries is weakened, and the order of common factors may switch among the posterior samples. To control this issue, we first permute the factors for each posterior sample of loading matrix. Suppose that a factor could have an either large or small effect on input components, and each component is highly associated with a single factor. For each posterior sample of loading matrix, we can reorder the columns based on the first large entry in the factor loadings. Specifically, for any posterior sample $\tilde{\underline{\Lambda}}$, let $s_h = \text{argmin}_j\{|\tilde{\underline{\lambda}}_{jh}| > \Delta\lambda\}$ denote the first large entry corresponding to factor h, where $\Delta\lambda$ represents a threshold to distinguish large factor loadings with small ones. In the empirical study, we set $\Delta\lambda = 0.5$. We permute columns of $\tilde{\underline{\Lambda}}$ such that $s_1 < \cdots < s_{\hat{k}}$. The permutation of factors can make the posterior samples of loading matrix having the same order of factors as long as the estimation error is not too large. Further, according to [42], to avoid the sign-switching identification issue, we switch the sign of each factor in $\tilde{\underline{\Lambda}}$ so that $\tilde{\underline{\lambda}}_{s_h,h} > 0$ for $h = 1, 2, \ldots, \hat{h}$.

5 EMPIRICAL STUDY

In this section, we first study the input model estimation for marginal distributions and the correlation matrix in Section 5.1. Then, we use a portfolio management example to demonstrate the impact of exploring the factor structure in the input correlation on the system risk performance estimation in Section 5.2. After that, we interpret the latent common factors and estimate their effects on the system quantile performance in Section 5.3.

ALGORITHM 4: Estimate the Effects of Factors on System Risk Performance

```
Input: The real-world data $\mathcal{X}_m^{(0)}$
Output: Percentile credible intervals for the effects from common factors on the quantile response: $\mathrm{CrI}(\Delta Q_0)$
  and $\mathrm{CrI}(\Delta Q_{-h})$ for $h = 1, 2, \ldots, \hat{k}$
Fix the number of common factors to be $\hat{k}$. Implement the procedure in Section 3.1.3 to generate posterior
  samples of the input model, $\tilde{F}^{(b)} \equiv (\tilde{F}_1^{(b)}, \tilde{F}_2^{(b)}, \ldots, \tilde{F}_d^{(b)}, \tilde{\mathbf{C}}^{(b)})$ and $\underline{\tilde{\Lambda}}^{(b)}$ for $b = 1, 2, \ldots, B_0$.
for $b = 1, 2, \ldots, B_0$ do
    Run simulations at input models $\tilde{F}^{(b)}$. Then, by following the procedure in Section 3.3, obtain posterior
      samples $\tilde{q}_{F,b}$.
    Run simulations at input models without dependence $\tilde{F}_0^{(b)} \equiv (\tilde{F}_1^{(b)}, \tilde{F}_2^{(b)}, \ldots, \tilde{F}_d^{(b)}, \mathbf{I}_d)$. Then, by following
      the procedure in Section 3.3, obtain posterior samples $\tilde{q}_{0,b}$.
    for $h = 1, 2, \ldots, \hat{k}$ do
        Suppose the $h$th factor is removed. Obtain posterior samples of the loading matrix without the $h$th
          common factor, denoted by $\underline{\tilde{\Lambda}}_{-h}^{(b)}$, through setting the $h$th column of $\underline{\tilde{\Lambda}}^{(b)}$ to be zero. Then, by
          applying Equation (1), obtain samples $\tilde{\mathbf{C}}_{-h}^{(b)}$.
        Run simulations at input models $\tilde{F}_{-h}^{(b)} \equiv (\tilde{F}_1^{(b)}, \tilde{F}_2^{(b)}, \ldots, \tilde{F}_d^{(b)}, \tilde{\mathbf{C}}_{-h}^{(b)})$. Then, by following the procedure
          in Section 3.3, obtain posterior samples $\tilde{q}_{-h,b}$.
        Obtain the posterior samples, $\Delta\tilde{q}_{0,b} = \tilde{q}_{0,b} - \tilde{q}_{F,b}$ and $\Delta\tilde{q}_{-h,b} = \tilde{q}_{-h,b} - \tilde{q}_{F,b}$ for $b = 1, 2, \ldots, B_0$
          and $h = 1, 2, \ldots, \hat{k}$.
    end
end
Record the $(1-\alpha)100\%$ percentile CrIs quantifying the overall uncertainty for $\Delta Q_0$ and $\Delta Q_{-h}$ with
  $h = 1, 2, \ldots, \hat{k}$
      $\mathrm{CrI}(\Delta Q_0) = [\Delta\tilde{q}_{0,(\lceil B_0\alpha/2\rceil)}, \Delta\tilde{q}_{0,(\lceil 1-B_0\alpha/2\rceil)}]$
      $\mathrm{CrI}(\Delta Q_{-h}) = [\Delta\tilde{q}_{-h,(\lceil B_0\alpha/2\rceil)}, \Delta\tilde{q}_{-h,(\lceil 1-B_0\alpha/2\rceil)}]$
  where $\Delta\tilde{q}_{0,(b)}$ and $\Delta\tilde{q}_{-h,(b)}$ denote the $b$th order statistics with $\Delta\tilde{q}_{0,(1)} \leq \Delta\tilde{q}_{0,(2)} \leq \cdots \leq \Delta\tilde{q}_{0,(B_0)}$ and
  $\Delta\tilde{q}_{-h,(1)} \leq \Delta\tilde{q}_{-h,(2)} \leq \cdots \leq \Delta\tilde{q}_{-h,(B_0)}$.
```

5.1 Input Model Estimation

Our study in [53] demonstrates that our DPM based Bayesian nonparametric approach has better finite-sample behaviors compared to many existing approaches on input modeling, such as finite mixture and empirical distribution. In this section, we first use two test examples to compare the performance of DPM with that of a flexible parametric approach, JTS, in modeling marginal inputs. Since we consider each marginal distribution separately, for notation simplification, we drop the subscript for input component temporarily. The first example is a Gumbel distribution, Gumbel(μ, β), with the location parameter $\mu = 0.1$ and the scale parameter $\beta = 0.2$. It has heavy-tails and large skewness. The second example is a skewed t distribution, skewed-$t(\xi, \omega, \alpha, df)$, with the location parameter $\xi = 0$, the scale parameter $\omega = 0.5$, the slant parameter $\alpha = 8$ controlling the skewness, and the degree of freedom $df = 10$; see [3, 4] for the detailed information about skewed-t distribution.

To compare the goodness of fit obtained from DPM and JTS, we record the Kolmogorov-Smirnov (KS) distance, defined by $D_m \equiv \sup_{x\in\Re} |F^c(x) - \hat{F}_m(x|\mathcal{X}_m^{(0)})|$, and Anderson-Darling (AD) distance, defined by $A_m^2 \equiv m\int |F^c(x) - \hat{F}_m(x|\mathcal{X}_m^{(0)})|^2 w(x) dF^c(x)$, measuring the difference between the underlying true distribution $F^c(\cdot)$ and the distribution estimated by m real-world data $\mathcal{X}_m^{(0)}$, denoted by $\hat{F}_m(\cdot)$, where the weight function is $w(x) = 1/(F^c(x)(1-F^c(x)))$. Since the AD distance places more weight on the tails of F^c, it can detect the discrepancies at the tails better. According to [26], the posterior predictive distribution with density, defined by $\hat{f}_m(x|\mathcal{X}_m^{(0)}) = \int f(x|F)dP(F|\mathcal{X}_m^{(0)})$, is used to assess the fit of input model. Since the true input from both examples has unbounded support, we use DPM with Gaussian kernel. Table 1 reports the mean and standard deviation (SD) (SD is inside the parenthesis) of KS and AD distances obtained by JTS and DPM when the sample size is $m = 30, 50, 200$. The results are estimated based on 100 macro-replications. Table 1 indicates that DPM has better fitting performance and the advantage is more obvious in the second example with multiple modes.

Table 1. The mean and SD of KS and AD distances obtained by using JTS and DPM

Example 1: Gumbel		$m = 30$	$m = 50$	$m = 200$
JTS	D_m	0.114 (0.043)	0.083 (0.034)	0.046 (0.017)
	A_m	12.429 (4.214)	9.250 (3.813)	4.695 (1.702)
DPM	D_m	0.095 (0.042)	0.076 (0.031)	0.042 (0.016)
	A_m	10.594 (4.375)	8.134 (3.206)	4.319 (1.658)
Example 2: Skewed-t		$m = 30$	$m = 50$	$m = 200$
JTS	D_m	0.166 (0.019)	0.114 (0.012)	0.081 (0.009)
	A_m	13.712 (1.780)	10.839 (1.243)	6.745 (0.584)
DPM	D_m	0.071 (0.021)	0.052 (0.017)	0.028 (0.007)
	A_m	7.216 (2.518)	5.330 (1.958)	2.806 (0.452)

Then, to study the impact of exploring factor structure on the inference of input correlation, we compare the mean and SD of the estimation error of correlation matrix, defined as Error(**C**) = E[$\|\tilde{\mathbf{C}} - \mathbf{C}^c\| |\mathcal{X}_m^{(0)}$], obtained by using GCF and GC models with and without exploring the factor structure in the underlying input correlation, where $\|\cdot\|$ denotes the Frobenius norm. We also study the finite-sample behavior of the estimated number of common factors.

In the test example, the marginal distributions are Gumble with different locations but scale equal to 0.1, the correlation matrix is generated by a factor model with $k^0 = 1, 3$ when $d = 10, 30$ and $k^0 = 3, 10$ when $d = 100$; see the appendix for the true marginal parameters and correlation matrices. To study the robust behavior of our approach, we set $m = 30, 50, 200$, and record the mean and SD (inside the parenthesis) of Error(**C**) in Table 2. The results are estimated based on 100 macro-replications. In each macro-replication, we first generate m real-world data $\mathcal{X}_m^{(0)}$ by using F^c to mimic the data collection. Then, we generate $B = 1000$ posterior samples of $\tilde{\mathbf{C}}$ by using the sampling procedure in Sections 3.1.3 and 3.1.4, and calculate the posterior mean of $\|\tilde{\mathbf{C}} - \mathbf{C}^c\|$ to obtain Error(**C**). Since the number of factors are usually far less than the dimension [42], in our empirical study, we set $k_{\max} = 5$ when $d = 10, 30$, and set $k_{\max} = 20$ when $d = 100$. We discuss the impact of choice for $k_{\max}$ later. In the preliminary MCMC analysis, for each $k \in \{1, 2, \ldots, k_{\max}\}$, we run 5000 iterations after 2000 burn-in iterations, and record a posterior sample for every 10 iterations to avoid serial dependence in the MCMC. The sample mean and variance of these posterior samples are used to determine the proposal distributions for the RJMCMC sampling procedure. The uniform prior is used for the number of factors k. For the loading matrix, non-informative priors are used, $\psi_{jh} \sim$ Inverse-Gamma$(\alpha_0/2, \beta_0/2)$ with $\alpha_0 = 1$ and $\beta_0 = 1$. We compare the performance of GCF with the GC; see the implementation of GC in [33].

Table 2. The mean and SD of Error($\mathbf{C}$) obtained by using GCF and GC.

	m	GC: Error($\mathbf{C}$)	GCF: Error($\mathbf{C}$)
$d = 10, k^0 = 1$	30	1.182 (0.257)	0.880 (0.216)
	50	0.813 (0.166)	0.735 (0.138)
	200	0.540 (0.073)	0.542 (0.085)
$d = 10, k^0 = 3$	30	2.241 (0.260)	1.794 (0.225)
	50	1.613 (0.238)	1.332 (0.185)
	200	0.827 (0.115)	0.824 (0.121)
$d = 30, k^0 = 1$	30	NA	3.894 (0.463)
	50	3.150 (0.477)	2.286 (0.683)
	200	1.194 (0.153)	1.189 (0.176)
$d = 30, k^0 = 3$	30	NA	6.538 (1.257)
	50	4.963 (0.782)	3.920 (0.709)
	200	2.415 (0.368)	2.362 (0.355)
$d = 100, k^0 = 3$	30	NA	11.485 (2.089)
	50	NA	8.623 (1.704)
	200	6.437 (1.278)	5.740 (1.136)
$d = 100, k^0 = 10$	30	NA	18.264 (3.820)
	50	NA	13.672 (2.749)
	200	10.531 (2.260)	8.652 (1.815)

Table 2 demonstrates that the mean and SD of estimation error of input correlation matrix obtained by using GCF is much smaller than that of GC when the amount of real-world data m is relatively small. Since the number of parameters in the correlation matrix increases dramatically as d increases, GC does not work when $d = 30, m = 30$ and $d = 100, m = 30, 50$, indicated by "NA" in Table 2. As m increases, the estimation errors from GC and GCF decrease and become close to each other.

Since the factor model could simplify the input model, it reduces the estimation error of the correlation matrix and also takes less computational time. The empirical study in this section is ran on one node of the DRP cluster with two eight-core 2.6 GHz Intel Xeon E5-2650 processors and 128GB of system memory. We select the case with $d = 30$ and $k^0 = 3$ to compare the running time of GCF and GC. When $m = 50$, the average running time of GC is 76.78 (0.092) seconds, while that of GCF is 51.63 (0.088). When $m = 200$, the average running time of GC is 347.16 (1.540) while that of GCF is 107.31 (0.292). *Therefore, by exploring the factor structure in the input correlation, GCF requires less computational effort to achieve better estimation accuracy. These advantages become more obvious as the dimension of input model increases.*

Theorem 3.1 shows that the estimated number of factors asymptotically converges to k^0. To study the finite-sample behavior, Table 3 records the average frequency of the number of factors with 1000 posterior samples. The results are estimated based on 100 macro-replications. We can observe that the estimated number of factor shows convergence to k^0 as m increases. *Even as m is relatively small, the posterior mode $\hat{k}$ tends to equal to the true value k^0.*

The choice of $k_{\max}$ typically depends on the expert's opinion. Here, we also study the effect of selecting different $k_{\max}$ on the estimation accuracy for the correlation matrix and the posterior mode of k. We use the above example with $d = 30, k^0 = 3$, and set $k_{\max} = 5, 10, 15$ respectively. In Table 4, the results of Error($\mathbf{C}$) and the frequency of k are estimated based on 100 macro-replications.

Table 3. The average frequency of estimated number of factors in 1000 posterior samples.

	m	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$d = 10, k^0 = 1$	30	662.3	171.5	96.2	46.8	23.2
	50	794.7	113.6	72.8	16.5	2.4
	200	918.2	52.3	19.0	10.5	0
$d = 10, k^0 = 3$	30	2.5	26.3	747.8	140.0	83.4
	50	0.8	8.7	851.6	125.5	13.4
	200	0	1.1	908.3	87.0	3.6
$d = 30, k^0 = 1$	30	562.4	243.7	118.2	52.5	23.2
	50	726.0	173.4	62.6	30.1	7.9
	200	880.5	69.3	34.1	13.8	2.3
$d = 30, k^0 = 3$	30	52.6	103.4	528.7	202.5	112.8
	50	28.7	96.2	650.9	147.3	76.9
	200	13.2	48.5	814.8	89.6	33.9

	m	$k < 3$	$k = 3$	$k = 4$	$k = 5$	$k > 5$
$d = 100, k^0 = 3$	30	14.1	378.5	265.2	167.3	174.9
	50	5.6	449	291.5	175.6	78.3
	200	1.4	511.8	321.4	146.9	18.5

	m	$k < 10$	$k = 10$	$k = 11$	$k = 12$	$k > 12$
$d = 100, k^0 = 10$	30	63.8	223.6	262.1	209.5	241.0
	50	25.0	273.7	279.6	223.5	198.2
	200	10.8	328.6	306.9	190.7	163.0

The choice of $k_{\max}$ has little effect on the estimation error for the correlation matrix, Error(**C**). Larger $k_{\max}$ could overestimate k. However, the posterior mode $\hat{k}$ tends to equal to the true value k^0, especially when m increases.

Table 4. The estimation error Error(**C**) and the average frequency of k estimate under different $k_{\max}$ when $d = 30$ and $k^0 = 3$

	$k_{\max}$	Error(**C**)	$k < 3$	$k = 3$	$k = 4$	$k = 5$	$k > 5$
$m = 30$	5	6.510 (1.162)	142.4	539.6	223.8	94.2	0
	10	6.858 (1.093)	57.2	452.7	286	123.5	80.6
	15	6.376 (0.788)	16.4	407.6	311.9	144.3	119.8
$m = 50$	5	4.149 (0.870)	117.2	664.7	152.1	66	0
	10	4.275 (0.803)	40.3	552.8	229.5	102.9	74.5
	15	4.227 (0.748)	6.1	426.2	256.6	146.7	164.4
$m = 200$	5	2.437 (0.503)	58.5	806.5	85.8	49.2	0
	10	2.425 (0.488)	15.8	728.9	126.5	67.2	61.6
	15	2.430 (0.409)	1.3	683.6	148.3	87.7	79.1

5.2 System Risk Performance Estimation

In this section, we use a Portfolio Management (PM) example to study the finite-sample performance of our Bayesian framework. An investigator buys $d = 10$ stocks with the return rates, denoted by

$\mathbf{X} = (X_1, X_2, \ldots, X_d)$. Denote the units of stocks held by the investigator by $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$. We want to estimate the quantiles of overall return rate $Y = \sum_{j=1}^{d} \theta_j X_j$.

The underlying input model is a Gaussian copula with Gumbel marginal distributions and the correlation matrix $\mathbf{C}^c$ having a factor structure with $k^0 = 3$. Suppose that the correlation between the components of $\mathbf{X}$ is induced by various industry indexes. The first three components are stocks return of automotive companies, the next four components are stocks return of health care companies, and the remaining three are stocks return of semiconductor companies. The true loading matrix is

$$\underline{\Lambda}^{c\top} = \begin{bmatrix} \mathbf{0.893} & \mathbf{0.870} & \mathbf{0.863} & 0.231 & 0.218 & 0.196 & 0.218 & 0.167 & 0.201 & 0.226 \\ 0.000 & 0.196 & 0.188 & \mathbf{0.835} & \mathbf{0.861} & \mathbf{0.873} & \mathbf{0.860} & 0.184 & 0.193 & 0.240 \\ 0.000 & 0.000 & 0.243 & 0.231 & 0.202 & 0.180 & 0.175 & \mathbf{0.878} & \mathbf{0.864} & \mathbf{0.828} \end{bmatrix}. \quad (14)$$

By applying Equation (1), we can calculate the underlying correlation matrix $\mathbf{C}^c$ whose elements give the true correlations between different stock returns. The correlations within each industry are around 0.8 while those across different industries are about 0.3. The unknown marginals are Gumbel distributions with location parameters equal to $0.245, -0.205, 0.095, 0.045, -0.155, 0.095, 0.195, 0.045, 0.095, -0.155$, and all scale parameters equal to 0.1.

The portfolio is defined by unknown parameters $\boldsymbol{\theta}^c = (4, 2, 3, 2, 3, 2, 4, 2, 2, 3)$. We run a side experiment to estimate the true quantiles of the portfolio return, q_1^c and q_2^c with probabilities $p_1 = 5\%$ and $p_2 = 10\%$, by using $n = 10^7$ samples of stock returns. The estimated quantiles are $q_1^c = -0.6228$ and $q_2^c = -0.0293$ with the 95% confidence intervals (CIs) equal to $[-0.6253, -0.6212]$ and $[-0.0311, -0.0276]$. The $(1-\alpha)100\%$ CI for the pth quantile is obtained by using the ℓ-th and u-th order statistics, where $\ell = np - z_{1-\frac{\alpha}{2}}\sqrt{np(1-p)}$ and $u = np + z_{1-\frac{\alpha}{2}}\sqrt{np(1-p)}$; see [20].

To study the performance of our approach, suppose that we do not know the true input model and the system response. We use the DPM with Gaussian kernel to model marginal distributions and use GCF to model the correlated input. The unknown input model is estimated by m "real-world data" generated from F^c. After that, we run simulations, estimate the quantiles of the overall return rate and construct CrIs for the quantiles. To further study the robustness of our approach, we set the size of real-world data $m = 30, 50, 200$ and the run length $L = 100, 1000$. Let $B = 1000$. Then, we perform Bayesian inference for the quantiles by using 500 MCMC iterations after 1000 burn-in iterations.

Tables 5 shows the mean and SD (in the parenthesis) of the quantile estimation error, defined by $\text{Error}(\bar{q}_\ell) = |\bar{q}_\ell - q_\ell^c|$ with $\bar{q}_\ell = \text{E}[q_\ell | \mathcal{X}_m^{(0)}, \text{Y}]$ for $\ell = 1, 2$, and the width of 95% CrI in Equation (13) obtained by using GC and GCF for the input model, where $\mathcal{X}_m^{(0)}$ and Y represent the real-world input data and simulation outputs. The results in Tables 5 are based on 100 macro-replications. As m and L increase, the system response quantile estimation becomes more accurate and less variable. Compared to GC, the advantage of GCF is more obvious as the sample size m becomes smaller.

5.3 Estimating the Effects of Factors

In this section, we interpret the underlying common factors and then estimate their effects on the quantile response. We use the PM example described in Section 5.2 with $d = 10$, $k^0 = 3$, $m = 50$ and $L = 1000$ to demonstrate the representative performance of our approach given the real-world data $\mathcal{X}_m^{(0)}$ and the simulation outputs Y.

According to Table 3, when $d = 10$, $k^0 = 3$ and $m = 50$, we have the posterior mode of the number of common factors $\hat{k} = 3$. However, to study the robustness of the factor interpretation over the estimation error of the number of factors, we analyze the models with $k = 3, 4$, which have

Table 5. The mean and SD of the quantile estimation error Error($\bar{q}_\ell$) and the CrI width $|\text{CrI}(q_\ell)|$ with $\ell = 1, 2$ obtained by using GC and GCF.

$m = 30, L = 100$	Error($\bar{q}_1$)	Error($\bar{q}_2$)	$\|\text{CrI}(q_1)\|$	$\|\text{CrI}(q_2)\|$
GCF	0.278 (0.243)	0.221 (0.186)	1.818 (0.233)	1.507 (0.184)
GC	0.396 (0.294)	0.320 (0.255)	1.823 (0.229)	1.514 (0.180)
$m = 30, L = 1000$	Error($\bar{q}_1$)	Error($\bar{q}_2$)	$\|\text{CrI}(q_1)\|$	$\|\text{CrI}(q_2)\|$
GCF	0.252 (0.217)	0.204 (0.193)	1.225 (0.117)	1.016 (0.088)
GC	0.367 (0.265)	0.283 (0.212)	1.324 (0.132)	1.089 (0.101)
$m = 50, L = 100$	Error($\bar{q}_1$)	Error($\bar{q}_2$)	$\|\text{CrI}(q_1)\|$	$\|\text{CrI}(q_2)\|$
GCF	0.237 (0.203)	0.194 (0.168)	1.277 (0.125)	0.983 (0.084)
GC	0.318 (0.239)	0.265 (0.220)	1.385 (0.140)	1.026 (0.097)
$m = 50, L = 1000$	Error($\bar{q}_1$)	Error($\bar{q}_2$)	$\|\text{CrI}(q_1)\|$	$\|\text{CrI}(q_2)\|$
GCF	0.208 (0.179)	0.170 (0.114)	0.942 (0.103)	0.788 (0.072)
GC	0.278 (0.203)	0.241 (0.135)	1.004 (0.090)	0.820 (0.075)
$m = 200, L = 100$	Error($\bar{q}_1$)	Error($\bar{q}_2$)	$\|\text{CrI}(q_1)\|$	$\|\text{CrI}(q_2)\|$
GCF	0.132 (0.087)	0.105 (0.066)	1.036 (0.093)	0.904 (0.081)
GC	0.137 (0.082)	0.108 (0.064)	1.075 (0.094)	0.933 (0.086)
$m = 200, L = 1000$	Error($\bar{q}_1$)	Error($\bar{q}_2$)	$\|\text{CrI}(q_1)\|$	$\|\text{CrI}(q_2)\|$
GCF	0.106 (0.069)	0.080 (0.053)	0.784 (0.062)	0.690 (0.053)
GC	0.105 (0.062)	0.078 (0.048)	0.776 (0.059)	0.682 (0.050)

relatively high selection probabilities. We use the posterior mean of the factor loading $\hat{\underline{\Lambda}}$ as the summary for interpretation. To avoid order-switching and sign-switching among posterior samples of the loading matrix, we reorder and transform $\tilde{\underline{\Lambda}}$ according to the description in Section 4, and set the threshold $\Delta\lambda = 0.5$. We first study the loading matrix for the model with $k = 3$ factors, denoted by

$$\hat{\underline{\Lambda}}_3^\top = \begin{bmatrix} \mathbf{0.880} & \mathbf{0.882} & \mathbf{0.862} & 0.311 & 0.269 & 0.225 & 0.232 & 0.255 & 0.323 & 0.237 \\ 0 & 0.048 & -0.046 & \mathbf{0.795} & \mathbf{0.814} & \mathbf{0.821} & \mathbf{0.780} & -0.026 & -0.076 & -0.015 \\ 0 & 0 & 0.048 & 0.285 & 0.251 & 0.252 & 0.259 & \mathbf{0.837} & \mathbf{0.802} & \mathbf{0.790} \end{bmatrix}.$$

The factor structure of $\hat{\underline{\Lambda}}_3$ is quite similar to that of $\underline{\Lambda}^c$ in Equation (14). It reveals the strong correlations among three groups of components: Components 1–3, Components 4-7, and Components 8-10 are associated with three underlying factors. This information could be used to identify the underlying factors. Then, we study the posterior mean of loading matrix for the model with $k = 4$ factors, denoted by

$$\hat{\underline{\Lambda}}_4^\top = \begin{bmatrix} \mathbf{0.872} & \mathbf{0.825} & \mathbf{0.804} & 0.165 & 0.175 & 0.214 & 0.197 & 0.107 & 0.153 & 0.253 \\ 0 & 0.274 & 0.183 & \mathbf{0.790} & \mathbf{0.843} & \mathbf{0.795} & \mathbf{0.859} & -0.194 & -0.231 & 0.031 \\ 0 & 0 & 0.235 & 0.331 & 0.301 & 0.286 & 0.170 & \mathbf{0.836} & \mathbf{0.766} & \mathbf{0.815} \\ 0 & 0 & 0 & 0.268 & 0.218 & 0.304 & 0.103 & -0.341 & -0.181 & -0.309 \end{bmatrix}.$$

Since all elements in the last column of $\hat{\underline{\Lambda}}_4$ are close to zero, the last factor is redundant. By removing this factor, we obtain the similar interpretation with that from $\hat{\underline{\Lambda}}_3$. *Thus, the interpretation of underlying factors is robust to the estimation uncertainty of the loading matrix and the number of factors.*

After identifying the underlying common factors, we further estimate their effects on the system quantiles for two portfolio investment strategies:

(1) Portfolio 1 with $\boldsymbol{\theta}^c = (4, 2, 3, 2, 3, 2, 4, 2, 2, 3)$,
(2) Portfolio 2 with $\boldsymbol{\theta}^c = (4, 2, 3, \underline{20, 30, 20, 40,} 2, 2, 3)$.

In Portfolio 1, we invest almost evenly in each stock, and in Portfolio 2, we invest much more in stocks 4-7 associated to health care companies; see Section 5.2.

By following the procedure in Section 4, we record the results of the effects of underlying factors in Table 6. For Portfolio 1, all the CrIs do not cover 0 and have positive lower bound. Ignoring any factors leads to overestimation on the quantile response. We also note that ignoring all the factors leads to much larger estimation bias, and factor 2 has the largest impact on the quantile response. In Portfolio 2, only the factor 2 has a significant impact on the quantile response, since the portfolio invests much more on the stocks associated with factor 2 than the others. In such situations, ignoring the other factors does not bring significant estimation bias. Notice that since the CrI accounts for both input and simulation estimation uncertainty, as m and L increase, the power to detect the effects of underlying factors increases.

Table 6. The CrIs quantifying the effects of underlying factors on the 5% and 10% quantiles q_1 and q_2 for Portfolios 1 and 2.

	Portfolio 1		Portfolio 2	
	$q = q_1$	$q = q_2$	$q = q_1$	$q = q_2$
CrI(ΔQ_0)	[1.405, 1.933]	[1.097, 1.492]	[6.228, 8.917]	[4.852, 7.108]
CrI(ΔQ_{-1})	[0.094, 0.651]	[0.064, 0.529]	[−1.239, 1.761]	[−1.056, 1.465]
CrI(ΔQ_{-2})	[0.364.0.919]	[0.273, 0.716]	[4.843, 7.641]	[3.801, 6.071]
CrI(ΔQ_{-3})	[0.043, 0.603]	[0.031, 0.492]	[−1.174, 1.824]	[−1.045, 1.562]

6 CONCLUSION

In this paper, we propose a flexible multivariate input model characterized by marginal distributions and a correlation matrix. Without strong prior information on the families of marginal distributions, a Bayesian nonparametric approach is used to capture the important properties in each marginal, including multi-modality and skewness. Since the input correlation could be induced by latent common factors in many situations, a factor model is proposed to efficiently explain the input correlation. Then, given finite real-world data and simulation resource, a Bayesian framework is developed to deliver a CrI quantifying the overall uncertainty of system risk performance estimates. Our approach can improve both computational and statistical efficiency. It allows us to interpret the underlying factors and provide insights of input correlation, especially for large-scale stochastic systems. We further propose a procedure to estimate the effects of factors on the system risk behaviors characterized by a vector of quantiles.

ACKNOWLEDGMENTS

The authors acknowledge helpful advice from Barry L. Nelson and Thomas R. Willemain. They thank the associate editor, three anonymous referees for helpful comments and corrections.

REFERENCES

[1] O. Aguilar and M. West. 2000. Bayesian Dynamic Factor Models and Portfolio Allocation. *Journal of Business & Economic Statistics* 18 (2000), 338–357.

[2] A. Akcay, B. Biller, and S. Tayur. 2011. Improved Inventory Targets in the Presence of Limited Historical demand data. *Manufacturing and Service Operations Management* 13, 3 (2011), 297–309.

[3] Adelchi Azzalini and Antonella Capitanio. 2003. Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-Distribution. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 65, 2 (2003), 367–389. http://www.jstor.org/stable/3647510

[4] A. Azzalini, A. with the collaboration of Capitanio. 2014. *The Skew-normal and Related Families*. Cambridge University Press.

[5] R. R. Barton. 2007. Presenting A More Complete Characterization of Uncertainty: Can It Be Done?. In *Proceedings of the 2007 INFORMS Simulation Society Research Workshop*. INFORMS Simulation Society, Fontainebleau.

[6] R. R. Barton. 2012. Tutorial: Input Uncertainty in Output Analysis. In *Proceedings of the 2012 Winter Simulation Conference*, C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher (Ed.). IEEE, 67–78.

[7] R. R. Barton, B. L. Nelson, and W. Xie. 2014. Quantifying input uncertainty via simulation confidence intervals. *Informs Journal on Computing* 26 (2014), 74–87.

[8] R. R. Barton and L. W. Schruben. 1993. Uniform And Bootstrap Resampling of Input Distributions. In *Proceedings of the 1993 Winter Simulation Conference*. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 503–508.

[9] B. Biller, A. Akcay, C. Corlu, and S. Tayur. 2014. A Simulation-based Support Tool for Data-driven Decision Making: Operational Testing for Dependence Modeling. In *Proceedings of the 2014 Winter Simulation Conference (WSC '14)*. 899–909.

[10] B. Biller and C. G. Corlu. 2011. Accounting for Parameter Uncertainty in Large-Scale Stochastic Simulations with Correlated Inputs. *Operations Research* 59 (2011), 661–673.

[11] B. Biller and C. G. Corlu. 2012. Copula-based Multivariate Input Modeling. *Surveys in Operations Research and Management Science* 17 (2012), 69–84.

[12] B. Biller and S. Ghosh. 2006. Multivariate Input Processes. In *Handbooks in Operations Research and Management Science: Simulation*, S. Henderson and B. L. Nelson (Eds.). Elsevier, Chapter 5.

[13] B. Biller and C. Gunes. 2010. Introduction to Simulation Input Modeling. In *Proceedings of the 2010 Winter Simulation Conference*, B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yucesan (Eds.). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

[14] B. Biller and B. L. Nelson. 2009. Modeling and Generating Multivariate Time-series Input Processes Using a Vector Autoregressive Technique. *ACM Transactions on Modeling and Computer Simulation* 13, 3 (2009), 211–237.

[15] C. M. Bishop. 2006. *Pattern Recognition and Machine Learning*. Springer, New York.

[16] M.C. Cario and B. L. Nelson. 1997. *Modeling and Generating Random Vectors with Arbitrary Marginal Distributions and Correlation Matrix*. Technical Report. Department of Industrial Engineering and Management Sciences, Northwestern University.

[17] R. C. H. Cheng and W. Holland. 1997. Sensitivity of Computer Simulation Experiments to Errors in Input Data. *Journal of Statistical Computation and Simulation* 57 (1997), 219–241.

[18] R. C. H. Cheng and W. Holland. 2004. Calculation of Confidence Intervals for Simulation Output. *ACM Transactions on Modeling and Computer Simulation* 14 (2004), 344–362.

[19] S. E. Chick. 2001. Input distribution selection for simulation experiments: Accounting for input uncertainty. *Operations Research* 49 (2001), 744–758.

[20] W. J. Conover. 1980. *Practical Nonparametric Statistics*. John Wiley and Sons, New York.

[21] M. Drton. 2009. Likelihood ratio tests and singularities. *The Annals of Statistics* 37 (2009), 979–1012.

[22] M. Drton and M. Plummer. 2013. A Bayesian information criterion for singular models. *arXiv preprint* (2013). arXiv: 1309.0911v3.

[23] M. Drton, B. Sturmfels, and S. Sullivant. 2007. Algebraic factor analysis: tetrads, pentads and beyond. *Probability Theory and Related Fields* 138 (2007), 463–493.

[24] D. B. Dunson and J. A. Taylor. 2005. Approximate Bayesian Inference for Quantiles. *Journal of Nonparametric Statistics* 17 (2005), 385–400.

[25] M. D. Escobar and M. West. 1995. Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association* 90, 430 (1995), 577–588.

[26] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. *Bayesian Data Analysis* (2nd ed.). Taylor and Francis Group, LLC, New York.

[27] J. F. Geweke and K. J. Singleton. 1980. Interpreting the Likelihood Ratio Statistic in Factor Models when Sample Size is Small. *Journal of the American Statistical Association* 75, 369 (1980), 133–137.

[28] S. Ghosal, J. K. Ghosh, and R. V. Ramamoorthi. 1999. Posterior consistency of Dirichlet mixtures in density estimation. *Annals of Statistics* 27, 1 (1999), 143–158.

[29] S. Ghosal and A. W. van der Vaart. 2007. Posterior convergence rates of Dirichlet mixtures at smooth densities. *Annals of Statistics* 35, 2 (2007), 697–723.

[30] J. K. Ghosh and R. V. Ramamoorthi. 2003. *Bayesian Nonparametrics*. Springer–Verlag, New York.

[31] S. Ghosh and S.G. Henderson. 2002. Properties of the NORTA Method in Higher Dimensions. In *Proceedings of the 2002 Winter Simulation Conference*, E. Yűcesan, C. H. Chen, J. L. Snowdon, and J. M. Charnes (Eds.). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 263–269.

[32] R. J. Hanson. 1975. Stably Updating Mean and Standard Deviation of Data. *Commun. ACM* 18, 1 (Jan. 1975), 57–58.

[33] P. D. Hoff. 2007. Extending the Rank Likelihood for Semiparametric Copula Estimation. *The Annals of Applied Statistics* 1 (2007), 265–283.

[34] L. J. Hong, Z. Hu, and G. Liu. 2014. Monte Carlo Methods for Value-at-Risk and Conditional Value-at-Risk: A Review. *ACM Trans. Model. Comput. Simul.* 24, 4, Article 22 (2014), 37 pages.

[35] K. Horiguchi, N. Raghavan, R. Uzsoy, and S. Venkateswaran. 2001. Finite-capacity Production Planning Algorithms for a Semiconductor Wafer Fabrication Facility. *International Journal of PRoduction Research* 39 (2001), 825–842.

[36] J. C. Hsu and B. L. Nelson. 1990. Control Variate for Quantile Estimation. *Management Science* 36, 7 (1990), 835–851.

[37] H. Jeffreys. 1961. *Theory of Probability* (3 ed.). Oxford: Claredon.

[38] H. Joe. 2005. Asymptotic Efficiency of the Two-stage Estimation Method for Copula-based Models. *Journal of Multivariate Analysis* 94 (2005), 401419.

[39] H. Lam. 2016. Advanced Tutorial: Input Uncertainty and Robust Analysis in Stochastic Simulation. In *Proceedings of the 2016 Winter Simulation Conference*, T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick (Eds.). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

[40] T. Lancaster and J. J. Sung. 2010. Bayesian quantile regression methods. *Journal of Applied Econometrics* 25, 2 (2010), 287–307.

[41] M. Lavine. 1995. On an approximate likelihood for quantiles. *Biometrika* 82, 1 (1995), 220–222.

[42] H. F. Lopes and M. West. 2004. Bayesian model assessment in factor analysis. *Statistica Sinica* 14 (2004), 41–67.

[43] Y. Ma. 2011. *Risk Management in Biopharmaceutical Supply Chains.* Ph.D. Dissertation. University of California, Berkeley.

[44] B. Masih-Tehrani, S. H. Xu, S. Kumara, and H. Li. 2011. A Single-Period Analysis of a Two-Echelon Inventory System with Dependent Supply Uncertainty. *Transportation Research* 45, 1128-1151 (2011).

[45] J. S. Murray, D. B. Dunson, L. Carin, and J. E. Lucas. 2013. Bayesian Gaussian Copula Factor Models for Mixed Data. *Journal of the American Statistical Association* 108 (2013), 656–665.

[46] B. L. Nelson. 2016. 'Some Tactical Problems in Digital Simulation' for the Next 10 Years. *Journal of Simulation* 10 (2016). 2–11.

[47] S. H. Ng and S. E. Chick. 2006. Reducing Parameter Uncertainty for Stochastic Systems. *ACM Transactions on Modeling and Computer Simulation* 16 (2006), 26–51.

[48] T. A. Severini. 2000. *Likelihood Methods in Statistic.* Oxford Statistical Science Series.

[49] M. S. Smith. 2011. *Bayesian Approaches to Copula Modeling.* Technical Report. University of Melbourne.

[50] E. Song, B. L. Nelson, and C. D. Pegden. 2014. Advanced Tutorial: Input Uncertainty Quantification. In *Proceedings of the 2014 Winter Simulation Conference*, A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller (Eds.). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

[51] S. M. Wagner, C. Bode, and P. Koziol. 2009. Supplier Default Dependencies: Empirical Evidence from the Automotive Industry. *European Journal of Operational Research* 199 (2009), 150–161.

[52] M. West. 1990. *Bayesian Kernel Density Estimation.* Technical Report. Duke University.

[53] W. Xie, C. Li, and P. Zhang. 2017. A Bayesian Nonparametric Hierarchical Framework for Uncertainty Quantification in Simulation. (2017). submitted.

[54] W. Xie, B. L. Nelson, and R. R. Barton. 2014. A Bayesian Framework for Quantifying Uncertainty in Stochastic Simulation. *Operations Research* 62, 6 (2014), 1439–1452.

[55] W. Xie, B. L. Nelson, and R. R. Barton. 2016. Multivariate Input Uncertainty in Output Analysis for Stochastic Simulation. *ACM Transactions on Modeling and Computer Simulation* 27, 5 (2016).

[56] F. Zouaoui and J. R. Wilson. 2003. Accounting for Parameter Uncertainty in Simulation Input Modeling. *IIE Transactions* 35 (2003), 781–792.

[57] F. Zouaoui and J. R. Wilson. 2004. Accounting for Input-Model and Input-Parameter Uncertainties in Simulation. *IIE Transactions* 36 (2004), 1135–1151.

7 APPENDIX (INTENDED FOR AN ONLINE COMPANION)

7.1 Derivation of Conditional Posteriors for the Parameters in GCF

In this section, we derive the conditional posteriors for the Gibbs sampler described in Section 3.1.3. We first consider the loading matrix with element λ_{jh}. If $j < h$, by applying constraints for identification, $\lambda_{jh} = 0$. If $j > h$, the posterior for λ_{jh} is

$$\begin{aligned}
&\lambda_{jh}|\psi_{jh}, \boldsymbol{Q}_{\cdot j}, \mathbf{H} \propto p(\lambda_{jh}|\psi_{jh}) \prod_{i=1}^{m} p(Q_{ij}|\lambda_{j\cdot}, \boldsymbol{\eta}_i) \\
&\quad\propto \quad \exp\left(-\frac{\lambda_{jh}^2}{2\psi_{jh}}\right) \prod_{i=1}^{m} \exp\left[-\frac{(Q_{ij} - \sum_{h'=1}^{k} \lambda_{jh'}\eta_{ih'})^2}{2}\right] \\
&\quad\propto \quad \exp\left(-\frac{\lambda_{jh}^2}{2\psi_{jh}}\right) \prod_{i=1}^{m} \exp\left\{-\frac{\left[(Q_{ij} - \sum_{h'\neq h} \lambda_{jh'}\eta_{ih'}) - \eta_{ih}\lambda_{jh}\right]^2}{2}\right\} \\
&\quad\propto \quad \exp\left\{-\frac{(\sum_{i=1}^{m} \eta_{ih}^2 + \psi_{jh}^{-1})\lambda_{jh}^2 - 2\left[\sum_{i=1}^{m}(Q_{ij} - \sum_{h'\neq h} \lambda_{jh'}\eta_{ih'})\eta_{ih}\right]\lambda_{jh} + \sum_{i=1}^{m}(Q_{ij} - \sum_{h'\neq h} \lambda_{jh'}\eta_{ih'})^2}{2}\right\} \\
&\quad\propto \quad \exp\left[-\frac{(\lambda_{jh} - v_{jh}(\sum_{i=1}^{m} a_{ijh}\eta_{ih}))^2}{2v_{jh}}\right] \\
&\quad\sim \quad \mathcal{N}\left(v_{jh}\sum_{i=1}^{m} a_{ijh}\eta_{ih}, v_{jh}\right)
\end{aligned}$$

where $v_{jh} = \left(\sum_{i=1}^{m} \eta_{ih}^2 + \psi_{jh}^{-1}\right)^{-1}$ and $a_{ijh} = Q_{ij} - \sum_{h'\neq h} \lambda_{jh'}\eta_{ih'}$. If $j = h$,

$$\lambda_{jh}|\psi_{jh}, \boldsymbol{Q}_{\cdot j}, \mathbf{H} \propto p(\lambda_{jh}|\psi_{jh}) \prod_{i=1}^{m} p(Q_{ij}|\lambda_{j\cdot}, \boldsymbol{\eta}_i) \sim \mathrm{TN}\left(v_{jh}\sum_{i=1}^{m} a_{ijh}\eta_{ih}, v_{jh}, 0\right)$$

where $\mathrm{TN}\left(v_{jh}\sum_{i=1}^{m} a_{ijh}\eta_{ih}, v_{jh}, 0\right)$ denotes Normal distribution with mean $v_{jh}\sum_{i=1}^{m} a_{ijh}\eta_{ih}$, variance v_{jh} and truncated to be strictly positive.

Then, we derive the conditional posterior for ψ_{ih}. Given the prior $\psi_{jh} \sim \text{Inverse-Gamma}\left(\frac{\alpha_0}{2}, \frac{\beta_0}{2}\right)$, by applying Bayes' rule,

$$\begin{aligned}
p(\psi_{jh}|\lambda_{jh}) &\propto p(\psi_{jh})p(\lambda_{jh}|\psi_{jh}) \\
&\propto \quad \psi_{jh}^{-\frac{\alpha_0}{2}-1} \exp\left(-\frac{\beta_0}{2\psi_{jh}}\right) \psi_{jh}^{-\frac{1}{2}} \exp\left(-\frac{\lambda_{jh}^2}{2\psi_{jh}}\right) \\
&\propto \quad \psi_{jh}^{-\frac{\alpha_0+3}{2}} \exp\left(-\frac{\beta_0 + \lambda_{jh}^2}{2\psi_{jh}}\right) \\
&\sim \quad \text{Inverse-Gamma}\left(\frac{\alpha_0+1}{2}, \frac{\beta_0 + \lambda_{jh}^2}{2}\right).
\end{aligned}$$

After that, we derive the conditional posterior for $\boldsymbol{\eta}_i$. Given the prior $\boldsymbol{\eta}_i \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$, by applying Bayes' rule,

$$\begin{aligned} p(\boldsymbol{\eta}_i|\mathrm{Q}_i, \Lambda) &\propto p(\boldsymbol{\eta}_i)p(\mathrm{Q}_i|\boldsymbol{\eta}_i, \Lambda) \\ &\propto \quad \exp\left(-\boldsymbol{\eta}_i^T\boldsymbol{\eta}_i/2\right)\exp\left[-\frac{(\mathrm{Q}_i - \Lambda\boldsymbol{\eta}_i)^T(\mathrm{Q}_i - \Lambda\boldsymbol{\eta}_i)}{2}\right] \\ &\propto \quad \exp\left(-\frac{\mathrm{Q}_i^T\mathrm{Q}_i - 2\mathrm{Q}_i^T\Lambda\boldsymbol{\eta}_i + \boldsymbol{\eta}_i^T\Lambda^T\Lambda\boldsymbol{\eta}_i + \boldsymbol{\eta}_i^T\boldsymbol{\eta}_i}{2}\right) \\ &\propto \quad \exp\left[-\frac{\boldsymbol{\eta}_i^T\boldsymbol{\eta}_i - 2\mathrm{Q}_i^T\Lambda(\Lambda^T\Lambda + \mathbf{I})^{-1}\boldsymbol{\eta}_i}{2(\Lambda^T\Lambda + \mathbf{I})^{-1}}\right] \\ &\sim \quad \mathcal{N}\left((\Lambda^T\Lambda + \mathbf{I})^{-1}\Lambda^T\mathrm{Q}_i, (\Lambda^T\Lambda + \mathbf{I})^{-1}\right). \end{aligned}$$

7.2 Proof for Theorem 3.1

Proof: We show that for the number of factors k, $p\left(k \neq k^0 \mid \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})\right) \longrightarrow 0$ in probability as $m \longrightarrow \infty$. We divide all the models into two separate groups: (i) $\mathcal{M}_k \supset \mathcal{M}_{k^0}$ with $k^0 < k \leq k_{\max}$; (ii) $\mathcal{M}_k \subset \mathcal{M}_{k^0}$ with $1 \leq k < k^0$. For models in the group (i), from (A1) we have

$$\begin{aligned} \log\frac{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)}{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_{k^0}\right)} &= \log\frac{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \widehat{\mathrm{C}}_k, \mathcal{M}_k\right)}{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \widehat{\mathrm{C}}_{k^0}, \mathcal{M}_{k^0}\right)} - (t_k - t_{k^0})\log m \\ &\quad + (s_k - s_{k^0})\log\log m + O_p(1) \\ &\overset{(*)}{=} O_p(1) - (t_k - t_{k^0})\log m + (s_k - s_{k^0})\log\log m + O_p(1) \\ &\overset{(**)}{\leq} -\frac{t_k - t_{k^0}}{2}\log m, \end{aligned}$$

where (*) comes from (A2), and (**) holds for sufficiently large m since $t_k > t_{k^0}$ and $\log\log m \ll \log m$. Let $\Delta t = \min_{k>k^0}(t_k - t_{k^0}) > 0$. Let the constant upper bound in (A4) be $c_1 > 0$. Then we have

$$\begin{aligned} \sum_{k^0<k\leq k_{\max}}\frac{p\left(\mathcal{M}_k \mid \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})\right)}{p\left(\mathcal{M}_{k^0} \mid \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})\right)} &\leq \sum_{k^0<k\leq k_{\max}}\frac{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)p(\mathcal{M}_k)}{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_{k^0}\right)p(\mathcal{M}_{k^0})} \\ &\leq c_1\sum_{k^0<k\leq k_{\max}} m^{-\frac{t_k-t_{k^0}}{2}} \leq c_1 k_{\max} m^{-\frac{\Delta t}{2}}. \end{aligned} \tag{15}$$

In (A3), let $\delta_0 = \min_{1\leq k<k^0}\delta_k$. For models in the group (ii), we have

$$\begin{aligned} \log\frac{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)}{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_{k^0}\right)} &= \log\frac{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \widehat{\mathrm{C}}_k, \mathcal{M}_k\right)}{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \widehat{\mathrm{C}}_{k^0}, \mathcal{M}_{k^0}\right)} - (t_k - t_{k^0})\log m \\ &\quad + (s_k - s_{k^0})\log\log m + O_p(1) \\ &\overset{(*)}{\leq} -\delta_0 m - (t_k - t_{k^0})\log m + (s_k - s_{k^0})\log\log m + O_p(1) \\ &\overset{(**)}{\leq} -\frac{\delta_0}{2}m, \end{aligned}$$

where (*) follows from (A3), and (**) holds for sufficiently large m since $\log m \ll m$ and $\log\log m \ll m$. Then we have

$$\begin{aligned}\sum_{1\le k<k^0}\frac{p\left(\mathcal{M}_k \mid \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})\right)}{p\left(\mathcal{M}_{k^0} \mid \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})\right)} &\le \sum_{1\le k<k^0}\frac{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)p(\mathcal{M}_k)}{p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_{k^0}\right)p(\mathcal{M}_{k^0})}\\ &\le c_1\sum_{1\le k<k^0} e^{-\frac{\delta_0}{2}m} \le c_1 k_{\max} e^{-\frac{\delta_0}{2}m}. \qquad (16)\end{aligned}$$

By combining (15) and (16), we obtain

$$\begin{aligned}p\left(k \neq k^0 \mid \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})\right) &= \frac{\sum_{k\neq k^0} p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)p(\mathcal{M}_k)}{\sum_{1\le k\le k_{\max}} p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)p(\mathcal{M}_k)}\\
&= \frac{\sum_{k^0<k\le k_{\max}} p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)p(\mathcal{M}_k)}{\sum_{1\le k\le k_{\max}} p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)p(\mathcal{M}_k)} + \frac{\sum_{1\le k<k^0} p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)p(\mathcal{M}_k)}{\sum_{1\le k\le k_{\max}} p\left(\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}) \mid \mathcal{M}_k\right)p(\mathcal{M}_k)}\\
&= \frac{1}{1+\left[\frac{\sum_{k^0<k\le k_{\max}} p\left(\mathcal{M}_k|\mathcal{Z}_m\in D(\mathcal{X}_m^{(0)})\right)}{\sum_{1\le k\le k^0} p\left(\mathcal{M}_k|\mathcal{Z}_m\in D(\mathcal{X}_m^{(0)})\right)}\right]^{-1}} + \frac{1}{1+\left[\frac{\sum_{1\le k<k^0} p\left(\mathcal{M}_k|\mathcal{Z}_m\in D(\mathcal{X}_m^{(0)})\right)}{\sum_{k^0\le k\le k_{\max}} p\left(\mathcal{M}_k|\mathcal{Z}_m\in D(\mathcal{X}_m^{(0)})\right)}\right]^{-1}}\\
&\le \frac{1}{1+\left[\sum_{k^0<k\le k_{\max}} \frac{p\left(\mathcal{M}_k|\mathcal{Z}_m\in D(\mathcal{X}_m^{(0)})\right)}{p\left(\mathcal{M}_{k^0}|\mathcal{Z}_m\in D(\mathcal{X}_m^{(0)})\right)}\right]^{-1}} + \frac{1}{1+\left[\sum_{1\le k<k^0} \frac{p\left(\mathcal{M}_k|\mathcal{Z}_m\in D(\mathcal{X}_m^{(0)})\right)}{p\left(\mathcal{M}_{k^0}|\mathcal{Z}_m\in D(\mathcal{X}_m^{(0)})\right)}\right]^{-1}}\\
&\le \frac{1}{1+(c_1 k_{\max})^{-1} m^{\frac{\Delta t}{2}}} + \frac{1}{1+(c_1 k_{\max})^{-1} e^{\frac{\delta_0}{2}m}}.\end{aligned}$$

Thus, as $m \longrightarrow \infty$, we have $p\left(k \neq k^0 \mid \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})\right) \longrightarrow 0$. □

7.3 Proof for Theorem 3.2

Proof: For any generic matrix $\mathbf{A}$, let $\|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^\top)}$ denote the Frobenius norm of $\mathbf{A}$. We first show that as $m \longrightarrow \infty$, the correlation matrix estimate is consistent in the Frobenius norm, $\|\tilde{\mathbf{C}} - \mathbf{C}^c\| \xrightarrow{p} 0$ as $m \longrightarrow \infty$. Suppose that the posterior sample $\tilde{\mathbf{C}}$ is generated by a $\tilde{k}$-factor model. For any $\delta > 0$, by Theorem 1 in [45], we have that almost surely under the true input model F^c,

$$\lim_{m\to\infty} \mathrm{P}(\|\tilde{\mathbf{C}} - \mathbf{C}^c\| \le \delta | \mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}), \tilde{k} \ge k^0) = 1. \qquad (17)$$

Then, by accounting for the finite sampling uncertainty for $\mathcal{X}_m^{(0)}$ and the model selection uncertainty of $\tilde{k}$, we have that for any $\delta > 0$, almost surely under F^c,

$$\begin{aligned}
& \lim_{m\to\infty} \mathrm{P}\left(\|\tilde{\mathbf{C}} - \mathbf{C}^c\| \le \delta \mid \mathcal{X}_m^{(0)}\right) \\
= \; & \lim_{m\to\infty} \left(\mathrm{E}_{\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})|\mathcal{X}_m^{(0)}}\right)\left\{\mathrm{E}_{\tilde{k}|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})}\left[\mathrm{P}(\|\tilde{\mathbf{C}} - \mathbf{C}^c\| \le \delta|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}), \tilde{k})\right]\right\} \\
\overset{(*)}{=} \; & \left(\mathrm{E}_{\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})|\mathcal{X}_m^{(0)}}\right)\left\{\lim_{m\to\infty} \mathrm{E}_{\tilde{k}|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})}\left[\mathrm{P}(\|\tilde{\mathbf{C}} - \mathbf{C}^c\| \le \delta|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}), \tilde{k})\right]\right\} \\
= \; & \left(\mathrm{E}_{\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})|\mathcal{X}_m^{(0)}}\right) \\
& \left\{\lim_{m\to\infty} \left[\mathrm{P}(\tilde{k} \ge k^0|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}))\mathrm{P}(\|\tilde{\mathbf{C}} - \mathbf{C}^c\| \le \delta|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}), \tilde{k} \ge k^0)\right.\right. \\
& \left.\left. + \mathrm{P}(\tilde{k} < k^0|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}))\mathrm{P}(\|\tilde{\mathbf{C}} - \mathbf{C}^c\| \le \delta|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}), \tilde{k} < k^0)\right]\right\} \\
\overset{(**)}{=} \; & \left(\mathrm{E}_{\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)})|\mathcal{X}_m^{(0)}}\right)\left[\lim_{m\to\infty} \mathrm{P}(\|\tilde{\mathbf{C}} - \mathbf{C}^c\| \le \delta|\mathcal{Z}_m \in D(\mathcal{X}_m^{(0)}), \tilde{k} \ge k^0)\right] \\
\overset{(***)}{=} \; & 1.
\end{aligned}$$

where (*) follows by applying the dominated convergence theorem, (**) follows by applying Theorem 3.1, and (***) follows by applying Equation (17). Therefore, the correlation matrix estimate converges to the true correlation matrix in probability, $\tilde{\mathbf{C}} \xrightarrow{p} \mathbf{C}^c$ as $m \to \infty$.

According to Assumption (B3), for any $\tilde{F}_j \sim p(F_j|\mathcal{X}_{jm}^{(0)})$ with $j = 1, 2, \ldots, d$, we have $\|\tilde{F}_j - F_j^c\|_\infty \xrightarrow{p} 0$ as $m \to \infty$. This implies that for any fixed $(x_1, x_2, \ldots, x_d)$ in the support of F^c, we have $\tilde{F}_j(x_j) \xrightarrow{p} F_j^c(x_j)$ as $m \to \infty$. Then, by applying the continuous mapping theorem, we have

$$\begin{aligned}
\tilde{F}(x_1, x_2, \ldots, x_d) &= \Phi_d\left(\Phi^{-1}\left[\tilde{F}_1(x_1)\right], \Phi^{-1}\left[\tilde{F}_2(x_2)\right], \ldots, \Phi^{-1}\left[\tilde{F}_d(x_d)\right]; \tilde{\mathbf{C}}\right) \\
&\xrightarrow{p} \Phi_d\left(\Phi^{-1}\left[F_1^c(x_1)\right], \Phi^{-1}\left[F_2^c(x_2)\right], \ldots, \Phi^{-1}\left[F_d^c(x_d)\right]; \mathbf{C}^c\right) = F^c(x_1, x_2, \ldots, x_d).
\end{aligned}$$

Since both $\tilde{F}$ and F^c are distribution functions and $F^c(x_1, x_2, \ldots, x_d)$ is continuous for all $(x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, the input distribution $\tilde{F}$ uniformly converges to F^c in probability, i.e. $\|\tilde{F} - F^c\|_\infty \xrightarrow{p} 0$ as $m \to \infty$. □

7.4 Proof for Theorem 3.4

Proof: Let $\tilde{F}^{(b)}$ to be a posterior sample from $p(F|\mathcal{X}_m^{(0)})$, $\mathrm{Y}(\tilde{F}^{(b)})$ to be the simulation outputs driven by $\tilde{F}^{(b)}$ with runlength L and replications R, and $\tilde{q}^{(b)} = \tilde{q}(\tilde{F}^{(b)})$ to be a sample from the posterior $p(q(\tilde{F}^{(b)})|\mathrm{Y}(\tilde{F}^{(b)}), \tilde{F}^{(b)})$. By applying Theorem 3.2, we have $\|\tilde{F}^{(b)} - F^c\|_\infty \xrightarrow{p} 0$ as $m \to \infty$. Then, under Condition (D1) on the continuity of $q_\ell(F)$ with respect to F for each $\ell = 1, 2, \ldots, \gamma$, by applying the continuous mapping theorem, we have $|q_\ell(\tilde{F}) - q_\ell(F^c)| \xrightarrow{p} 0$ as $m \to \infty$. Theorem 3 implies that at any $\tilde{F}^{(b)}$, the quantiles drawn from the approximate posterior $p(q_\ell|\mathrm{Y}(\tilde{F}^{(b)}), \tilde{F}^{(b)})$ converge in probability to the exact quantile, i.e.,

$$|\tilde{q}_\ell(\tilde{F}^{(b)}) - q_\ell(\tilde{F}^{(b)})| \xrightarrow{p} 0 \text{ as } RL \to \infty,$$

for $\ell = 1, 2, \ldots, \gamma$, where R and L depend on the input model $\tilde{F}^{(b)}$. Then by the triangular inequality

$$|\tilde{q}_\ell(\tilde{F}^{(b)}) - q_\ell(F^c)| \le |\tilde{q}_\ell(\tilde{F}^{(b)}) - q_\ell(\tilde{F}^{(b)})| + |q_\ell(\tilde{F}^{(b)}) - q_\ell(F^c)|,$$

we have $|\tilde{q}_\ell(\tilde{F}) - q_\ell(F^c)| \xrightarrow{p} 0$ as $RL \longrightarrow \infty$ and $m \longrightarrow \infty$ given the sampled input model $F^{(1)}, \ldots, F^{(B)}$, where R and L depend on the finitely many input models $F^{(1)}, \ldots, F^{(B)}$. Since this relation holds for each $\ell = 1, 2, \ldots, \gamma$, it implies that both $\tilde{q}_{\ell,(\lceil B\alpha/2 \rceil)}$ and $\tilde{q}_{\ell,(\lceil B(1-\alpha/2) \rceil)}$ converge to $q_\ell(F^c)$ in probability. Thus, $\mathrm{CrI}(q_\ell)$ shrinks to $q_\ell(F^c)$ in probability as $RL \longrightarrow \infty$ and $m \longrightarrow \infty$, conditional on the sampled input models $F^{(1)}, \ldots, F^{(B)}$ and the simulation outputs $\mathrm{Y}(F^{(1)}), \ldots, \mathrm{Y}(F^{(B)})$. □

7.5 Marginal Distributions, Loading and Correlation Matrices Λ^c and C^c Used in Section 5.1

In this section, we provide the underlying marginal distributions F_j for $j = 1, \ldots, d$, loading matrices Λ^c and C^c used in the empirical study in Section 5.1. All marginals are Gumbel distributions with the scale parameter equal to 0.1 but different location parameters. When $d = 10$, the location parameters are $[0.245, -0.305, 0.095, -0.355, -0.155, 0.095, 0.195, -0.155, 0.095, -0.255]$. When $d = 30$, the location parameters are $[0.245, -0.305, 0.095, -0.355, -0.155, 0.095, 0.195, -0.155, 0.095, -0.255, 0.145, 0.045, -0.355, 0.095, -0.155, 0.095, -0.255, 0.045, 0.245, -0.305, -0.205, 0.195, 0.045, -0.255, 0.195, 0.095, -0.155, -0.205, -0.255, 0.245]$. Since it is difficult to show all location parameters for $d = 100$, we present the procedure to generate them. We generate 100 values from $\mathcal{N}(0.045, 0.15)$ as the location parameters.

Suppose that Λ^c has d rows and k columns. Let λ^c_{jh} with $j = 1, 2, \ldots, d$ and $h = 1, 2, \ldots, k$ denote the elements of Λ^c. The underlying correlation matrix C^c can be obtained based on Equation (1). The loading matrix of the examples when $d = 10$,

$$\text{as } k^0 = 1, \Lambda^c = \begin{bmatrix} 1.98 \\ 1.93 \\ 2.04 \\ 1.82 \\ 1.73 \\ 1.88 \\ 1.71 \\ 1.91 \\ 1.88 \\ 1.60 \end{bmatrix}; \quad \text{as } k^0 = 3, \Lambda^c = \begin{bmatrix} 1.98 & 0 & 0 \\ 2.13 & 0.06 & 0 \\ 2.04 & 0.04 & 0.08 \\ 0.02 & 1.88 & 0.12 \\ 0.03 & 2.09 & 0.09 \\ 0.08 & 2.14 & 0.04 \\ 0.01 & 2.01 & 0.01 \\ 0.01 & 0.05 & 2.15 \\ 0.08 & 0.06 & 2.06 \\ 0.1 & 0.03 & 1.83 \end{bmatrix}.$$

The loading matrix of the examples when $d = 30$,

$$\text{as } k^0 = 1, \text{ we have } \Lambda^c = \begin{bmatrix} 1.98 \\ 1.93 \\ 2.04 \\ 1.82 \\ 1.73 \\ 1.88 \\ 1.71 \\ 1.91 \\ 1.88 \\ 1.60 \\ 1.93 \\ 1.95 \\ 1.88 \\ 2.01 \\ 1.96 \\ 1.89 \\ 2.13 \\ 1.86 \\ 1.94 \\ 1.97 \\ 2.01 \\ 2.02 \\ 1.87 \\ 1.92 \\ 1.90 \\ 2.00 \\ 1.78 \\ 1.99 \\ 2.08 \\ 1.87 \end{bmatrix}; \quad \text{as } k^0 = 3, \text{ we have } \Lambda^c = \begin{bmatrix} 1.98 & 0 & 0 \\ 2.13 & 0.01 & 0 \\ 2.04 & 0.01 & 0.01 \\ 2.02 & 0.01 & 0.01 \\ 2.03 & 0.01 & 0.01 \\ 2.08 & 0.01 & 0.01 \\ 2.01 & 0.01 & 0.01 \\ 2.01 & 0.01 & 0.01 \\ 1.98 & 0.01 & 0.01 \\ 1.90 & 0.01 & 0.01 \\ 0.01 & 2.06 & 0.01 \\ 0.01 & 2.04 & 0.01 \\ 0.01 & 1.88 & 0.01 \\ 0.01 & 2.09 & 0.01 \\ 0.01 & 2.14 & 0.01 \\ 0.01 & 2.01 & 0.01 \\ 0.01 & 2.05 & 0.01 \\ 0.01 & 2.06 & 0.01 \\ 0.01 & 2.03 & 0.01 \\ 0.01 & 1.96 & 0.01 \\ 0.01 & 0.01 & 1.97 \\ 0.01 & 0.01 & 2.00 \\ 0.01 & 0.01 & 2.08 \\ 0.01 & 0.01 & 1.92 \\ 0.01 & 0.01 & 2.09 \\ 0.01 & 0.01 & 2.04 \\ 0.01 & 0.01 & 2.01 \\ 0.01 & 0.01 & 2.15 \\ 0.01 & 0.01 & 2.06 \\ 0.01 & 0.01 & 1.83 \end{bmatrix}.$$

Since it is difficult to show the loading matrices for the examples with $d = 100$, we present the procedure to generate them. For cases with $k^0 = 3$, we let the first 30 components to be highly associated with the first factor, components 31 to 70 highly associated with the second factor and the last 30 components highly associated with the third factor. For cases with $k^0 = 10$, we let that component j is highly associated with factor $\lceil j/10 \rceil$. For loading elements λ^c_{jh} where the j-th component is highly associated with the h-th factor, we generate values from $\mathcal{N}(1.8, 0.05)$. Otherwise, we generate values from $\text{Uniform}(0, 0.2)$.

accepted October 2017