Essays on Operations Management

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Operations Management and Manufacturing

by

Canan Gunes

Tepper School of Business

Carnegie Mellon University

December 2010

Dissertation Committee:

Bahar Biller (Chair)

Sridhar Tayur

Willem-Jan van Hoeve

Mustafa Akan

Abstract

This thesis focuses on the design and analysis of discrete-event stochastic simulations involving correlated inputs, input modeling for stochastic simulations, and application of OM/OR techniques to the operations of food banks. This thesis contributes to the stochastic simulation theory by describing how to jointly represent stochastic and parameter uncertainties in stochastic simulations with correlated inputs, and decompose the variance of the simulation output into terms related to stochastic uncertainty and parameter uncertainty. Such a decomposition would be beneficial for developing data collection schemas to reduce the parameter uncertainty in stochastic simulations. Furthermore, this thesis contributes to the vehicle routing theory by being the first work to rigorously study the 1-Commodity Pickup and Delivery Vehicle Routing Problem (1-PDVRP) that arises in the context of food rescue programs of food banks. A synopsis of the three chapters of the thesis follows.

Chapter 1: "Accounting for Parameter Uncertainty in Large-Scale Stochastic Simulations with Correlated Inputs"

This chapter considers large-scale stochastic simulations with correlated inputs having Normal-To-Anything (NORTA) distributions with arbitrary continuous marginal distributions. Examples of correlated inputs include processing times of workpieces across several workcenters in manufacturing facilities and product demands and exchange rates in global supply chains. Our goal is to obtain mean performance measures and confidence intervals for simulations with such correlated inputs by accounting for the uncertainty around the NORTA distribution parameters estimated from finite historical input data. This type of uncertainty is known as the parameter uncertainty in the discrete-event stochastic simulation literature. We demonstrate how to capture parameter uncertainty with a Bayesian model that uses Sklar's marginal-copula representation and Cooke's copula-vine specification for sampling the parameters of the NORTA distribution. The development of such a Bayesian model well suited for handling many correlated inputs is the primary contribution of this chapter. We incorporate the Bayesian model into the simulation replication algorithm and the Bayesian simulation replication algorithm for the joint representation of stochastic uncertainty (i.e., the uncertainty that is due to the random numbers used in the simulation) and parameter uncertainty in the mean performance estimate and the confidence interval. Using the Bayesian

simulation replication algorithm, we further decompose the variance of the simulation output into terms related to stochastic uncertainty and parameter uncertainty. Such a variance decomposition might be beneficial for developing data-collection schemas to reduce parameter uncertainty in stochastic simulations. A comprehensive numerical analysis shows that our model improves both the consistency of the mean line-item fill-rate estimates and the coverage of the confidence intervals in multi-product inventory simulations with correlated demands.

Chapter 2: "Comparison of Least-Squares and Bayesian Inferences for Johnson's S_B and S_L Distributions"

Johnson translation system is a flexible system of distributions with the ability to match any finite first four moments of a random variable. This chapter considers the problem of fitting lognormal and bounded distributions of the Johnson translation system to finite sets of stationary, independent and identically distributed input data. The focus on the Johnson translation system is due to the flexibility it provides in comparison to the standard input models that are built in commercial input-modeling software. Specifically, Johnson's lognormal family is a positively skewed distribution with one-sided bounded support, while Johnson's bounded family contains two-sided bounded distributions capturing a wide variety of unimodal and bimodal distributional shapes. The bounded supports of these distributions make it difficult to obtain robust parameter estimates via the use of the maximum likelihood estimation method. However, the least-squares estimation method used for fitting Johnson translation system does not suffer from the existence of bounded supports; and it outperforms well-known fitting methods of matching moments and percentiles. Another fitting method that is frequently used for distributions with bounded supports as an alternative to the maximum likelihood estimation method is the Bayesian method. The main goal of this chapter is to investigate the relative performance of the least-squares estimation method and the Bayesian method in fitting the parameters of the distributions from Johnson's lognormal and bounded families, and provide guidelines to the simulation practitioner on when to use each fitting method.

Chapter 3: "Food Banks Can Enhance Their Operations with OR/OM Tools: A Pilot Study with Greater Pittsburgh Community Food Bank"

Food assistance programs have been challenged to serve an increasing number of low-income families in the recent economic downturn. Soaring demand is combined with diminishing

supply (donations) attributed to both recession and donors' improved inventory management. As a remedy for food demand-supply mismatch, many food banks, whose primary goal is to reach as many needy people as possible, are trying to purchase more food by reducing their operational costs and by improving fundraising. In this study, we work together with our local food bank, Greater Pittsburgh Community Food Bank (GPCFB), in order to achieve the following two goals: First, by using the limited available data we illustrate the extent to which GPCFB is being affected by the recent economic downturn. We identify how they can collect better data for future use. We believe this will help GPCFB in their fundraising efforts. Second, we focus on the 1-PDVRP that arises in the context of the food rescue program of GPCFB. We present a thorough study on the state-of-the-art solution methods for the 1-PDVRP, utilizing technologies Mixed Integer Programming, Constraint Programming, and Constraint-based Local Search, and evaluate potential cost savings with respect to the current practice of GPCFB. The results indicate substantial cost savings, being at least 10% on the largest instance, which can be used to purchase more food and to reach more people in need. In addition to the practical value of this work for GPCFB, this study contributes to the theory by being the first academic work to provide a rigorous treatment of the 1-PDVRP. Overall, this study not only seeks to help GPCFB, but it is intended as a good starting place for other food banks around the U.S as they struggle with similar issues.

To my parents Hatice and Mustafa Gunes
& and my husband Alper Corlu

Acknowledgements

I would like to express my deepest gratitude to my advisor Dr. Bahar Biller for her wonderful guidance over the years, her patience, and her genereous help with my professional development. This would not be possible without her endless support.

I would like to thank my thesis committee members Dr. Sridhar Tayur, Dr. Willem-Jan van Hoeve, and Dr. Mustafa Akan for their guidance and invaluable advices through this process.

I would like to extend my special thanks to my beloved friends Serife Genc and Elvin Coban for their friendship, hospitality, and generosity. I also would like to thank Emre Nadar and Erkut Sonmez for their friendship and enjoyable discussions in our office. My special thanks go to Sefanur Karalti, Sevcan Bilir Kimukin, Elif Incekara Hafalir, and Esra Kaya for their warmth and for making Pittsburgh a joyful place.

I also would like to thank our PhD coordinator Lawrence Rapp for making life easier for me at CMU.

I would like to express my gratitude to my dear family. I am very thankful to my mother Hatice Gunes, my father Mustafa Gunes, and my brother Ali Ihsan Gunes who were always there to listen to me and encourage me with their continuous love and support not only in my PhD studies but also in my entire life. I am also indebted to my grandmothers and my grandfathers for their unconditional support during my entire education.

I am very appreciative to my husband Alper Corlu for his continuous support, encouragement, and love. I feel very lucky to have him in my life and I owe him more gratefulness than I could ever express.

Contents

[1] This chapter is a forthcoming article in the journal *Operations Research* with co-author Bahar Biller.

[2] This chapter is submitted to the journal *INFORMS Journal on Computing* with co-author Bahar Biller.

[3]Part of Section 3.4 of this chapter is published in the Proceedings of the 2010 International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems (CPAIOR), LNCS 6140, Springer, pp. 176 – 180 with co-authors Willem-Jan van Hoeve and Sridhar Tayur.

List of Figures

List of Tables

Introduction

This thesis focuses on the following three areas of Operations Management: design and analysis of stochastic simulations involving correlated inputs, input modeling for stochastic simulations, and application of OM/OR tools to the operations of food banks. There are three main chapters, which are self-contained and stand on their own. The first chapter demonstrates how to account for both stochastic uncertainty and parameter uncertainty in the estimation of mean performance measures and confidence intervals of stochastic simulations with correlated inputs. Furthermore, this chapter describes how to decompose the variance of the simulation output into terms related to stochastic uncertainty and parameter uncertainty. The second chapter investigates the performance of the least-squares estimation method and the Bayesian method in estimating the parameters of the distributions from the lognormal and bounded families of the Johnson translation system. The third chapter is as a result of a collaborative study with Greater Pittsburgh Community Food Bank (GPCFB) focusing on the analysis of demand and supply data of GPCFB, and a thorough study of the 1-Commodity Pickup and Delivery Problem (1-PDVRP), which arises in the context of the food rescue program of food banks.

The common practice in performing stochastic simulations involving correlated inputs is to estimate multivariate input models using historical data sets of finite length and to drive the simulations with the random variates generated from the estimated models. However, this practice ignores the parameter uncertainty (i.e., uncertainty around the estimated parameter values of a given probability distribution) and accounts only for stochastic uncertainty (i.e., uncertainty that is due to the dependence of the simulation output on the simulation's random input streams) in the simulation output analysis. Consequently, it provides not only an inaccurate estimate for the mean performance measure but also undercoverage of the confidence interval. The first chapter, "Accounting for Parameter Uncertainty in Large-Scale Stochastic Simulations with Correlated Inputs" develops a Bayesian model for the joint representation of stochastic uncertainty and parameter uncertainty in the estimation of mean performance measures and confidence intervals of stochastic simulations involving correlated inputs. Additionally, this chapter decomposes the variance of the simulation output into terms related to stochastic uncertainty and parameter uncertainty. We

illustrate that our model improves both the consistency of the mean line-item fill-rate estimates and the coverage of the confidence intervals in multi-product inventory simulations with correlated demands.

In the second chapter, "Comparison of Least-Squares and Bayesian Inferences for Johnson's S_B and S_L Distributions," we consider the fitting of the lognormal and bounded families of the Johnson translation system to independent and identically distributed input data. We describe the use of the Bayesian method for Johnson's lognormal and bounded distributions, and compare the goodness of the Bayesian fits to those obtained from the least-squares estimation method.

Finally, in the third chapter, "Food Banks Can Enhance Their Operations with OR/OM Tools: A Pilot Study with Greater Pittsburgh Community Food Bank," we collaborate with GPCFB, and perform a comprehensive analysis of their demand and supply data to validate the belief that GPCFB is facing increasing challenges to meet its demand due to the recent economic crises. Our analysis shows that the demand data is obscured; we identify the reasons and propose solution methods that will help GPCFB to gather more accurate data. Additionally, this chapter presents a rigorous study of the 1-PDVRP, utilizing technologies Mixed Integer Programming (MIP), Constraint Programming (CP), and Constraint-Based Local Search (CBLS), and finds that CBLS provides significant cost savings with respect to the current practice of GPCFB. This chapter can serve as a template for others, who are willing to help their local food banks with similar issues.

Chapter 1

Accounting for Parameter Uncertainty in Large-Scale Stochastic Simulations with Correlated Inputs [1]

1.1. Introduction

In recent years, large-scale discrete-event stochastic simulation has become a tool that is used routinely for the design and analysis of manufacturing and service systems. Two important components of the large-scale stochastic simulation are multivariate input modeling and output analysis. Multivariate input modeling is the estimation of an appropriate multivariate probability distribution that characterizes the stochastic behavior of the system inputs. Output analysis is the study of the simulation output data to estimate the distributional properties (e.g., mean, probability, or quantile) of the performance measure.

In this paper, we are interested in the case where the objective of the output analysis is to predict a mean performance measure and a confidence interval. There are three main sources of uncertainty to account for in the output analysis: stochastic uncertainty, model uncertainty, and parameter uncertainty. Stochastic uncertainty arises from the dependence of the output on the simulation's random input streams (Helton 1997). Model uncertainty arises due to the uncertainty around the selection of an appropriate family of distributions for the system inputs, while parameter uncertainty arises due to the uncertainty around the parameter values of a given probability distribution (Raftery et al. 1996).

The goal of this paper is to account for the stochastic uncertainty and the parameter uncertainty in the estimation of the mean performance measure and the confidence interval of the stochastic simulation with correlated inputs. Accounting for parameter uncertainty

[1] This chapter is a forthcoming article in the journal *Operations Research* with co-author Bahar Biller.

in a stochastic simulation is not common practice. The simulation often starts with fitting a probability distribution to the historical input data of finite length. Although the parameters of the fitted distribution are shown to have asymptotical properties (e.g., consistency and normality) for the number of historical data points approaching infinity, the simulation is driven with the probability distribution estimated from the finite historical input data. The output data obtained from the simulation are analyzed for predicting the mean performance measure and constructing the confidence interval. This practice of using the estimated probability distribution for driving the simulation ignores both the model uncertainty and the parameter uncertainty, and accounts only for the stochastic uncertainty in the output analysis. Consequently, the simulation analyst obtains not only an inconsistent estimate for the mean performance measure, but also an inconsistent coverage for the confidence interval.

The problem of accounting for model uncertainty and/or parameter uncertainty in stochastic simulations has been studied by a number of researchers including Cheng and Holland (1997, 1998, 2004), Chick (1997, 1999, 2001), Barton and Schruben (2001), and Zouaoui and Wilson (2003, 2004). Cheng and Holland (1997) made the first attempt to show the dependence of the simulation output on stochastic and parameter uncertainties. They continued the study of this problem in Cheng and Holland (1998) and Cheng and Holland (2004). However, using frequentist techniques to represent parameter uncertainty did not allow the incorporation of any relevant information other than the historical input data into the simulation output. Furthermore, the problem of accounting for model uncertainty remained unsolved. As a result of following the Bayesian Model Averaging (BMA) approach, Chick (1999, 2001) captured not only stochastic uncertainty and parameter uncertainty, but also model uncertainty. Although the BMA approach had been used in a number of different settings to capture model uncertainty and/or parameter uncertainty (e.g., Draper 1995, Hoeting et al. 1999, and George 1999), it was Chick (1997, 1999, 2001) who outlined the basic methodology for implementing the BMA approach in discrete-event stochastic simulations.

Chick's formulation led to the simulation replication algorithm that allowed the simulation analyst to capture both model uncertainty and parameter uncertainty by sampling input distributions and their parameters from Bayesian posterior density functions before each replication. This made it possible to drive the simulation without using a single distribution and a single set of parameter values. Consequently, the use of the simulation replication algorithm improved the consistency of the mean performance estimate and the coverage of the confidence interval. However, the simulation replication algorithm did not provide separate

estimates for the variances due to different sources of uncertainty. The Bayesian simulation replication algorithm introduced by Zouaoui and Wilson (2003) provided a solution for the problem of decomposing the variance in the simulation output data into variances due to stochastic and parameter uncertainties. They extended their framework to also account for model uncertainty in Zouaoui and Wilson (2004).

The focus in these papers has been on discrete-event stochastic simulations with independent inputs. However, building large-scale simulations may require the development of multivariate input models. Examples of multivariate inputs include the processing times of a workpiece across several workcenters in a manufacturing facility (Xu 1999) and the product demands and exchange rates in a global supply chain (Kouvelis and Su 2007). In this paper, we focus on simulations with multivariate inputs that have stochastic dependencies among them, and we describe how to account for both stochastic uncertainty and parameter uncertainty in their output analyses. Although the BMA approach formulated in Chick (2001) can accommodate the stochastic dependencies among the inputs, the simulation analyst still needs a Bayesian model that works for simulations with many correlated inputs. Also, it is not clear how to sample the parameters of the joint distribution of the correlated inputs before each replication so that the Bayesian model is easily incorporated into the simulation replication algorithm. The development of a Bayesian model, which overcomes these challenges and leads to a fast sampling algorithm well suited for handling a large number of correlated inputs, is the primary contribution of our paper to the discrete-event stochastic simulation literature. We also decompose the simulation output variance into variances due to stochastic and parameter uncertainties. Since the simulation replication algorithm does not allow us to obtain such a decomposition of the simulation output variance, we use the Bayesian simulation replication algorithm introduced by Zouaoui and Wilson (2003) for this purpose.

We represent multivariate inputs using random vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)'$ denoting a collection of k correlated components, each of which is a real-valued input random variable. We characterize the joint stochastic behavior of these correlated inputs with the flexible Normal-To-Anything (NORTA) distribution of Cario and Nelson (1997). Thus, we construct our $k-$dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)'$ by first taking Z_i as the i^{th} component of a $k-$dimensional standard normal random vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_k)'$ with positive definite $k \times k$ correlation matrix $\mathbf{\Sigma}_k$. Then we obtain $X_i = F_i^{-1}(\Phi(Z_i); \mathbf{\Psi}_i)$ for $i = 1, 2, \ldots, k$, where F_i is the arbitrary continuous marginal cumulative distribution func-

tion (cdf) of component i with parameter vector $\mathbf{\Psi}_i$ and Φ is the cdf of the standard normal random variable. A random vector constructed in this way is said to have a $k-$dimensional NORTA distribution. From this point on, we call the parameters of this distribution, $\mathbf{\Psi}_i$, $i = 1, 2, \ldots, k$, and $\mathbf{\Sigma}_k$, the NORTA parameters.

Considering a stochastic simulation with inputs having a $k-$dimensional NORTA distribution, we demonstrate how to account for parameter uncertainty (i.e., the uncertainty around the values of the NORTA parameters estimated from finite historical input data) in the mean performance estimate and the confidence interval. More specifically, we develop a Bayesian model that samples NORTA parameters from their Bayesian posterior density functions before each replication of the simulation replication algorithm. This allows us to drive the stochastic simulation without using a single set of NORTA parameters. However, it is a challenging task to develop such a Bayesian model as the number of NORTA parameters needed to sample increases very quickly with k, the number of components. For example, if each of the k components are exponentially distributed, we need to sample k different parameters for the component marginal distributions and $2k$ different parameters for the gamma distributed components. In addition to the parameters of the component marginal distributions, we need to sample $k(k-1)/2$ correlations of $\mathbf{\Sigma}_k$ in a way that the resulting correlation matrix is positive definite.

We overcome the challenge of sampling a large number of NORTA parameters using Sklar's marginal-copula representation together with Cooke's copula-vine specification. Sklar's marginal-copula representation allows us to write the joint posterior density function as the multiplication of the marginal posterior density functions and the posterior normal copula density function. Thus, we separate the problem of sampling the parameters of the component marginals from the problem of sampling the dependence parameters of the NORTA distribution. Furthermore, Cooke's copula-vine specification enables us to represent the $k-$dimensional posterior normal copula density function as the product of $k(k-1)/2$ two-dimensional posterior normal copula density functions. Therefore, we do not need to satisfy any algebraic constraints for positive definiteness.

We organize the remainder of the paper as follows. In Section 1.2, we describe the NORTA distribution and present a copula-based representation for its joint density function. We use this representation in Section 1.3 for developing a Bayesian model that samples NORTA parameters. In Section 1.4, we describe how to incorporate the Bayesian model into the simulation replication algorithm (Section 1.4.1) and the Bayesian simulation replication

algorithm (Section 1.4.2) for estimating the mean performance measure and the confidence interval that accounts for both stochastic uncertainty and parameter uncertainty. In Section 1.5, we decompose the variance of the simulation output into terms related to stochastic uncertainty and the parameter uncertainty using the Bayesian simulation replication algorithm of Section 1.4.2. In Section 1.6, we show that our model allows the simulation analyst to improve both the consistency of the mean line-item fill-rate estimates and the coverage of the confidence intervals in multi-product inventory simulations with correlated demands. We conclude with a summary of results in Section 1.7. For clarity in the presentation of the results, we moved the implementation details of sampling NORTA parameters to an online companion.

1.2. The NORTA Distribution and Its Copula-Based Representation

We introduce the $k-$dimensional NORTA distribution in Section 1.2.1 and provide a copula-based representation for its joint density function in Section 1.2.2.

1.2.1. The $k-$dimensional NORTA Distribution

We characterize the joint stochastic behavior of correlated inputs X_i, $i = 1, 2, \ldots, k$ using the NORTA distribution of Cario and Nelson (1997). The central idea is to transform a standard multivariate normal random vector into the random vector referred to as having a NORTA distribution. Specifically, we let

$$\mathbf{X} \;=\; \left(F_1^{-1}\left(\Phi(Z_1); \boldsymbol{\Psi}_1\right), F_2^{-1}\left(\Phi(Z_2); \boldsymbol{\Psi}_2\right), \ldots, F_k^{-1}\left(\Phi(Z_k); \boldsymbol{\Psi}_k\right)\right)',$$

where $F_i\left(\cdot; \boldsymbol{\Psi}_i\right)$, $i = 1, 2, \ldots, k$ are arbitrary marginal cdfs with parameter vectors $\boldsymbol{\Psi}_i$, $i = 1, 2, \ldots, k$ and the base vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_k)'$ is a $k-$dimensional standard normal random vector with $k \times k$ positive definite correlation matrix $\boldsymbol{\Sigma}_k \equiv [\rho(i, j); i, j = 1, 2, \ldots, k]$. In this characterization, $\rho(i, j)$ is the Pearson product-moment correlation between Z_i (i.e., $\Phi^{-1}[F_i(X_i; \boldsymbol{\Psi}_i)]$) and Z_j (i.e., $\Phi^{-1}[F_j(X_j; \boldsymbol{\Psi}_j)]$), while the transformation $X_i = F_i^{-1}(\Phi(Z_i); \boldsymbol{\Psi}_i)$ ensures that X_i has cdf $F_i(\cdot; \boldsymbol{\Psi}_i)$. A more detailed description of the NORTA distribution is available in Biller and Ghosh (2006).

1.2.2. A Copula-Based Representation for the NORTA Distribution

We first present a brief review of the copula theory. Next, we derive a copula-based representation for the NORTA distribution using Sklar's marginal-copula representation and Cooke's copula-vine specification.

Copula Theory

We begin this section with the definition of a $k-$dimensional copula (Nelsen 1999, Definition 2.10.6). The first condition of this definition provides the lower bound on the distribution function and insures that the marginal distributions of the copula are uniform, while the second condition insures that the probability of observing a point in a $k-$box is non-negative:

Definition 1 *A $k-$dimensional copula is a function C_k: $[0,1]^k \to [0,1]$ with the following properties: (1) For every $\mathbf{u} = (u_1, u_2, \ldots, u_k)$ in $[0,1]^k$, $C_k(\mathbf{u}) = 0$ if at least one coordinate of $\mathbf{u}$ is 0, and if all coordinates of $\mathbf{u}$ are 1 except u_ℓ, then $C_k(\mathbf{u}) = u_\ell$ for $\ell = 1, 2, \ldots, k$. (2) For every $\mathbf{a} = (a_1, a_2, \ldots, a_k)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_k)$ in $[0,1]^k$ such that $\mathbf{a} \leq \mathbf{b}$, i.e., $a_\ell \leq b_\ell$, $\ell = 1, 2, \ldots, k$, and for every $k-$box $[\mathbf{a}, \mathbf{b}]$, i.e., $[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_k, b_k]$, the C_k-volume given by $\Delta_{\mathbf{a}}^{\mathbf{b}} C_k(\mathbf{t}) = \Delta_{a_k}^{b_k} \Delta_{a_{k-1}}^{b_{k-1}} \cdots \Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} C_k(\mathbf{t})$ with $\Delta_{a_\ell}^{b_\ell} C_k(\mathbf{t}) = C_k(t_1, \ldots, t_{\ell-1}, b_\ell, t_{\ell+1}, \ldots, t_k) - C_k(t_1, \ldots, t_{\ell-1}, a_\ell, t_{\ell+1}, \ldots, t_k)$ is ≥ 0.*

The use of a copula for understanding the joint distribution of a random vector has been studied extensively for the last two decades (Schweizer 1991, Joe 1997, Nelsen 1999). In this paper, we restrict our attention to Sklar's theorem, which describes how to extract the dependence structure of a random vector from its joint distribution with arbitrary continuous marginal distributions (Nelsen 1999, Theorem 2.10.9):

Theorem 1 *Let H_k be a $k-$dimensional distribution function with continuous marginal cdfs $F_i(\cdot; \mathbf{\Psi}_i)$, $i = 1, 2, \ldots, k$. Then there exists a $k-$dimensional unique copula C_k such that for all x_i, $i = 1, 2, \ldots, k$ in $\Re$,*

$$H_k(x_1, x_2, \ldots, x_k) \;=\; C_k(F_1(x_1; \mathbf{\Psi}_1), F_2(x_2; \mathbf{\Psi}_2), \ldots, F_k(x_k; \mathbf{\Psi}_k)). \tag{1.1}$$

Conversely, if C_k is a $k-$dimensional copula and $F_i(\cdot; \mathbf{\Psi}_i)$, $i = 1, 2, \ldots, k$ are continuous distribution functions with parameter vectors $\mathbf{\Psi}_i$, $i = 1, 2, \ldots, k$, then the function H_k defined in (1.1) is a $k-$dimensional distribution function with marginal cdfs $F_i(\cdot; \mathbf{\Psi}_i)$, $i = 1, 2, \ldots, k$.

The major implication of this theorem is that copula C_k is the joint distribution function of $U_i \equiv F_i(X_i; \mathbf{\Psi}_i)$, $i = 1, 2, \ldots, k$, where U_i, $i = 1, 2, \ldots, k$ are the probability integral transforms of X_i, $i = 1, 2, \ldots, k$. Thus, each of the random variables U_i, $i = 1, 2, \ldots, k$ follows a uniform distribution in $[0, 1]$, regardless of the distributions of the input random variables X_i, $i = 1, 2, \ldots, k$. More importantly, C_k can be interpreted as the dependence structure of the joint cdf H_k and written as $C_k(u_1, u_2, \ldots, u_k) = H_k(F_1^{-1}(u_1; \mathbf{\Psi}_1), F_2^{-1}(u_2; \mathbf{\Psi}_2), \ldots, F_k^{-1}(u_k; \mathbf{\Psi}_k))$, where $F_i^{-1}(\cdot; \mathbf{\Psi}_i)$ is the generalized inverse of marginal cdf $F_i(\cdot; \mathbf{\Psi}_i)$ (Nelsen 1999, Corollary 2.10.10).

Another important implication of the representation in (1.1) is that a joint probability density function (pdf) can be written as a product of marginal pdfs and copula density function, which encodes all of the information about the stochastic dependencies among the components. More specifically, for differentiable marginal cdfs $F_i(\cdot; \mathbf{\Psi}_i)$, $i = 1, 2, \ldots, k$ and differentiable copula C_k, the $k-$dimensional pdf, which is denoted by h_k below, can be written as

$$\begin{aligned} h_k(x_1, x_2, \ldots, x_k) &= \prod_{i=1}^{k} f_i(x_i; \mathbf{\Psi}_i) \\ &\times \; c_k\Big(F_1(x_1; \mathbf{\Psi}_1), F_2(x_2; \mathbf{\Psi}_2), \ldots, F_k(x_k; \mathbf{\Psi}_k)\Big). \end{aligned}$$

In this representation, $f_i(\cdot; \mathbf{\Psi}_i)$ is the pdf of X_i, i.e., $f_i(x; \mathbf{\Psi}_i) \equiv \partial F_i(x; \mathbf{\Psi}_i)/\partial x$, and c_k is the $k-$dimensional copula density function given by $\partial^k C_k(u_1, u_2, \ldots, u_k)/(\partial u_1 \partial u_2 \ldots \partial u_k)$. This copula density function takes the value of 1 when X_i, $i = 1, 2, \ldots, k$ are independent and therefore, the joint density function reduces to the product of only the marginal pdfs.

Sklar's Marginal-Copula Representation and NORTA

The use of Sklar's theorem for representing a $k-$dimensional NORTA distribution shows that the $k-$dimensional random vector with the NORTA distribution and the $k-$dimensional normal random vector share the same copula:

Corollary 1 *Let $\mathbf{X} = (X_1, X_2, \ldots, X_k)'$ correspond to a $k-$dimensional random vector with a NORTA distribution characterized by continuous marginal cdfs $F_i(\cdot; \mathbf{\Psi}_i)$, $i = 1, 2, \ldots, k$ and positive definite correlation matrix $\mathbf{\Sigma}_k$. Then there exists a $k-$dimensional unique normal copula that represents the dependence structure of $\mathbf{X}$.*

The key to the proof of this corollary is that the joint cdf H_k of X_i, $i = 1, 2, \ldots, k$ is given by

$$H_k\left(x_1, x_2, \ldots, x_k\right) = \Phi_k\left(\Phi^{-1}[F_1(x_1; \boldsymbol{\Psi}_1)], \Phi^{-1}[F_2(x_2; \boldsymbol{\Psi}_2)], \ldots, \Phi^{-1}[F_k(x_k; \boldsymbol{\Psi}_k)]; \boldsymbol{\Sigma}_k\right),$$

where Φ^{-1} is the functional inverse of Φ and $\Phi_k(\cdot\,; \boldsymbol{\Sigma}_k)$ is the $k-$dimensional standard normal cdf with correlation matrix $\boldsymbol{\Sigma}_k$. Since the normal copula is the dependence function implicitly assumed whenever the multivariate normal distribution is used, the dependence structure of a $k-$dimensional NORTA distribution is represented by a $k-$dimensional normal copula.

The joint pdf of the $k-$dimensional NORTA distribution, h_k can now be written as the multiplication of the k component marginal density functions and the $k-$dimensional normal copula density function; i.e.,

$$\begin{aligned} h_k\left(x_1, x_2, \ldots, x_k\right) &= \prod_{i=1}^{k} f_i(x_i; \boldsymbol{\Psi}_i) \\ &\times \phi_k\left(\Phi^{-1}[F_1(x_1; \boldsymbol{\Psi}_1)], \Phi^{-1}[F_2(x_2; \boldsymbol{\Psi}_2)], \ldots, \Phi^{-1}[F_k(x_k; \boldsymbol{\Psi}_k)]; \boldsymbol{\Sigma}_k\right). \end{aligned} \tag{1.2}$$

The normal copula density function ϕ_k is further given by

$$\begin{aligned} \phi_k\left(\Phi^{-1}[u_1], \Phi^{-1}[u_2], \ldots, \Phi^{-1}[u_k]; \boldsymbol{\Sigma}_k\right) &\equiv \frac{\partial^k \Phi_k\left(\Phi^{-1}[u_1], \Phi^{-1}[u_2], \ldots, \Phi^{-1}[u_k]; \boldsymbol{\Sigma}_k\right)}{\partial u_1 \partial u_2 \ldots \partial u_k}, \\ &\equiv |\boldsymbol{\Sigma}_k|^{-1/2} \exp\left\{-\frac{1}{2}\varsigma'\left(\boldsymbol{\Sigma}_k^{-1} - \mathbf{I}_k\right)\varsigma\right\}, \end{aligned} \tag{1.3}$$

where $u_i \equiv F_i(x_i; \boldsymbol{\Psi}_i)$ for $i = 1, 2, \ldots, k$, $\varsigma \equiv (\Phi^{-1}[u_1], \Phi^{-1}[u_2], \ldots, \Phi^{-1}[u_k])'$, and $\mathbf{I}_k$ is the $k-$dimensional identity matrix. Thus, the copula density function ϕ_k captures all of the information about the dependence structure of $\mathbf{X}$ using correlations $\rho(i, j)$, $i, j = 1, 2, \ldots, k$.

Cooke's Copula-Vine Specification and NORTA

A vine is a graphical model for constructing high-dimensional joint distributions using a series of two-dimensional (conditional) distributions. It has been introduced in Cooke (1997), studied extensively in Bedford and Cooke (2001, 2002) and Kurowicka and Cooke (2003), and described comprehensively in Kurowicka and Cooke (2006). In this paper, we use a vine for representing NORTA's normal copula density function in (1.3).

More specifically, we represent the joint distribution of the base random variables of the $k-$dimensional NORTA distribution (i.e., $Z_i \equiv \Phi^{-1}[F_i(X_i; \boldsymbol{\Psi}_i)]$, $i = 1, 2, \ldots, k$) with a regular vine defined as follows (Definition 4.4, Kurowicka and Cooke 2006):

Definition 2 *$\mathcal{V}$ is a vine on k elements under the following conditions: (1) $\mathcal{V} = (\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_{k-1})$. (2) $\mathcal{T}_1$ is a connected tree with nodes $\mathcal{N}_1 = \{1, 2, \ldots, k\}$ and edges $\mathcal{E}_1$; for $i = 2, 3, \ldots, k-1$, $\mathcal{T}_i$ is a connected tree with nodes $\mathcal{N}_i = \mathcal{E}_{i-1}$. $\mathcal{V}$ is a regular vine on k elements if additionally the following condition is satisfied: (3) For $i = 2, 3, \ldots, k-1$, if $\{a, b\} \in \mathcal{E}_i$, then $\#a \Delta b = 2$, where Δ denotes the symmetric difference. In other words, if a and b are nodes of $\mathcal{T}_i$ connected by an edge in $\mathcal{T}_i$, where $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$, then exactly one of the a_i equals one of the b_i.*

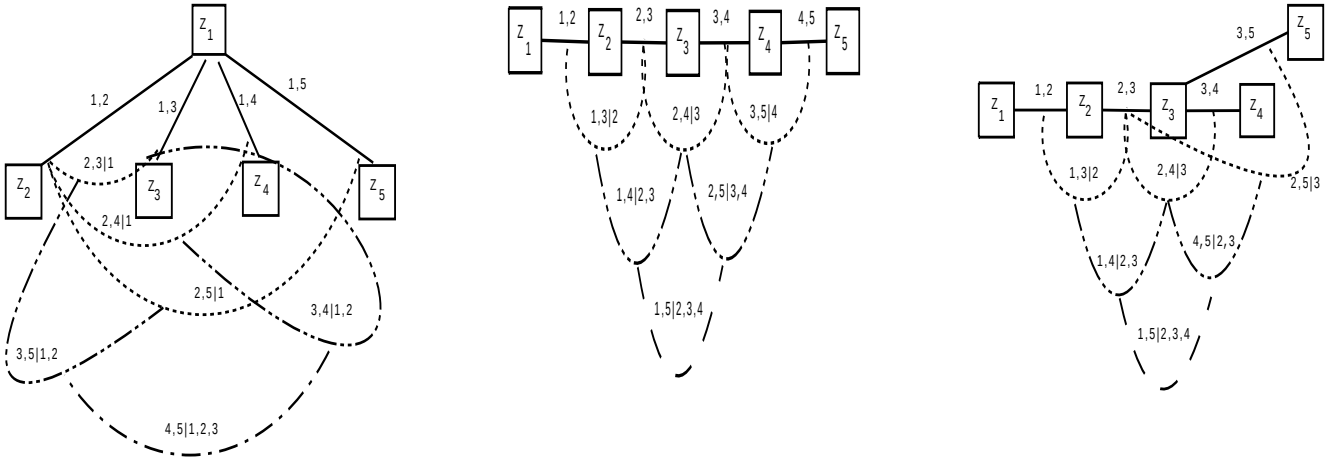

Figure 1.1: Three different regular vine specifications for the 5−dimensional NORTA distribution

No unique regular vine exists for representing the dependence structure of the NORTA distribution. Figure 1.1 provides examples of three different regular vines, each of which can be used for representing the dependence structure of the 5−dimensional NORTA distribution. The first two vines are known as the $\mathcal{C}-$vine and the $\mathcal{D}-$vine, introduced in Bedford and Cooke (2002). Our solution approach works for any of these vines as well as any regular vine constructed as described in Definition 2. Since using different regular vine specifications leads to different sampling algorithms for NORTA's dependence parameters, we use a $\mathcal{C}-$vine for describing our solution approach in the remainder of the paper due to the ease of its implementation (Section 6.4.2, Kurowicka and Cooke 2006).

We represent NORTA's $k-$dimensional normal copula density function in (1.3) with a $\mathcal{C}-$vine on $\Phi^{-1}[F_i(X_i; \mathbf{\Psi}_i)]$, $i = 1, 2, \ldots, k$. This vine has the following unconditional and conditional correlations assigned to its edges: $\rho(1, i)$, $i = 2, 3, \ldots, k$ and $\rho(j-1, i; 1, 2, \ldots, j-2)$, $i = j, j+1, \ldots, k$, $j = 3, 4, ..., k$. While $\rho(1, i)$ is the (unconditional) correlation between

random variables $\Phi^{-1}[F_1(X_1;\,\mathbf{\Psi}_1)]$ and $\Phi^{-1}[F_i(X_i;\mathbf{\Psi}_i)]$, $\rho(j-1,i;1,2,\ldots,j-2)$ is the (conditional) correlation between conditional random variables $\Phi^{-1}[F_{j-1}(X_{j-1};\mathbf{\Psi}_{j-1})]|\Phi^{-1}[F_\ell(X_\ell;\,\mathbf{\Psi}_\ell)]$, $\ell=1,2,\ldots,j-2$ and $\Phi^{-1}[F_i(X_i;\,\mathbf{\Psi}_i)]\ |\Phi^{-1}[F_\ell(X_\ell;\mathbf{\Psi}_\ell)]$, $\ell=1,2,\ldots,j-2$.

Since the dependence structure of the NORTA distribution is represented by a normal copula (Corollary 1), the conditional correlation $\rho(j-1,i;1,2,\ldots,j-2)$ is also the partial correlation (i.e., the correlation between the orthogonal projections of $\Phi^{-1}[F_{j-1}(X_{j-1};\,\mathbf{\Psi}_{j-1})]$ and $\Phi^{-1}[F_i(X_i;\,\mathbf{\Psi}_i)]$ on the plane orthogonal to the space spanned by $\Phi^{-1}[F_\ell(X_\ell;\,\mathbf{\Psi}_\ell)]$, $\ell=1,2,\ldots,j-2$) (Morales et al. 2006). Recursive formulas exist that allow the identification of the partial correlations from the correlations of $\mathbf{\Sigma}_k$ (Yule and Kendall 1965).

All of the (partial) correlations in the $\mathcal{C}-$vine specification of the $k-$dimensional NORTA distribution are algebraically independent. Therefore, they do not need to satisfy any algebraic constraints for positive definiteness. Furthermore, the resulting copula-vine specification uniquely determines the correlation matrix $\mathbf{\Sigma}_k$:

Corollary 2 *For the $\mathcal{C}-$vine on $\Phi^{-1}[F_i(X_i;\,\mathbf{\Psi}_i)]$, $i=1,2,\ldots,k$, there is a one-to-one correspondence between the set of $k\times k$ positive definite correlation matrices of the form $\mathbf{\Sigma}_k$ and the set of correlations $\rho(1,i)$, $i=2,3,\ldots,k$ and partial correlations $\rho(j-1,i;1,2,\ldots,j-2)$, $i=j,j+1,\ldots,k$, $j=3,4,...,k$ of the $k-$dimensional NORTA distribution.*

The proof of this corollary is from the application of Theorem 4.4 of Kurowicka and Cooke (2006) to the copula-vine specification of the NORTA distribution. Therefore, all assignments of the numbers between -1 and 1 to the edges of the $\mathcal{C}-$vine (and to the edges of any arbitrary regular vine) are consistent in the sense that there is a NORTA distribution realizing these (partial) correlations.

We are now ready to replace $h_k(x_1,x_2,\ldots,x_k)$, the joint pdf of the $k-$dimensional NORTA distribution in (1.2), with the product of the k component marginal density functions and the $k(k-1)/2$ two-dimensional normal copula density functions, each of which is associated with an edge of the $\mathcal{C}-$vine:

$$
\begin{aligned}
&= \prod_{i=1}^{k} f_i(x_i;\,\mathbf{\Psi}_i)\\
&\times \prod_{i=2}^{k} \phi_2\Big(F_1(x_1;\,\mathbf{\Psi}_1),F_i(x_i;\,\mathbf{\Psi}_i);\mathbf{\Sigma}_2(1,i)\Big)\\
&\times \prod_{j=3}^{k}\prod_{i=j}^{k} \phi_2\Big(\Phi_{j-1|1,2,\ldots,j-2}\left(\Phi^{-1}[F_{j-1}(x_{j-1};\,\mathbf{\Psi}_{j-1})]|\Phi^{-1}[F_\ell(x_\ell;\,\mathbf{\Psi}_\ell)],\ell=1,2,\ldots,j-2\right),
\end{aligned}
$$

$$\Phi_{i|1,2,\ldots,j-2}\left(\Phi^{-1}[F_i(x_i;\boldsymbol{\Psi}_i)]|\Phi^{-1}[F_\ell(x_\ell;\boldsymbol{\Psi}_\ell)],\ell=1,2,\ldots,j-2\right);$$
$$\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)\Big) \tag{1.4}$$

In this representation, $\boldsymbol{\Sigma}_2(1,i)$ is the two-dimensional correlation matrix with correlation $\rho(1,i)$ as its $(1,2)^{\text{th}}$ entry (i.e., $\boldsymbol{\Sigma}_2(1,i) \equiv [1\ \rho(1,i);\rho(1,i)\ 1]$). Similarly, $\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2) \equiv [1\ \rho(j-1,i;1,2,\ldots,j-2);\rho(j-1,i;1,2,\ldots,j-2)\ 1]$. Additionally, $\Phi_{s|1,2,\ldots,j-2}$ is the marginal cdf of the conditional random variable $\Phi^{-1}[F_s(X_s;\boldsymbol{\Psi}_s)]|\Phi^{-1}[F_\ell(X_\ell;\boldsymbol{\Psi}_\ell)],\ell = 1,2,\ldots,j-2$ with mean $\mu_{s|1,2,\ldots,j-2}$ and variance $\sigma^2_{s|1,2,\ldots,j-2}$ that can be obtained using Theorem 3.3.4 of Tong (1990) with the recursive formulas of Yule and Kendall (1965). Appendix B describes how to do this for a 5−dimensional NORTA distribution.

1.3. Bayesian Model for Sampling NORTA Parameters

The key to the development of our Bayesian model is to separate the sampling of the component parameter vectors from the sampling of the (partial) correlations using Sklar's marginal-copula representation and Cooke's copula-vine specification. Therefore, in Section 1.3.1 we first focus on the i^{th} component of the NORTA vector and describe how to sample parameter vector $\boldsymbol{\Psi}_i$. Then we discuss the sampling of correlation $\rho(1,i)$ and partial correlation $\rho(j-1,i;1,2,\ldots,j-2)$ in Section 1.3.2. Finally, in Section 1.3.3 we describe how to sample all parameters of the k−dimensional NORTA distribution (i.e., $\boldsymbol{\Psi}_i$, $i=1,2,\ldots,k$, $\rho(1,i)$, $i=2,3,\ldots,k$, and $\rho(j-1,i;1,2,\ldots,j-2)$, $i=j,j+1,\ldots,k$, $j=3,4,\ldots,k$) using the sampling algorithms of Sections 1.3.1 and 1.3.2.

1.3.1. Sampling the Parameters of the Component Marginal Distributions

Well-established literature exists on Bayesian probability theory for sampling the parameters of the standard families of distributions (Gelman et al. 2000, Carlin and Louis 2000).

Assuming the availability of the historical data of finite length, this section describes how to use this literature to sample the parameters of the i^{th} component having the exponential distribution or the gamma distribution of the standard input-modeling packages.

First, we choose the distribution of the i^{th} component X_i as exponential with scale parameter β_i; i.e., $f_i(x_i;\beta_i) = \beta_i^{-1}\exp(-x_i\beta_i^{-1})$. Therefore, we can write the likelihood function $\prod_{t=1}^{n} f_i(x_{i,t};\beta_i)$, which describes the joint pdf of the historical data $x_{i,t}$, $t=1,2,\ldots,n$ of

length n, as $\beta_i^{-n} \exp(-\beta_i^{-1} \sum_{t=1}^{n} x_{i,t})$. The next step is to construct a prior density function $\pi_i(\beta_i)$ on scale parameter β_i. We do this with the conjugate[2], inverse gamma prior with shape parameter θ_i and scale parameter ν_i; i.e., $\pi_i(\beta_i) = \nu_i^{\theta_i} \Gamma^{-1}(\theta_i) \beta_i^{-(\theta_i+1)} \exp(-\nu_i \beta_i^{-1})$. Finally, we denote the vector of the historical data available for the i^{th} component with $\mathbf{x}_i$ and combine the prior density function with the likelihood function using Bayes' rule to obtain the posterior density function $p_i(\beta_i|\mathbf{x}_i)$ of parameter β_i (Bernardo and Smith 1994):

$$p_i(\beta_i|\mathbf{x}_i) \quad \propto \quad \pi_i(\beta_i) \prod_{t=1}^{n} f_i(x_{i,t};\beta_i) \propto \beta_i^{-(n+\theta_i+1)} \exp\left\{-\frac{\nu_i + \sum_{t=1}^{n} x_{i,t}}{\beta_i}\right\} \tag{1.5}$$

Thus, representing parameter uncertainty for component X_i reduces to the sampling of β_i^{-1} from a gamma distribution with shape parameter $n+\theta_i$ and scale parameter $(\nu_i+\sum_{t=1}^{n} x_{i,t})^{-1}$, for which an efficient algorithm is available in Appendix A.1.

Next, we consider a gamma component with shape parameter α_i and scale parameter β_i; i.e., $f_i(x_i;\alpha_i,\beta_i) = x_i^{\alpha_i-1}\Gamma^{-1}(\alpha_i)\beta_i^{-\alpha_i}\exp(-x_i\beta_i^{-1})$. We use Bayes' rule for combining Jeffreys' prior density function $\pi_i(\alpha_i,\beta_i) \propto \beta_i^{-1}$ with the likelihood function of $\mathbf{x}_i$ and obtain the following joint posterior density function for the parameters of the gamma component (Son and Oh 2006):

$$p_i(\alpha_i,\beta_i|\mathbf{x}_i) \quad \propto \quad \pi_i(\alpha_i,\beta_i) \prod_{t=1}^{n} f_i(x_{i,t};\alpha_i,\beta_i) = \frac{\beta_i^{-\alpha_i n-1}}{[\Gamma(\alpha_i)]^n} \left(\prod_{t=1}^{n} x_{i,t}^{\alpha_i-1}\right) \exp\left\{-\frac{\sum_{t=1}^{n} x_{i,t}}{\beta_i}\right\}$$

Parameters α_i and β_i can be sampled from this joint posterior density function using the Markov Chain Monte Carlo (MCMC) method. The idea behind the MCMC method is to simulate a random walk in the space of (α_i,β_i) that converges to the joint posterior density function $p_i(\alpha_i,\beta_i|\mathbf{x}_i)$ (Gilks et al. 1996). A widely used MCMC method is the Gibbs sampler algorithm (Geman and Geman 1984, Gelfand and Smith 1990). We describe the implementation of this algorithm for the parameters of the gamma distribution in Appendix A.2.

1.3.2. Sampling the (Partial) Correlations

First, we describe the sampling of correlation $\rho(1,i)$. Since the focus is on the correlation between random variables $\Phi^{-1}[F_1(X_1;\mathbf{\Psi}_1)]$ and $\Phi^{-1}[F_i(X_i;\mathbf{\Psi}_i)]$, we provide an explicit

[2]A prior density function is said to be conjugate to a likelihood function if the resulting posterior density function is in the same family of distributions as the prior density function (Bernardo and Smith 1994).

representation of their joint density function $\phi_2(F_1(x_1;\boldsymbol{\Psi}_1), F_i(x_i;\boldsymbol{\Psi}_i);\boldsymbol{\Sigma}_2(1,i))$:

$$|\boldsymbol{\Sigma}_2(1,i)|^{-1/2}\exp\left\{-\frac{1}{2}\begin{pmatrix}\Phi^{-1}[F_1(X_1;\boldsymbol{\Psi}_1)]\\ \Phi^{-1}[F_i(X_i;\boldsymbol{\Psi}_i)]\end{pmatrix}'\left(\boldsymbol{\Sigma}_2^{-1}(1,i)-\mathbf{I}_2\right)\begin{pmatrix}\Phi^{-1}[F_1(X_1;\boldsymbol{\Psi}_1)]\\ \Phi^{-1}[F_i(X_i;\boldsymbol{\Psi}_i)]\end{pmatrix}\right\}$$

Thus, defining

$$\mathbf{S}_2(1,i|\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i) \equiv \sum_{t=1}^{n}\begin{pmatrix}\Phi^{-1}[F_1(x_{1,t};\boldsymbol{\Psi}_1)]\\ \Phi^{-1}[F_i(x_{i,t};\boldsymbol{\Psi}_i)]\end{pmatrix}\begin{pmatrix}\Phi^{-1}[F_1(x_{1,t};\boldsymbol{\Psi}_1)]\\ \Phi^{-1}[F_i(x_{i,t};\boldsymbol{\Psi}_i)]\end{pmatrix}'$$

and using the trace operator (tr), we can represent the likelihood function associated with the dependence structure (i.e., $\prod_{t=1}^{n}\phi_2(F_1(x_{1,t};\boldsymbol{\Psi}_1), F_i(x_{i,t};\boldsymbol{\Psi}_i);\boldsymbol{\Sigma}_2(1,i))$) as follows:

$$|\boldsymbol{\Sigma}_2(1,i)|^{-n/2}\ \exp\left\{\mathrm{tr}\left(-\frac{1}{2}\mathbf{S}_2(1,i|\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i)\left(\boldsymbol{\Sigma}_2^{-1}(1,i)-\mathbf{I}_2\right)\right)\right\} \tag{1.6}$$

The form of this likelihood function suggests the use of the inverse Wishart density function as a conjugate prior for $\boldsymbol{\Sigma}_2(1,i)$ (Rossi et al. 2006). Therefore, we follow Jeffreys' invariance principle (Kass and Wasserman 1996) and choose prior density function $\pi(\boldsymbol{\Sigma}_2(1,i)) \propto |\boldsymbol{\Sigma}_2(1,i)|^{-3/2}$ for correlation matrix $\boldsymbol{\Sigma}_2(1,i)$. The right-hand side of this representation corresponds to the inverse Wishart density function with zero degrees of freedom and serves as a diffuse prior density function. Furthermore, it coincides with the beta prior density function of Barnard et al. (2000) with $\nu = 0$.

Combining the prior density function $\pi(\boldsymbol{\Sigma}_2(1,i))$ with the likelihood function in (1.6) using Bayes' rule leads to the following conditional posterior copula density function:

$$p\left(\boldsymbol{\Sigma}_2(1,i)|\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i\right) \propto |\boldsymbol{\Sigma}_2(1,i)|^{-(n+3)/2}\exp\left\{\mathrm{tr}\left(-\frac{1}{2}\mathbf{S}_2(1,i|\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i)\boldsymbol{\Sigma}_2^{-1}(1,i)\right)\right\}$$

Thus, sampling $\rho(1,i)$ reduces to the sampling of the correlation matrix $\boldsymbol{\Sigma}_2(1,i)$ from the inverse Wishart density function with parameters n and $\mathbf{S}_2(1,i|\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i)$. An algorithm for sampling $\boldsymbol{\Sigma}_2(1,i)$ is provided in Appendix A.3.

Next, we consider the partial correlation $\rho(j-1,i;1,2,\ldots,j-2)$ between conditional random variables $\Phi^{-1}[F_{j-1}(X_{j-1};\boldsymbol{\Psi}_{j-1})]|\Phi^{-1}[F_\ell(X_\ell;\boldsymbol{\Psi}_\ell)]$, $\ell = 1,2,\ldots,j-2$ and $\Phi^{-1}[F_i(X_i;\boldsymbol{\Psi}_i)]|\Phi^{-1}[F_\ell(X_\ell;\ \boldsymbol{\Psi}_\ell)]$, $\ell = 1,2,\ldots,j-2$. Similarly, we choose the conjugate, inverse Wishart prior density function

$$\pi\left(\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)\right) \propto |\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)|^{-3/2}$$

for partial correlation matrix $\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)$. This leads to the conditional posterior copula density function $p(\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)|\boldsymbol{\Lambda}_j,\mathbf{x})$ of the form

$$\begin{aligned}&\propto |\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)|^{-(n+3)/2}\\ &\times \exp\left\{\mathrm{tr}\left(-\frac{1}{2}\mathbf{S}_2(j-1,i;1,2,\ldots,j-2|\boldsymbol{\Lambda}_j,\mathbf{x})\,\boldsymbol{\Sigma}_2^{-1}(j-1,i;1,2,\ldots,j-2)\right)\right\},\end{aligned}$$

where $\mathbf{x}$ is the $kn-$dimensional vector of the historical data; $\mathbf{\Lambda}_j$ is the vector of NORTA parameters $\mathbf{\Psi}_m$, $m = 1, 2, \ldots, k$, $\rho(1, m)$, $m = 2, 3, \ldots, k$, and $\rho(\ell - 1, m; 1, 2, \ldots, \ell - 2)$, $m = \ell, \ell + 1, \ldots, k$, $\ell = 3, 4, \ldots, j - 1$; and $\mathbf{S}_2(j - 1, i; 1, 2, \ldots, j - 2|\mathbf{\Lambda}_j, \mathbf{x})$ is the two-dimensional matrix defined by

$$\sum_{t=1}^{n} \begin{pmatrix} \frac{\Phi^{-1}[F_{j-1}(x_{j-1,t};\mathbf{\Psi}_{j-1})]-\mu_{j-1|1,2,\ldots,j-2}}{\sigma_{j-1|1,2,\ldots,j-2}} \\ \frac{\Phi^{-1}[F_i(x_{i,t};\mathbf{\Psi}_i)]-\mu_{i|1,2,\ldots,j-2}}{\sigma_{i|1,2,\ldots,j-2}} \end{pmatrix} \begin{pmatrix} \frac{\Phi^{-1}[F_{j-1}(x_{j-1,t};\mathbf{\Psi}_{j-1})]-\mu_{j-1|1,2,\ldots,j-2}}{\sigma_{j-1|1,2,\ldots,j-2}} \\ \frac{\Phi^{-1}[F_i(x_{i,t};\mathbf{\Psi}_i)]-\mu_{i|1,2,\ldots,j-2}}{\sigma_{i|1,2,\ldots,j-2}} \end{pmatrix}'.$$

Therefore, sampling $\rho(j-1, i; 1, 2, \ldots, j-2)$ reduces to the sampling of the partial correlation matrix $\mathbf{\Sigma}_2(j-1, i; 1, 2, \ldots, j-2)$ from the inverse Wishart density function with parameters n and $\mathbf{S}_2(j - 1, i; 1, 2, \ldots, j - 2|\mathbf{\Lambda}_j, \mathbf{x})$.

An alternative to the use of conjugate, inverse Wishart priors for the (partial) correlations of the copula-vine specification is to use non-informative priors (Liechty et al. 2004). These priors include Jeffreys' prior (Jeffreys 1961), log-matrix prior (Leonard and Hsu 1992), reference prior (Berger and Sun 2008), uniform shrinkage prior (Daniels 1999), and uniform prior (Barnard et al. 2000). Barnard et al. (2000) further note that $\pi(\rho) = (1 - \rho^2)^{(\nu-3)/2}$, which is a beta density function with shape parameters $(\nu-1)/2$ as well as a uniform density function for $\nu = 3$, can be chosen as a prior density function for $\rho \in [-1, 1]$. Selecting any of these prior density functions for the (partial) correlation requires the use of the MCMC method. Therefore, more computational effort is needed than that required by the sampling of the (partial) correlation from the inverse Wishart density function.

1.3.3. Sampling All NORTA Parameters

Motivated by the decomposition of the joint pdf in (1.4) into separate terms associated with the component marginal distributions and the (partial) correlations, we independently choose a prior density function for each of the NORTA parameters. Specifically, we select prior density functions $\pi_i(\mathbf{\Psi}_i)$, $i = 1, 2, \ldots, k$ for component parameter vectors $\mathbf{\Psi}_i$, $i = 1, 2, \ldots, k$, utilizing the well-established Bayesian literature on standard families of distribution as described in Section 1.3.1. Assuming the probabilistic independence of the (partial) correlations, we choose prior density functions $\pi(\mathbf{\Sigma}_2(1, i))$, $i = 2, 3, \ldots, k$ for correlation matrices $\mathbf{\Sigma}_2(1, i)$, $i = 2, 3, \ldots, k$, and $\pi(\mathbf{\Sigma}_2(j - 1, i; 1, 2, \ldots, j - 2))$, $i = j, j + 1, \ldots, k$, $j = 3, 4, \ldots, k$ for partial correlation matrices $\mathbf{\Sigma}_2(j - 1, i; 1, 2, \ldots, j - 2)$, $i = j, j + 1, \ldots, k$, $j = 3, 4, \ldots, k$. The selection of the prior density functions for the (partial) correlation matrices is discussed in Section 1.3.2.

Since different regular vine specifications are characterized by different (partial) correlations, the probabilistic independence of the partial correlations of the $\mathcal{C}-$vine does not imply the probabilistic independence of the partial correlations of any other regular vine. Therefore, if the simulation analyst chooses to represent the dependence structure of the NORTA distribution with a different regular vine, then she must assume the probabilistic independence of the partial correlations of that vine. This assumption not only provides flexibility in choosing prior density functions for NORTA's dependence parameters, but also allows the use of the existing literature on Bayesian inference for correlation matrices without being challenged by the high-dimensional nature of the large-scale stochastic simulations.

Assuming the availability of the $k-$dimensional historical data of length n (i.e., $x_{i,t}$, $i = 1, 2, \ldots, k$, $t = 1, 2, \ldots, n$), we use Bayes' rule for combining the prior density functions with the likelihood function $\prod_{t=1}^{n} h_k(x_{1,t}, x_{2,t}, \ldots, x_{k,t})$, i.e.,

$$
\begin{aligned}
& \prod_{t=1}^{n}\prod_{i=1}^{k} f_i(x_{i,t};\boldsymbol{\Psi}_i) \\
\times\ & \prod_{t=1}^{n}\prod_{i=2}^{k} \phi_2\Big(F_1(x_{1,t};\boldsymbol{\Psi}_1), F_i(x_{i,t};\boldsymbol{\Psi}_i); \boldsymbol{\Sigma}_2(1,i)\Big) \\
\times\ & \prod_{t=1}^{n}\prod_{j=3}^{k}\prod_{i=j}^{k} \phi_2\Bigg(\Phi_{j-1|1,2,\ldots,j-2}\left(\Phi^{-1}[F_{j-1}(x_{j-1,t};\boldsymbol{\Psi}_{j-1})]|\Phi^{-1}[F_\ell(x_{\ell,t};\boldsymbol{\Psi}_\ell)], \ell = 1, 2, \ldots, j-2\right), \\
& \qquad\qquad \Phi_{i|1,2,\ldots,j-2}\left(\Phi^{-1}[F_i(x_{i,t};\boldsymbol{\Psi}_i)]|\Phi^{-1}[F_\ell(x_{\ell,t};\boldsymbol{\Psi}_\ell)], \ell = 1, 2, \ldots, j-2\right); \\
& \qquad\qquad \boldsymbol{\Sigma}_2(j-1, i; 1, 2, \ldots, j-2)\Bigg),
\end{aligned}
$$

and obtain the following joint posterior density function:

$$
\begin{aligned}
\propto\ & \prod_{i=1}^{k} \overbrace{\left[\pi_i(\boldsymbol{\Psi}_i)\prod_{t=1}^{n} f_i(x_{i,t};\boldsymbol{\Psi}_i)\right]}^{p_i(\boldsymbol{\Psi}_i|\mathbf{x}_i)} \\
\times\ & \prod_{i=2}^{k} \overbrace{\left[\pi\left(\boldsymbol{\Sigma}_2(1,i)\right)\prod_{t=1}^{n} \phi_2\Big(F_1(x_{1,t};\boldsymbol{\Psi}_1), F_i(x_{i,t};\boldsymbol{\Psi}_i); \boldsymbol{\Sigma}_2(1,i)\Big)\right]}^{p(\boldsymbol{\Sigma}_2(1,i)|\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i)}
\end{aligned}
$$

$$\overbrace{\times \prod_{j=3}^{k}\prod_{i=j}^{k}\Bigg[\pi\left(\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)\right)}^{p(\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)|\boldsymbol{\Lambda}_j,\mathbf{x})}$$

$$\prod_{t=1}^{n}\phi_2\Big(\Phi_{j-1|1,2,\ldots,j-2}\left(\Phi^{-1}[F_{j-1}(x_{j-1,t};\boldsymbol{\Psi}_{j-1})]|\Phi^{-1}[F_\ell(x_{\ell,t};\boldsymbol{\Psi}_\ell)],\ell=1,2,\ldots,j-2\right),$$

$$\Phi_{i|1,2,\ldots,j-2}\left(\Phi^{-1}[F_i(x_{i,t};\boldsymbol{\Psi}_i)]|\Phi^{-1}[F_\ell(x_{\ell,t};\boldsymbol{\Psi}_\ell)],\ell=1,2,\ldots,j-2\right);$$

$$\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)\Big)\Bigg] \tag{1.7}$$

Thus, the joint posterior density function of the NORTA parameters is the product of the k marginal posterior density functions $p_i(\boldsymbol{\Psi}_i|\mathbf{x}_i)$, $i=1,2,\ldots,k$; the two-dimensional posterior copula density functions $p(\boldsymbol{\Sigma}_2(1,i)|\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i)$, $i=2,3,\ldots,k$ associated with the first tree of the $\mathcal{C}-$vine; and the two-dimensional posterior copula density functions $p(\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)|\boldsymbol{\Lambda}_j,\mathbf{x})$, $i=j,j+1,\ldots,k$ associated with the $(j-1)^{\text{th}}$ tree of the $\mathcal{C}-$vine for $j=3,4,\ldots,k$. Appendix B provides an explicit representation of this posterior density function for the $5-$dimensional NORTA random vector with exponentially distributed components.

The form of the joint posterior density function in (1.7) allows us to develop a fast, $(k(k+1)/2)-$stage algorithm for sampling the NORTA parameters. We provide the resulting algorithm in Figure 1.2. In the first k stages, we sample parameter vectors $\boldsymbol{\Psi}_i$, $i=1,2,\ldots,k$ (i.e., $\tilde{\boldsymbol{\Psi}}_i$, $i=1,2,\ldots,k$) from posterior density functions $p_i(\boldsymbol{\Psi}_i|\mathbf{x}_i)$, $i=1,2,\ldots,k$, as described in Section 1.3.1 for exponentially and gamma distributed components. This allows us to account for the uncertainty around the parameters of the marginal distributions of the components. The next $k-1$ stages sample $\boldsymbol{\Sigma}_2(1,i)$, $i=2,3,\ldots,k$ (i.e., $\tilde{\boldsymbol{\Sigma}}_2(1,i)$, $i=2,3,\ldots,k$) from posterior density functions $p(\boldsymbol{\Sigma}_2(1,i)|\tilde{\boldsymbol{\Psi}}_1,\tilde{\boldsymbol{\Psi}}_i,\mathbf{x}_1,\mathbf{x}_i)$, $i=2,3,\ldots,k$ using $\tilde{\boldsymbol{\Psi}}_i$, $i=1,2,\ldots,k$ obtained in the first k stages of the algorithm. We sample each of these $k-1$ correlation matrices as described in Section 1.3.2 and set the $(1,2)^{\text{th}}$ entry of $\tilde{\boldsymbol{\Sigma}}_2(1,i)$ to $\tilde{\rho}(1,i)$ for $i=2,3,\ldots,k$. This allows us to account for the uncertainty around the (unconditional) correlations of the $\mathcal{C}-$vine specification. In the remaining stages, we capture the uncertainty around the partial correlations. To do this, we first construct vector $\tilde{\boldsymbol{\Lambda}}_j$ with the sampled NORTA parameters; i.e., $\tilde{\boldsymbol{\Psi}}_m$, $m=1,2,\ldots,k$, $\tilde{\rho}(1,m)$, $m=2,3,\ldots,k$, and $\tilde{\rho}(\ell-1,m;1,2,\ldots,\ell-2)$, $m=\ell,\ell+1,\ldots,k$, $\ell=3,4,\ldots,j-1$. Then, we characterize the conditional normal cdfs $\Phi_{i|1,2,\ldots,j-2}$ and $\Phi_{j-1|1,2,\ldots,j-2}$ by obtaining their means $\mu_{i|1,2,\ldots,j-2}$ and $\mu_{j-1|1,2,\ldots,j-2}$ and variances $\sigma^2_{i|1,2,\ldots,j-2}$ and $\sigma^2_{j-1|1,2,\ldots,j-2}$ from $\tilde{\boldsymbol{\Lambda}}_j$ via Theorem 3.3.4 of

Tong (1990) and the recursive formulas of Yule and Kendall (1965). We use the resulting cdfs for determining the posterior density function $p(\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)|\tilde{\boldsymbol{\Lambda}}_j,\mathbf{x})$. We sample $\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)$ (i.e., $\tilde{\boldsymbol{\Sigma}}_2(j-1,i;1,2,\ldots,j-2)$) from this posterior density function as described in Section 1.3.2 and set the $(1,2)^{\text{th}}$ entry of $\tilde{\boldsymbol{\Sigma}}_2(j-1,i;1,2,\ldots,j-2)$ to $\tilde{\rho}(j-1,i;1,2,\ldots,j-2)$. Repeating this for $i=j,j+1,\ldots,k$ and $j=3,4,\ldots,k$ allows us to account for the uncertainty around the partial correlations of the $\mathcal{C}-$vine specification. Appendix B provides a detailed implementation of this NORTA parameter sampling algorithm for the 5−dimensional NORTA distribution with exponentially distributed components.

Figure 1.2: An algorithm that samples NORTA parameters for capturing parameter uncertainty in a stochastic simulation with k correlated inputs having a $k-$dimensional NORTA distribution.

for $i=1,2,\ldots,k$ **do**
  sample the i^{th} component parameter vector $\boldsymbol{\Psi}_i$ (i.e., $\tilde{\boldsymbol{\Psi}}_i$) from $p_i(\boldsymbol{\Psi}_i|\mathbf{x}_i)$
end loop
for $i=2,3,\ldots,k$ **do**
  sample correlation matrix $\boldsymbol{\Sigma}_2(1,i)$ (i.e., $\tilde{\boldsymbol{\Sigma}}_2(1,i)$) from $p(\boldsymbol{\Sigma}_2(1,i)|\tilde{\boldsymbol{\Psi}}_1,\tilde{\boldsymbol{\Psi}}_i,\mathbf{x}_1,\mathbf{x}_i)$
end loop
for $j=3,4,\ldots,k$ **do**
  construct vector $\tilde{\boldsymbol{\Lambda}}_j$ consisting of the sampled NORTA parameters
  for $i=j,j+1,\ldots,k$ **do**
    (a) compute the means and variances of cdfs $\Phi_{i|1,2,\ldots,j-2}$ and $\Phi_{j-1|1,2,\ldots,j-2}$ using $\tilde{\boldsymbol{\Lambda}}_j$, Theorem 3.3.4 of Tong (1990), and recursive formulas of Yule and Kendall (1965), and insert them into the density function $p(\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)|\tilde{\boldsymbol{\Lambda}}_j,\mathbf{x})$
    (b) sample partial correlation matrix $\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)$ (i.e., $\tilde{\boldsymbol{\Sigma}}_2(j-1,i;1,2,\ldots,j-2)$) from the posterior density function $p(\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)|\tilde{\boldsymbol{\Lambda}}_j,\mathbf{x})$
  end loop
end loop

1.4. Estimation of the Mean Performance Measure and the Confidence Interval

In this section, we let Y be the performance measure whose mean is relevant to the decision-making process. Our goal is to incorporate the Bayesian model of Section 1.3 into the simulation replication algorithm of Chick (2001) in Section 1.4.1 and into the Bayesian simulation replication algorithm of Zouaoui and Wilson (2003) in Section 1.4.2 with the purpose of generating a point estimate and a confidence interval of $\mathrm{E}_{Y|\mathbf{x}}(Y|\mathbf{x})$ (i.e., the posterior mean of the output random variables Y_r, $r = 1, 2, \ldots, R$ of a stochastic simulation with R replications, given the historical input data vector $\mathbf{x}$ and the prior information about the NORTA parameters).

1.4.1. Simulation Replication Algorithm

We present the simulation replication algorithm for our $k-$dimensional input process in Figure 1.3. Step 1 uses our Bayesian model to sample NORTA parameters for each of the R replications of Step 2. We denote the parameters sampled in the r^{th} replication by $\tilde{\mathbf{\Psi}}_i^r$, $i = 1, 2, \ldots, k$, $\tilde{\rho}^r(1, i)$, $i = 2, 3, \ldots, k$, and $\tilde{\rho}^r(j-1, i; 1, 2, \ldots, j-2)$, $i = j, j+1, \ldots, k$, $j = 3, 4, \ldots, k$. Step 2 captures both stochastic uncertainty and parameter uncertainty by generating k correlated input variates $\tilde{x}_i^r$, $i = 1, 2, \ldots, k$ from the $k-$dimensional NORTA distribution with parameters $\tilde{\mathbf{\Psi}}_i^r$, $i = 1, 2, \ldots, k$, $\tilde{\rho}^r(1, i)$, $i = 2, 3, \ldots, k$, and $\tilde{\rho}^r(j-1, i; 1, 2, \ldots, j-2)$, $i = j, j+1, \ldots, k$, $j = 3, 4, \ldots, k$ for $r = 1, 2, \ldots, R$. Thus, the key difference of this algorithm from the one presented in Chick (2001) is its first step where the NORTA parameters are sampled from their Bayesian posterior density functions.

The proper implementation of the simulation replication algorithm requires the consideration of two important issues: the independent sampling of the NORTA parameters for each of the R replications and the analysis of the simulation output data y_r, $r = 1, 2, \ldots, R$ for estimating a point estimate and a confidence interval of $\mathrm{E}_{Y|\mathbf{x}}(Y|\mathbf{x})$. The first issue arises when the prior density functions chosen for the component parameter vectors and/or the (partial) correlations are not conjugate. In this case, we sample the NORTA parameters using the Gibbs sampler algorithm. When this algorithm is used for generating a distribution parameter from its posterior density function, there often appears auto-correlations between the values sampled within the chain. There might also appear cross-correlations between different parameters sampled in different chains. However, the simulation replica-

Figure 1.3: The simulation replication algorithm that captures stochastic uncertainty and parameter uncertainty, and generates a point estimate and a confidence interval of $\mathrm{E}_{Y|\mathbf{x}}(Y|\mathbf{x})$.

Step 1

for $i = 1, 2, \ldots, k$ **do**

sample R independent variates of component parameter vector $\mathbf{\Psi}_i$ (i.e., $\tilde{\mathbf{\Psi}}_i^r$, $r = 1, 2, \ldots, R$) from $p_i(\mathbf{\Psi}_i|\mathbf{x}_i)$, independent of the parameters generated for other components

end loop

for $i = 2, 3, \ldots, k$ **do**

for $r = 1, 2, \ldots, R$ replications **do**

sample $\mathbf{\Sigma}_2(1, i)$ (i.e., $\tilde{\mathbf{\Sigma}}_2^r(1, i)$) from $p(\mathbf{\Sigma}_2(1, i)|\tilde{\mathbf{\Psi}}_1^r, \tilde{\mathbf{\Psi}}_i^r, \mathbf{x}_1, \mathbf{x}_i)$, independent of the correlation matrices generated in other replications, and set $\tilde{\mathbf{\Sigma}}_2^r(1, i)[1, 2]$ to $\tilde{\rho}^r(1, i)$

end loop

end loop

for $j = 3, 4, \ldots, k$ **do**

for $i = j, j + 1, \ldots, k$ **do**

for $r = 1, 2, \ldots, R$ replications **do**

sample $\mathbf{\Sigma}_2(j - 1, i; 1, 2, \ldots, j - 2)$ (i.e., $\tilde{\mathbf{\Sigma}}_2^r(j - 1, i; 1, 2, \ldots, j - 2)$) from $p(\mathbf{\Sigma}_2(j - 1, i; 1, 2, \ldots, j - 2)|\tilde{\mathbf{\Lambda}}_j^r, \mathbf{x})$, independent of the partial correlation matrices of other replications, and set $\tilde{\mathbf{\Sigma}}_2^r(j - 1, i; 1, 2, \ldots, j - 2)[1, 2]$ to $\tilde{\rho}^r(j - 1, i; 1, 2, \ldots, j - 2)$

end loop

end loop

end loop

Step 2

for $r = 1, 2, \ldots, R$ replications **do**

sample input random variates (i.e., $\tilde{x}_i^r$, $i = 1, 2, \ldots, k$) given NORTA parameters of the r^{th} replication (i.e., $\tilde{\mathbf{\Psi}}_i^r$, $i = 1, 2, \ldots, k$, $\tilde{\rho}^r(1, i)$, $i = 2, 3, \ldots, k$, $\tilde{\rho}^r(j - 1, i; 1, 2, \ldots, j - 2)$, $i = j, j + 1, \ldots, k$, $j = 3, 4, \ldots, k$) and calculate the output y_r as a function of the input random variates $\tilde{x}_i^r$, $i = 1, 2, \ldots, k$

end loop

analyze the simulation output data y_r, $r = 1, 2, \ldots, R$ and generate a point estimate and a confidence interval of $\mathrm{E}_{Y|\mathbf{x}}(Y|\mathbf{x})$

tion algorithm requires the independent sampling of the NORTA parameters for each of its replications. Therefore, it is important to implement the Gibbs sampler algorithm in a way that it provides statistically independent values of a NORTA parameter for each of the R replications. Appendix C describes how to do this using the method of batching.

There are also cases in which the sampling of the NORTA parameters does not require the use of an MCMC method and thus, the independent sampling of the NORTA parameters is easy to achieve. One such case occurs when the components of the NORTA vector are exponentially distributed and conjugate inverse gamma density functions are chosen as the priors for the scale parameters of the components, while conjugate inverse Wishart prior density functions are used for the (partial) correlations. Therefore, we can easily generate R independent sets of NORTA parameters using well-known algorithms for sampling from gamma and Wishart density functions (Appendix A).

Despite obtaining statistically independent sets of NORTA parameters in the first step of the simulation replication algorithm, this is an approximation to the independent sampling of the NORTA parameters when the prior density functions chosen for the component parameter vectors and/or the (partial) correlations are not conjugate. The failure to independently sample NORTA parameters often leads to dependent simulation output data. Therefore, the second issue that might arise in the implementation of the simulation replication algorithm is related to the lack of independence in the output data y_r, $r = 1, 2, \ldots, R$. Law (2007) provides an excellent overview of the methods that have been proposed for the analysis of dependent simulation output data. In Appendix D, we describe how to use the method of batching for analyzing the (dependent) output data y_r, $r = 1, 2, \ldots, R$ to obtain a point estimate and a confidence interval of $\mathrm{E}_{Y|\mathbf{x}}(Y|\mathbf{x})$.

1.4.2. Bayesian Simulation Replication Algorithm

We provide the Bayesian simulation replication algorithm for our $k-$dimensional input process in Figure 1.4. We use $\mathbf{\Psi}_i$, $i = 1, 2, \ldots, k$ to represent the vectors of the marginal distribution parameters, and $\mathbf{\Psi}_i^{m_i}$ to denote the marginal-distribution parameters sampled in the m_i^{th} replication. Similarly, $\mathbf{\Lambda}$ represents all (conditional and unconditional) two-dimensional correlation matrices of the C-vine specification; i.e., $\mathbf{\Lambda}$ is composed of $\mathbf{\Sigma}_2(1, i), i = 2, 3, \ldots, k$, and $\mathbf{\Sigma}_2(j-1, i; 1, 2, \ldots, j-2), i = j, j+1, \ldots, k, j = 3, 4, \ldots, k$. $\mathbf{\Lambda}^d$, on the other hand, denotes all (conditional and unconditional) correlation matrices sampled in the d^{th} replication. Furthermore, we use notation $\mathbf{x} = \{x_{i,t}; i = 1, 2, \ldots, k, t = 1, 2, \ldots, n\}$ for the

$k-$dimensional historical input data of length n, and $y_{m_1,m_2,\ldots,m_k,d,r}$ for the output response from the r^{th} simulation run using the random-number input $\mathbf{u}^r$ and the sampled input parameters $\mathbf{\Psi}_i^{m_i}$, $i = 1, 2, \ldots, k$, and $\mathbf{\Lambda}^d$. The first (i.e., the most outer) loop estimates the uncertainty around the parameter vector $\mathbf{\Psi}_1$ of the marginal distribution of the first component, while the second loop estimates the uncertainty around the parameter vector $\mathbf{\Psi}_2$ of the marginal distribution of the second component. Similarly, the k^{th} loop estimates the uncertainty around the parameter vector $\mathbf{\Psi}_k$ of the marginal distribution of the k^{th} component. The $(k+1)^{\text{th}}$ loop, on the other hand, estimates the uncertainty around the dependence parameters, while the $(k+2)^{\text{th}}$ (i.e., the most inner) loop estimates the stochastic uncertainty. Due to the use of the copula-vine specification to represent the uncertainty around the dependence parameter $\mathbf{\Lambda}$, it is possible to decompose this uncertainty into parts associated with individual correlations and partial correlations. However, for ease of presentation, we represent the uncertainty around $\mathbf{\Lambda}$ in a single loop in Figure 1.4.

The analysis of the next section assumes the independence of the simulation output data $y_{m_1,m_2,\ldots,m_k,d,r}$, $m_i = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, k$, $d = 1, 2, \ldots, D$, and $r = 1, 2, \ldots, R$; we refer the reader to Appendix D for the analysis of the possibly dependent simulation output data for the proper execution of the Bayesian simulation replication algorithm.

1.5. Output Variance Decomposition

In this section, we use the output data obtained from the execution of the Bayesian simulation replication algorithm, and estimate both a point estimate for the mean posterior response $E_{Y|\mathbf{x}}(Y|\mathbf{x})$ and a posterior response variance as a function of stochastic uncertainty (λ^2), dependence-parameter uncertainty ($\theta^2_{\mathbf{\Lambda}}$), and marginal-distribution uncertainty for each simulation input ($\theta^2_{\mathbf{\Psi}_i}$, $i = 1, 2, \ldots, k$). Following Zouaoui and Wilson (2003), we express the output response from the r^{th} simulation run as follows:

$$\begin{aligned} y_{m_1,m_2,\ldots,m_k,d,r} &= y\left(\mathbf{u}^r, \mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_k^{m_k}, \mathbf{\Lambda}^d\right) \\ &= \eta\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_k^{m_k}, \mathbf{\Lambda}^d\right) + e_r\left(\mathbf{u}^r, \mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_k^{m_k}, \mathbf{\Lambda}^d\right) \quad (1.8) \end{aligned}$$

The error $e_r(\mathbf{u}^r, \mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_k^{m_k}, \mathbf{\Lambda}^d)$ is the deviation of the simulation output $y_{m_1,m_2,\ldots,m_k,d,r}$ from the response-surface $\eta(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_k^{m_k}, \mathbf{\Lambda}^d)$ due to the stochastic uncertainty whose source is the random-number input $\mathbf{u}^r$ for that run. Under the assumptions of

$$\mathrm{E}_{\mathbf{u}^r}\left[e_r\left(\mathbf{u}^r, \mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_k^{m_k}, \mathbf{\Lambda}^d\right) \mid \mathbf{x}, \mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_k^{m_k}, \mathbf{\Lambda}^d\right] = 0 \quad (1.9)$$

Figure 1.4: Bayesian simulation replication algorithm for the $k-$dimensional input process.

for $m_1 = 1, 2, \ldots, M_1$ replications **do**

generate the m_1^{th} sample parameter vector $\mathbf{\Psi}_1^{m_1}$ from $p_1(\mathbf{\Psi}_1|\mathbf{x}_1)$;

set the parameter vector $\mathbf{\Psi}_1 \leftarrow \mathbf{\Psi}_1^{m_1}$;

for $m_2 = 1, 2, \ldots, M_2$ replications **do**

generate the m_2^{th} sample parameter vector $\mathbf{\Psi}_2^{m_2}$ from $p_2(\mathbf{\Psi}_2|\mathbf{x}_2)$;

set the parameter vector $\mathbf{\Psi}_2 \leftarrow \mathbf{\Psi}_2^{m_2}$;

$\vdots$ $\vdots$ $\vdots$

for $m_k = 1, 2, \ldots, M_k$ replications **do**

generate the m_k^{th} sample parameter vector $\mathbf{\Psi}_k^{m_k}$ from $p_k(\mathbf{\Psi}_k|\mathbf{x}_k)$;

set the parameter vector $\mathbf{\Psi}_k \leftarrow \mathbf{\Psi}_k^{m_k}$;

for $d = 1, 2, \ldots, D$ replications **do**

generate the d^{th} sample dependence vector $\mathbf{\Lambda}^d$ from $p(\mathbf{\Lambda}|\mathbf{\Psi}_i, i = 1, 2, \ldots, k, \mathbf{x})$;

set the dependence vector $\mathbf{\Lambda} \leftarrow \mathbf{\Lambda}^d$;

for $r = 1, \ldots, R$ **do**

set the random number input $\mathbf{u} \leftarrow \mathbf{u}^r$;

perform the r^{th} simulation run using $\mathbf{u}$, $\mathbf{\Psi}_i$, $i = 1, 2, \ldots, k$, and $\mathbf{\Lambda}$;

calculate the output response $y_{m_1,m_2,\ldots,m_k,d,r} = y(\mathbf{u}, \mathbf{\Psi}_1, \mathbf{\Psi}_2, \ldots, \mathbf{\Psi}_k, \mathbf{\Lambda})$;

end for

compute $y_{m_1,m_2,\ldots,m_k,d} = \frac{1}{R}\sum_{r=1}^{R} y_{m_1,m_2,\ldots,m_k,d,r}$;

end for

compute $y_{m_1,m_2,\ldots,m_k} = \frac{1}{D}\sum_{d=1}^{D} y_{m_1,m_2,\ldots,m_k,d}$;

end for

$\vdots$ $\vdots$ $\vdots$

compute $y_{m_1,m_2} = \frac{1}{M_3}\sum_{m_3=1}^{M_3} y_{m_1,m_2,m_3}$;

end for

compute $y_{m_1} = \frac{1}{M_2}\sum_{m_2=1}^{M_2} y_{m_1,m_2}$;

end for

compute $\bar{y} = \frac{1}{M_1}\sum_{m_1=1}^{M_1} y_{m_1}$, which is an unbiased estimate of $E_{Y|\mathbf{x}}(Y \mid \mathbf{x})$

and

$$\mathrm{Var}_{\mathbf{u}^r}\left[e_r\left(\mathbf{u}^r, \boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right) \mid \mathbf{x}, \boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right] \;=\; \lambda^2, \quad (1.10)$$

it holds that

$$\mathrm{E}_{\mathbf{u}^r}\left[y_{m_1,m_2,\ldots,m_k,d,r} \mid \mathbf{x}, \boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right] \;=\; \eta\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right)$$

and

$$\mathrm{Var}_{\mathbf{u}^r}\left[y_{m_1,m_2,\ldots,m_k,d,r} \mid \mathbf{x}, \boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right] \;=\; \lambda^2$$

for $m_i = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, k$, $d = 1, 2, \ldots, D$, and $r = 1, \ldots, R$, where λ^2 represents the stochastic uncertainty. This response surface model is known as the classical random effects model in the statistics literature (Rao 1997). Its use allows us to estimate stochastic uncertainty λ^2 under the assumption of constant error variance; i.e., λ^2 does not depend on the parameters of the multivariate input distribution. Next, we assume that

$$\begin{aligned} \eta\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right) &= \omega\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}\right) \\ &+ \varpi_d\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right) \end{aligned} \quad (1.11)$$

for $m_i = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, k$, and $d = 1, 2, \ldots, D$, where $\varpi_d\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right)$ is the deviation of $\eta\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right)$ from the response surface $\omega\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}\right)$ due to the uncertainty associated with the underlying dependence structure captured in $\boldsymbol{\Lambda}^d$. Under the assumptions of

$$\mathrm{E}_{\boldsymbol{\Lambda}^d}\left[\varpi_d\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right) \mid \mathbf{x}, \boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}\right] \;=\; 0 \quad (1.12)$$

and

$$\mathrm{Var}_{\boldsymbol{\Lambda}^d}\left[\varpi_d\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right) \mid \mathbf{x}, \boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}\right] \;=\; \theta_{\boldsymbol{\Lambda}}^2, \quad (1.13)$$

it holds that

$$\mathrm{E}_{\boldsymbol{\Lambda}^d}\left[\eta\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right) \mid \mathbf{x}, \boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}\right] \;=\; \omega\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}\right)$$

and

$$\mathrm{Var}_{\boldsymbol{\Lambda}^d}\left[\eta\left(\boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}, \boldsymbol{\Lambda}^d\right) \mid \mathbf{x}, \boldsymbol{\Psi}_1^{m_1}, \boldsymbol{\Psi}_2^{m_2}, \ldots, \boldsymbol{\Psi}_k^{m_k}\right] \;=\; \theta_{\boldsymbol{\Lambda}}^2$$

for $m_i = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, k$, and $d = 1, 2, \ldots, D$, where $\theta^2_{\mathbf{\Lambda}}$ represents the uncertainty in the dependence structure of the multivariate input process.

Finally, we express the output variability due to the uncertainty associated with the marginal distribution of the i^{th} simulation input using

$$\begin{aligned} \varphi^i\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_i^{m_i}\right) &= \varphi^{i-1}\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_{i-1}^{m_{i-1}}\right) \\ &+ \kappa^i_{m_i}\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_i^{m_i}\right) \end{aligned} \tag{1.14}$$

for $m_i = 1, 2, \ldots, M_i$. Specifically, $\kappa^i_{m_i}\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_i^{m_i}\right)$ is the deviation of $\varphi^i\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_i^{m_i}\right)$ from the response surface $\varphi^{i-1}\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_{i-1}^{m_{i-1}}\right)$ due to the uncertainty in the i^{th} marginal distribution parameters. Under the assumptions of

$$\mathrm{E}_{\mathbf{\Psi}_i}\left[\kappa^i_{m_i}\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_i^{m_i}\right) \mid \mathbf{x}, \mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_{i-1}^{m_{i-1}}\right] = 0 \tag{1.15}$$

and

$$\mathrm{Var}_{\mathbf{\Psi}_i}\left[\kappa^i_{m_i}\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_i^{m_i}\right) \mid \mathbf{x}, \mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_{i-1}^{m_{i-1}}\right] = \theta^2_{\mathbf{\Psi}_i}, \tag{1.16}$$

where $\beta = \mathrm{E}_{\mathbf{\Psi}_1}\left[\varphi^1\left(\mathbf{\Psi}_1\right)|\mathbf{x}\right]$ is an unbiased estimator of the mean posterior response $E_{Y|\mathbf{x}}\left(Y \mid \mathbf{x}\right)$, it holds that

$$\mathrm{E}_{\mathbf{\Psi}_i}\left[\varphi^i\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_i^{m_i}\right) \mid \mathbf{x}, \mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_{i-1}^{m_{i-1}}\right] = \varphi^{i-1}\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_{i-1}^{m_{i-1}}\right)$$

and

$$\mathrm{Var}_{\mathbf{\Psi}_i}\left[\varphi^i\left(\mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_i^{m_i}\right) \mid \mathbf{x}, \mathbf{\Psi}_1^{m_1}, \mathbf{\Psi}_2^{m_2}, \ldots, \mathbf{\Psi}_{i-1}^{m_{i-1}}\right] = \theta^2_{\mathbf{\Psi}_i},$$

where $\sigma^2_{\mathbf{\Psi}_i}$ stands for the uncertainty associated with the marginal distribution parameters of the i^{th} input.

Based on the assumptions (1.8)-(1.16), the posterior response variance can be written as $\mathrm{Var}\left(y|\mathbf{x}\right) = \lambda^2 + \theta^2_{\mathbf{\Lambda}} + \sum_{i=1}^{k} \theta^2_{\mathbf{\Psi}_i}$; i.e., the sum of the three variance components that quantify, respectively, the stochastic uncertainty, the uncertainty in the parameters of the dependence structure, and the uncertainty in the marginal distribution parameters.

Using the simulation output data obtained from the Bayesian simulation replication algorithm of Figure 1.4 and the well-established theory on the classical random-effects model (Rao 1997), we estimate β, λ^2, $\theta^2_{\mathbf{\Lambda}}$, and $\theta^2_{\mathbf{\Psi}_i}$, $i = 1, 2, \ldots, k$ as follows:

$$\hat{\beta} = \bar{y}$$

$$
\begin{aligned}
\hat{\lambda}^2 &= \frac{1}{\prod_{\ell=1}^{k} M_\ell D\,(R-1)} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \cdots \sum_{m_k=1}^{M_k} \sum_{d=1}^{D} \sum_{r=1}^{R} \left(y_{m_1,m_2,\ldots,m_k,d,r} - y_{m_1,m_2,\ldots,m_k,d}\right)^2 \\
\hat{\theta}^2_{\mathbf{\Lambda}} &= \frac{1}{D-1} \sum_{d=1}^{D} \left(y_{m_1,m_2,\ldots,m_k,d} - y_{m_1,m_2,\ldots,m_k}\right)^2 - \frac{\hat{\lambda}^2}{R} \\
\hat{\theta}^2_{\mathbf{\Psi}_k} &= \frac{1}{M_k} \sum_{m_k=1}^{M_k} \left(y_{m_1,m_2,\ldots,m_k} - y_{m_1,m_2,\ldots,m_{k-1}}\right)^2 - \frac{\hat{\theta}^2_{\mathbf{\Lambda}}}{D} - \frac{\hat{\lambda}^2}{DR} \\
\hat{\theta}^2_{\mathbf{\Psi}_i} &= \frac{1}{M_i - 1} \sum_{m_i=1}^{M_i} \left(y_{m_1,m_2,\ldots,m_i} - y_{m_1,m_2,\ldots,m_{i-1}}\right)^2 \\
&\quad - \sum_{\ell=i+1}^{k} \frac{\hat{\theta}^2_{\mathbf{\Psi}_\ell}}{\prod_{s=i+1}^{k} M_s} - \frac{\hat{\theta}^2_{\mathbf{\Lambda}}}{\prod_{s=i+1}^{k} M_s D} - \frac{\hat{\lambda}^2}{\prod_{s=i+1}^{k} M_s DR},\ i = k-1, k-2, \ldots, 1
\end{aligned}
$$

Following Zouaoui and Wilson (2003), we construct the $100(1-\varphi)\%$ confidence interval for β as $[y_{(\lceil M_1\varphi/2 \rceil)}, y_{(\lceil M_1(1-\varphi/2) \rceil)}]$, where the quantities $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(M_1)}$ denote the order statistics of the output data $\{y_{m_1}; m_1 = 1, 2, \ldots, M_1\}$ defined in Figure 1.4.

Such decomposition of the simulation output variance is valuable for the simulation practitioner as it provides him/her a guideline on how to proceed to reduce the uncertainty in the simulation output. Specifically, if the stochastic uncertainty is high compared to the parameter uncertainty, then this suggests to make more replications in the simulation. Conversely, if the parameter uncertainty is high, then the simulation practitioner needs to collect more field data to reduce the parameter uncertainty.

1.6. An Inventory Simulation Example

This section performs a numerical study demonstrating the importance of the joint representation of stochastic and parameter uncertainties in the estimation of the mean line-item fill rates[3] and the confidence intervals of multi-product inventory simulations with correlated demands. We refer the reader to Section 1.6.1 for the experimental design and Section 1.6.2 for the results.

[3]The line-item fill rate compares the number of different products shipped complete to the number of different products demanded. The use of the line-item fill rate, which is joint across products, is common in settings where the demands for the items can be correlated as they are frequently used in sets.

1.6.1. Experimental Design

We consider a periodic-review inventory system with $k \geq 1$ different products whose demands follow a $k-$dimensional NORTA distribution. We assume the following properties for the true NORTA distribution: (i) The i^{th} product demand has an exponential distribution with a mean of $10(k+1-i)/k$ units for $i = 1, 2, \ldots, k$. (ii) Each (partial) correlation in the $\mathcal{C}-$vine specification of the $k-$dimensional NORTA demand distribution is 0.30. We let the number of different products, k take the values of 1, 2, 3, 5, and 10 and manage the inventories of the products with the base-stock policy assuming zero ordering cost and zero lead time. More specifically, we identify the base-stock levels I_i, $i = 1, 2, \ldots, k$ via the use of the single-product models, each of which has a non-stockout probability of 0.90; i.e., $I_i \equiv F_i^{-1}(0.90; 10(k+1-i)/k)$ for $i = 1, 2, \ldots, k$. This results in a true mean line-item fill rate of 0.90 in each of the $k-$product inventory simulations.

We assume the availability of the historical demand data of length 100 generated from the true NORTA distribution. We let Y be the line-item fill rate whose mean is relevant to the inventory manager and use $\mathbf{x}$ for denoting the vector of the historical demand data. We implement the simulation replication algorithm as described in Section 1.4.1 for generating a point estimate and a 95% confidence interval of $\mathrm{E}_{Y|\mathbf{x}}[Y|\mathbf{x}]$. We note that the Bayesian simulation replication algorithm of Section 1.4.2 can also be used, and it produces similar results. Our goal is to compare the performances of the point estimates and the confidence intervals obtained from the implementation of our approach to those obtained from stochastic simulations that consider only stochastic uncertainty. We assess the performance of the point estimate using the mean absolute percentage error ($\mathcal{MAPE}$) and the mean square error ($\mathcal{MSE}$), while we evaluate the performance of the confidence interval using the average confidence-interval half-width ($\mathcal{HW}$) and the average coverage probability ($\mathcal{CP}$) (Zouaoui and Wilson 2003).

1.6.2. Results

Table 1.1 presents the results obtained when the view of a frequentist is taken and the exponentially distributed product demands are assumed to be independent. Table 1.2 uses the Bayesian model and presents the results obtained assuming independent product demands, despite the increasing strength of dependence with the number of products. Finally, Table 1.3 presents the results obtained when both the view of the Bayesian is taken and the

stochastic dependencies among the product demands are considered. Each of these tables reports the results for three different values of R (i.e., run length) of the simulation replication algorithm: 1000, 5000, and $10,000$. As the run length increases, we observe that the average confidence-interval half-width approaches zero, while the accuracy in the estimation of the mean absolute percentage error and the mean square error increases. We also observe that the use of our model allows the simulation analyst to obtain point estimates with lower mean absolute percentage errors and confidence intervals with higher coverage than those of the stochastic simulations that account only for stochastic uncertainty.

Specifically, the comparison of the results tabulated in Table 1.1 to those in Table 1.2 and Table 1.3 shows that the point estimator accuracy for the Bayesian approach is better than the point estimator accuracy for the frequentist approach. The mean absolute percentage error is 1.23% in the $2-$product setting, while it is 2.12% in the $5-$product setting for the Bayesian approach with a run length of $10,000$ replications (Table 1.3). On the other hand, for the frequentist approach the mean absolute percentage errors are 1.73% and 4.12% in the $2-$product and $5-$product settings (Table 1.1). Although the average confidence-interval half-widths are tighter than their Bayesian counterparts, the frequentist approach delivers decreasing coverage probabilities with increasing number of products. Since the confidence intervals of the frequentist approach are centered on biased estimates of the mean line-item fill rate, the confidence interval coverage eventually drops to zero as the number of products increases.

On the other hand, the confidence intervals based on the Bayesian approach, even under the assumption of independent product demands, show much higher coverage probabilities as they account for the uncertainty around the parameters of the component marginal distributions as well as the stochastic uncertainty. We find that the mean absolute percentage error is 3.01% and the coverage probability is 72.96% in the $5-$product setting under the assumption of independent demands (Table 1.2). However, the mean absolute percentage error increases to 5.14% and the coverage probability decreases to 64.24% in the $10-$product setting, while accounting for the correlations among the product demands results in a mean absolute percentage error of 3.41% and a coverage probability of 77.83%. Thus, the consideration of the demand correlations further improves both the mean absolute percentage error of the point estimates and the coverage probability of the confidence intervals, especially as the number of products increases.

Table 1.1: The results obtained via frequentist approach assuming independent demands.

R=1000	mean fill rate		95% confidence interval	
k	$\mathcal{MAPE}$	$\mathcal{MSE}$	$\mathcal{HW}$	$\mathcal{CP}$
1	1.64%	3.44×10^{-4}	4.12×10^{-2}	79.90%
2	2.01%	6.60×10^{-4}	2.93×10^{-2}	73.60%
3	2.88%	1.21×10^{-3}	2.44×10^{-2}	70.90%
5	4.45%	2.56×10^{-3}	1.19×10^{-2}	67.10%
10	8.98%	5.98×10^{-3}	1.35×10^{-2}	49.40%
R=5000	mean fill rate		95% confidence interval	
k	$\mathcal{MAPE}$	$\mathcal{MSE}$	$\mathcal{HW}$	$\mathcal{CP}$
1	1.53%	3.40×10^{-4}	1.34×10^{-2}	59.50%
2	1.94%	6.30×10^{-3}	9.43×10^{-3}	55.20%
3	2.60%	1.12×10^{-3}	7.78×10^{-3}	52.30%
5	4.15%	2.13×10^{-3}	6.05×10^{-3}	44.80%
10	8.41%	5.73×10^{-3}	4.27×10^{-3}	31.70%
R=10000	mean fill rate		95% confidence interval	
k	$\mathcal{MAPE}$	$\mathcal{MSE}$	$\mathcal{HW}$	$\mathcal{CP}$
1	1.52%	3.31×10^{-4}	4.26×10^{-3}	57.10%
2	1.73%	3.61×10^{-4}	3.02×10^{-3}	53.20%
3	2.37%	8.91×10^{-4}	2.47×10^{-3}	48.70%
5	4.12%	2.09×10^{-3}	1.91×10^{-3}	41.90%
10	8.36%	5.29×10^{-3}	1.32×10^{-3}	30.20%

Table 1.2: The results obtained via Bayesian approach assuming independent demands.

R=1000	mean fill rate		95% confidence interval	
k	$\mathcal{MAPE}$	$\mathcal{MSE}$	$\mathcal{HW}$	$\mathcal{CP}$
1	1.48%	3.06×10^{-4}	4.86×10^{-2}	88.35%
2	1.77%	5.70×10^{-4}	3.05×10^{-2}	83.12%
3	2.12%	9.70×10^{-4}	2.97×10^{-2}	79.88%
5	3.24%	1.85×10^{-3}	1.82×10^{-2}	76.10%
10	5.63%	3.47×10^{-3}	2.01×10^{-2}	68.24%
R=5000	mean fill rate		95% confidence interval	
k	$\mathcal{MAPE}$	$\mathcal{MSE}$	$\mathcal{HW}$	$\mathcal{CP}$
1	1.38%	2.90×10^{-4}	3.26×10^{-2}	87.29%
2	1.72%	5.30×10^{-4}	2.41×10^{-2}	82.18%
3	1.94%	6.00×10^{-4}	9.03×10^{-3}	78.57%
5	3.08%	1.52×10^{-3}	6.82×10^{-3}	73.36%
10	5.34%	2.99×10^{-3}	5.94×10^{-3}	66.87%
R=10000	mean fill rate		95% confidence interval	
k	$\mathcal{MAPE}$	$\mathcal{MSE}$	$\mathcal{HW}$	$\mathcal{CP}$
1	1.35%	2.73×10^{-4}	5.12×10^{-3}	82.83%
2	1.68%	3.46×10^{-4}	4.02×10^{-3}	79.52%
3	1.88%	5.87×10^{-4}	2.98×10^{-3}	78.04%
5	3.01%	1.50×10^{-3}	1.62×10^{-3}	72.96%
10	5.14%	2.82×10^{-3}	2.62×10^{-3}	64.24%

Table 1.3: The results obtained via Bayesian approach assuming correlated demands.

R=1000	mean fill rate		95% confidence interval	
k	$\mathcal{MAPE}$	$\mathcal{MSE}$	$\mathcal{HW}$	$\mathcal{CP}$
2	1.32%	3.27×10^{-4}	5.20×10^{-2}	88.76%
3	2.08%	6.54×10^{-4}	3.98×10^{-2}	84.02%
5	2.67%	1.17×10^{-3}	3.01×10^{-2}	81.45%
10	4.34%	2.32×10^{-3}	2.62×10^{-2}	80.00%
R=5000	mean fill rate		95% confidence interval	
k	$\mathcal{MAPE}$	$\mathcal{MSE}$	$\mathcal{HW}$	$\mathcal{CP}$
2	1.26%	3.01×10^{-4}	3.65×10^{-2}	87.93%
3	1.83%	5.78×10^{-4}	9.87×10^{-3}	84.00%
5	2.19%	6.67×10^{-4}	8.05×10^{-3}	80.43%
10	3.92%	2.07×10^{-3}	6.13×10^{-3}	78.12%
R=10000	mean fill rate		95% confidence interval	
k	$\mathcal{MAPE}$	$\mathcal{MSE}$	$\mathcal{HW}$	$\mathcal{CP}$
2	1.23%	2.92×10^{-4}	7.02×10^{-3}	84.72%
3	1.82%	5.76×10^{-4}	4.72×10^{-3}	82.61%
5	2.12%	6.54×10^{-4}	3.61×10^{-3}	79.28%
10	3.41%	2.02×10^{-3}	3.06×10^{-3}	77.83%

1.7. Conclusion

We consider a large-scale stochastic simulation whose correlated inputs have a NORTA distribution with arbitrary continuous marginal distributions. We investigate how to account for stochastic and parameter uncertainties in the estimation of the mean performance measure and the confidence interval of this simulation. Utilizing Sklar's marginal-copula representation together with Cooke's copula-vine specification, we develop a Bayesian model for the fast sampling of the parameters of the NORTA distribution. The development of such a Bayesian model, which enables simulation analysts to capture parameter uncertainty in stochastic simulations with correlated inputs, is the primary contribution of this paper to the discrete-event stochastic simulation literature. We incorporate the Bayesian model into the simulation replication algorithm of Chick (2001) and the Bayesian simulation replication algorithm of Zouaoui and Wilson (2003) for the joint representation of stochastic uncertainty and parameter uncertainty in the computation of the mean performance estimate and the confidence interval. We also decompose the variance of the simulation output obtained using the Bayesian simulation replication algorithm into variances associated with the stochastic uncertainty and the parameter uncertainty.

We demonstrate the effectiveness of the Bayesian model in decreasing the mean absolute percentage error of the mean line-item fill-rate estimate and increasing the coverage of the confidence interval in a multi-product inventory simulation with correlated stochastic demands. The variance decomposition can help the simulation practitioner to decide whether he/she needs to collect more data to reduce the parameter uncertainty or to make more replications in the simulation to reduce the stochastic uncertainty.

Chapter 2

Comparison of Least-Squares and Bayesian Inferences for Johnson's S_B and S_L Distributions [1]

2.1. Introduction

The common approach in building an input model for a discrete-event stochastic simulation is to use a standard distribution such as beta, exponential, gamma, or normal (Law 2007). However, the shapes represented by the standard family of distributions are limited and therefore, they might fail to adequately capture the distributional characteristics of the historical input data. For example, although widely used in practice, normal distribution is characterized by its first two moments; it sets the coefficients of skewness and kurtosis to zero and three, respectively, for any first two moments. Therefore, normal distribution falls short of capturing the skewness in the historical input data. It is also possible that, despite its statistical validity, a goodness-of-fit test rejects or accepts all candidate distributions, depending on the number of available data points. Furthermore, the simulation analyst might not be familiar with all the standard distributions on a lengthy list that can be used for input modeling.

The shortcomings of using standard distributions for input modeling are often overcome by the construction of flexible distributions that are capable of representing a wide variety of distributional shapes. The well-known flexible distributions of the simulation input-modeling literature include the curves proposed by Pearson (1895), the Johnson translation system (Johnson 1949), the generalized lambda distribution (Ramberg and Schmeiser 1974), the four-parameter distribution introduced by Schmeiser and Deutsch (1977), and the generalized

[1] This chapter is submitted to the journal *INFORMS Journal on Computing* with co-author Bahar Biller.

beta family of distributions (Kuhl et al. 2009). The focus of this paper is on the S_L and S_B distributions of the translation system developed by Johnson (1949). Due to the flexibility it offers in modeling and the ease of its implementation, the Johnson translation system (JTS) has been one of the popular flexible distribution systems in recent years. It is easy to use, adjust, and understand. The four-parameter distribution of Schmeiser and Deutsch (1977) is also easy to use, but the one-to-one relationship between the distribution parameters and the moments is lost, and the distributional characteristics such as bimodality and heavy tails are not captured. The distributional shapes that can be represented by the generalized beta family are also limited in the sense that they are unimodal. The lambda family, on the other hand, matches any first two moments but limited third and fourth moments, while the JTS can capture any pair of finite third and fourth moments. Although the curves of the JTS generally agree with Pearson curves having the same (or nearly the same) first four moments, the mathematical structure of the distribution functions from the JTS provides a convenient aid to developing input models for stochastic simulations.

A close look at the existing literature shows that JTS has been used in a variety of fields including forestry, epidemiology, hydrology, and bioinformatics. A variety of fitting methods has also been developed for estimating the Johnson parameters, including the method of matching moments, the method of matching percentiles, the maximum likelihood estimation (MLE) method, the least-squares estimation (LSE) method, and the Bayesian method. The method of matching moments is known to lead to highly variable parameter estimates for small data sets (Slifker and Shapiro 1980), while the estimates obtained by the method of matching percentiles depend on the choice of percentile points (Chou et al. 1999). Also, the (threshold) parameters associated with the bounded supports of Johnson's S_L and S_B distributions violate the standard regularity conditions of the MLE method (Hill 1963, Lambert 1970). However, the LSE method introduced by Swain et al. (1988) for fitting target Johnson distributions to independent data does not suffer from this particular limitation of the MLE method. Furthermore, Swain et al. (1988) show that the LSE method is superior to both the method of matching moments and the method of matching percentiles, while Biller and Nelson (2005) prove the asymptotic consistency of the estimates obtained from the LSE method. A frequently used method for estimating the threshold parameters of the bounded distributions is the Bayesian method. This method has been used by a number of researchers for Johnson's S_L and S_B distributions (e.g., Hill 1963, Upadhyay and Peshwani 2001, and Tsionas 2001), and it has been shown to be superior to both the MLE method

and the method of matching moments. Although the review of the literature suggests that both the LSE method and the Bayesian method are promising fitting methods for Johnson's S_L and S_B distributions, it is not clear how the goodness of their fits compare to each other.

The goal of this paper is to investigate the relative performance of the Bayesian method and the LSE method of Swain et al. (1988) in estimating the parameters of Johnson's S_L and S_B distributions, and to provide insights about when to use each fitting method. We do this by assuming the availability of finite, independent, and identically distributed data from the S_L and S_B families of the JTS. We compare the goodness of the fits using the Kolmogorov-Smirnov and Anderson-Darling test statistics as well as the quantile-quantile plots. The Bayesian method allows the incorporation of expert opinion into the estimation procedure via the use of a joint prior density function on the distribution parameters. To perform a fair comparison between the Bayesian and LSE methods in this paper, we obtain the Bayesian fits by using a joint noninformative prior density function for the parameters of the JTS. Our results suggest that the LSE method performs better than the Bayesian method for highly skewed unimodal distributions of the S_B family and the long-tailed distributions of the S_L family. The comparison of the goodness of the fits obtained for other cases does not suggest any statistically significant difference between the performances of the Bayesian and LSE methods.

We organize the remainder of the paper as follows. In Section 2.2, we introduce the JTS, describe the application areas in which it has been found useful, and discuss the fitting methods proposed for the JTS. We present the LSE method of Swain et al. (1988) and the statistical properties of its parameter estimates in Section 2.3. We describe the Bayesian method and provide its implementation details in Section 2.4. We present our experimental results and discuss our major findings in Section 2.5. Finally, we conclude with future research directions in Section 2.6.

2.2. Johnson Translation System

Section 2.2.1 introduces the Johnson translation system (JTS) and provides examples of its application, while Section 2.2.2 reviews the fitting methods used for estimating the parameters of the JTS.

2.2.1. Description and Applications

A random variable X from the JTS has a cumulative distribution function (cdf) of the form

$$F(x) = \Phi\left\{\gamma + \delta r\left[\frac{x-\xi}{\lambda}\right]\right\},$$

where Φ is the cdf of the standard normal random variable, γ and δ are shape parameters, ξ is a location parameter, λ is a scale parameter, and $r(\cdot)$ is one of the following transformations:

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

Within each family, a distribution is completely specified by the values of the parameters γ, δ, λ, and ξ, and the range of X depends on the family of interest; e.g., $X > \xi$ and $\lambda = 1$ for the S_L family, $\xi < X < \xi + \lambda$ for the S_B family, and $-\infty < X < \infty$ for the S_U and S_N families of the JTS. There is a unique family (choice of r) for each feasible combination of the coefficient of skewness and the coefficient of kurtosis that determines parameters γ and δ. Also, any mean and (positive) variance can be attained by each of the families by further manipulation of the parameters ξ and λ. The wide range of shapes the probability density functions (pdfs) of the JTS may represent can be found in Johnson (1987).

JTS has been used successfully for a variety of applications. In this paper, we consider the S_B and S_L families of the JTS. Johnson's S_B family has been found useful for applications of forestry and epidemiology research, while the S_L family has been used for the fields of hydrology and bioinformatics. Specifically, Hafley and Schreuder (1977) show that the S_B family provides a good fit to the distribution of the tree diameter in forest stands. Monness (1982), Von Gadow (1984), and Parresol (2003) also use the S_B family for solving similar estimation problems. Mage (1980) is the first to use the S_B family for modeling environmental airborne concentration measurements. Flynn (2004) discusses that the S_B family is a good model for fitting occupational exposures to airborne contaminants as compared to the normal distribution and the two-parameter lognormal distribution. Examples of papers that discuss the applications of the S_L family, which is also known as the three-parameter lognormal distribution, are Stedinger (1980), Singh (1987), Kosugi (1994), and Li et al. (2006). Stedinger (1980), Singh (1987), and Kosugi (1994) use the S_L family to fit the hydrological

data, while Li et al. (2006) demonstrate the use of the S_L family for analyzing electromigration data. A more detailed presentation of the application areas of the JTS can be found in Biller and Gunes (2010b).

2.2.2. Fitting Methods

In this section, we first review the method of matching moments and percentiles, then the MLE method, and finally the LSE and Bayesian methods, which are the two fitting methods of interest in this paper.

The Methods of Matching Moments and Percentiles

Although the method of matching moments (Hill et al. 1976) and the method of matching percentiles (Slifker and Shapiro 1980, Bowman and Shenton 1988, and Bowman and Shenton 1989) are often used for practical applications, they suffer from a number of shortcomings. The method of matching moments requires the computation of the third and fourth sample moments that are highly biased especially when the size of the historical data set is small (Johnson and Lowe 1979). Another problem with this method is that the variances of the third and fourth moment estimators can be quite high. Also, the existence of the outliers in the data has a significant impact on the performance of the moment estimators (Slifker and Shapiro 1980). The method of matching percentiles, on the other hand, requires the selection of four percentiles of the standard normal distribution as well as the matching of these percentiles with the corresponding percentiles of the historical data. Therefore, the estimates obtained by this method depend on the percentiles selected by the user; this is the major disadvantage of the method of matching percentiles.

The MLE Method

Despite the use of the MLE method for estimating the parameters of many distributions, threshold parameters violate the standard regularity conditions that ensure the asymptotical properties of the maximum likelihood estimates (Billingsley 1961). Specifically, the range of values the random variable from Johnson's S_L family takes (i.e., (ξ, ∞)) depends on the location parameter ξ. Similarly, the range of values the random variable from Johnson's S_B family takes (i.e., $(\xi, \xi + \lambda)$) depends on the location parameter ξ and the scale parameter λ.

Hill (1963) shows that the likelihood function describing the joint distribution of the input data x_i, $i = 1, 2, \ldots, n$ of length n from Johnson's S_L distribution approaches infinity as the threshold parameter ξ approaches $x_{(1)}$, where $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ stand for the order statistics of the input data. Furthermore, the likelihood function approaches a positive constant as ξ approaches $-\infty$. Therefore, the likelihood function suggests that the maximum likelihood estimates of the parameters ξ, γ, and δ are $\hat{\xi} = x_{(1)}$, $\hat{\gamma} = -\infty$, and $\hat{\delta} = 0$, even though the likelihood function evaluated at these parameter estimates is zero.

Lambert (1970) argues that the findings of Hill (1963) extend to the analysis of Johnson's S_B distribution. More specifically, the likelihood function approaches ∞ as the threshold parameters ξ and $\xi + \lambda$ approach $x_{(1)}$ and $x_{(n)} - x_{(1)}$, respectively, even though the likelihood function evaluated at these parameter estimates is zero. Also, Siekierski (1992) finds that the maximum likelihood estimates of Johnson's S_B distribution can take preposterous values when the distribution is skewed and/or the length of the input data is small. Tsionas (2001) explains the findings of Siekierski (1992) by the flat portion of the likelihood function that causes the optimization to proceed slowly and even to be trapped in a neighborhood of a suboptimal solution.

The difficulty of applying the MLE method to the S_L and S_B families of the JTS has been long recognized by Johnson (1949). A number of researchers including Cohen (1951), Calitz (1973), Cheng and Amin (1983), Ranneby (1984), Cheng and Iles (1987), and Komori and Hirose (2001) suggest methods to overcome the drawbacks of the MLE method. In the following section, we consider the use of the LSE method and the Bayesian method for this purpose.

The LSE Method of Swain et al. (1988) and the Bayesian Method

The LSE method of interest in this paper has been introduced by Swain et al. (1988) for fitting a Johnson distribution function to independent and identically distributed data via the minimization of the distance between a vector of uniformized order statistics and their expected values. The LSE method does not suffer from the shortcomings discussed in the previous sections for the MLE method and methods of matching moments and percentiles. We provide a detailed description of the LSE method in Section 2.3.

Another method for distributions with threshold parameters is the Bayesian method. Unlike the frequentist MLE and LSE methods, the Bayesian method requires the determination of a joint prior density function for quantifying the initial uncertainty about the

distribution parameters. The inference is based on the posterior density function, which is obtained by multiplying the likelihood function with the prior density function and dividing it by the normalization constant. In many cases, however, it is not possible to calculate the normalization constant due to the high-dimensional integrals involved in the computation. This, indeed, limited the use of the Bayesian method in the past for solving estimation problems. However, the advancement of the Markov Chain Monte Carlo (MCMC) method in the last decade has made it possible to estimate distribution parameters with any joint posterior density function. Consequently, the Bayesian method has gained popularity and is used by an increasing number of researchers; e.g., Smith and Naylor (1987), Green et al. (1994), Desmond and Yang (1998), and Bermudez and Turkman (2003). A close look at this literature reveals that Hill (1963), Upadhyay and Peshwani (2001), and Tsionas (2001) have used the Bayesian method for the JTS. Specifically, Upadhyay and Peshwani (2001) consider the S_L family of distributions and describe the use of the MCMC method for parameter estimation. This is also the method we use for Johnson's S_L distributions in this paper. Tsionas (2001), on the other hand, uses the Bayesian method for estimating the parameters of Johnson's S_B family. He shows that the use of a joint noninformative prior density function leads to smooth posterior density functions from which the Johnson parameters are estimated in two consecutive steps. First, the posterior mode estimates of the shape parameters are obtained from their corresponding marginal posterior density function. Then, these estimates are inserted into the joint posterior density function and the resulting joint posterior density function is maximized to estimate the threshold parameters. One shortcoming of this method is the bias in the estimation of one of the shape parameters. In this paper, we, for the first time, provide the implementation details of the Bayesian method via an MCMC method for Johnson's S_B distributions.

The review of the fitting methods for the JTS suggests that the LSE method of Swain et al. (1988) and the Bayesian method are two promising methods for estimating the parameters of Johnson's S_L and S_B distributions. However, to the best of our knowledge, the performance of these two estimation methods has not been evaluated with respect to each other. In the remainder of the paper, we describe the LSE method and the Bayesian method in detail and compare their performance of fitting Johnson's S_L and S_B distributions to limited input data.

2.3. The LSE Method of Swain et al. (1988)

In this section, we assume the availability of the independent and identically distributed input data x_i, $i = 1, 2, \ldots, n$ of length n, use $\boldsymbol{\psi} = (\gamma, \delta, \lambda, \xi)$ for the vector of Johnson parameters, and define $V_i(\boldsymbol{\psi}) = \gamma + \delta r[(x_i - \xi)/\lambda]$ as the transformation of the i^{th} data point x_i to a standard normal random variate. We also assume that random variable X_i has a Johnson distribution with parameter vector $\boldsymbol{\psi}^*$ composed of γ^*, δ^*, λ^*, and ξ^*. If all of the parameter values are correct (i.e., $\boldsymbol{\psi} = \boldsymbol{\psi}^*$), then $V_i(\boldsymbol{\psi}^*)$, $i = 1, 2, \ldots, n$ are independent and identically distributed standard normal random variables; i.e., the transformed random variable $R_{(i)}(\boldsymbol{\psi}^*) = \Phi\{V_{(i)}(\boldsymbol{\psi}^*)\}$ has the distribution of the i^{th} order statistic in a random sample of size n from the uniform distribution on the unit interval $(0, 1)$. Since $R_{(i)}(\boldsymbol{\psi}^*)$ has mean $\rho_i = i/(n+1)$, $R_{(i)}(\boldsymbol{\psi}^*)$ can be written as equal to $\rho_i + \varepsilon_i(\boldsymbol{\psi}^*)$ so that the $\varepsilon_i(\boldsymbol{\psi}^*)$, $i = 1, 2, \ldots, n$ are translated uniform order statistics with mean zero and covariance

$$\text{Cov}\Big(\varepsilon_j(\boldsymbol{\psi}^*), \varepsilon_k(\boldsymbol{\psi}^*)\Big) \;=\; \frac{\rho_j(1-\rho_k)}{n+2}$$

for $1 \le j \le k \le n$ (Kendall and Stuart 1979).

Next, we let $\mathbf{R_o}(\boldsymbol{\psi}) \equiv (R_{(1)}(\boldsymbol{\psi}), R_{(2)}(\boldsymbol{\psi}), \ldots, R_{(n)}(\boldsymbol{\psi}))'$, $\boldsymbol{\rho} \equiv (\rho_1, \rho_2, \ldots, \rho_n)'$, and $\boldsymbol{\varepsilon}(\boldsymbol{\psi}) \equiv (\varepsilon_1(\boldsymbol{\psi}), \varepsilon_2(\boldsymbol{\psi}), \ldots, \varepsilon_n(\boldsymbol{\psi}))'$, so that $\boldsymbol{\varepsilon}(\boldsymbol{\psi}^*) \equiv \mathbf{R_o}(\boldsymbol{\psi}^*) - \boldsymbol{\rho}$. Since the first and second moments of the translated uniform order statistics are known and easily computed, Swain et al. (1988) exploit this fact to develop a single, distribution-free formulation for the fitting problem. Specifically, the distance between $\boldsymbol{\rho}$ and $\mathbf{R_o}(\boldsymbol{\psi})$ is minimized as a function of $\boldsymbol{\psi}$ with respect to a metric defined by a quadratic form in the $n-$dimensional Euclidean space. If $\mathbf{W}$ is the $n \times n$ matrix associated with this quadratic form (i.e., $\mathbf{W} = [\text{Cov}(\varepsilon_j(\boldsymbol{\psi}^*), \varepsilon_k(\boldsymbol{\psi}^*)); j, k = 1, 2, \ldots, n]$), then the Johnson parameters can be estimated by solving the following least-squares fitting problem:

$$\begin{array}{ll} \min_{\boldsymbol{\psi}} & \boldsymbol{\varepsilon}(\boldsymbol{\psi})' \mathbf{W} \boldsymbol{\varepsilon}(\boldsymbol{\psi}) \\ \text{subject to} & \boldsymbol{\psi} \in \boldsymbol{\Psi} \end{array}$$

The feasible region $\boldsymbol{\Psi}$, which ensures the feasibility of Johnson's S_B and S_L parameters, is given by

$$\begin{array}{lll} \boldsymbol{\Psi} = \{(\gamma, \delta, \lambda, \xi)' : & \delta > 0 & \\ & \lambda \begin{cases} > x_{(n)} - \xi & \text{for } f = S_B, \\ = 1 & \text{for } f = S_L, \end{cases} & \\ & \xi < x_{(1)} & \text{for } f = S_L \text{ and } S_B\}. \end{array}$$

The least-squares fitting problem gives rise to different estimators depending on the form of the weight matrix $\mathbf{W}$. If it is the $n \times n$ identity matrix $\mathbf{I}$, which assumes that the error terms $\varepsilon_i(\boldsymbol{\psi}^*)$, $i = 1, 2, \ldots, n$ are independent and identically distributed with equal variances, we obtain the ordinary least-squares estimators for the Johnson parameters. We obtain the weighted least-squares (WLS) estimators for $\mathbf{W} \neq \mathbf{I}$. The statistical theory recommends the selection of $\mathbf{W}$ as a weight matrix as $\varepsilon_i(\boldsymbol{\psi}^*)$, $i = 1, 2, \ldots, n$ are neither independent nor homoscedastic. Furthermore, the inverse of $\mathbf{W}$ can be easily computed and such a weight matrix yields the minimum variance linear unbiased estimators for the Johnson parameters (Seber 1977). Nevertheless, we use the diagonal weight matrix defined by

$$\mathbf{D} = \operatorname{diag}\left\{1/\operatorname{Var}\left(\varepsilon_1(\boldsymbol{\psi}^*)\right), 1/\operatorname{Var}\left(\varepsilon_2(\boldsymbol{\psi}^*)\right), \ldots, 1/\operatorname{Var}\left(\varepsilon_n(\boldsymbol{\psi}^*)\right)\right\}$$

and obtain the diagonally-weighted least-squares (DWLS) estimators. Our choice of $\mathbf{D}$ as the weight matrix is based on the empirical comparison/analysis of Swain et al. (1988), in which the DWLS estimators are found to be superior to the WLS estimators. Kuhl and Wilson (1999) explain the poor performance of the WLS estimators by the cancellations that occur in the objective function $\boldsymbol{\varepsilon}(\boldsymbol{\psi})' \mathbf{W} \boldsymbol{\varepsilon}(\boldsymbol{\psi})$. The resulting objective function contains relatively little information about the discrepancy between $\mathbf{R_o}(\boldsymbol{\psi})$ and $\boldsymbol{\rho}$; the WLS estimators do not provide good estimates of the parameters in the sense that they do not adequately capture the characteristics of the historical data. Furthermore, Theorem 2 of Biller and Nelson (2005) indicates that the Johnson parameter estimates obtained from the LSE method with the diagonal weight matrix are strongly consistent.

2.4. The Bayesian Method

The Bayesian model development starts with the selection of a joint prior density function quantifying the prior information about the parameters of Johnson's S_L and S_B distributions. The prior density function is then updated with the available input data to obtain the posterior density function from which the inference is made. Specifically, Section 2.4.1 derives the joint prior density function of the Johnson parameters and Section 2.4.2 combines the joint prior density function with the likelihood function of the input data and obtains the joint posterior density function. Due to the form of the joint posterior density function, we resort to an MCMC method for obtaining estimates of the Johnson parameters. The idea behind any MCMC method is to simulate a random walk in the entire parameter space that

converges to the joint posterior density function of the parameters (Gilks et al. 1996). Then, the parameters sampled in the simulation are averaged to obtain a final estimate of each parameter. In this paper, we use the Gibbs sampler; i.e., a widely used MCMC method that requires the sampling of the Johnson parameters from their conditional posterior density functions (Geman and Geman 1984, Gelfand and Smith 1990). We present the conditional posterior density functions and the Gibbs sampler, and discuss the distributional properties of the Johnson parameter estimates obtained from the Gibbs sampler in Section 2.4.2.

2.4.1. Joint Prior Density Function of the Johnson Parameters

The key to the selection of a joint prior density function for the Johnson parameters is that random variable $r[(X-\xi)/\lambda]$ has a normal distribution with mean $-\gamma/\delta$ and standard deviation $1/\delta$. Therefore, we treat parameter $-\gamma/\delta$ as the location parameter and denote it by α in the remainder of the paper. Similarly, we consider $1/\delta^2$ as the scale parameter and denote it by β. Furthermore, we separate the selection of prior density functions for parameters α and β from the selection of prior density functions for ξ and λ, and use Jeffreys' prior density function for each of these parameters. Specifically, Jeffreys' prior is a noninformative prior density function that is often used when little is known about the distribution parameters; the goal is to extract as much information as possible from the available input data (Kass and Wasserman 1996). The resulting joint prior density functions for the parameters of Johnson's S_L and S_B distributions are as follows:

Proposition 1 *Jeffreys' joint prior density function for parameters α, β, and ξ of Johnson's S_L distribution is $\pi(\alpha,\beta,\xi) \propto \beta^{-1}$, while Jeffreys' joint prior density function for parameters α, β, ξ, and λ of Johnson's S_B distribution is $\pi(\alpha,\beta,\xi,\lambda) \propto \beta^{-1}\lambda^{-2}$.*

Proof 1 *Using the existing theory on Jeffreys' well-known priors for location and scale parameters, we set $\pi(\alpha) = 1$ for parameter α and $\pi(\beta) = \beta^{-1}$ for parameter β. Thus, we obtain $\pi(\alpha,\beta) = \beta^{-1}$ for the joint prior density function of α and β.*

We define the joint prior density function $\pi(\xi)$ on ξ of Johnson's S_L distribution as proportional to $|I(\xi)|^{1/2}$, where $I(\xi) \equiv -\mathrm{E}\left[\partial^2 \log f(X)/\partial^2 \xi\right]$ is the expected Fisher information matrix with E denoting the expectation operator and f denoting the pdf of the random variable from the S_L family. Since the corresponding pdf with $\lambda = 1$ and $X > \xi$ is given by

$$f(x) \;=\; \frac{1}{\sqrt{2\pi\beta}(x-\xi)} \exp\left\{ -\frac{1}{2}\left[-\frac{\alpha}{\sqrt{\beta}} + \frac{1}{\sqrt{\beta}} \log\left(x-\xi\right) \right]^2 \right\},$$

Jeffreys' prior for ξ is proportional to 1; i.e., $\pi(\xi) \propto 1$. This results in a joint prior density function of the form $\pi(\alpha, \beta, \xi) = \pi(\alpha, \beta)\pi(\xi) \propto \beta^{-1}$ for the parameters of the S_L distribution.

We define the joint prior density function $\pi(\xi, \lambda)$ on (ξ, λ) of Johnson's S_B distribution as proportional to $|I(\xi, \lambda)|^{1/2}$, where $I(\xi, \lambda) \equiv -\mathrm{E}\left[\partial^2 \log f(X) / \left(\partial(\xi, \lambda)\partial(\xi, \lambda)'\right)\right]$ is the expected Fisher information matrix. Because we separate the selection of $\pi(\alpha, \beta)$ from the selection of (ξ, λ), we take $\alpha = 0$ and $\beta = 1$ without loss of generality (Tsionas 2001). Since the pdf of random variable X from the S_B family with $\xi < X < \xi + \lambda$ is given by

$$f(x) = \frac{\lambda}{\sqrt{2\pi}(x-\xi)(\xi+\lambda-x)} \exp\left\{-\frac{1}{2}\left[log\left(\frac{x-\xi}{\xi+\lambda-x}\right)\right]^2\right\},$$

the identification of the joint prior density function $\pi(\xi, \lambda)$ requires the evaluation of the expectations of the following second-order derivatives:

$$\begin{aligned}
\frac{\partial^2 \log f(x)}{\partial \lambda^2} &= -\lambda^{-2} - (\xi+\lambda-x)^{-2} \log\left(\frac{x-\xi}{\xi+\lambda-x}\right) \\
\frac{\partial^2 \log f(x)}{\partial \lambda \partial \xi} &= \lambda^{-2} - (x-\xi)^{-1}(\xi+\lambda-x)^{-1} \\
\frac{\partial^2 \log f(x)}{\partial \xi^2} &= -\lambda^{-2} + (x-\xi)^{-2} \log\left(\frac{x-\xi}{\xi+\lambda-x}\right)
\end{aligned}$$

To evaluate the expectations of these expressions, we use the following results from Tsionas (2001), where κ_1, κ_2, and κ_3 denote constants: (i) The S_B random variable X has the representation $X = [(\xi+\lambda)\exp(Z) + \xi]/[1+\exp(Z)]$, where Z is a standard normal random variable.

(ii)

$$\mathrm{E}\left[(\xi+\lambda-X)^{-2} \log\left(\frac{X-\xi}{\xi+\lambda-X}\right)\right] = \lambda^{-2}\mathrm{E}\left[Z\left(1+\exp(Z)\right)^2\right] = \kappa_1\lambda^{-2} \propto \lambda^{-2}$$

(iii)

$$\mathrm{E}\left[(X-\xi)^{-1}(\xi+\lambda-X)^{-1}\right] = \lambda^{-2}\mathrm{E}\left[Z\left(1+\exp(Z)\right)^2 / \exp(Z)\right] = \kappa_2\lambda^{-2} \propto \lambda^{-2}$$

(iv)

$$\mathrm{E}\left[(X-\xi)^{-2} \log\left(\frac{X-\xi}{\xi+\lambda-X}\right)\right] = \lambda^{-2}\mathrm{E}^2\left[\left(1+\exp(Z)\right)/\exp(Z)\right] = \kappa_3\lambda^{-2} \propto \lambda^{-2}$$

Consequently, we find that $|I(\xi, \lambda)| \propto \lambda^{-4}$; i.e., $\pi(\xi, \lambda) \propto \lambda^{-2}$. Therefore, the joint prior density function for Johnson's S_B distribution parameters is given by $\pi(\alpha, \beta, \xi, \lambda) \propto \beta^{-1}\lambda^{-2}$.

The use of these joint prior density functions for Johnson's S_L and S_B parameters allows us to perform a fair comparison between the performance of the Bayesian method and the performance of the LSE method.

2.4.2. Joint Posterior Density Function and Sampling Algorithms

We start our presentation with the likelihood function $\varphi(\mathbf{x}|\alpha,\beta,\xi)$ describing the joint distribution of the input data vector $\mathbf{x} = (x_1,\ x_2,\ldots,\ x_n)'$ of dimension n from Johnson's S_L distribution:

$$\varphi(\mathbf{x}|\alpha,\beta,\xi) = \frac{\beta^{-n/2}}{\prod_{i=1}^{n}(x_i-\xi)} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^{n}\left[\log(x_i-\xi)-\alpha\right]^2\right\}$$

Similarly, the likelihood function $\varphi(\mathbf{x}|\alpha,\beta,\xi,\lambda)$ for Johnson's S_B distribution is given by

$$\varphi(\mathbf{x}|\alpha,\beta,\xi,\lambda) = \frac{\beta^{-n/2}\lambda^n}{\prod_{i=1}^{n}(x_i-\xi)(\lambda+\xi-x_i)} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^{n}\left[\log\left(\frac{x_i-\xi}{\lambda+\xi-x_i}\right)-\alpha\right]^2\right\}.$$

Multiplication of these likelihood functions with the joint prior density functions of Proposition 1 leads to a joint posterior density function of the form

$$h(\alpha,\beta,\xi|\mathbf{x}) \propto \frac{\beta^{-n/2-1}}{\prod_{i=1}^{n}(x_i-\xi)} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^{n}\left[\log(x_i-\xi)-\alpha\right]^2\right\}$$

for Johnson's S_L parameters and

$$h(\alpha,\beta,\xi,\lambda|\mathbf{x}) \propto \frac{\beta^{-n/2-1}\lambda^{n-2}}{\prod_{i=1}^{n}(x_i-\xi)(\lambda+\xi-x_i)} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^{n}\left[\log\left(\frac{x_i-\xi}{\lambda+\xi-x_i}\right)-\alpha\right]^2\right\}$$

for Johnson's S_B parameters.

We obtain Johnson parameter estimates from these joint posterior density functions via the use of a Gibbs sampler. Specifically, the Gibbs sampler produces a chain of each parameter and the estimate of the parameter is obtained by taking the ergodic average of the chain after proper analysis. We provide the Gibbs sampler for Johnson's S_B distribution in Figure 2.1. It proceeds through iterated sampling from the conditional posterior density functions $p(\beta|\alpha,\xi,\lambda,\mathbf{x})$, $p(\alpha|\beta,\xi,\lambda,\mathbf{x})$, $p(\xi|\beta,\alpha,\lambda,\mathbf{x})$, and $p(\lambda|\beta,\alpha,\xi,\mathbf{x})$ for parameters β, α, ξ, and λ, respectively. The Gibbs sampler works similarly for Johnson's S_L distribution except that $\lambda = 1$. In both cases, the implementation of the Gibbs sampler requires solutions to the selection of an appropriate sampling plan, the choice of an appropriate warm-up period (i.e., a value for $\mathcal{L}^*$ of Figure 2.1), and the determination of the length of the chain (i.e., a value for $\mathcal{L}$ of Figure 2.1) for the convergence of the chain to the joint posterior density function. A detailed discussion of these implementation issues can be found in Biller and Gunes (2010a).

set α, β, ξ, and λ to α^0, β^0, ξ^0, and λ^0, respectively, by fitting Johnson curves via the method of matching moments.

for $\ell = 1, 2, \ldots, \mathcal{L}^*, \mathcal{L}^* + 1, \ldots, \mathcal{L}$ replications **do**

generate β^ℓ from conditional posterior density function $p(\beta|\alpha^{\ell-1}, \xi^{\ell-1}, \lambda^{\ell-1}, \mathbf{x})$

generate α^ℓ from conditional posterior density function $p(\alpha|\beta^\ell, \xi^{\ell-1}, \lambda^{\ell-1}, \mathbf{x})$

generate ξ^ℓ from conditional posterior density function $p(\xi|\beta^\ell, \alpha^\ell, \lambda^{\ell-1}, \mathbf{x})$

generate λ^ℓ from conditional posterior density function $p(\lambda|\beta^\ell, \alpha^\ell, \xi^\ell, \mathbf{x})$

end loop

estimate the Johnson parameters as follows:

$\hat{\beta} = \sum_{\ell=\mathcal{L}^*+1}^{\mathcal{L}} \beta^\ell/(\mathcal{L} - \mathcal{L}^*)$ and $\hat{\alpha} = \sum_{\ell=\mathcal{L}^*+1}^{\mathcal{L}} \alpha^\ell/(\mathcal{L} - \mathcal{L}^*)$

$\hat{\xi} = \sum_{\ell=\mathcal{L}^*+1}^{\mathcal{L}} \xi^\ell/(\mathcal{L} - \mathcal{L}^*)$ and $\hat{\lambda} = \sum_{\ell=\mathcal{L}^*+1}^{\mathcal{L}} \lambda^\ell/(\mathcal{L} - \mathcal{L}^*)$

Figure 2.1: Gibbs sampler for Johnson's S_B parameters α, β, ξ, and λ.

Next, we provide the conditional posterior density functions necessary for the implementation of the Gibbs sampler. Specifically, $p(\beta|\alpha, \xi, \mathbf{x})$ of Johnson's S_L distribution and $p(\beta|\alpha, \xi, \lambda, \mathbf{x})$ of Johnson's S_B distribution have the following functional forms:

$$\propto \begin{cases} \beta^{-n/2-1} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^n \left[\log(x_i - \xi) - \alpha\right]^2\right\} & \text{for the } S_L \text{ distribution} \\ \beta^{-n/2-1} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^n \left[\log\left(\frac{x_i-\xi}{\lambda+\xi-x_i}\right) - \alpha\right]^2\right\} & \text{for the } S_B \text{ distribution} \end{cases}$$

Similarly, $p(\alpha|\beta, \xi, \mathbf{x})$ of Johnson's S_L distribution and $p(\alpha|\beta, \xi, \lambda, \mathbf{x})$ of Johnson's S_B distribution are given by

$$\propto \begin{cases} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^n \left[\log(x_i - \xi) - \alpha\right]^2\right\} & \text{for the } S_L \text{ distribution,} \\ \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^n \left[\log\left(\frac{x_i-\xi}{\lambda+\xi-x_i}\right) - \alpha\right]^2\right\} & \text{for the } S_B \text{ distribution,} \end{cases}$$

while $p(\xi|\beta, \alpha, \mathbf{x})$ and $p(\xi|\beta, \alpha, \lambda, \mathbf{x})$ are defined by

$$\propto \begin{cases} \frac{1}{\prod_{i=1}^n (x_i-\xi)} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^n \left[\log(x_i - \xi) - \alpha\right]^2\right\} & \text{for the } S_L \text{ distribution,} \\ \frac{1}{\prod_{i=1}^n (\lambda+\xi-x_i)} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^n \left[\log\left(\frac{x_i-\xi}{\lambda+\xi-x_i}\right) - \alpha\right]^2\right\} & \text{for the } S_B \text{ distribution.} \end{cases}$$

Finally, the conditional posterior density function $p(\lambda|\beta, \alpha, \xi, \mathbf{x})$ of Johnson's S_B distribution is given by

$$\frac{\lambda^{n-2}}{\prod_{i=1}^n (\lambda + \xi - x_i)} \exp\left\{-\frac{1}{2\beta}\sum_{i=1}^n \left[\log\left(\frac{x_i - \xi}{\xi + \lambda - x_i}\right) - \alpha\right]^2\right\}.$$

The sampling of parameter β^{-1} from $p(\beta|\alpha,\xi,\mathbf{x})$ reduces to the sampling of a gamma variate with shape parameter $n/2$ and scale parameter $2/\sum_{i=1}^{n}(\log(x_i-\xi)-\alpha)^2$ for Johnson's S_L distribution. Similarly, the sampling of parameter β^{-1} from $p(\beta|\alpha,\xi,\lambda,\mathbf{x})$ reduces to the sampling of a gamma variate with shape parameter $n/2$ and scale parameter $2/\sum_{i=1}^{n}(\log((x_i-\xi)/(\lambda+\xi-x_i))-\alpha)^2$ for Johnson's S_B distribution. Efficient procedures for sampling random variates from a gamma density function can be found in Law (2007).

The conditional posterior density function of parameter α is logconcave and therefore, parameter α can be sampled using the adaptive rejection sampling (ARS) algorithm for the Gibbs sampler (Gilks and Wild 1992). However, neither the conditional posterior density function of ξ nor the conditional posterior density function of λ is logconcave with a standard functional form. Therefore, we use the adaptive rejection metropolis sampling (ARMS) algorithm proposed by Gilks et al. (1995) for sampling parameters ξ and λ. We present the ARS and the ARMS algorithms in the next sections.

A natural question to ask is whether the Johnson parameter estimates obtained by the Gibbs sampler are asymptotically consistent. There exists a well established theory on the asymptotical properties of the Bayesian estimates; i.e., they converge to the true parameter values as the length of the input data approaches infinity and the joint posterior density function of the parameters converges in distribution to a multivariate normal density function (Bernardo and Smith 1994). However, these asymptotical properties cannot be proven for the improper posterior density functions (Gelman et al. 1995, Section 4.3). In this paper, we use Jeffreys' (noninformative) joint prior density function for the parameters of Johnson's S_L and S_B distributions. This allows us to perform a fair comparison between the LSE method and the Bayesian method, but noninformative priors are improper and they lead to improper joint posterior density functions. Therefore, we cannot prove the asymptotical consistency of the Johnson parameter estimates obtained by the Gibbs sampler in Figure 2.1. Nevertheless, Johnson parameter estimates converge to their true parameter values in each experiment performed in Section 2.5.

Sampling α from $p(\alpha|\beta,\xi,\mathbf{x})$ via the ARS Algorithm

In this section, we focus on sampling parameter α of Johnson's S_L distribution from $p(\alpha|\beta,\xi,\mathbf{x})$ via the ARS algorithm. However, our discussion is readily applicable to parameter α of Johnson's S_B distribution with the modification of the posterior density function.

For ease of presentation, we denote the posterior density function $p(\alpha|\beta,\xi,\mathbf{x})$ of Johnson's S_L distribution with $h(\alpha)$, its logarithm with $M(\alpha)$; i.e.,

$$M(\alpha) \;=\; -\frac{1}{2\beta}\sum_{i=1}^{n}\left[\log\left(x_i-\xi\right)-\alpha\right]^2,$$

and the domain of $h(\alpha)$; i.e., the set of α for which $h(\alpha)>0$, with D. We represent the set of ℓ different α values with $T_\ell=\{\alpha_p; p=1,2,\ldots,\ell\}$, where α_p, $p=1,2,\ldots,\ell$ are arranged in an ascending order. We use $\alpha_{(p)}$ for the p^{th} lowest element of T_ℓ. We define $L_{p,q}(\alpha;T_\ell)$ as the straight line passing through points $(\alpha_p,M(\alpha_p))$ and $(\alpha_q,M(\alpha_q))$ for $1\le p\le q\le n$. Additionally, we use $u_\ell(\alpha)$ for the piecewise linear function that is defined as the minimum of $L_{p-1,p}(\alpha;T_\ell)$ and $L_{p+1,p+2}(\alpha;T_\ell)$, where $\alpha_p\le\alpha<\alpha_{p+1}$. Since the posterior density function $h(\alpha)$ is logconcave, $u_\ell(\alpha)$ is an envelope (i.e., an upper hull) for $M(\alpha)$; i.e., $u_\ell(\alpha)\ge M_i(\alpha)$ for every $\alpha\in D$. Finally, we define the normalized exponential hull of the upper hull as $s_\ell(\alpha)\equiv\exp\left(u_\ell(\alpha)\right)/\int_D\exp\left(u_\ell(\alpha)\right)d\alpha$. Since $s_\ell(\alpha)$ is a piecewise exponential distribution, α can be easily sampled from $s_\ell(\alpha)$ (Law 2007).

The ARS algorithm is composed of steps of initialization, sampling, evaluation, rejection, and acceptance, and it proceeds as follows. The algorithm starts with the initialization of ℓ to 2 and the construction of T_2 with $\alpha_1=-1$ and $\alpha_2=+1$. These two values have been reported to provide good starting values for high computational efficiency (Gilks and Wild 1992). In the sampling step, $\alpha^\dagger$ and w are generated independently from $s_\ell(\alpha)$ and the uniform distribution on the unit interval. If w is found to be greater than $h(\alpha^\dagger)/\exp(u_\ell(\alpha^\dagger))$ in the evaluation step, then the algorithm moves to the rejection step. Otherwise, the algorithm moves to the acceptance step. In the rejection step, $\alpha^\dagger$ is used for forming $T_{\ell+1}$ whose elements are to be arranged in an ascending order, and the algorithm goes back to the sampling step. The acceptance step, on the other hand, returns $\alpha^\dagger$ as the value sampled for parameter α.

Sampling ξ from $p(\xi|\alpha,\beta,\mathbf{x})$ via the ARMS Algorithm

The ARMS algorithm of this section is obtained by appending the Hastings-Metropolis algorithm (Hastings 1970) to the end of the ARS algorithm of the previous section. We describe implementation details of this algorithm by focusing on sampling parameter ξ of Johnson's S_L distribution from conditional posterior density function $p(\xi|\alpha,\beta,\mathbf{x})$. However, the approach applies to the sampling of parameters ξ and λ of Johnson's S_B distribution via the proper modification of the posterior density functions.

For notational convenience, we denote posterior density function $p(\xi|\alpha,\beta,\mathbf{x})$ with $h(\xi)$, its logarithm with $M(\xi)$, and the domain of $h(\xi)$ with D. Similarly, we represent the set of ℓ different ξ values with $T_\ell = \{\xi_p; p = 1, 2, \ldots, \ell\}$, where ξ_p, $p = 1, 2, \ldots, \ell$ are arranged in an ascending order. We use $\xi_{(p)}$ for the p^{th} lowest element of T_ℓ and $L_{p,q}(\xi; T_\ell)$ for the straight line passing through points $(\xi_p, M(\xi_p))$ and $(\xi_q, M(\xi_q))$ for $1 \leq p \leq q \leq n$. We define piecewise linear function $u_\ell(\xi)$ as the maximum of $L_{p,p+1}(\xi, T_\ell)$ and the minimum of $L_{p-1,p}(\xi, T_\ell)$ and $L_{p+1,p+2}(\xi, T_\ell)$ for $\xi_p \leq \xi \leq \xi_{p+1}$. We define $s_\ell(\xi)$ as equivalent to $\exp(u_\ell(\xi)) / \int_D \exp(u_\ell(\xi))\, d\xi$. Finally, we let $\xi^\ddagger$ denote the current value of ξ sampled in the previous iteration of the Gibbs sampler. Our goal is to replace $\xi^\ddagger$ with $\xi^\dagger$ to be sampled from $h(\xi)$ in the current iteration.

The ARMS algorithm starts with the initialization of ℓ to 2 and the construction of T_2 with $\xi_1 = -1$ and $\xi_2 = +1$. Then, $\xi^\dagger$ and w are sampled independently from $s_\ell(\xi)$ and the uniform distribution on the unit interval. If w is greater than $h(\xi^\dagger)/\exp(u_\ell(\xi^\dagger))$, then the algorithm moves to the ARS rejection step; i.e., $\xi^\dagger$ is used for forming $T_{\ell+1}$ whose elements are to be arranged in an ascending order and the algorithm goes back to the independent sampling of $\xi^\dagger$ and w from $s_\ell(\xi)$ and the uniform distribution on the unit interval. Otherwise, the algorithm moves to the ARS acceptance step; i.e., a uniform random variate, which we denote by u, is sampled from the unit interval and if u is greater than the minimum of 1 and

$$\frac{h(\xi^\dagger)\min\{h(\xi^\ddagger), \exp(u_\ell(\xi^\ddagger))\}}{h(\xi^\ddagger)\min\{h(\xi^\dagger), \exp(u_\ell(\xi^\dagger))\}},$$

then $\xi^\dagger$ is set to $\xi^\ddagger$. Otherwise, $\xi^\dagger$ is returned as the value sampled for ξ.

2.5. Numerical Study

Our objective is to evaluate the performance of the LSE method of Section 2.3 with respect to the performance of the Bayesian method of Section 2.4 for Johnson's various S_L and S_B distributions. We present the design of the experiments in Section 2.5.1 and the results in Section 2.5.2.

2.5.1. Design of the Experiments

Figures 2.2 and 2.3 present the pdfs of the S_L and S_B distributions we experiment with in Section 2.5.2. The title of each plot is an abbreviation for the index of the experiment performed for the distribution of interest. Tables 2.1 and 2.2 provide the properties of these

distributions; i.e., the coefficient of variation (σ/μ), the coefficient of skewness ($\sqrt{\beta_1}$), the coefficient of kurtosis (β_2), the 90$^{\text{th}}$ quantile of the Johnson random variable ($X_{0.90}$), the 95$^{\text{th}}$ quantile ($X_{0.95}$), and the 99$^{\text{th}}$ quantile ($X_{0.99}$). Specifically, the S_B distributions of Figure 2.2 and Table 2.1 are obtained for $\xi = 0$ and $\lambda = 1$, while the S_L distributions of Figure 2.3 and Table 2.2 are obtained for $\xi = 0$, $\lambda = 1$, $\gamma = 0$, and three different values of δ that control the level of skewness and the length of the right tails in the pdfs.

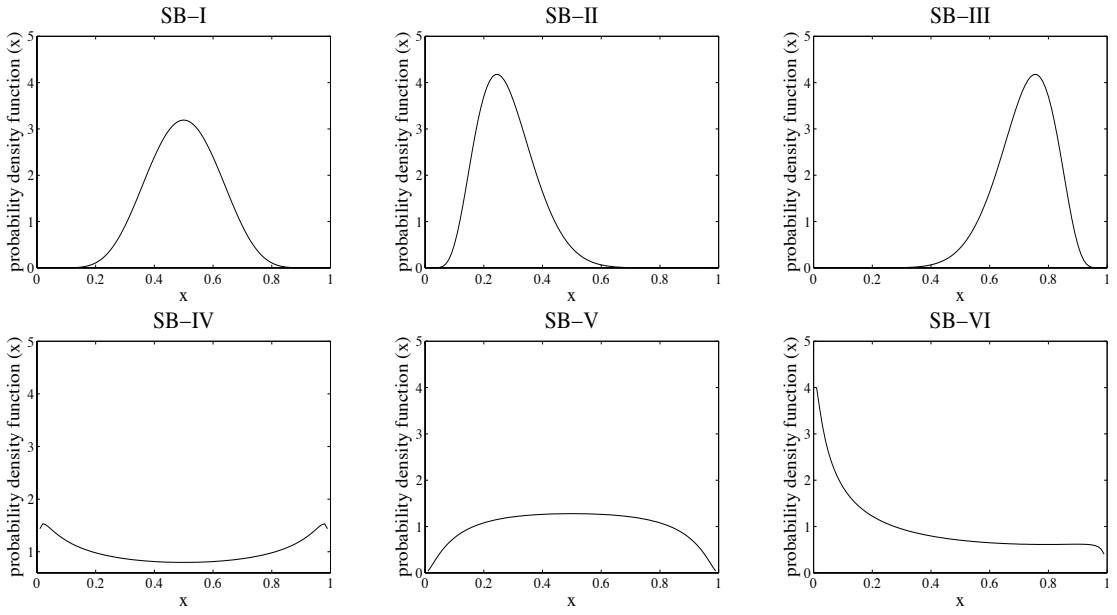

Figure 2.2: Pdfs of unimodal and bimodal S_B distributions

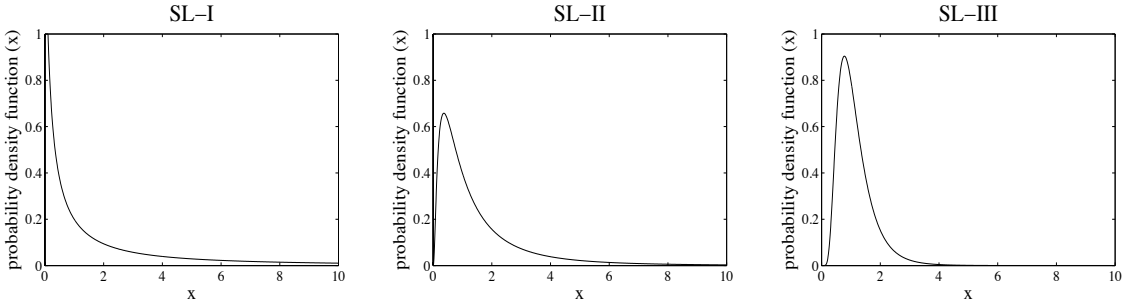

Figure 2.3: Pdfs of Johnson's S_L distributions

In each experiment, we let the number of input data points, n take the values of 30, 50, 100, and 1000. To calculate the goodness of the estimated Johnson cdf $\hat{F}$, we use the Kolmogorov-Smirnov (KS) test statistic (Chakravant et al. 1967) and the Anderson-Darling (AD) test statistic (Anderson and Darling 1954). Both of these tests compare $\hat{F}$ to the empirical cdf F_n. More specifically, the KS test statistic corresponds to the largest distance

Table 2.1: Properties of the S_B distributions illustrated in Figure 2.2

Type	Parameters	σ/μ	$\sqrt{\beta_1}$	β_2	$X_{0.90}$	$X_{0.95}$	$X_{0.99}$
SB-I	$\gamma = 0, \delta = 2$	0.2	-8.4 $\times 10^{-9}$	2.6	0.6	0.7	0.8
SB-II	$\gamma = 2, \delta = 2$	0.3	0.5	3.2	0.4	0.5	0.6
SB-III	$\gamma = -2, \delta = 2$	0.3	-0.5	3.2	0.8	0.9	0.9
SB-IV	$\gamma = 0, \delta = 0.5$	0.6	3.1 $\times 10^{-9}$	1.6	0.9	0.9	1.0
SB-V	$\gamma = 0, \delta = 0.8$	0.5	2.1 $\times 10^{-9}$	2.0	0.8	0.9	0.9
SB-VI	$\gamma = 0.5, \delta = 0.5$	0.6	0.4	2.2	0.7	0.8	0.9

Table 2.2: Properties of the S_L distributions illustrated in Figure 2.3

Type	Parameters	σ/μ	$\sqrt{\beta_1}$	β_2	$X_{0.90}$	$X_{0.95}$	$X_{0.99}$
SL-I	$\delta = 0.5$	7.3	414.4	9.2×10^6	13.0	26.8	105.0
SL-II	$\delta = 1$	1.3	6.2	113.9	3.6	5.2	10.2
SL-III	$\delta = 2$	0.5	1.8	8.9	1.9	2.3	3.2

between $F_n(x)$ and $\hat{F}(x)$; i.e., $\sup_x\{|F_n(x) - \hat{F}(x)|\}$, while the AD test statistic is the weighted average of the squared differences $[F_n(x) - \hat{F}(x)]^2$, where the weights are the largest for the values of $\hat{F}(x)$ that are closest to zero and one. Therefore, the KS test statistic emphasizes the discrepancies in the middle, while the AD test statistic indicates the discrepancies in the tails of the distribution functions.

The results of the next section are averaged over the number of replications necessary for an absolute error of no more than 0.1 on the KS and AD test statistics. We also provide the quantile-quantile (Q-Q) plots comparing the i^{th} quantile of the input data $x_{(i)}$ to the i^{th} quantile of the estimated Johnson cdf $\hat{F}^{-1}((i - 1/2)/n)$ for $i = 1, 2, \ldots, n$. The Q-Q plot is expected to be approximately linear with a slope of 1 if $\hat{F}(x)$ is a good fit for $F_n(x)$. Any departure from linearity indicates a discrepancy between the distribution functions $F_n(x)$ and $\hat{F}(x)$. Furthermore, we obtain the Bayesian estimates of the Johnson parameters by performing a single run of the Gibbs sampler, determining the length of the run (i.e., $\mathcal{L}$ of Figure 2.1) as 60000 iterations using the convergence diagnostic of Heidelberger and Welch (1983) and the method of batching (Law 2007), and identifying the warm-up period (i.e., $\mathcal{L}^*$ of Figure 2.1) as 10000 iterations using the convergence diagnostics of Heidelberger and Welch (1983) and Geweke (1992).

2.5.2. Results

First, we discuss our findings for Johnson's S_B distributions. The KS and AD test statistics of the fits obtained for the unimodal distributions via the LSE method and Bayesian method are presented in Table 2.3. The comparison of the KS statistics reveals no conclusion about

whether the LSE method or the Bayesian method performs better; the results favor the LSE method in some cases, while they favor the Bayesian method in others. However, the AD

Table 2.3: The KS and AD test statistics for the unimodal S_B distributions

SB-I	$n = 30$	$n = 50$	$n = 100$	$n = 1000$
LSE (KS)	0.57	0.52	0.47	0.47
Bayesian (KS)	0.48	0.51	0.45	0.46
LSE (AD)	0.77	0.25	0.18	0.17
Bayesian (AD)	0.78	0.25	0.23	0.23
SB-II	$n = 30$	$n = 50$	$n = 100$	$n = 1000$
LSE (KS)	0.64	0.54	0.45	0.45
Bayesian (KS)	0.57	0.58	0.53	0.51
LSE (AD)	0.22	0.23	0.12	0.10
Bayesian (AD)	0.36	0.28	0.23	0.21
SB-III	$n = 30$	$n = 50$	$n = 100$	$n = 1000$
LSE (KS)	0.60	0.55	0.46	0.44
Bayesian (KS)	0.58	0.52	0.44	0.43
LSE (AD)	0.24	0.21	0.14	0.14
Bayesian (AD)	0.32	0.30	0.24	0.22

test statistics obtained for the LSE method consistently appear to be smaller than those obtained for the Bayesian method. This leads us to conclude that the LSE method performs better than the Bayesian method in capturing the tail behavior of Johnson's unimodal S_B distributions. The Q-Q plots presented in Figures 2.4, 2.5, and 2.6 also indicate that the LSE method performs better in capturing especially the right tail behavior of the skewed data; e.g., the curves of Figure 2.5 corresponding to the Bayesian method exhibit more deviation from linearity in the right tails. As the number of historical data points increases and the parameter estimates approach their true values, the Q-Q plots get more linear for both the LSE and the Bayesian method. Therefore, we present the Q-Q plots of SB-I distribution for $n = 30$, $n = 50$, $n = 100$, and $n = 1000$ in Figure 2.4, and the Q-Q plots of the SB-II and SB-III distributions in Figures 2.5, and 2.6, respectively, for $n = 30$ and $n = 50$.

The similar analysis of the test statistics in Table 2.4 and the Q-Q plots in Figures 2.7, 2.8, and 2.9 obtained for the bimodal S_B distributions indicate almost no difference between the performances of the LSE method and the Bayesian method. Therefore, our conclusion about the superiority of the LSE method over the Bayesian method in fitting skewed unimodal S_B distributions does not extend to bimodal S_B distributions.

We are now ready to present the results for Johnson's S_L distributions. We provide the KS and AD test statistics in Table 2.5 and the Q-Q plots in Figures 2.10, 2.11, and 2.12.

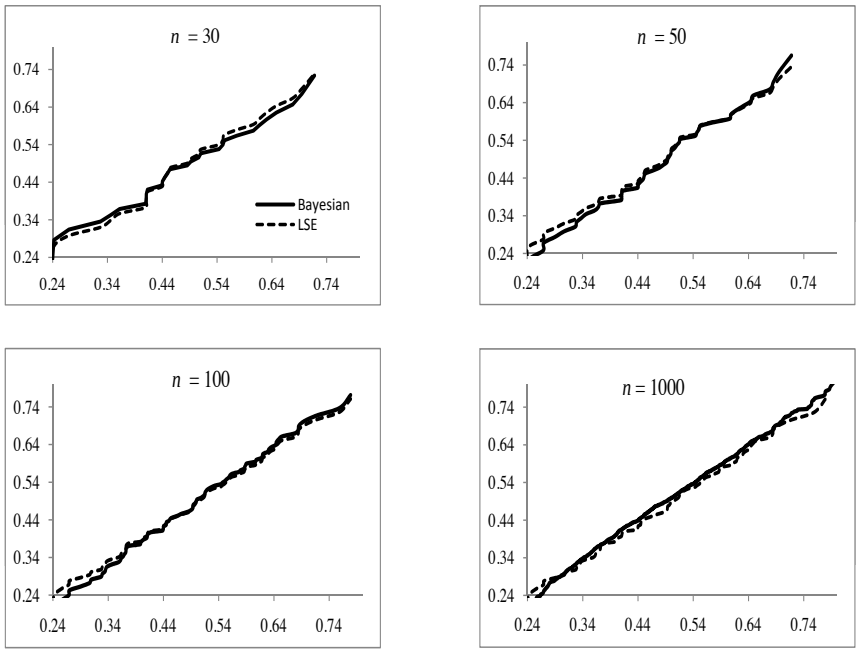

Figure 2.4: Q-Q plots for the SB-I distribution

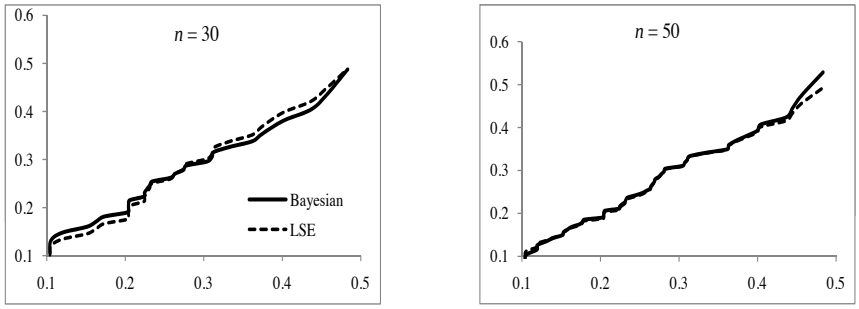

Figure 2.5: Q-Q plots for the SB-II distribution

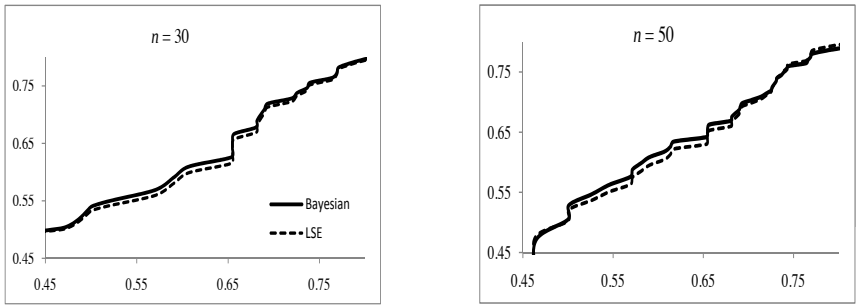

Figure 2.6: Q-Q plots for the SB-III distribution

Table 2.4: The KS and AD test statistics for the bimodal S_B distributions

SB-IV	$n = 30$	$n = 50$	$n = 100$	$n = 1000$
LSE (KS)	0.58	0.46	0.39	0.36
Bayesian (KS)	0.53	0.50	0.42	0.40
LSE (AD)	0.47	0.25	0.12	0.20
Bayesian (AD)	0.47	0.22	0.21	0.10
SB-V	$n = 30$	$n = 50$	$n = 100$	$n = 1000$
LSE (KS)	0.55	0.49	0.44	0.43
Bayesian (KS)	0.63	0.56	0.52	0.50
LSE (AD)	0.23	0.20	0.15	0.12
Bayesian (AD)	0.27	0.22	0.17	0.18
SB-VI	$n = 30$	$n = 50$	$n = 100$	$n = 1000$
LSE (KS)	0.64	0.50	0.42	0.41
Bayesian (KS)	0.66	0.58	0.53	0.53
LSE (AD)	0.67	0.22	0.12	0.10
Bayesian (AD)	0.39	0.36	0.21	0.20

Table 2.5: The KS and AD test statistics for Johnson's S_L distributions

SL-I	$n = 30$	$n = 50$	$n = 100$	$n = 1000$
LSE (KS)	1.99	2.67	5.02	17.50
Bayesian (KS)	2.02	2.54	5.00	17.80
LSE (AD)	3.71	6.66	23.05	277.50
Bayesian (AD)	3.59	6.58	22.80	267.30
SL-II	$n = 30$	$n = 50$	$n = 100$	$n = 1000$
LSE (KS)	0.90	1.13	1.52	3.43
Bayesian (KS)	0.90	0.90	1.50	3.50
LSE (AD)	1.06	1.58	2.69	16.20
Bayesian (AD)	1.04	1.15	2.50	15.80
SL-III	$n = 30$	$n = 50$	$n = 100$	$n = 1000$
LSE (KS)	0.67	0.71	0.74	0.77
Bayesian (KS)	0.61	0.69	0.70	0.72
LSE (AD)	0.32	0.36	0.40	0.70
Bayesian (AD)	0.29	0.34	0.38	0.65

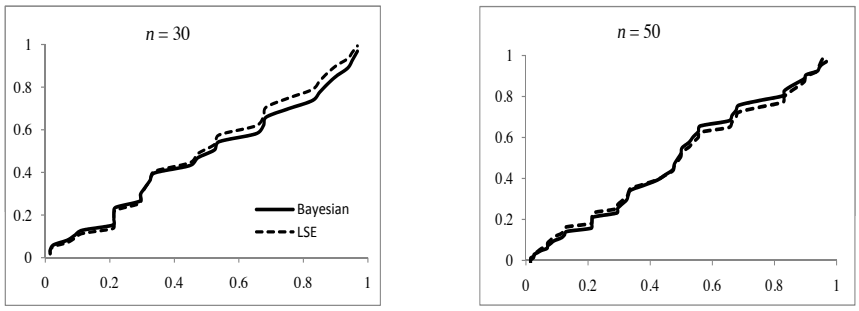

Figure 2.7: Q-Q plots for the SB-IV distribution

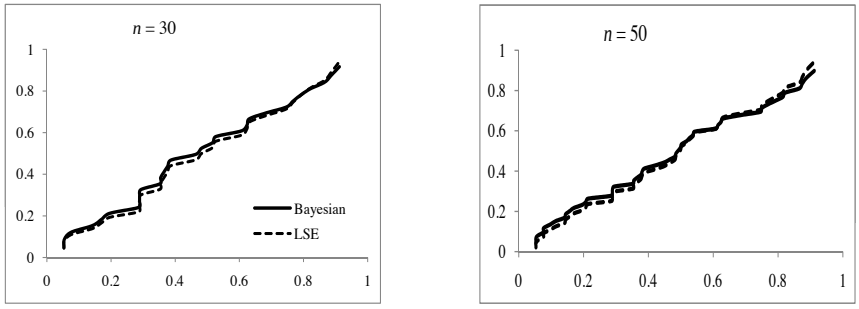

Figure 2.8: Q-Q plots for the SB-V distribution

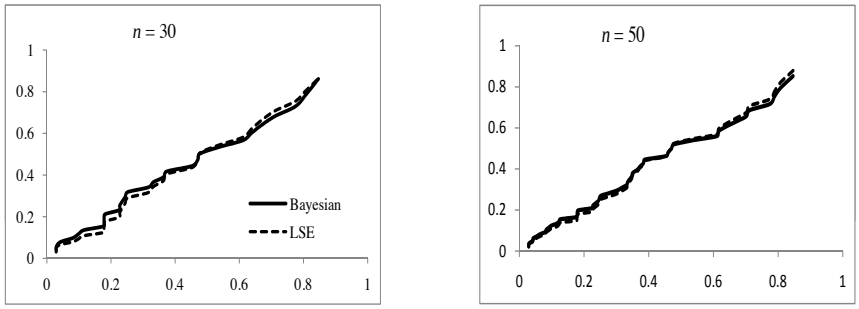

Figure 2.9: Q-Q plots for the SB-VI distribution

First, we note that the KS and AD test statistics obtained for Johnson's S_L distributions are larger than those obtained for the S_B distributions. Similarly, the Q-Q plots of the S_L distributions are less linear than the Q-Q plots of the S_B distributions. These observations can be explained by the tail behavior of the data from the S_L family, which produces outliers in the data set. We find that the AD test statistics obtained for the Bayesian method are slightly smaller than the AD test statistics of the LSE method. However, a closer look at the Q-Q plots shows that the LSE method performs better than the Bayesian method for the long-tailed SL-I and SL-II distributions, while the Bayesian method performs slightly better than the LSE method for the short-tailed SL-III distribution. Therefore, we conclude that the LSE method performs better than the Bayesian method in fitting long-tailed distributions of the S_L family.

2.6. Conclusion

In this paper, we consider the problem of estimating the parameters of Johnson's S_L and S_B distributions from historical input data of finite length. A close look at the existing literature suggests that two promising data-fitting methods for Johnson's S_L and S_B distributions are the LSE method of Swain et al. (1988) and the Bayesian method. We contribute to the discrete-event stochastic simulation literature by describing the use of the Bayesian method for Johnson's S_L and S_B distributions and comparing the performances of these two methods for historical data sets of different lengths and distributional characteristics. An important feature of the Bayesian method is its ability to incorporate expert opinion into the estimation procedure using a joint prior density function on the distribution parameters. Since our goal is to perform a fair comparison between the Bayesian and LSE methods in this paper, we assume a joint noninformative prior density function for the distribution parameters; i.e., little is known about the distribution parameters and the goal is to extract as much information as possible from the available input data. Our numerical study suggests that the LSE method performs better than the Bayesian method in capturing the characteristics of the data generated from skewed unimodal distributions of the S_B family and long-tailed distributions of the S_L family.

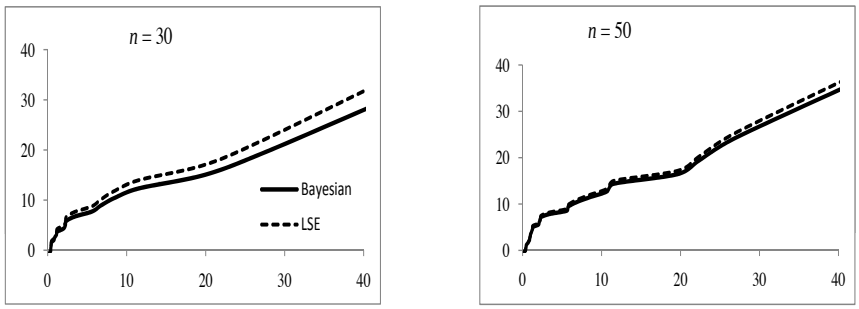

Figure 2.10: Q-Q plots for the SL-I distribution

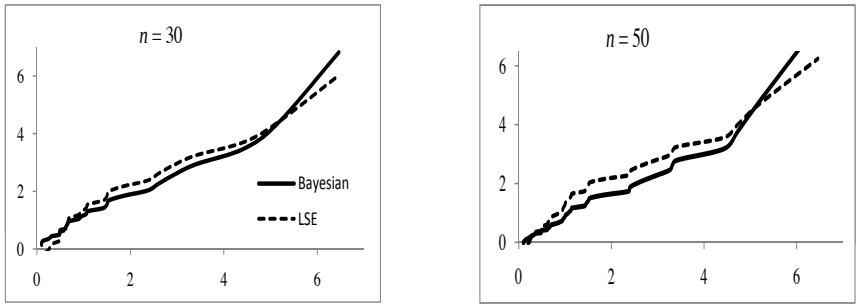

Figure 2.11: Q-Q plots for the SL-II distribution

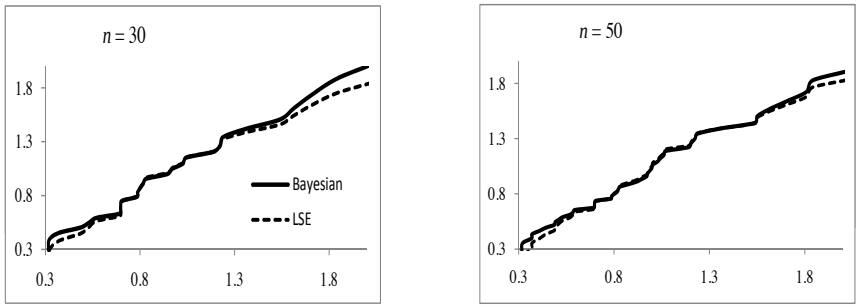

Figure 2.12: Q-Q plots for the SL-III distribution

Chapter 3

Food Banks Can Enhance Their Operations with OR/OM Tools: A Pilot Study with Greater Pittsburgh Community Food Bank [1]

3.1. Introduction

Food banks are private, non-profit organizations that procure food through several channels and make it available to people in need via agencies. These supply channels are mainly comprised of food manufacturers and retailers that have been forced to reduce their charity contributions due to the current market conditions. In addition, they have been holding fewer defective and surplus items - items that are traditionally considered for donation - with their ever-improving inventory management systems.

On the demand side, greater number of low-income people are turning to food assistance programs. In fact, a December 2008 survey conducted by Feeding America, the largest hunger relief organization in the U.S. affiliated with more than 200 food banks, reports a 30% increase in demand over last year. Among the 160 participating food banks, three-fourths have had to reduce the amount of food distributed per capita.[2] Below we provide a short summary of these surveys reported by two small-sized (Cincinnati and Cleveland) and two large-sized (Chicago and New York City) food banks:

- FreeStore Food Bank located at Cincinnati: Reports 25% increase in demand over the

[1] Part of Section 3.4 of this chapter is published in the Proceedings of the 2010 International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems (CPAIOR), LNCS 6140, Springer, pp. 176 – 180 with co-authors Willem-Jan van Hoeve and Sridhar Tayur.

[2] The results of the survey are available from the web site http://feedingamerica.org/newsroom/local-impact-study.aspx.

past year. Increasing unemployment, rising food, fuel and mortgage and rent costs are blamed for the rise.

- Cleveland Food Bank: Reports 7% demand increase in 2008 compared to 2007. In the beginning of 2009 they are 24% ahead of where they are in 2008. In order to meet the current demand, they believe they either need more donated food or more dollars to purchase food.

- Greater Chicago Food Depository: Reports 32% increase during July-October 2008 compared to the same period in 2007. In order to meet rising demand, their member agencies have made several changes in their operations including distributing smaller bags to clients, offering less variety of foods, and limiting business hours.

- Food Bank for New York City: Reports 24% increase in the number of people turning to food assistance from 2007 to 2008 with 2 million people in 2008.

In this environment of rising demand and diminishing donations, in order to meet their demand food banks have to purchase more food and are in greater need of fundraising calls. The former requires minimizing operational costs and for the latter more compelling local data needs to be collected.

We have been working closely with our local food bank, Greater Pittsburgh Community Food Bank (GPCFB) in two areas: First, we perform a critical review of their demand and supply data, and help to validate the common belief that due to economic slowdown, GPCFB is facing increasing challenges. In the demand analysis, we find that the available demand data is sketchy and potentially erroneous. We identify possible sources of inaccuracy and provide recommendations to acquire more precise data for future use. Furthermore, in order to assess the recent changes in demand, we prepare a survey to be distributed to GPCFB's member agencies. The results of our survey support the claim that GPCFB's demand is increasing. In the supply analysis, we study the capital supply and the product supply separately. Our findings alert a significant drop in both dimensions. Second, we consider the 1-Commodity Pickup and Delivery Vehicle Routing Problem (1-PDVRP) that arises in the context of the food rescue program of GPCFB. We present a thorough study on the state-of-the-art solution methods for the 1-PDVRP, utilizing technologies Mixed Integer Programming (MIP), Constraint Programming (CP), and Constraint-based Local Search (CBLS), and evaluate potential cost savings with respect to the current practice of GPCFB.

Our results indicate that the CBLS provides solutions of good quality with at least 10% cost savings on the largest instance (weekly schedule) of GPCFB in a reasonable time. Savings can be used by GPCFB to purchase more food and to reach more people in need.

Beyond serving GPCFB, the results of this paper are also valuable for many other food banks around the U.S. with similar issues. Also, this is the first paper, to the best of our knowledge, that presents a theoretical analysis of the 1-PDVRP.

The remainder of the paper is organized as follows. In Section 3.2, GPCFB is introduced and its food procurement and distribution processes are explained. The operational aspects outlined in this section can be generalized to other food banks in the U.S. In Section 3.3, we present our data analysis. Specifically, Section 3.3.1 analyzes the demand data while Section 3.3.2 analyzes the supply data. In Section 3.4, the 1-PDVRP that arises in the context of GPCFB's food rescue program is studied. Section 3.4.1 provides a brief review of the related literature, while Section 3.4.2 presents different approaches used to model the 1-PDVRP. Finally, we conclude in Section 3.5.

3.2. A Brief Introduction to GPCFB

Greater Pittsburgh Community Food Bank, located in Duquesne, Pennsylvania, has been serving to 11 counties including Allegheny, Beaver, Butler and Lawrence since 1980. Like many others, it is a member of Feeding America (the nation's food bank network), which is the major source of supply for many food banks. In Sections 3.2.1 and 3.2.2, respectively, we present GPCFB's major supply and distribution channels, while in Section 3.2.3 we review GPCFB's daily operations.

3.2.1. GPCFB's Supply Channels

Feeding America: Feeding America collects food from several manufacturers and retailers, and makes it available to its network members such as GPCFB. Food banks can either have some items free of charge or use their bidding shares to purchase popular products. In both cases, transportation costs add up.

Regional and Local Food Donors: Regional and local manufacturers, wholesalers, retailers donate food that otherwise would be discarded due to damaged packaging or excess production.

Three Rivers Table: Three Rivers Table is GPCFB's food rescue program. Restaurants, hos-

pitals, hotels, and similar organizations have good-quality excess food which, if not collected, would be disposed at the end of the day. GPCFB's trucks pick up prepared and perishable food from these organizations and deliver it to the agencies (particularly to onsite programs) along their way. The vehicle routing problem that is studied in Section 3.4 arises as a result of this practice.

Food Drives: These are typically organized at picnics and parties where the event attendants bring a can of food with them for donation to GPCFB. Food drives are vital for GPCFB and every year they contribute a huge portion of the donated food.

Healthy Harvest: Volunteers harvest excess produce from farmers' fields for GPCFB. The harvested vegetables and fruits are distributed to low-income people living nearby, ensuring their access to fresh produce.

Purchased Food: GPCFB wants to maintain a steady inventory on nutritious but rarely donated items such as fresh produce, milk, meat and eggs. GPCFB purchases these items with funds provided by Pennsylvania's State Food Purchase Program.

Government: Government supplies food to GPCFB through the Commodity Supplemental Food Program and the Emergency Food Assistance Program. Both programs aim at improving the health of low-income women, children and seniors by providing nutritious United States Department of Agriculture food to them.

3.2.2. GPCFB's Distribution Channels

GPCFB distributes food to agencies directly or through other smaller food banks known as partner distribution organizations (PDO). Each agency demands food from its food bank based on the program it runs. In return, GPCFB or PDO provides the donated items for free and the purchased items for a small handling fee.

An agency could be running an onsite or a pantry program or both. Pantries provide a bag of food and non-food items that will be enough for 2 to 4 weeks whereas onsite programs serve cooked meal. They include soup kitchens, homeless shelters, after school programs, senior centers, community centers, etc.

3.2.3. GPCFB's Daily Operations

Most of the donated and purchased food are *picked up* by GPCFB trucks. The wholesale buying program allows bulk purchases, which should be *repacked* into smaller bags to make them ready for distribution. The repacked items are then *distributed* to the agencies

according to their orders. However, not all operations follow the pick-up $\longrightarrow$ repack $\longrightarrow$ distribute path. For example, some wholesalers and manufacturers bring their donations to GPCFB eliminating the need to pickup. Likewise, repacking is not necessary for the Tree Rivers Table program, where the prepared food is picked up and delivered consecutively. In some instances, GPCFB will not have to distribute if agencies prefer to collect their orders themselves.

3.3. Data Analysis

The goal of this section is to perform a comprehensive data review and analysis to show the extent to which GPCFB is being affected by the recent economic downturn. We present the analysis of demand data in Section 3.3.1 and the analysis of supply data in Section 3.3.2.

3.3.1. Demand Analysis

GPCFB collects demand information via agencies. Each agency, depending on the program it runs, collects statistics, and then reports to the Agency Relations Department of GPCFB through surveys each month. Among the two types of agencies, those running onsite programs and those operating pantry programs, the former types report the number of individuals served and the number of meals served. The latter types provide extensive reports including number of households served, number of individuals served, age decomposition of the individuals, the number of disabled individuals, and the number of new households (households that did not use the pantry in the last 12 months). This monthly data is used to determine the trend in demand by GPCFB.

In order to observe the trend in demand over the past few years, one has to separately identify the number of individuals served by the pantries, the number of individuals served by the onsite programs and the number of served onsite meals. Among these, we were able to observe a reliable trend only in the onsite meal quantity (see Figure 3.1) as the rest of the data is obscured. Next, we explain the problems in data collection and propose solutions to mitigate them.

Problem: An individual accessing service on a pantry is asked to fill out a "Household Information Card". Ideally, he/she should be able to use the same card for his/her next visits. For most pantries, however, multiple cards are filled out and it becomes impossible to identify the number of unique individuals served.

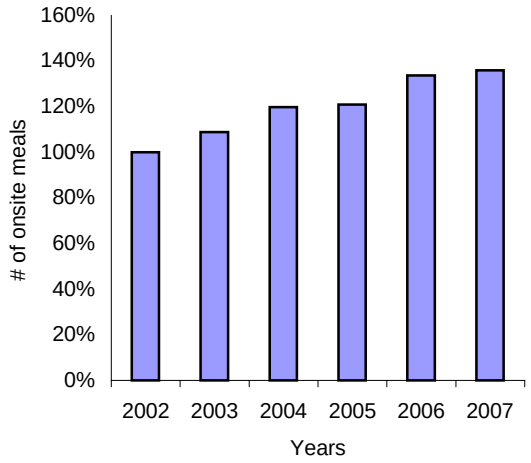

Figure 3.1: Onsite programs are serving more meals.

Solution: Improving data collection We think the pantries should prepare "Household Information Cards" tailored to each month and each individual. A participant requiring service should fill out the card that has his/her name on it. At the end of the month, by counting only the number of cards, agencies can identify the unique number of individuals served. This process is easy as it does not require any digitalization. It is very important to keep processes simple for the agencies as most of them lack the necessary technology, i.e., computers.

Problem: GPCFB's survey, conducted monthly to retrieve the demand information from agencies, can be confusing and hard to interpret. Statistics relying on this survey can possibly yield inaccurate estimates.

Solution: Improving GPCFB's survey The questions in the survey should be prepared such that all staff members working in different agencies, should understand them in the same way. Therefore, we have made several modifications in the GPCFB's survey.

Problem: At the beginning of each month, agencies are expected to report the number of individuals served, otherwise GPCFB uses the most recent agency data to calculate the total demand. It is not uncommon that some of the agencies fail to report for four to five consecutive months. In some months, almost 40% of the total demand relies on GPCFB's estimates based on the recent data. In addition, observed diverging trends in the actual and estimate numbers further deteriorate the use of GPCFB's total demand data as a good indicator of real demand.

Solution: Improving demand estimation We recommend GPCFB to analyze the estimated and the actual parts of the data separately as opposed to adding them up and reporting the

sum. This will help GPCFB to see the trend in the actual numbers more clearly.

Due to the above problems in the demand data gathering process of GPCFB, our analysis in this section does not provide an accurate picture of the trend in demand. In order to assess a more reliable estimate of demand, we conducted a survey to be filled out by the agencies. We discuss our findings in the next section.

Agency Survey

We conducted a survey in order to assess the trend in agencies' demand and to have a better understanding of agencies' challenges (see Appendix E). We mailed two kinds of surveys, a pantry survey and an onsite program survey, to each agency (in total 330 agencies) and requested each agency to fill out one or both of the surveys depending on the program it runs. We obtained a 40% response rate with 132 participating agencies. Detailed analysis of the survey has been shared with GPCFB and we present our major findings here. We believe this survey in the first place will help GPCFB in its fundraising efforts as it clearly supports the claim that food assistance programs have been visited more frequently.

Demand of Pantry Programs:

- Of the 92 pantries participated in the survey, 82% report that their demand has been increasing over the years with a sharp increase in the recent past. The reported rises range from 15% to 300%.
- 21% have been turning needy people away due to the following reasons enumerated by their importance:
    1. People requesting service are out of service area.
    2. Participants come more often than program rules allowed.
    3. People request help with their rent and transportation.
    4. Participants requesting help are above the income limits.
- Among the pantries that turn people away, 73% suggest people to go to another pantry. However, the majority of the pantries cannot make sure that those people find service in the second agency.

- 46% observe no change in their demand during summer months while 24% claim increase and other 24% report decrease.
- 78% believe that in holidays their demand is increasing.
- 83% think that events (such as superbowl) do not cause any change in their demand.
- 57% report that their clients' preferences are not affected by TV/branding.

Supply of Pantry Programs:

- 30% report GPCFB as their only source of supply; i.e., 30% completely rely on GPCFB in their food distribution.
- 74% report GPCFB as their major supplier. As a secondary supplier, 22% benefit from individual donations, and 17% purchase food through donated money.
- Almost all pantries wish to offer a rich set of food and non-food items to their clients. However, 25% report shortages in meat, 14% are short in cereal, 13%, 10% and 6.5% are short in fresh produce, dairy products and toiletries, respectively.

Demand of Onsite Programs:

- Among 59 onsite programs participated in the survey, 60% think that their demand has increased while 32% believe it is steady.
- Only 8.5% of them turn their clients away.
- 46% think that there is no change in their demand during summer time whereas 24% report increase and other 24% report decrease in their demand.
- 46% report increase in their demand during holidays.
- 73% believe events like superbowl have no effect on their demand.
- 68% think that their clients' preferences are not affected by TV/branding.

Supply of Onsite Programs:

- 36% of the onsite programs report GPCFB as their major supplier whereas for 20% GPCFB is the only source of supply.

3.3.2. Supply Analysis

We analyze item supply (food and non-food in poundages) and capital supply (in dollars) separately.

Supply (in poundages):

We compare the trend in GPCFB's received and distributed products and demonstrate that received poundages have decreased in 2007 as opposed to the increase in distributed poundages in that year (see Figure 3.2). While there is a steady trend in the received and distributed poundages for years 2004-2006, the imbalance in 2007 indicates not only a higher demand (given that the number of agencies served over the years is steady) but also a lower supply. We also note that the increase in distributed poundages throughout the years 2004-2008 is remarkable and personal communication suggests that it reveals the rise in demand. In order to communicate more recent data, we compare GPCFB's received

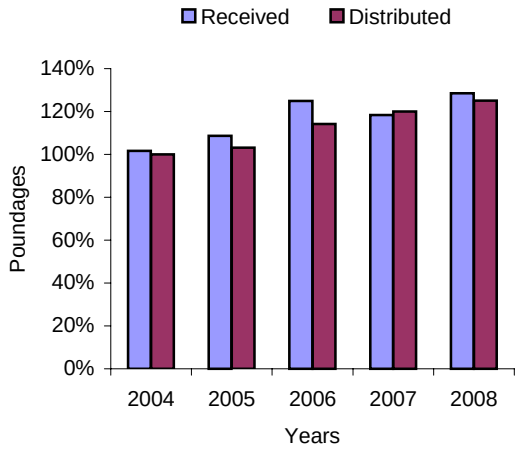

Figure 3.2: There is a clear increase in distributed poundages, however the received poundages do not follow the same trend.

products in July 2007-January 2008 (their fiscal calendar) to the received products in July 2008-January 2009 and conclude with a 11% decrease in July 2008-January 2009 (see Figure 3.3). A "received" product belongs to one of the categories: "donation" , "government commodity" or "purchase". The decrease in "received" products from 2006 to 2007 is best explained with the diminishing "government commodity" and "donated" items (see Figure 3.4). The recovery in 2008 comes both from the increase in "donated" products by 4% as a result of fundraising efforts and the rise in the "purchased" products. This, indeed, supports the claim that many food banks now have to purchase more products in order to meet

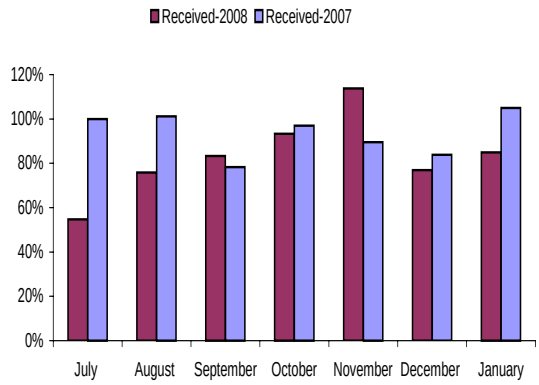

Figure 3.3: The poundages obtained by GPCFB in July 2008-January 2009 is 11% less than those obtained in July 2007-January 2008.

rising demand. Although there is a clear increasing trend in the poundage of "purchased" products over the years that might depend on several reasons, the rise in 2008 is mainly attributed to the inappropriate donations (low in nutrition). Moreover, GPCFB notes the further increasing purchases in the last months of 2008 and beginning of 2009 due to the demand-supply mismatch.

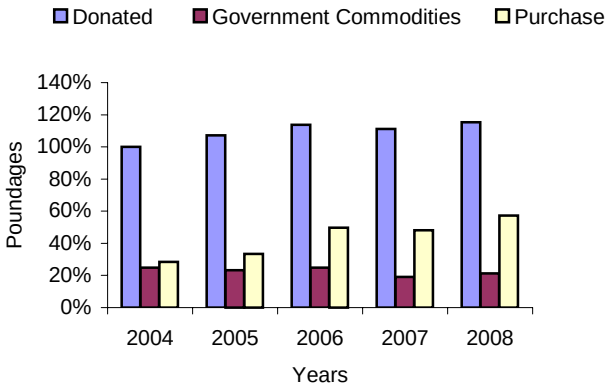

Figure 3.4: A received product is categorized as a "donation" or "government commodity" or "purchase". We note that GPCFB has had to purchase more products in the recent past.

Supply (in dollars):

In this section we compare GPCFB's total revenue and expenses over the years and show that both have been rising (see Figure 3.5). However, we detect a greater rate of increase in expenses compared to revenues for recent past (see Table 3.1). We point out the diminishing growth rate of the revenue from 2007 to 2008 and the adjustment of expenses accordingly in 2008.

We also decompose GPCFB's revenues and expenses (see Figures 3.6 and 3.7). "Fundrais-

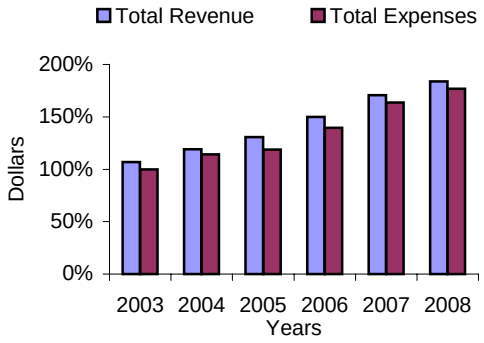

Figure 3.5: GPCFB's revenues and expenses have both been increasing.

Table 3.1: Expenses are growing at a higher pace than revenues. Moreover, 2008 shows a clear decrease in growth rates of both revenues and expenses.

Year	2004	2005	2006	2007	2008
% Growth in Revenue	12	10	15	14	7
% Growth in Expenses	14	4	18	18	8

ing and grants" contribute a significant portion of the revenues, and they are steadily increasing over the years. On the other hand, "employee benefits and salaries" is the number one expense for GPCFB followed by "purchased food" and "operating expenses" (operating expenses include transportation costs).

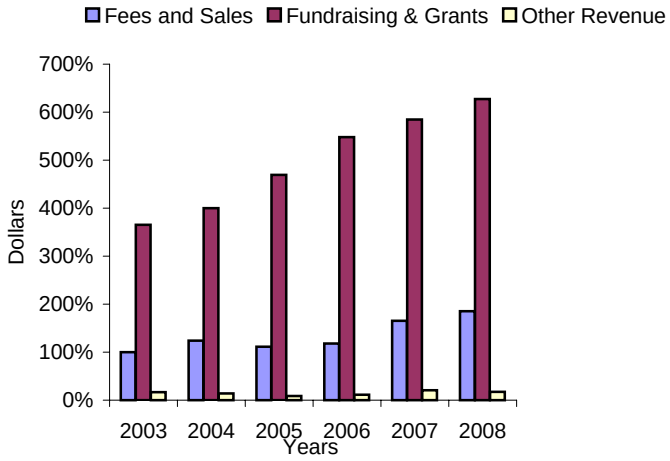

Figure 3.6: Fundraising and grants contribute to a significant portion of GPCFB's revenues.

Among the operating expenses the cost of fuel is worth looking at in more detail. GPCFB's fuel cost follows the general trend depicted by the gasoline prices (Figure 3.8). Fuel costs from 2006 to 2008, however, were steady or decreasing in the midst of an oil rally due to cancellation of several scheduled routes. Optimizing transportation schedules,

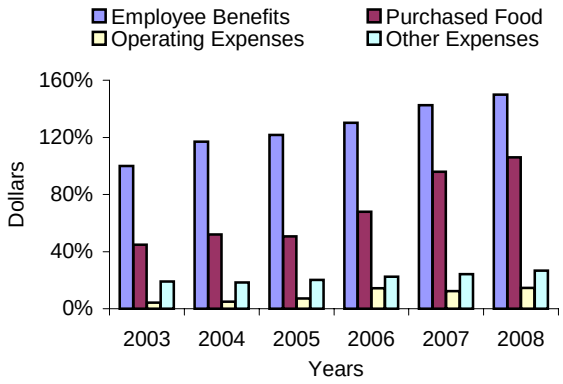

Figure 3.7: Although expenses are increasing at all categories, we point out the rise in expenditure for food purchase.

detailed in the next section, offers potential advantages that will enable GPCFB to operate at full schedule.

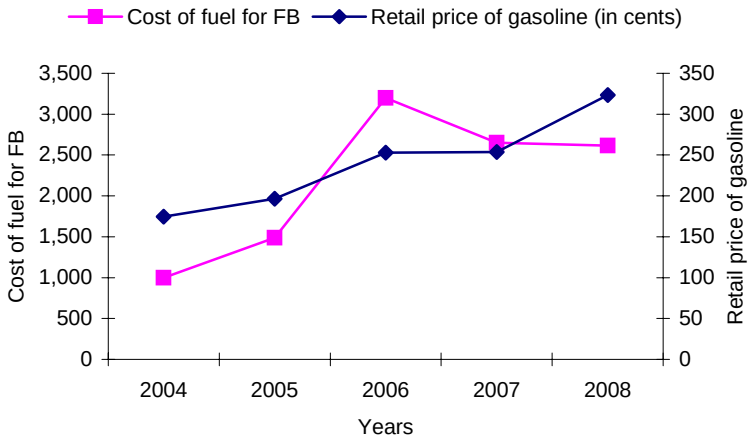

Figure 3.8: Cost of fuel for GPCFB and retail gasoline prices.

3.4. GPCFB's Pickup and Delivery Problem

As part of the Tree Rivers Table program, which is the food rescue program of GPCFB, GPCFB's fixed capacity vehicles pick up prepared food from donors and deliver it along their way to the onsite programs. The decision faced by GPCFB is to decide on the order of the pickup and delivery points such that the vehicle's capacity is never exceeded nor deceeded along with some additional constraints. The objective is to minimize the total cost of all routes generated under this program. In its most general form, this problem is related to pickup and delivery vehicle routing problems. In Section 3.4.1, we provide a

review of the pickup and delivery vehicle routing problems and conclude that our problem is named 1-Commodity Pickup and Delivery Vehicle Routing Problem (1-PDVRP), which has received limited attention. Section 3.4.2 compares off-the-shelf solution methods for 1-PDVRP, including Mixed Integer Programming (MIP), Constraint Programming (CP), and Constraint-Based Local Search (CBLS), and compare these approaches by a comprehensive numerical study that is performed on data provided by GPCFB.

3.4.1. Related Work

Pickup and delivery vehicle routing problems have been studied extensively in the literature. According to Parragh et al. (2008), two problem classes can be identified: Vehicle Routing Problem with Backhauls (VRPB) and Vehicle Routing Problem with Pickups and Deliveries (VRPPD). VRPB considers the transportation of goods from the depot to the delivery customers and from pickup customers to the depot. VRPPD, on the other hand, deals with all those problems where goods are transported between pickup and delivery locations. Figure 3.9[3] illustrates the classification scheme of Parragh et al (2008). As GPCFB's problem considers the transportation between pickup and delivery points, it falls under the category of VRPB. Furthermore, as each unit picked up can be used to satisfy the demand of each delivery point, the pickup and delivery points are unpaired, therefore we have an unpaired-VRPPD. When only one vehicle is considered, the problem is named 1-Commodity Pickup and Delivery Traveling Salesman Problem (1-PDTSP); in the multi-vehicle case it is named 1-Commodity Pickup and Delivery Vehicle Routing Problem (1-PDVRP).

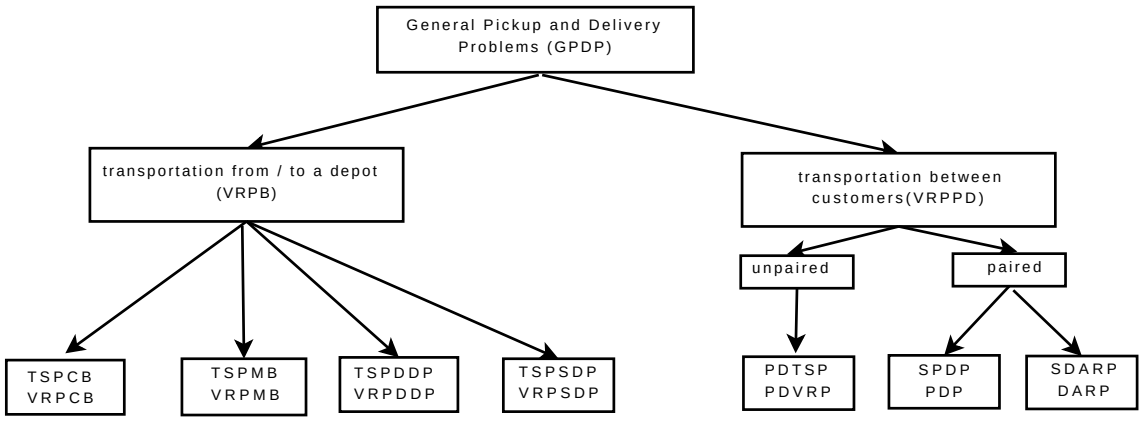

Figure 3.9: Pickup and Delivery problems.

[3]This figure is adapted from Parragh et al. (2008).

Several solution methods are developed for 1-PDTSP including exact methods, heuristics, and metaheuristics. A special case of 1-PDTSP in which the number of goods picked up is equal to the number of goods delivered and the demand (supply) at delivery (pickup) location is equal to one is studied in Chalasani and Motwani (1999) and Anily and Bramel (1999). Chalasani and Motwani (1999) propose an approximation algorithm with a worst case bound of 9.5, while Anily and Bramel (1999) devise a polynomial time iterated tour matching algorithm. Hernández-Pérez and Salazar-González (2003, 2004a, 2007) develop a branch-and-cut algorithm for 1-PDTSP, while Hernández-Pérez and Salazar-González (2003, 2004b) propose two heuristic methods that can solve instances up to 500 customers. Recently, Hernández-Pérez et al. (2009) propose a hybrid heuristic method, which combines a greedy randomized adaptive search procedure with variable neighborhood search. The proposed algorithm can solve instances up to 1000 customers.

The only paper that considers the 1-PDVRP, to the best of our knowledge, is Dror et al. (1998). The authors present different approaches, including MIP, CP and Local Search, which are applied to instances involving nine locations. The approaches we consider for the 1-PDVRP in the next section are similar in spirit to those of Dror et al. (1998). Our MIP model is quite different, however. Further, although the CP and CBLS models are based on the same modeling concepts, the underlying solver technology has been greatly improved over the years.

3.4.2. Different Approaches to the 1-PDVRP

Input Data and Parameters

Let the set V denote the set of locations, and let $O \in V$ denote the origin (or depot) from which the vehicles depart and return. With each location i in V we associate a number $q_i \in R$ representing the quantity to be picked up ($q_i > 0$) or delivered ($q_i < 0$) at i. The distance between two locations i and j in V will be denoted by d_{ij}. Distance can be represented by length or time units.

Let T denote the set of vehicles (or trucks). For simplicity, we assume that all vehicles have an equal 'volume' capacity Q of the same unit as the quantities q to be picked up (e.g., pounds). In addition, all vehicles are assumed to have an equal 'horizon' capacity H of the same unit as the distances d.

Mixed Integer Programming

Our MIP model is based on column generation. The master problem of our column generation procedure consists of a set of 'columns' S representing feasible routes. The routes are encoded as binary vectors on the index set V of locations; that is, the actual order of the route is implicitly encoded. The columns are assumed to be grouped together in a matrix A of size V by S. The length of the routes is represented by a 'cost' vector $c \in R^{|S|}$. We let $z \in \{0,1\}^{|S|}$ be a vector of binary variables representing the selected routes. The master problem can then be encoded as the following set covering model:

$$\begin{array}{ll} \text{Minimize} & c^{\mathsf{T}} z \\ \text{subject to} & Az = \vec{1} \end{array} \tag{3.1}$$

For our column generation procedure, we will actually solve the continuous relaxation of (3.1), which allows us to use the shadow prices corresponding to the constraints. We let λ_j denote the shadow price of constraint j in (3.1), where $j \in V$. The steps of the column generation scheme can be summarized as follows.

Column Generation Scheme

Step 1. Solve the master problem and determine the dual variables λ_j, $j \in V$.

Step 2. Use the dual variables to solve the subproblem, i.e., in computing the objective function of the subproblem.

Step 3. If the objective function of the subproblem is less than zero, add the generated column to the master problem and revisit Step 1. Otherwise, stop. Recall that the objective function of the subproblem is nothing but the reduced cost of the new generated column. In a minimization problem, any column with negative reduced cost is a candidate to enter the basis as it can improve the current solution.

Step 4. Solve the master problem as an integer program. Determine the optimality gap.

The subproblem for generating new feasible routes uses a model that employs a flow-based representation on a layered graph, where each layer consists of nodes representing all

locations. The new route comprises M steps, where each step represents the next location to be visited. We can safely assume that M is the minimum of $|V|+1$ and (an estimate on) the maximum number of locations that 'fit' in the horizon H for each vehicle. We present the layered graph used for generation new routes in Figure 3.10.

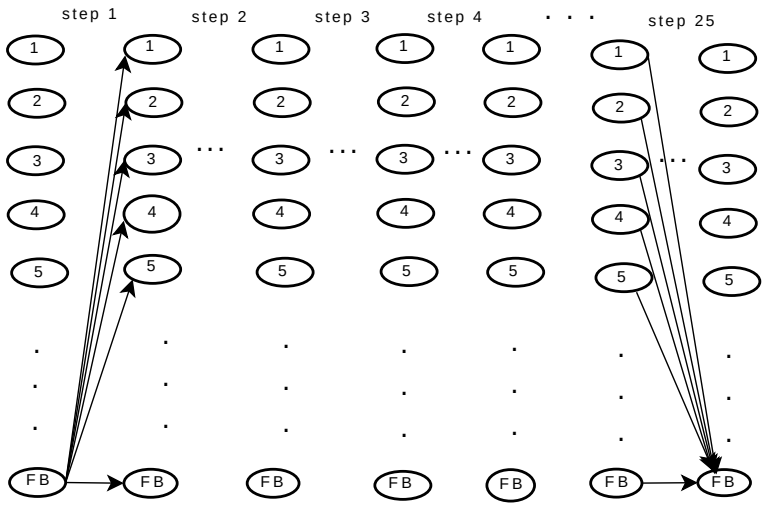

Figure 3.10: Subproblem route generation schema

We let x_{ijk} be a binary variable that represents whether we travel from location i to location j in step k. We further let y_j be a binary variable representing whether we visit location j at any time step. The vector of variables y will represent the column to be generated. Further, variable I_k represents the inventory of the vehicle, while variable D_k represents the total distance traveled up to step k, where $k = 0, \ldots, M$. We let $D_0 = 0$, while $0 \leq I_0 \leq Q$. The problem of finding an improving route can then be modeled as follows.

$$
\begin{aligned}
\text{Minimize} \quad & \textstyle\sum_{i\in V}\sum_{j\in V}\sum_{k=1}^{M} d_{ij}x_{ijk} - \sum_{j\in V}\lambda_j y_j & \\
\text{subject to} \quad & \textstyle\sum_{j\in V} x_{O,j,1} = 1 & \\
& \textstyle\sum_{j\in V} x_{i,j,1} = 0 & \forall i \in V \setminus O \\
& \textstyle\sum_{i\in V} x_{i,O,M} = 1 & \\
& \textstyle\sum_{i\in V} x_{i,j,M} = 0 & \forall j \in V \setminus O \\
& x_{O,j,k} = 0 & \forall j \in V \setminus O, \forall k \in [1..M] \\
& \textstyle\sum_{i\in V} x_{ijk} = \sum_{l\in V} x_{j,l,k+1} & \forall j \in V, \forall k \in [1..M-1]
\end{aligned}
$$

$$\sum_{j \in V \setminus O} \sum_{k=1}^{M} x_{ijk} \leq 1 \qquad \forall j \in V \setminus O$$
$$I_k = I_{k-1} + \sum_{i \in V} \sum_{j \in V} q_i x_{ijk} \qquad \forall k \in [1..M]$$
$$0 \leq I_k \leq Q \qquad \forall k \in [0..M]$$
$$D_k = D_{k-1} + \sum_{i \in V} \sum_{j \in V} d_{ij} x_{ijk} \qquad \forall k \in [1..M]$$
$$0 \leq D_k \leq H \qquad \forall k \in [0..M]$$
$$\sum_{i \in V} \sum_{k=1}^{M} x_{ijk} = y_j \qquad \forall j \in V$$

In this model, the first four sets of constraints ensure that we leave from and finish at the origin. The fifth set of constraints enforce that we can enter the origin at any time, but not leave it again. The sixth set of constraints model the flow conservation at each node, while the seventh set of constraints prevent the route from visiting a location more than once. The following four sets of constraints represent the capacity constraints of the vehicle in terms of quantities picked up and delivered, and in terms of distance. The last set of constraints link together the 'flow' variables x with the new column represented by the variables y.

Constraint Programming

Our CP model is based on a well-known interpretation of the VRP as a multi-machine job scheduling problem with sequence-dependent setup times. In the CP literature, this is usually modeled using alternative resources (the machines) and activities (the jobs). That is, each visit to a location corresponds to an activity, and each vehicle corresponds to two (linked) resources: one 'unary resource' modeling the distance constraint, and one 'reservoir' modeling the inventory of the vehicle. With each activity we associate variables representing its start time and end time, as well as a fixed duration (this can be 0 if we assume that the (un-)loading time is negligible). Further, each activity either depletes or replenishes the inventory reservoir of a vehicle. The distance between two locations is modeled as the 'transition time' between the corresponding activities. We minimize the sum of the completion times of all vehicles.

All these concepts are readily available in most industrial CP solvers. We have implemented the model in ILOG Solver 6.6 (which includes ILOG Scheduler). A snapshot of the ILOG model for a single vehicle is provided in Figure 3.11. It shows that the concepts presented above can almost literally be encoded as a CP model.

```
IloReservoir truckReservoir(ReservoirCapacity, 0);
truckReservoir.setLevelMax(0, TimeHorizon, ReservoirCapacity);

IloUnaryResource truckTime();
IloTransitionTime T(truckTime, Distances);

vector<IloActivity> visit;
visit = vector<IloActivity>(N);

for (int i=0; i<N; i++) {
  visit[i].requires(truckTime);
  if (supply[i] > 0)
    visit[i].produces(truckReservoir, supply[i]);
  else
    visit[i].consumes(truckReservoir, -1*supply[i]);
}
```

Figure 3.11: Snapshot of the ILOG Scheduler model for a single vehicle

Constraint-Based Local Search

Our final approach uses CBLS. With CBLS we can express the problem similar to a CP model, which will then be used to automatically derive the neighborhoods and penalty function needed to define a local search procedure. Our CBLS is based on the semantics offered by ILOG Dispatcher (included in ILOG Solver 6.6). These semantics are specifically designed to model routing problems. A snapshot of the ILOG model for a single vehicle is provided in Figure 3.12.

ILOG Dispatcher uses the concepts *nodes*, *vehicles*, and *visits*. The nodes are defined by the coordinates of the locations, and contain as an attribute the amount to be picked up or delivered. The vehicles contain several attributes, including time, distance, and weight (load). Vehicles also contain, by default, a 'unary resource' constraint with respect to time, and a 'capacity' constraint with respect to the load, similar to the resources in ILOG Scheduler. The attributes of visits include the location, the quantity to be picked up (positive) or delivered (negative), a time window, and possibly other problem-specific constraints.

In a first phase, we create a feasible solution. ILOG Dispatcher uses various heuristics for this, including a nearest-neighbourhood heuristic that we applied in our experiments. Where applicable, we started from the current schedule that we extracted from the data.

The second phase improves upon the starting solution using various local search meth-

```
class RoutingModel {
   ...
   IloDimension2 _time;
   IloDimension2 _distance;
   IloDimension1 _weight;
   ...
}

IloNode node( <read coordinates from file> );

IloVisit visit(node);
visit.getTransitVar(_weight) == Supply);
minTime <= visit.getCumulVar(_time) <= maxTime;
visit.getCumulVar(_weight) >= 0);

IloVehicle vehicle(firstNode, lastNode);
vehicle.setCapacity(_weight, Capacity);
vehicle.setCost(_distance);
```

Figure 3.12: Snapshot of the ILOG Dispatcher model for a single vehicle

ods. We applied successively the methods IloTwoOpt, IloOrOpt, IloRelocate, IloCross and IloExchange. Within each method, we take the first legal cost-decreasing move encountered.

Evaluation

Our experimental results are performed on data provided by Greater Pittsburgh Community Food Bank. Their food rescue program visits 130 locations per week. The provided data allowed us to extract a fairly accurate estimate on the expected pickup amount for the donor locations. The precise delivery amounts were unknown, and we therefore approximate the demand based on the population served by each location (which is known accurately), scaled by the total supply. We allow the total demand to be slightly smaller than the total supply, to avoid pathological behavior of the algorithm. We note however, that although this additional 'slack' influences the results, the qualitative behavior of the different techniques remains the same. The MIP model is solved using ILOG CPLEX 11.2, while the CP and CBLS model are solved using ILOG Solver 6.6, all on a 2.33GHz Intel Xeon machine.

The first set of instances are for individual vehicles, on routes serving 13 to 18 locations (corresponding to a daily schedule). The second set of instances group together schedules over multiple days, ranging from 30 to 130 locations. The results are presented in Table 3.2.

We report for each instance the cost savings (in terms of total distance traveled) with respect to the current operational schedule. Here, $|V|$ and $|T|$ denote the number of locations and vehicles, respectively. The optimal solutions found with MIP and CP took several (2–3) minutes to compute, while the solutions found with CBLS took several seconds or less. The time limit was set to 30 minutes.

Table 3.2: Savings obtained with different approaches.

$\|V\|$	$\|T\|$	MIP	CP	CBLS
13	1	12%	12%	12%
14	1	15%	15%	14%
15	1	-	7%	6%
16	1	-	5%	3%
18	1	-	16%	15%
30	2	-	-	4%
60	4	-	-	8%
130	9	-	-	10%

Our experimental results indicate that on this problem domain, our MIP model is outperformed by our CP model to find an optimal solution (we note that a specialized 1-PDTSP MIP approach such as Hernández-Pérez (2007) might perform better than our 'generic' MIP model on the single-vehicle instances). Further, the CP model is able to find optimal solutions for up to 18 locations and one vehicle; for a higher number of locations or vehicles, the CP model is unable to find even a single solution. Lastly, the CBLS approach is able to handle large-scale instances, up to 130 locations and 9 vehicles. The expected savings are substantial, being at least 10% on the largest instance.

3.5. Conclusion

Our conclusions can be summarized as follows.

1. Data Analysis: The review of GPCFB's demand and supply data shows that the demand data is obscured. The possible sources of inaccuracy are identified and we recommend ways to help GPCFB to gather more reliable demand data for future use. By collecting more precise data, GPCFB can improve its fundraising as well as its planning and execution of purchases.

2. Transportation: The 1-PDVRP, which appears in the context of the food rescue program of GPCFB, received very limited attention in the vehicle routing literature. We

model this problem using three different approaches including MIP, CP and CBLS. CBLS provides solutions of good quality with at least 10% cost savings compared to the current practice of GPCFB in designing the weekly schedule of the food rescue program. Savings can be used to purchase more food and to reach more people in need.

Conclusions

This thesis has presented a Bayesian model for the representation as well as the quantification of the stochastic uncertainty and the parameter uncertainty in large-scale stochastic simulations with correlated inputs, the comparison of least-squares and Bayesian inferences for Johnson's lognormal and bounded distributions, and a collaborative work with the Greater Pittsburgh Community Food Bank focusing on their demand and supply data analysis and the 1-Commodity Pickup and Delivery Vehicle Routing Problem that arises as a result of their food rescue program.

In the first chapter, we present a Bayesian model for capturing the parameter uncertainty in addition to the stochastic uncertainty in the mean performance measure and the confidence interval of a stochastic simulation with correlated inputs. We further decompose the variance of the simulation output into terms related to stochastic uncertainty and the parameter uncertainty. Such decomposition of the simulation output variance can be used to reduce the parameter uncertainty by developing data-collection schemas. We show that our Bayesian model improves the consistency of the mean line-item fill-rate estimates and the coverage of the confidence intervals in the simulation of a multi-product inventory system with correlated demands.

In the second chapter, we compare the least-squares estimation method and the Bayesian method in fitting the lognormal and the bounded families of the Johnson translation system, and provide insights for the simulation practitioners on when to use each fitting method. Our comprehensive experimental analysis shows that the least-squares estimation method performs better than the Bayesian method for the skewed unimodal distributions of the bounded family and long-tailed distributions of the lognormal family.

Finally, in the third chapter, we investigate how GPCFB is being affected by the recent economic downturn. We perform a critical review of their demand and supply data, and suggest methods to improve their demand data gathering process. Our analysis clearly shows the diminishing supply and increasing expenses of GPCFB, which forced them to cancel some of the routes in designing the schedules of their food rescue program. We focus on the vehicle routing problem that arises in the context of the food rescue program, the 1-PDVRP, with the purpose of investigating potential cost savings. To this end, we model

the 1-PDVRP using IP, CP, and CBLS, and our results show that CBLS provides at least 10% cost savings over the current practice of GPCFB. This chapter may be a starting point for others willing to help their local food banks.

Appendix A

Section A.1 presents an algorithm that generates random variates from a gamma density function, Section A.2 provides the Gibbs sampler algorithm that samples shape and scale parameters of a component with a gamma distribution, and Section A.3 describes how to sample a two-dimensional correlation matrix from an inverse Wishart density function. After the presentation of each algorithm, we describe how to use the algorithm for supporting our Bayesian model development.

A.1. Generating a random variate, x of the gamma random variable X with shape parameter α and scale parameter β

We consider a random variable, X that has a gamma distribution with shape parameter α (> 1) and scale parameter β. First, we present the procedure **GammaVariate**(α,β) that generates a random variate, x from this gamma distribution. Then, we describe how to use this procedure for supporting the sampling of the NORTA parameters in Section 1.3 of the paper.

GammaVariate(α,β) The generation of the gamma random variate x with shape parameter α and scale parameter β starts with the generation of the gamma random variate x' with shape parameter α and scale parameter 1. Then x' is multiplied by the scale parameter β to obtain x. Cheng (1977) recommends the following algorithm for doing this (Law 2007):

(1) Prespecify constants $a = 1/\sqrt{2\alpha - 1}$, $b = \alpha - \ln 4$, $q = \alpha + a^{-1}$, $\theta = 4.5$, and $d = 1 + \ln \theta$.

(2) Generate independent and identically distributed random variates u_1 and u_2 from the uniform distribution on the unit interval.

(3) Let $v = a \ln[u_1/(1 - u_1)]$, $y = \alpha \exp(v)$, $z = u_1^2 u_2$, and $w = b + qv - y$.

(4) If $w + d - \theta z \geq 0$, then let $x' = y$. Otherwise, proceed to the next step.

(5) If $w \geq \ln z$, then let $x' = y$. Otherwise, go back to the first step.

(6) Return $x = \beta x'$.

We use this generation procedure in Section 1.3.1 of the paper as follows:

- We choose the distribution of the i^{th} component X_i as exponential with scale parameter β_i. Assuming the availability of the historical input data $x_{i,t}$, $t = 1, 2, \ldots, n$ of length n, we select the conjugate, inverse gamma density function with shape parameter θ_i and scale parameter ν_i as the prior density function on the scale parameter β_i. We show that the i^{th} component parameter uncertainty can be captured by sampling β_i^{-1} from the gamma distribution with shape parameter $n + \theta_i$ and scale parameter $(\nu_i + \sum_{t=1}^{n} x_{i,t})^{-1}$. This can be easily done by using the procedure **GammaVariate** $(n + \theta_i, (\nu_i + \sum_{t=1}^{n} x_{i,t})^{-1})$; i.e.,

$$\tilde{\beta}_i^{-1} \quad = \quad \textbf{GammaVariate}\left(n + \theta_i, \left(\nu_i + \sum_{t=1}^{n} x_{i,t}\right)^{-1}\right).$$

- We also consider the case component X_i has the gamma distribution with shape parameter α_i and scale parameter β_i. We use Jeffreys' prior for the parameters of this gamma component and obtain a posterior density function that allows the sampling of the parameters via the Gibbs sampler algorithm. The resulting Gibbs sampler algorithm is provided in Appendix A.2. One of the steps of this algorithm is to sample scale parameter β_i^{-1} from the gamma distribution with shape parameter $n\alpha_i$ and scale parameter $(\sum_{t=1}^{n} x_{i,t})^{-1}$. This can be done by using the procedure **GammaVariate**$(n\alpha_i, (\sum_{t=1}^{n} x_{i,t})^{-1})$; i.e.,

$$\tilde{\beta}_i^{-1} \quad = \quad \textbf{GammaVariate}\left(n\alpha_i, \left(\sum_{t=1}^{n} x_{i,t}\right)^{-1}\right).$$

A.2. Sampling the gamma distribution parameters via the Gibbs sampler algorithm

Figure A.1 presents the Gibbs sampler algorithm, a widely used Markov Chain Monte Carlo method that samples shape parameter α_i and scale parameter β_i of the gamma component X_i from the parameters' conditional posterior density functions. Specifically, the conditional

Figure A.1: Gibbs sampler algorithm for parameters β_i and α_i of the gamma component.

set β_i^0 and α_i^0 to their maximum likelihood estimates

for $\ell = 1, 2, \ldots, \mathcal{L}^*, \mathcal{L}^* + 1, \ldots, \mathcal{L}$ replications **do**

generate β_i^ℓ from $p_i(\beta_i|\alpha_i^{\ell-1}, \mathbf{x}_i) \propto \beta_i^{-\alpha_i^{\ell-1} n - 1} \exp\{-\sum_{t=1}^n x_{i,t}/\beta_i\}$

generate α_i^ℓ from $p_i(\alpha_i|\beta_i^\ell, \mathbf{x}_i) \propto \left[\beta_i^\ell\right]^{-\alpha_i n} / \left[\Gamma(\alpha_i)\right]^n \prod_{t=1}^n x_{i,t}^{\alpha_i}$

end loop

analyze the output data β_i^ℓ, $\ell = \mathcal{L}^*, \mathcal{L}^* + 1, \ldots, \mathcal{L}$ and α_i^ℓ, $\ell = \mathcal{L}^*, \mathcal{L}^* + 1, \ldots, \mathcal{L}$ for generating point estimates and confidence intervals of parameters β_i and α_i

posterior density function $p_i(\beta_i|\alpha_i, \mathbf{x}_i)$ of parameter β_i is inverse gamma with shape parameter $n\alpha_i$ and scale parameter $\sum_{t=1}^n x_{i,t}$. Therefore, β_i^{-1} can be sampled using the generation procedure of Appendix A.1; i.e.,

$$\tilde{\beta}_i^{-1} \quad = \quad \textbf{GammaVariate}\left(n\alpha_i, \left(\sum_{t=1}^n x_{i,t}\right)^{-1}\right).$$

Although the conditional posterior density function $p_i(\alpha_i|\beta_i, \mathbf{x}_i)$ of parameter α_i does not have a standard form, it is a logconcave function; i.e., $\partial^2 \log p_i(\alpha_i|\beta_i, \mathbf{x}_i)/\partial\alpha_i^2 \leq 0$. Therefore, parameter α_i can be sampled using the adaptive rejection sampling (ARS) algorithm of Gilks and Wild (1992). We present this algorithm in the remainder of the section. For ease of presentation, we denote the conditional posterior density function $p_i(\alpha_i|\beta_i, \mathbf{x}_i)$ with $h_i(\alpha_i)$, its logarithm with $M_i(\alpha_i)$, i.e.,

$$M_i(\alpha_i) \quad = \quad -n\alpha_i \log(\beta_i) - n\log(\Gamma(\alpha_i)) + \alpha_i \sum_{t=1}^n \log(x_{i,t}),$$

and the domain of $h_i(\alpha_i)$, i.e., the set of α_i for which $h_i(\alpha) > 0$, with D_i. We represent the set of ℓ different α_i values with $T_\ell(i) = \{\alpha_{i,p}; p = 1, 2, \ldots, \ell\}$, where $\alpha_{i,p}$, $p = 1, 2, \ldots, \ell$ are arranged in an ascending order. We use $\alpha_{i,(p)}$ for the p^{th} lowest element of $T_\ell(i)$. We define $L_{i,p,q}(\alpha_i; T_\ell(i))$ as the straight line passing through points $(\alpha_{i,p}, M_i(\alpha_{i,p}))$ and $(\alpha_{i,q}, M_i(\alpha_{i,q}))$ for $1 \leq p \leq q \leq n$. Additionally, we use $u_\ell(\alpha_i)$ for the piecewise linear function that is defined as the minimum of $L_{i,p-1,p}(\alpha_i; T_\ell(i))$ and $L_{i,p+1,p+2}(\alpha_i; T_\ell(i))$, where $\alpha_{i,p} \leq \alpha_i < \alpha_{i,p+1}$. Since $h_i(\alpha)$ is logconcave, $u_\ell(\alpha_i)$ is an envelope (i.e., an upper hull) for $M_i(\alpha_i)$; i.e., $u_\ell(\alpha_i) \geq M_i(\alpha_i)$ $\forall \alpha_i \in D_i$. Finally, we define the normalized exponential hull of the upper

hull as $s_\ell(\alpha_i) \equiv \exp\left(u_\ell(\alpha_i)\right) / \int_{D_i} \exp\left(u_\ell(\alpha_i)\right) d\alpha_i$. Since $s_\ell(\alpha_i)$ is a piecewise exponential density function, α_i can be easily sampled from $s_\ell(\alpha_i)$ (Law 2007).

The ARS algorithm is composed of the steps of initialization, sampling, evaluation, rejection, and acceptance:

Initialization The algorithm starts with the initialization of ℓ to 2 and the construction of $T_2(i)$ with $\alpha_{i,1} = -1$ and $\alpha_{i,2} = +1$. These two values have been reported to provide good starting values for high computational efficiency (Gilks and Wild 1992).

Sampling $\alpha_i^\dagger$ and w are sampled independently from $s_\ell(\alpha_i)$ and the uniform distribution function on the unit interval.

Evaluation If w is greater than $h_i(\alpha_i^\dagger)/\exp(u_\ell(\alpha_i^\dagger))$, then the algorithm moves to the rejection step. Otherwise, the algorithm moves to the acceptance step.

Rejection $\alpha_i^\dagger$ is used for forming $T_{\ell+1}(i)$ whose elements are to be arranged in an ascending order. The algorithm goes back to the sampling step.

Acceptance Return $\alpha_i^\dagger$ as the value sampled for α_i.

This completes our description of sampling parameters β_i and α_i from conditional posterior density functions $p_i(\beta_i|\alpha_i, \mathbf{x}_i)$ and $p_i(\alpha_i|\beta_i, \mathbf{x}_i)$, respectively. We provide further details of implementing the Gibbs sampler algorithm in Appendix C.

A.3. Sampling a two-dimensional correlation matrix, $\mathbf{\Sigma}_2$ from the inverse Wishart density function with parameters n and $\mathbf{S}_2$

The procedure **InverseWishartVariate**(n,$\mathbf{S}_2$) samples correlation matrix $\mathbf{\Sigma}_2$ from the inverse Wishart density function with parameters n and $\mathbf{S}_2$ (Rossi et al. 2006):

InverseWishartVariate(n,$\mathbf{S}_2$)

(1) Generate random variates a, b, and c, respectively, from a chi-square probability density function with n degrees of freedom, a standard normal probability density function, and a chi-square probability density function with $n-1$ degrees of freedom.

(2) Use a, b, and c for constructing the following two-dimensional matrix:

$$\mathbf{T}_2 = \begin{pmatrix} \sqrt{a} & 0 \\ b & \sqrt{c} \end{pmatrix}$$

(3) Determine matrix $\mathbf{W}_2$ that satisfies $\mathbf{W}_2'\mathbf{W}_2 = \mathbf{S}_2^{-1}$.

(4) Define matrix $\mathbf{C}_2$ as the multiplication of the transpose of $\mathbf{T}_2$ and $\mathbf{W}_2$; i.e., $\mathbf{C}_2 = \mathbf{T}_2'\mathbf{W}_2$.

(5) Return $\mathbf{\Sigma}_2 = \mathbf{C}_2^{-1}(\mathbf{C}_2^{-1})'$.

We use this procedure for sampling each (partial) correlation matrix of the $\mathcal{C}-$vine specification of the NORTA random vector:

- In Section 1.3.2, we assume a conjugate, inverse Wishart density function for the (unconditional) correlation matrix $\mathbf{\Sigma}_2(1,i)$ associated with the correlation $\rho(1,i)$ between random variables $\Phi^{-1}[F_1(X_1;\mathbf{\Psi}_1)]$ and $\Phi^{-1}[F_i(X_i;\mathbf{\Psi}_i)]$. We show that sampling correlation $\rho(1,i)$ reduces to the sampling of $\mathbf{\Sigma}_2(1,i)$ from the inverse Wishart density function with parameters n and $\mathbf{S}_2(1,i|\mathbf{\Psi}_1,\mathbf{\Psi}_i,\mathbf{x}_1,\,\mathbf{x}_i)$. Therefore, we sample $\mathbf{\Sigma}_2(1,i)$ using the procedure **InverseWishartVariate**$(n,\mathbf{S}_2(1,i|\mathbf{\Psi}_1,\mathbf{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i))$, i.e.,

$$\tilde{\mathbf{\Sigma}}_2(1,i) \;=\; \textbf{InverseWishartVariate}\left(n,\mathbf{S}_2\left(1,i|\mathbf{\Psi}_1,\mathbf{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i\right)\right),$$

and set the $(1,2)^{\text{th}}$ entry of $\tilde{\mathbf{\Sigma}}_2(1,i)$ to $\tilde{\rho}(1,i)$.

- Section 1.3.2 shows that sampling partial correlation $\rho(j-1,i;1,2,\ldots,j-2)$ reduces to the sampling of the partial correlation matrix $\mathbf{\Sigma}_2(j-1,i;1,2,\ldots,j-2)$ from the inverse Wishart density function with parameters n and $\mathbf{S}_2(j-1,i;1,2,\ldots,j-2|\mathbf{\Lambda}_j,\mathbf{x})$. Therefore, we use the procedure **InverseWishartVariate**$(n,\mathbf{S}_2(j-1,i;1,2,\ldots,j-2|\mathbf{\Lambda}_j,\mathbf{x}))$, i.e.,

$$\tilde{\mathbf{\Sigma}}_2(j-1,i;1,2,\ldots,j-2) \;=\; \textbf{InverseWishartVariate}\left(n,\mathbf{S}_2\left(j-1,i;1,2,\ldots,j-2|\mathbf{\Lambda}_j,\mathbf{x}\right)\right),$$

and set the $(1,2)^{\text{th}}$ entry of $\tilde{\mathbf{\Sigma}}_2(j-1,i;1,2,\ldots,j-2)$ to $\tilde{\rho}(j-1,i;1,2,\ldots,j-2)$.

Appendix B

In this section, we consider a 5−dimensional NORTA distribution and assume that component X_i is exponentially distributed with scale parameter β_i for $i = 1, 2, \ldots, 5$. Our objective is to describe the implementation of the sampling algorithm presented in Figure 1.2 of the paper for this NORTA distribution. First, we present the NORTA parameters and their joint posterior density function in Appendix B.1. Then, we present the recursive formulas of Yule and Kendall (1965) in Appendix B.2 and Theorem 3.3.4 of Tong (1990) in Appendix B.3. Finally, we present the NORTA parameter sampling algorithm in Appendix B.4.

B.1. Joint posterior density function of the NORTA parameters

The first plot in Figure 1.1 of the paper illustrates the $\mathcal{C}-$vine of the 5−dimensional NORTA distribution. We conclude from this illustration that the NORTA parameters are composed of the following component scale parameters and the (partial) correlations assigned to the edges of the vine:

$$\begin{array}{ccccc} \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \\ & \rho(1,2) & \rho(1,3) & \rho(1,4) & \rho(1,5) \\ & & \rho(2,3;1) & \rho(2,4;1) & \rho(2,5;1) \\ & & & \rho(3,4;1,2) & \rho(3,5;1,2) \\ & & & & \rho(4,5;1,2,3) \end{array} \tag{B.1}$$

The joint posterior density function of the NORTA parameters in (B.1) is given by

$$\begin{aligned} &\propto \prod_{i=1}^{5} \beta_i^{-(n+\theta_i+1)} \exp\left\{-\frac{\nu_i + \sum_{t=1}^{n} x_{i,t}}{\beta_i}\right\} \qquad \text{(B.2)} \\ &\times \prod_{i=2}^{5} |\boldsymbol{\Sigma}_2(1,i)|^{-(n+3)/2} \exp\left\{\text{tr}\left(-\frac{1}{2}\mathbf{S}_2(1,i|\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_i,\mathbf{x}_1,\mathbf{x}_i)\boldsymbol{\Sigma}_2^{-1}(1,i)\right)\right\} \\ &\times \prod_{j=3}^{5}\prod_{i=j}^{5}\Bigg(|\boldsymbol{\Sigma}_2(j-1,i;1,2,\ldots,j-2)|^{-(n+3)/2} \\ &\qquad \exp\left\{\text{tr}\left(-\frac{1}{2}\mathbf{S}_2(j-1,i;1,2,\ldots,j-2|\boldsymbol{\Lambda}_j,\mathbf{x})\,\boldsymbol{\Sigma}_2^{-1}(j-1,i;1,2,\ldots,j-2)\right)\right\}\Bigg). \end{aligned}$$

The two-dimensional matrices $\mathbf{S}_2(1, i|\mathbf{\Psi}_1, \mathbf{\Psi}_i, \mathbf{x}_1, \mathbf{x}_i)$ and $\mathbf{S}_2(j-1, i; 1, 2, \ldots, j-2|\mathbf{\Lambda}_j, \mathbf{x})$ of this density function are defined in Section 1.3.2 of the paper.

A natural question to ask is how to determine the parameters θ_i, $i = 1, 2, \ldots, 5$ and ν_i, $i = 1, 2, \ldots, 5$ of the component prior density functions. A way to do this is to use the moment-matching method (Berger 1985). This method starts with asking the decision maker what the likely values are for certain moments of the unknown parameters (e.g., mean and variance). Then, the parameters of the prior density function are obtained by equating the functional forms of the moments of the unknown parameters to the answers gathered from the decision maker. For example, the following questions can be asked for the i^{th} component of the NORTA random vector: (1) What is the likely value for the unknown mean of the scale parameter β_i? (2) What is the most likely range for β_i? It follows from the definition of the inverse gamma distribution function that $\mathrm{E}(\beta_i) = \nu_i/(\theta_i - 1)$ for $\theta_i > 1$ and $\mathrm{Var}(\beta_i) = \nu_i^2/(\theta_i - 1)^2(\theta_i - 2)$ for $\theta_i > 2$. Therefore, if we assume that the decision maker answers the first question as a_i ($\equiv \nu_i/(\theta_i - 1)$) and the second question as b_i ($\equiv \nu_i^2/(\theta_i - 1)^2(\theta_i - 2)$), then the parameters of the component prior density functions are identified as $\theta_i = (a_i^2 - 2b_i)/b_i$, $i = 1, 2, \ldots, 5$ and $\nu_i = a_i(a_i^2 - 3b_i)/b_i$, $i = 1, 2, \ldots, 5$.

B.2. Recursive formulas of Yule and Kendall (1965)

The implementation of the algorithm that samples the NORTA parameters in (B.1) from the joint posterior density function in (B.2) requires the identification of the unconditional correlations $\rho(2, i)$, $i = 3, 4, 5$ and $\rho(3, i)$, $i = 4, 5$ to characterize the conditional distributions of the $\mathcal{C}-$vine specification. Fortunately, there exist recursive formulas that allow us to easily obtain the correlations that are not included in the $\mathcal{C}-$vine specification from the (partial) correlations of the $\mathcal{C}-$vine specification. For example, the formula

$$\rho(2, 3; 1) = \frac{\rho(2, 3) - \rho(1, 2)\rho(1, 3)}{\sqrt{1 - \rho^2(1, 2)}\sqrt{1 - \rho^2(1, 3)}}$$

allows us to obtain $\rho(2, 3)$ from $\rho(1, 2)$, $\rho(1, 3)$, and $\rho(2, 3; 1)$ of the $\mathcal{C}-$vine specification so as to characterize the cdfs of the conditional random variables $\Phi^{-1}[F_i(X_i; \mathbf{\Psi}_i)]|\Phi^{-1}[F_1(X_1; \mathbf{\Psi}_1)]$, $\Phi^{-1}[F_2(X_2; \mathbf{\Psi}_2)]$, $i = 3, 4, 5$ in step (a) of the NORTA parameter sampling algorithm (see Appendix B.4).

More generally, the partial correlation $\rho(1, 2; 3, \ldots, k)$ of the $k-$dimensional NORTA distribution can be interpreted as the correlation between the orthogonal projections of

random variables $\Phi^{-1}[F_1(X_1;\boldsymbol{\Psi}_1)]$ and $\Phi^{-1}[F_2(X_2;\boldsymbol{\Psi}_2)]$ on the plane orthogonal to the space spanned by $\Phi^{-1}[F_i(X_i;\boldsymbol{\Psi}_i)]$, $i = 3, 4, \ldots, k$. The recursive formula that relates this partial correlation to the correlations is given by

$$\rho(1,2;3,4,\ldots,k) = \frac{\rho(1,2;3,4,\ldots,k-1) - \rho(1,k;3,4,\ldots,k-1)\rho(2,k;3,4,\ldots,k-1)}{\sqrt{1-\rho^2(1,k;3,4,\ldots,k-1)}\sqrt{1-\rho^2(2,k;3,4,\ldots,k-1)}}$$

(Yule and Kendall 1965).

B.3. Theorem 3.3.4 of Tong (1990)

Using the correlations obtained from the recursive formulas of Yule and Kendall (1965) together with Theorem 3.3.4 of Tong (1990), we characterize the cdfs of the conditional random variables that appear in the joint posterior density function of the NORTA parameters. We present Theorem 3.3.4 of Tong (1990) below:

Theorem 3.3.4 (Tong 1990) Let $\mathbf{Z}$ be a k-dimensional random vector with mean $\boldsymbol{\mu}$ and correlation matrix $\boldsymbol{\Sigma}$. For fixed $\ell < k$, consider the partitions of $\mathbf{Z}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ given by

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{Z}_1 &= (Z_1, Z_2, \ldots, Z_\ell)', \quad \mathbf{Z}_2 = (Z_{\ell+1}, Z_{\ell+2}, \ldots, Z_k)', \\ \boldsymbol{\mu}_1 &= (\mu_1, \mu_2, \ldots, \mu_\ell)', \quad \boldsymbol{\mu}_2 = (\mu_{\ell+1}, \mu_{\ell+2}, \ldots, \mu_k)', \end{aligned}$$

$\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are the correlation matrices of $\mathbf{Z}_1$ and $\mathbf{Z}_2$, and $\boldsymbol{\Sigma}_{12}$ is the matrix of correlations between $\mathbf{Z}_1$ and $\mathbf{Z}_2$.

If $\mathbf{Z}$ is a $k-$dimensional normal random vector with mean $\boldsymbol{\mu}$ and correlation matrix $\boldsymbol{\Sigma}$, i.e., $\mathbf{Z} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$, then for any fixed $\ell < k$ the conditional distribution of $\mathbf{Z}_1$, given $\mathbf{Z}_2 = \mathbf{z}_2$, is $\mathcal{N}_\ell(\boldsymbol{\mu}_{1\cdot2}, \boldsymbol{\Sigma}_{11\cdot2})$ where $\boldsymbol{\mu}_{1\cdot2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{z}_2 - \boldsymbol{\mu}_2)$ and $\boldsymbol{\Sigma}_{11\cdot2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

B.4. Sampling the parameters of the $5-$dimensional NORTA distribution

We are now ready to present a detailed description of the algorithm that samples parameters β_i, $i = 1, 2, 3, 4, 5$; $\rho(1, i)$, $i = 2, 3, 4, 5$; and $\rho(j-1, i; 1, 2, \ldots, j-2)$, $i = j, j+1, \ldots, 5$,

$j = 3, 4, 5$ to capture parameter uncertainty in a stochastic simulation with 5 correlated inputs having the NORTA distribution described in Appendix B.1:

Sample the component parameters

The goal is to sample parameters β_i, $i = 1, 2, 3, 4, 5$.

for $i = 1, 2, 3, 4, 5$ **do**

- Sample parameter β_i from the inverse gamma posterior density function with shape parameter $n + \theta_i$ and scale parameter $\nu_i + \sum_{t=1}^{n} x_{i,t}$:

$$\tilde{\beta}_i^{-1} \;=\; \textbf{GammaVariate}\left(n + \theta_i, \left(\nu_i + \sum_{t=1}^{n} x_{i,t}\right)^{-1}\right)$$

end loop

Sample the correlations associated with the first tree of the $\mathcal{C}-$vine

The goal is to sample correlations $\rho(1, i)$, $i = 2, 3, 4, 5$.

for $i = 2, 3, 4, 5$ **do**

- Define the two-dimensional matrix $\mathbf{S}_2(1, i|\tilde{\beta}_1, \tilde{\beta}_i, \mathbf{x}_1, \mathbf{x}_i)$ as

$$\mathbf{S}_2(1, i|\tilde{\beta}_1, \tilde{\beta}_i, \mathbf{x}_1, \mathbf{x}_i) \;\equiv\; \sum_{t=1}^{n} \begin{pmatrix} \Phi^{-1}\left[1 - \exp\left\{-\tilde{\beta}_1^{-1} x_{1,t}\right\}\right] \\ \Phi^{-1}\left[1 - \exp\left\{-\tilde{\beta}_i^{-1} x_{i,t}\right\}\right] \end{pmatrix} \begin{pmatrix} \Phi^{-1}\left[1 - \exp\left\{-\tilde{\beta}_1^{-1} x_{1,t}\right\}\right] \\ \Phi^{-1}\left[1 - \exp\left\{-\tilde{\beta}_i^{-1} x_{i,t}\right\}\right] \end{pmatrix}',$$

  where $\mathbf{x}_1 = (x_{1,1}, x_{1,2}, \ldots, x_{1,n})'$ and $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,n})'$.

- Sample the two-dimensional correlation matrix $\mathbf{\Sigma}_2(1, i)$ from the inverted Wishart density function with parameters n and $\mathbf{S}_2(1, i|\tilde{\beta}_1, \tilde{\beta}_i, \mathbf{x}_1, \mathbf{x}_i)$:

$$\tilde{\mathbf{\Sigma}}_2(1, i) \;=\; \textbf{InverseWishartVariate}\Big(n, \mathbf{S}_2(1, i|\tilde{\beta}_1, \tilde{\beta}_i, \mathbf{x}_1, \mathbf{x}_i)\Big)$$

- Set $\tilde{\rho}(1, i)$ to the $(1, 2)^{\text{th}}$ entry of $\tilde{\mathbf{\Sigma}}_2(1, i)$.

end loop

Sample the partial correlations associated with the second tree of the $\mathcal{C}-$vine

The goal is to sample partial correlations $\rho(2, 3; 1)$, $\rho(2, 4; 1)$, and $\rho(2, 5; 1)$.

Construct vector $\tilde{\mathbf{\Lambda}}_3$ using $\tilde{\beta}_m$, $m = 1, 2, 3, 4, 5$ and $\tilde{\rho}(1, m)$, $m = 2, 3, 4, 5$.

for $i = 3, 4, 5$ **do**

- Determine conditional means $\mu_{i|1}$ and $\mu_{2|1}$ and conditional variances $\sigma^2_{i|1}$ and $\sigma^2_{2|1}$ of cdfs $\Phi_{i|1}$ and $\Phi_{2|1}$ using Theorem 3.3.4 of Tong (1990):

$$\begin{aligned}\mu_{i|1} &= \rho(1,i)\,\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_1^{-1}x_{1,t}\right\}\right] \qquad \sigma^2_{i|1}=1-\rho^2(1,i)\\ \mu_{2|1} &= \rho(1,2)\,\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_1^{-1}x_{1,t}\right\}\right] \qquad \sigma^2_{2|1}=1-\rho^2(1,2)\end{aligned}$$

- Use these conditional distribution functions for defining matrix $\mathbf{S}_2(2,i;1|\tilde{\mathbf{\Lambda}}_3,\mathbf{x})$ as

$$\sum_{t=1}^{n}\begin{pmatrix}\frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_2^{-1}x_{2,t}\right\}\right]-\mu_{2|1}}{\sigma_{2|1}}\\ \frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_i^{-1}x_{i,t}\right\}\right]-\mu_{i|1}}{\sigma_{i|1}}\end{pmatrix}\begin{pmatrix}\frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_2^{-1}x_{2,t}\right\}\right]-\mu_{2|1}}{\sigma_{2|1}}\\ \frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_i^{-1}x_{i,t}\right\}\right]-\mu_{i|1}}{\sigma_{i|1}}\end{pmatrix}'.$$

- Sample the partial correlation matrix $\mathbf{\Sigma}_2(2,i;1)$ from the inverted Wishart density function with parameters n and $\mathbf{S}_2(2,i;1|\tilde{\mathbf{\Lambda}}_3,\mathbf{x})$:

$$\tilde{\mathbf{\Sigma}}_2(2,i;1) = \textbf{InverseWishartVariate}(n,\mathbf{S}_2(2,i;1|\tilde{\mathbf{\Lambda}}_3,\mathbf{x}))$$

- Set $\tilde{\rho}(2,i;1)$ to the $(1,2)^{\text{th}}$ entry of $\tilde{\mathbf{\Sigma}}_2(2,i;1)$.

end loop

Sample the partial correlations associated with the third tree of the $\mathcal{C}-$vine

The goal is to sample partial correlations $\rho(3,4;1,2)$ and $\rho(3,5;1,2)$.

Construct vector $\tilde{\mathbf{\Lambda}}_4$ using $\tilde{\beta}_m$, $m=1,2,3,4,5$; $\tilde{\rho}(1,m)$, $m=2,3,4,5$; and $\tilde{\rho}(2,m;1)$, $m=3,4,5$.

for $i=4,5$ **do**

- Obtain correlations $\rho(2,3)$ and $\rho(2,i)$ using the recursive formulas of Yule and Kendall (1965):

$$\begin{aligned}\rho(2,3) &= \tilde{\rho}(2,3;1)\sqrt{1-\tilde{\rho}^2(1,2)}\sqrt{1-\tilde{\rho}^2(1,3)}+\tilde{\rho}(1,2)\tilde{\rho}(1,3)\\ \rho(2,i) &= \tilde{\rho}(2,i;1)\sqrt{1-\tilde{\rho}^2(1,2)}\sqrt{1-\tilde{\rho}^2(1,i)}+\tilde{\rho}(1,2)\tilde{\rho}(1,i)\end{aligned}$$

- Determine conditional means $\mu_{3|1,2}$ and $\mu_{i|1,2}$ and conditional variances $\sigma^2_{3|1,2}$ and $\sigma^2_{i|1,2}$ of cdfs $\Phi_{3|1,2}$ and $\Phi_{i|1,2}$ using Theorem 3.3.4 of Tong(1990):

$$\mu_{3|1,2} = \frac{[\tilde{\rho}(1,3)-\tilde{\rho}(1,2)\rho(2,3)]\,\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_1^{-1}x_{1,t}\right\}\right]}{1-\tilde{\rho}^2(1,2)}$$

$$
\begin{aligned}
&+ \quad \frac{[\rho(2,3)-\tilde{\rho}(1,2)\tilde{\rho}(1,3)]\,\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_2^{-1}x_{2,t}\right\}\right]}{1-\tilde{\rho}^2(1,2)} \\
\sigma^2_{3|1,2} &= 1-\frac{\tilde{\rho}^2(1,3)-2\tilde{\rho}(1,2)\tilde{\rho}(1,3)\rho(2,3)+\rho^2(2,3)}{1-\tilde{\rho}^2(1,2)} \\
\mu_{i|1,2} &= \frac{[\tilde{\rho}(1,i)-\tilde{\rho}(1,2)\rho(2,i)]\,\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_1^{-1}x_{1,t}\right\}\right]}{1-\tilde{\rho}^2(1,2)} \\
&+ \quad \frac{[\rho(2,i)-\tilde{\rho}(1,2)\tilde{\rho}(1,i)]\,\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_2^{-1}x_{2,t}\right\}\right]}{1-\tilde{\rho}^2(1,2)} \\
\sigma^2_{i|1,2} &= 1-\frac{\tilde{\rho}^2(1,i)-2\tilde{\rho}(1,2)\tilde{\rho}(1,i)\rho(2,i)+\rho^2(2,i)}{1-\tilde{\rho}^2(1,2)}
\end{aligned}
$$

- Use these conditional distribution functions for defining matrix $\mathbf{S}_2(3,i;1,2|\mathbf{\Lambda}_4,\mathbf{x})$ as

$$
\sum_{t=1}^{n}\begin{pmatrix} \frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_3^{-1}x_{3,t}\right\}\right]-\mu_{3|1,2}}{\sigma_{3|1,2}} \\ \frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_i^{-1}x_{i,t}\right\}\right]-\mu_{i|1,2}}{\sigma_{i|1,2}} \end{pmatrix}\begin{pmatrix} \frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_3^{-1}x_{3,t}\right\}\right]-\mu_{3|1,2}}{\sigma_{3|1,2}} \\ \frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_i^{-1}x_{i,t}\right\}\right]-\mu_{i|1,2}}{\sigma_{i|1,2}} \end{pmatrix}'.
$$

- Sample the partial correlation matrix $\mathbf{\Sigma}_2(3,i;1,2)$ from the inverted Wishart density function with parameters n and $\mathbf{S}_2(3,i;1,2|\mathbf{\Lambda}_4,\mathbf{x})$:

$$
\tilde{\mathbf{\Sigma}}_2(3,i;1,2) \;=\; \textbf{InverseWishartVariate}\left(n,\mathbf{S}_2(3,i;1,2|\mathbf{\Lambda}_4,\mathbf{x})\right)
$$

- Set $\tilde{\rho}(3,i;1,2)$ to the $(1,2)^{\text{th}}$ entry of $\tilde{\mathbf{\Sigma}}_2(3,i;1,2)$.

end loop

Sample the partial correlation associated with the last (i.e., fourth) tree of the $\mathcal{C}-$vine

The goal is to sample partial correlation $\rho(4,5;1,2,3)$.

- Construct vector $\tilde{\mathbf{\Lambda}}_5$ using $\tilde{\beta}_m$, $m=1,2,3,4,5$; $\tilde{\rho}(1,m)$, $m=2,3,4,5$; and $\tilde{\rho}(\ell-1,m;1,2,\ldots,\ell-2)$, $m=\ell,\ell+1,\ldots,5$, $\ell=3,4$.

- Obtain correlations $\rho(3,4)$ and $\rho(3,5)$ using the formulas of Yule and Kendall (1965) via

$$
\begin{aligned}
\rho(3,4) &= \rho(3,4;1)\sqrt{1-\tilde{\rho}^2(1,3)}\sqrt{1-\tilde{\rho}^2(1,4)}+\tilde{\rho}(1,3)\tilde{\rho}(1,4), \\
\rho(3,5) &= \rho(3,5;1)\sqrt{1-\tilde{\rho}^2(1,3)}\sqrt{1-\tilde{\rho}^2(1,5)}+\tilde{\rho}(1,3)\tilde{\rho}(1,5),
\end{aligned}
$$

where

$$\begin{aligned}\rho(3,4;1) &= \tilde{\rho}(3,4;1,2)\sqrt{1-\tilde{\rho}^2(2,3;1)}\sqrt{1-\tilde{\rho}^2(2,4;1)}+\tilde{\rho}(2,3;1)\tilde{\rho}(2,4;1),\\ \rho(3,5;1) &= \tilde{\rho}(3,5;1,2)\sqrt{1-\tilde{\rho}^2(2,3;1)}\sqrt{1-\tilde{\rho}^2(2,5;1)}+\tilde{\rho}(2,3;1)\tilde{\rho}(2,5;1).\end{aligned}$$

- Determine conditional means $\mu_{4|1,2,3}$ and $\mu_{5|1,2,3}$ and conditional variances $\sigma^2_{4|1,2,3}$ and $\sigma^2_{5|1,2,3}$ of cdfs $\Phi_{4|1,2,3}$ and $\Phi_{5|1,2,3}$ using Theorem 3.3.4 of Tong(1990):

$$\begin{aligned}\mu_{4|1,2,3} &= \begin{pmatrix}\tilde{\rho}(1,4)\\ \rho(2,4)\\ \rho(3,4)\end{pmatrix}' \begin{pmatrix}1 & \tilde{\rho}(1,2) & \tilde{\rho}(1,3)\\ \tilde{\rho}(1,2) & 1 & \rho(2,3)\\ \tilde{\rho}(1,3) & \rho(2,3) & 1\end{pmatrix}^{-1} \begin{pmatrix}\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_1^{-1}x_{1,t}\right\}\right]\\ \Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_2^{-1}x_{2,t}\right\}\right]\\ \Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_3^{-1}x_{3,t}\right\}\right]\end{pmatrix}\\ \mu_{5|1,2,3} &= \begin{pmatrix}\tilde{\rho}(1,5)\\ \rho(2,5)\\ \rho(3,5)\end{pmatrix}' \begin{pmatrix}1 & \tilde{\rho}(1,2) & \tilde{\rho}(1,3)\\ \tilde{\rho}(1,2) & 1 & \rho(2,3)\\ \tilde{\rho}(1,3) & \rho(2,3) & 1\end{pmatrix}^{-1} \begin{pmatrix}\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_1^{-1}x_{1,t}\right\}\right]\\ \Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_2^{-1}x_{2,t}\right\}\right]\\ \Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_3^{-1}x_{3,t}\right\}\right]\end{pmatrix}\\ \sigma^2_{4|1,2,3} &= 1-\begin{pmatrix}\tilde{\rho}(1,4)\\ \rho(2,4)\\ \rho(3,4)\end{pmatrix}' \begin{pmatrix}1 & \tilde{\rho}(1,2) & \tilde{\rho}(1,3)\\ \tilde{\rho}(1,2) & 1 & \rho(2,3)\\ \tilde{\rho}(1,3) & \rho(2,3) & 1\end{pmatrix}^{-1} \begin{pmatrix}\tilde{\rho}(1,4)\\ \rho(2,4)\\ \rho(3,4)\end{pmatrix}\\ \sigma^2_{5|1,2,3} &= 1-\begin{pmatrix}\tilde{\rho}(1,5)\\ \rho(2,5)\\ \rho(3,5)\end{pmatrix}' \begin{pmatrix}1 & \tilde{\rho}(1,2) & \tilde{\rho}(1,3)\\ \tilde{\rho}(1,2) & 1 & \rho(2,3)\\ \tilde{\rho}(1,3) & \rho(2,3) & 1\end{pmatrix}^{-1} \begin{pmatrix}\tilde{\rho}(1,5)\\ \rho(2,5)\\ \rho(3,5)\end{pmatrix}\end{aligned}$$

- Use these conditional distribution functions for defining matrix $\mathbf{S}_2(4,5;1,2,3|\tilde{\mathbf{\Lambda}}_5,\mathbf{x})$ as

$$\sum_{t=1}^{n}\begin{pmatrix}\frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_4^{-1}x_{4,t}\right\}\right]-\mu_{4|1,2,3}}{\sigma_{4|1,2,3}}\\ \frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_5^{-1}x_{5,t}\right\}\right]-\mu_{5|1,2,3}}{\sigma_{5|1,2,3}}\end{pmatrix}\begin{pmatrix}\frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_4^{-1}x_{4,t}\right\}\right]-\mu_{4|1,2,3}}{\sigma_{4|1,2,3}}\\ \frac{\Phi^{-1}\left[1-\exp\left\{-\tilde{\beta}_5^{-1}x_{5,t}\right\}\right]-\mu_{5|1,2,3}}{\sigma_{5|1,2,3}}\end{pmatrix}'.$$

- Sample the partial correlation matrix $\mathbf{\Sigma}_2(4,5;1,2,3)$ from the inverted Wishart density function with parameters n and $\mathbf{S}_2(4,5;1,2,3|\tilde{\mathbf{\Lambda}}_5,\mathbf{x})$:

$$\tilde{\mathbf{\Sigma}}_2(4,5;1,2,3) = \mathbf{InverseWishartVariate}(n,\mathbf{S}_2(4,5;1,2,3|\tilde{\mathbf{\Lambda}}_5,\mathbf{x}))$$

- Set $\tilde{\rho}(4,5;1,2,3)$ to the $(1,2)^{\text{th}}$ entry of $\tilde{\mathbf{\Sigma}}_2(4,5;1,2,3)$.

This completes the sampling of the parameters of the 5−dimensional NORTA distribution using our Bayesian model. Section 1.4 of our paper describes how to incorporate this sampling algorithm into the simulation replication algorithm of Chick (2001) for the estimation of the mean performance measure and the confidence interval.

Appendix C

Appendix A.2 presents the Gibbs sampler algorithm that samples the shape and scale parameters of the gamma distributed component. The objective of this section is to discuss the issues that arise in the implementation of this algorithm. Specifically, Section C.1 discusses the selection of the sampling plan, while Section C.2 elaborates on the choice of the warm-up period. Finally, Section C.3 describes the identification of the run length that leads to the convergence of the Markov chain.

C.1. Choosing the sampling plan

There are two major approaches to the use of Markov chains for sampling from posterior density functions. The first approach generates n independent realizations from the posterior density function using n separate runs, each of length m, and retains the final state of the chain. The second approach, on the other hand, uses a single long run. This is also the approach that has been favored by simulation analysts due to the computational burden associated with the first approach (Law and Kelton 1984, Bratley et al. 1987). Another shortcoming of the first approach is the inefficient use of the data. More specifically, only the last value sampled from each chain is used, while the second approach discards the values collected during the warm-up period (Smith and Roberts 1983, Tierney 1994). Therefore, we follow the second approach and sample the values of a parameter from its posterior density function using a single long chain. Two immediate implementation questions that arise are how to determine the warm-up period (i.e., a value for $\mathcal{L}^*$ in Figure A.1 of Appendix A.2) and how to identify the run length of the Markov chain (i.e., a value for $\mathcal{L}$ in Figure A.1 of Appendix A.2). We answer the first question in Section C.2 and provide a solution to the second question in Section C.3.

C.2. Determining the warm-up period

The determination of the warm-up period concerns the question of how many of the initially sampled values should be discarded. The common practice is to answer this question by a visual inspection of the plots of the values sampled from a single chain (Gilks et al. 1996). Another method of determining the warm-up period is the computation of the auto-covariances among the sampled values. The central idea is that it is not necessary to discard more values than those required for the auto-covariances to decay to a negligible level. Motivated by this idea, Geyer (1992) suggests that it is generally sufficient to discard the initial 1% or 2% of the sampled values. Another practical way of determining the warm-up period is the use of the convergence diagnostics (Cowles and Carlin 1996). We recommend the use of Heidelberger and Welch (1983) convergence diagnostic and Geweke (1992) convergence diagnostic for a sampling plan based on a single long chain. Both of these diagnostics are built in the public domain software Bayesian Output Analysis available in S-PLUS and R.

C.3. Determining the chain length

Another practical issue that arises in the implementation of the Gibbs sampler algorithm is the determination of the chain length. The goal is to run the chain long enough to obtain a precise estimate of the distribution parameter by taking the average of the values sampled from the posterior density function. An informal method suggested by Gilks et al. (1996) for determining the run length is to run several chains in parallel, with different starting points, and increase the run length as long as the estimates obtained from the chains are not sufficiently close to each other. Convergence diagnostics such as the Heidelberger and Welch (1983) convergence diagnostic that explicitly reports whether the chain is long enough are also used in practice for determining a value for the chain length. However, the computation of the variances of the parameter estimates is complicated by the correlation between the successive values of the parameters obtained from the Gibbs sampler algorithm. An informal approach is to subsample the chain to obtain (nearly) independent samples. Despite the wide use of this approach, it has been shown that systematic subsampling often increases the variance of the sample mean estimators (MacEachern and Berliner 1994). Therefore, it is a better approach to use all of the sampled values together with the utilization of the literature on correctly calculating the variance of an estimator in the presence of correlated

data (Carlin and Louis 2000).

A well-known method that can be used for obtaining the variance of an estimator when the sampled parameter values are correlated is the batching method. Specifically, the (classical) batching method divides a (single) run of length $\mathcal{L} - \mathcal{L}^*$ into m batches of size s (i.e., $\mathcal{L} - \mathcal{L}^* = ms$). If the batch means are represented by $B_1, B_2, \ldots, B_m$, then the overall mean is given by $\bar{B}_m = \sum_{i=1}^{m} B_i/m$, while its variance is obtained from $\sum_{i=1}^{m}(B_i - \bar{B}_m)^2/(m(m-1))$ for s large enough so that the correlation between the batch means is negligible. Thus, it is important to check whether the correlation between the batch means is statistically zero. There exist a variety of statistical tests that can be used for this purpose; e.g., the Portmanteau test (Wei 1990). If the correlation between the batch means is found to be statistically significant, then the batch size s and therefore, the chain length $\mathcal{L}$ are increased and the procedure is repeated. An alternative method that can be implemented for the proper choice of m and s is the ABATCH algorithm of Fishman and Yarberry (1997). Also, the discrete-event stochastic simulation literature has proposed a variety of batching methods that are alternative to the classical batching method. We refer the reader to Song and Chih (2008) for a comprehensive review of these batching methods.

We conclude this section with noting that the use of the Gibbs sampler algorithm might also lead to cross-correlations between the chains of different parameters. For example, the sampled parameter values $\{\beta^\ell, \ell = 1, 2, \ldots, \mathcal{L}\}$ and $\{\alpha^\ell, \ell = 1, 2, \ldots, \mathcal{L}\}$ of Figure A.1 might be correlated. It is easy to check the cross-correlations using the descriptive statistics menu of Bayesian Output Analysis, which simply creates a correlation matrix. High cross-correlations lead to slow convergence of the Markov chain to the joint posterior density function. In case of high correlations between the parameter samples, one can try to reparameterize the model and run the Gibbs sampler algorithm again. Also, the Bayesian literature has proposed a number of methods for breaking the correlations, including auxiliary variable technique (Besag and Green 1993, Higdon 1998), hierarchical centering (Gelfand et al. 1995, Roberts and Sahu 1997, Papaspiliopoulos et al. 2003, and Papaspiliopoulos et al. 2007), orthogonal parameterization (Hills and Smith 1992), and parameter expansion (Liu et al. 1998, Liu and Wu 1999). We refer the reader to these references and the references therein for examples of reparameterization techniques. However, the simulation analyst should be cautious in the implementation of these methods as many of them are problem specific, and therefore the success of the method depends on the problem being considered.

Appendix D

In this section, we assume the availability of the (dependent) output data y_r, $r = 1, 2, \ldots, R$ obtained from the simulation replication algorithm and describe how to use the method of batching for computing a point estimate and a confidence interval of $\mathrm{E}_{Y|\mathbf{x}}(Y|\mathbf{x})$.

First, we perform R replications of the simulation replication algorithm and delete the first ℓ observations to eliminate the initial-condition bias, which might arise from the dependence of the output on the initial NORTA parameter values sampled in the first step of the simulation replication algorithm. Mahajan and Ingalls (2004) present a variety of methods that can be used for choosing a value for ℓ. We implement the crossing of the means rule (Fishman 1973) using the algorithm developed by Gafarian et al. (1978). Then, we divide the remaining $R - \ell$ output data into m batches of size s (i.e., $R - \ell = ms$) and represent the batch means by $B_1, B_2, \ldots, B_m$. For a value of s that is large enough for the correlation between the batch means to be negligible, we generate a point estimate of $\mathrm{E}_{Y|\mathbf{x}}(Y|\mathbf{x})$ via $\hat{y} = \sum_{i=1}^{m} B_i/m$, and construct a $100(1-\alpha)\%$ confidence interval using $\hat{y} \pm z_{1-\alpha/2} \sum_{i=1}^{m} (B_i - \hat{y})^2/(m(m-1))$, where $z_{1-\alpha/2}$ is the upper $1-\alpha/2$ critical point for a standard normal random variable. Carlin and Louis (2000) recommend the replacement of $z_{1-\alpha/2}$ with $t_{m-1,1-\alpha/2}$, which is the upper $1-\alpha/2$ critical point for the t distribution with $m-1$ degrees of freedom, for a batch size less than 30.

Although the classical batching method is easy to implement, the improper choice of m and s may lead to inconsistent variance estimates. We obtain consistent confidence intervals using the ABATCH algorithm of Fishman and Yarberry (1997) and refer the reader to Alexopoulos (2006) for a survey of the methods for the consistent estimation of the batch means.

Appendix E

Survey of GPCFB's Partner Agencies

Agency Name:

Location:

Survey for pantries

.....

1. How many days in a week is your agency open?
   ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7

2. How many hours does your agency operate in a day?
   ☐2 − 3 ☐3 − 4 ☐4 − 5 ☐5 − 6 ☐6 − 7 ☐7 − 8 ☐ Other (Please specify)

3. Approximately how many people does your agency serve per week?

4. How many people do you turn away per week?
   ☐0 ☐1 − 5 ☐5 − 10 ☐10 − 15 ☐15 − 20 ☐ Other (Please specify).

5. What is primary reason you turn participants away? (If you have more than one reason, please number it with (1) being the most primary reason).
   ☐ Agency serves a limited number of pantry bags.
   ☐ Participants came more often than program rules allowed.
   ☐ Services needed were not provided by agency.
   ☐ Other (Please specify).

6. How frequently do needy people come to the agency? In other words, how many days will the provided food be enough for the needy people?
   ☐ Once a week ☐ Once a month

☐ Twice a week ☐ Twice a month

☐ Other (Please specify).

7. What do you suggest needy people to do when you turn them away?

   ☐ Nothing

   ☐ Go to another agency.

   ☐ Other (Please specify).

8. If your answer to the previous question is "Go to another agency" then of those people going to the second agency what percentage of them finds service in the second agency?

   ☐ % 0 ☐ %1-25 ☐ % 25-50 ☐ %50-75 ☐ %75-99 ☐ %100.

9. Among the participants what percentages are children, adults and elder?

   Children: Adults : Elder:

10. Among the participants what percentages are women and men?

    Women: Men:

11. Which day of the week is the busiest for the agency in terms of people coming to the agency?

12. If you can not distribute a crucial item such as bread, what do you do?

13. What is the main source of supply for your agency? (If you have more than one source of supply please order them with (1) being the most primary source).

    ☐ Greater Pittsburgh Community Food Bank.

    ☐ Individual donations.

    ☐ Corporate donations.

    ☐ Purchasing, through donated money.

    ☐ Other (Please specify).

14. What kind of goods (food & non-food) does your agency serve to needy people? Please specify the food type.

15. What are the 5 items that your agency wishes to distribute most? Please rank them with (1) being the top.

16. For the specified 5 items, approximately how short is your agency (in terms of lbs/ week)?

17. Do you realize any change in your demand in (please specify the amount of change in percentages):
    Summers: ☐ No change ☐ Increase Decrease
    Holidays: ☐ No change ☐ Increase Decrease
    Events (like superbowl): ☐ No change ☐ Increase Decrease

18. Do you think participants are effected by TV /Branding? In other words, do you have participants asking for a particular branded product? If yes, how frequent does that happen?

19. Do you see any increase in demand in recent years?

Survey for onsite programs

.....

1. How many days in a week is your agency open?
   ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7

2. How many hours does your agency operate in a day?
   ☐$2-3$ ☐$3-4$ ☐$4-5$ ☐$5-6$ ☐$6-7$ ☐$7-8$ ☐ Other (Please specify).

3. Approximately how many people does your agency serve per week?

4. How many people do you turn away per week?
   ☐0 ☐$1-5$ ☐$5-10$ ☐$10-15$ ☐$15-20$ ☐ Other (Please specify).

5. What is primary reason you turn participants away? (If you have more than one reason, please number it with (1) being the most primary reason).
   ☐ Agency serves a limited number of pantry bags.
   ☐ Participants came more often than program rules allowed.
   ☐ Services needed were not provided by agency.
   ☐ Other (Please specify).

6. What do you suggest needy people to do when you turn them away?
   ☐ Nothing
   ☐ Go to another agency.
   ☐ Other (Please specify).

7. If your answer to the previous question is "Go to another agency" then of those people going to the second agency what percentage of them finds service in the second agency?
   ☐ % 0 ☐ %1-25 ☐ % 25-50 ☐ %50-75 ☐ %75-99 ☐ %100.

8. Among the participants what percentages are children, adults and elder?
   Children: Adults : Elder:

9. Among the participants what percentages are women and men?
   Women: Men:

10. Which day of the week is the busiest for the agency in terms of people coming to the agency?

11. What is the main source of supply for your agency? (If you have more than one source of supply please order them with (1) being the most primary source).
    ☐ Greater Pittsburgh Community Food Bank.
    ☐ Individual donations.
    ☐ Corporate donations.
    ☐ Purchasing, through donated money.
    ☐ Other (Please specify).

12. Do you realize any change in your demand in (please specify the amount of change in percentages):
    Summers: ☐ No change ☐ Increase Decrease
    Holidays: ☐ No change ☐ Increase Decrease
    Events (like superbowl): ☐ No change ☐ Increase Decrease

13. Do you think participants are effected by TV /Branding? In other words, do you have participants asking for a particular branded product? If yes, how frequent does that happen?

14. Do you see any increase in demand in recent years?

Bibliography

Alexopoulos, C. 2006. A comprehensive review of methods for simulation output analysis. *Proceedings of the 2006 Winter Simulation Conference*, L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds., IEEE, Piscataway, New Jersey, 168 – 178.

Anderson, T. W., D. A. Darling. 1954. A test of goodness of fit. *Journal of American Statistical Association*, *49*, 765 – 769.

Anily, S. and J. Bramel. 1999. Approximation algorithms for the capacitated traveling salesman problem with pickups and deliveries. *Naval Res. Logist.*, *46*, 654 – 670.

Barnard, J., R. McCulloch, and X. Meng. 2000. Modeling covariance matrices in terms of standard deviations and correlations with applications to shrinkage. *Statistica Sinica*, *10*, 1281 – 1311.

Barton, R. R. and L. W. Schruben. 2001. Resampling methods for input modeling. In *Proceedings of the 2001 Winter Simulation Conference*, B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds., IEEE, Piscataway, New Jersey, 372 – 378.

Bedford, T. and R. M. Cooke. 2001. *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press, U.K.

Bedford, T. and R. M. Cooke. 2002. Vines – a new graphical model for dependent random variables. *Annals of Statistics*, *30*, 1031 – 1068.

Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics.

Berger, J. O. and D. Sun. 2008. Objective priors for the bivariate normal model. *The Annals of Statistics*, *36*, 963 – 982.

Bermudez, P. de Z., M. A. A. Turkman. 2003. Bayesian approach to parameter estimation of the generalized Pareto distribution. *Sociedad de Estadistica e Investigacion Opevativa Test* **12** 259–277.

Bernardo, J. M. and A. F. M. Smith. 1994. *Bayesian Theory*. New York: John Wiley and Sons, Inc.

Besag, J. and P. J. Green. 1993. Spatial statistics and Bayesian computation (with discussion). *Journal of the Royal Statistical Society*, *Ser.B*, *16*, 395 – 407.

Biller, B. and B. L. Nelson. 2005. Fitting time-series input processes for simulation. *Operations Research*, *53*, 549 – 559.

Biller, B. and C. Gunes. 2010a. Accounting for parameter uncertainty in large-scale stochastic simulations with correlated inputs. Accepted for publication in *Operations Research*.

Biller, B. and C. Gunes. 2010b. Solving high-dimensional simulation input-modeling problems. Working Paper, Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA.

Biller, B. and S. Ghosh. 2006. Multivariate input processes. In *Handbooks in Operations Research and Management Science: Simulation*, B. L. Nelson and S. G. Henderson, eds., Elsevier Science, Amsterdam.

Billingsley, P. 1961. *Statistical Inference for Markov Processes*. University of Chicago Press, Chicago.

Bowman, K. O. and L. R. Shenton. 1988. Solutions to Johnson's S_B and S_U. *Communications in Statistics: Simulation and Computation*, *17*, 343 – 348.

Bowman, K. O. and L. R. Shenton. 1989. S_B and S_U distributions fitted by percentiles: A general criterion. *Communications in Statistics: Simulation and Computation*, *18*, 1 – 13.

Bratley P., B. L. Fox, and L. E. Schrage. 1987. *A Guide to Simulation*. Second Edition. Springer, New York.

Calitz, F. 1973. Maximum likelihood estimation of the parameters of the three-parameter lognormal distribution – a reconsideration. *Australian Journal of Statistics*, *15*, 185 – 190.

Cario, M. C. and B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Working Paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

Carlin, B. P. and T. A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London, UK.

Chakravant, I. M., R. G. Laha and J. Roy. 1967. *Handbook of Methods of Applied Statistics, Volume I*. John Wiley and Sons, 392 – 394.

Chalasani, P. and R. Motwani. 1999. Approximating capacitated routing and delivery

problems. *SIAM J Comput*, *28*, 2133 – 2149.

Cheng, R. C. H. 1977. The generation of gamma variables with non-integral shape parameter. *Applied Statistics*, *26*, 71 – 75.

Cheng, R. C. H. and N. A. K. Amin. 1983. Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society: Series B*, *45*, 394 – 403.

Cheng, R. C. H. and T. C. Iles. 1987. Corrected maximum likelihood in nonregular problems. *Journal of the Royal Statistical Society: Series B*, *49*, 95 – 101.

Cheng, R. C. H. and W. Holland. 1997. Sensitivity of computer simulation experiments to errors in input data. *Journal of Statistical Computation and Simulation*, *57*, 327 – 335.

Cheng, R. C. H. and W. Holland. 1998. Two-point methods for assessing variability in simulation output. *Journal of Statistical Computation and Simulation*, *60*, 183 – 205.

Cheng, R. C. H. and W. Holland. 2004. Calculation of confidence intervals for simulation output. *ACM Transactions on Modeling and Computer Simulation*, *14*, 344 – 362.

Chick, S. E. 1997. Bayesian analysis for simulation input and output. In *Proceedings of the 1997 Winter Simulation Conference*, S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson, eds., IEEE, Piscataway, New Jersey, 253 – 260.

Chick, S. E. 1999. Steps to implement Bayesian input distribution selection. In *Proceedings of the 1999 Winter Simulation Conference*, P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds., IEEE, Piscataway, New Jersey, 317 – 324.

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

Chou, Y. M., A. M. Polansky, and R. L. Mason. 1999. An algorithm for fitting Johnson transformations to non-normal data. *Journal of Quality Technology*, *31*, 354 – 350.

Cohen, A. C. 1951. Estimating parameters of logarithmic-normal distributions by maximum likelihood. *Journal of the American Statistical Association*, *46*, 206 – 212.

Cooke, R. M. 1997. Markov and entropy properties of tree- and vine-dependent variables. In *Proceedings of the ASA Section on Bayesian Statistical Science*, American Statistical Association, Alexandria, VA, 166 – 175.

Cowles, M. K. and B. P. Carlin. 1996. Markov chain Monte Carlo convergence diagnostics:

a comprehensive review. *Journal of the American Statistical Association*, *91*, 883 – 904.

Daniels, M. J. 1999. A prior for the variance in hierarchical models. *Canadian Journal of Statistics*, *27*, 567 – 578.

Desmond, A. F. and Z. Yang. 1998. A comparison of likelihood and Bayesian inference for the threshold parameter in the inverse Gaussian distribution. *Commonications in Statististics-Theory and Methods*, *27*, 2173 – 2183.

Draper, D. 1995. Assessment of propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society*, *Ser B*, *56*, 501 – 514.

Dror M., D. Fortin, and C. Roucairol. 1998. Redistribution of self-service electric cars: A case of pickup and delivery. Tech. Rep. W.P.3543, INRIA-Rocquencourt, Rocquencourt, France.

Fishman, G. S. 1973. *Concepts and Methods in Discrete Event Digital Simulation.* John Wiley, New York.

Fishman, G. S. and L. S. Yarberry. 1997. An implementation of the batch means method. *INFORMS Journal on Computing*, *9*, 296 – 310.

Flynn, M. R. 2004. The four-parameter lognormal (S_B) model of human exposure. *Ann. Occup. Hyg.*, *48*, 617 – 622.

Gafarian, A. V., C.J. Ancker, and T. Morisaku. 1978. Evaluation of commonly used rules for detecting steady-state in computer simulation. *Naval Research Logistics Quarterly*, *25*, 511 – 529.

Gelfand, A. and A. Smith. 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398 – 409.

Gelfand, A. E., S. K. Sahu, and B. P. Carlin. 1995. Efficient parameterizations for normal linear mixed models. *Biometrika*, *83*, 479 – 488.

Gelman, A. J., B. Carlin, H. S. Stern, D. R. Rubin. 1995. *Bayesian Data Analysis.* Chapman & Hall, London, UK.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2000. *Bayesian Data Analysis.* Chapman & Hall, London, UK.

Geman, S. and A. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

6, 721 – 740.

George, E. I. 1999. Bayesian model selection. *Encyclopedia of Statistical Sciences, Volume 3*, S. Kotz, C. Read, and D. Banks, eds. Wiley New York, 39 – 46.

Geweke, J. 1992. Evaluating the accuracy of the sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*, J. M. Bernardo, J. Berger, A. P. David, and A. F. M. Smith, eds., Oxford, U.K., Oxford University Press, 169 – 193.

Geyer, C. J. 1992. Practical Markov chain Monte Carlo. *Statistical Science*, *7*, 473 – 483.

Gilks, W. R. and P. Wild. 1992. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, *41*, 337 – 348.

Gilks, W. R., N. G. Best, and K. K. C. Tan. 1995. Adaptive rejection metropolis sampling within Gibbs sampling. *Applied Statistics*, *44*, 455 – 472.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, UK.

Green, E. J., F. A. Roesch, A. F. M. Smith, and W. E. Strawderman. 1994. Bayesian estimation for the three-parameter Weibull distribution with tree diameter data. *Biometrics*, *50*, 254 – 269.

Hafley, W. L. and H. T. Schreuder. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. *Canadian Journal of Forest Research*, *7*, 481 – 487.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, *57*, 97 – 109.

Heidelberger, P. and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research*, *31*, 1109 – 1144.

Helton, J. C. 1997. Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *Journal of Statistical Computation and Simulation*, *57*, 3 – 76.

Hernández-Pérez, H. and J. J. Salazar-González. 2003. The one-commodity pickup-and-delivery traveling salesman problem. In: Junger M, Reinelt G, Rinaldi G (eds.) *Combinatorial Optimization- Eureka, You Shrink Springer, LNCS, vol. 2570*, 89 – 104.

Hernández-Pérez, H. and J. J. Salazar-González. 2004a. A branch-and cut algorithm for the traveling salesman problem with pickup and delivery. *Discrete Applied Mathematics*, *145*, 126 – 139.

Hernández-Pérez, H. and J. J. Salazar-González. 2004b. Heuristics for the one-commodity pickup-and- delivery traveling salesman problem. *Transport Science*, *38*, 245 – 255.

Hernández-Pérez, H. and J. J. Salazar-González. 2007. The one-commodity pickup-and-delivery traveling salesman problem: Inequalities and algorithms. *Networks*, *50*, 258 – 272.

Hernández-Pérez, H., I. Rodríguez-Martín, and J. J. Salazar-González. 2009. A hybrid GRASP/VND heuristic for the one-commodity pickup-and-delivery traveling salesman problem. *Computers and Operations Research*, *36*, 1639 – 1645.

Higdon, D. M. 1998. Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association*, *93*, 585 – 595.

Hill, B. M. 1963. The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *Journal of American Statistical Assocation*, *58*, 72 – 84.

Hill, I. D., R. Hill, and R. L. Holder. 1976. Fitting johnson curves by moments. *Applied Statististics*, *25*, 180 – 189.

Hills, S. E. and A. F. M. Smith. 1992. Parameterization issues in Bayesian inference. In *Bayesian Statistics 4*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., Oxford University Press, UK, 227 – 246.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky. 1999. Bayesian model averaging: A tutorial. Technical Report 9814, Department of Statistics, Colorado State University, Fort Collins, CO.

Jeffreys, H. 1961. *Theory of Probability*. Oxford University Press.

Joe, H. 1997. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

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

Johnson, M. E. and V. W. Lowe. 1979. Bounds on the sample skewness and kurtosis. *Technometrics*, *21*, 377 – 378.

Johnson, M. E. 1987. *Multivariate Statistical Simulation*. John Wiley & Sons, New York.

Kass, R. E. and L. Wasserman. 1996. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, *91*, 1343 – 1370.

Kendall, M. G., A. Stuart. 1979. *The Advanced Theory of Statistics*. Macmillan, New York.

Komori, Y., H. Hirose. 2001. Easy estimation by a new parameterization in the three-parameter lognormal distribution. Technical Report, Department of Control Engineering and Science, Kyushu Institute of Technology, Kawazu, Japan.

Kosugi, K. 1994. Three-parameter lognormal distribution model for soil water retention. *Water Resour. Res.*, *30*, 891 – 901.

Kouvelis, P. and P. Su. 2007. *The Structure of Global Supply Chains*. Research Monograph for Book Series on *Foundations and Trends@ in Technology, Information, and Operations Management*. NOW, Boston-Delft.

Kuhl, M. E., J. R. Wilson. 1999. Least-squares estimation of nonhomogeneous Poisson processes. *J. Statist. Comput. Simulation*, *67*, 75 – 108.

Kuhl, M. E., N. M. Steiger, E. K. Lada, M. A. Wagner, and J. R. Wilson. 2009. Introduction to modeling and generating probabilistic input processes for simulation. *Proceedings of the 2009 Winter Simulation Conference*. eds, M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls. 184 – 202. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Kurowicka, D. and R. Cooke. 2003. A parameterization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and Its Applications*, *372*, 225 – 251.

Kurowicka, D. and R. Cooke. 2006. *Uncertainty Analysis with High Dimensional Dependence Modeling*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York.

Lambert, J. A. 1970. Estimation of parameters in the four-parameter lognormal distribution. *Australian Journal of Statistics*, *12*, 33 – 43.

Law, A. M. 2007. *Simulation Modeling and Analysis*. Fourth Edition. New York: McGraw-Hill.

Law, A. M. and W. D. Kelton. 1984. Confidence intervals for steady-state simulations: I: A survey of fixed sample size procedures. *Operations Research*, *32*, 1221 – 1239.

Leonard, T. and J. S. Hsu. 1992. Bayesian inference for a covariance matrix. *Annals of Statistics*, *20*, 1669 – 1696.

Li, B., E. Yashchin, C. Chrisriansen, J. Gill, R. Filippi, and T. Sullivan. 2006. Application of three-parameter lognormal distribution in EM data analysis. *Microelectronics and Reliability*, *46*, 2049 – 2055.

Liechty, J. C., M. W. Liechty, and P. Muller. 2004. Bayesian correlation estimation. *Biometrika*, *91*, 1 – 14.

Liu, C., D. B. Rubin, and Y. N. Wu. 1998. Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika*, *85*, 755 – 770.

Liu, J. S. and Y. N. Wu. 1999. Parameter expansion for data augmentation. *Journal of the American Statistical Association*, *94*, 1264 – 1274.

MacEachern, S. N. and L. M. Berliner. 1994. Subsampling the Gibbs Sampler. *American Statistician*, *48*, 188 – 190.

Mage, D. T. 1980. An explicit solution for S_B parameters using four percentile points. *Technometrics*, *22*, 247 – 251.

Mahajan, P.S. and R.G. Ingalls. 2004. Evaluation of methods used to detect warm-up period in steady-state simulation. In *Proceedings of the 2004 Winter Simulation Conference*, R.G. Ingalls, M.D. Rossetti, J.S. Smith and B.A. Peters, eds., 663 – 671.

Monness, E. N. 1982. Diameter distributions and height curves in even-aged stands of Pinus sylvestris L. *Medd. Norsk Inst. Skogforsk.*, *36*, 1 – 43.

Morales, O., D. Kurowicka, and A. Roelen. 2006. Elicitation procedures for conditional and unconditional rank correlations. In *Resources for the future, Expert Judgement Policy Symposium and Technical Workshop, Washington, DC, March 13-14.*

Nelsen, R. B. 1999. *An Introduction to Copulas*. New York: Springer-Verlag.

Niermann, S. 2006. Evolutionary estimation of parameters of Johnson distributions. *Journal of Statistical Computation and Simulation*, *76*, 185 – 193.

Papaspiliopoulos, O., G. O. Roberts, and M. Skold. 2003. Non-centered parameterizations for hierarchical models and data augmentation. In *Bayesian Statistics 7*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, eds., Oxford University Press, UK, 307 – 326.

Papaspiliopoulos, O., G. O. Roberts, and M. Skold. 2007. A general framework for the parametrization of hierarchical models. *Statistical Science*, *22*, 59 – 73.

Parragh, S. N., K. F. Doerner, and R. F. Hartl. 2008. A survey on pickup and delivery problems. Part I: Transportation between customers and depot. *JFB*, *58*, 21 – 51.

Parresol, B. R. 2003. Recovering parameters of Johnson's S_B distribution. *U.S. Department*

of Agriculture, Forest Service, Southern Research Station, Paper SRS–31.

Pearson, K. 1895. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, *91*, 343 – 358.

Raftery, A. E., D. Madigan, and C. T. Volinsky. 1996. Accounting for model uncertainty in survival analysis improves predictive performance (with discussion). In *Bayesian Statistics 5*, J. M. Bernardo, J. O. Berger, A. P. David, and A. F. M. Smith, eds., Oxford Press, 323 – 349.

Ramberg, J. S. and B. W. Schmeiser. 1974. An approximate method for generating asymmetric random variables. *Communications of the Association for Computing Machinery*, *17*, 78 – 82.

Ranneby, B. 1984. The maximum spacing method: an estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics*, *11*, 93 – 112.

Rao, P. S. R. S. 1997. *Variance Components Estimation: Mixed Models, Methodologies, and Applications*. Chapman and Hall, London, UK.

Roberts, G. O. and S. K. Sahu. 1997. Updating schemes, correlation structure, blocking, and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society*, *Series B*, *59*, 291 – 317.

Rossi, P. E., G. M. Allenby, and R. McCulloch. 2006. *Bayesian Statistics and Marketing*. Wiley Series in Probability and Statistics, New Jersey.

Schmeiser, B. and S. Deutsch. 1977. A versatile four-parameter family of probability distributions suitable for simulation. *IIE Transactions*, *9*, 176 – 182.

Schweizer, B. 1991. Thirty years of copulas. In *Advances in Probability Distributions with Given Marginals*, G. Dall'Aglio, S. Kotz, and G. Salinetti, eds., 13 – 50. Dordrecht, Netherlands: Kluwer.

Seber, G. A. F. 1977. *Linear Regression Analysis*. John Wiley, New York.

Siekierski, K. 1992. Comparison and evaluation of three methods of estimation of the Johnson S_B distribution. *Biometrical Journal*, *34*, 879 – 895.

Singh, V. P. 1987. *Hydrologic Frequency Modeling*. Springer.

Slifker, J. F. and S. S. Shapiro. 1980. The Johnson System: Selection and parameter estimation. *Technometrics*, *22*, 239 – 246.

Smith, A. F. M. and G. O. Roberts. 1983. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B (Methodological)*, *55*, 3 – 23.

Smith, R. L. and J. C. Naylor. 1987. A comparison of maximum likelihood and Bayesian estimators for the three-parameter weibull distribution. *Applied Statistics*, *36*, 358 – 369.

Son, Y. S. and M. Oh. 2006. Bayesian estimation of the two-parameter gamma distribution. *Communications in Statistics-Simulation and Computation*, *35*, 285 – 293.

Song, W. T. and M. Chih. 2008. Implementable MSE-optimal dynamic partial-overlapping batch mean estimators for steady-state simulations. In *Proceedings of the 2008 Winter Simulation Conference*, S. J. Mason, R. R. Hill, L. Monch, O. Rose, T. Jefferson, and J. W. Fowler, eds., 426 – 435.

Stedinger, J. R. 1980. Fitting lognormal distributions to hydrologic data. *Water Resources Research*, *16*, 481 – 490.

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

Tierney, L. 1994. Markov chains for exploring posterior distributions. *The Annals of Statistics*, *22*, 1701 – 1728.

Tong, Y. L. 1990. *The Multivariate Normal Distribution*. Springer-Verlag, New Jersey.

Tsionas, E. G. 2001. Likelihood and posterior shapes in Johnson's S_B system. textit-Sankhya: The Indian Journal of Statistics, Series B, *63*, 3 – 9.

Upadhyay, S. K. and M. Peshwani. 2001. Full posterior analysis of three parameter lognormal distribution using Gibbs sampler. *Journal of Statistical Computation and Stimulation*, *71*, 215 – 230.

Von Gadow, K. 1984. Die Erfassung von Durchmesserverteilung in gleichartigen Kiefernbeständen. *Forstw. Cbl.*, *103*, 360 – 374.

Wei, W. W. S. 1990. *Time Series Analysis, Univariate and Multivariate Methods*. Addison-Wesley Publishing Company, Redwood City.

Yule, G. and M. G. Kendall. 1965. *An Introduction to the Theory of Statistics*. Charles Griffin & Company, Belmont, CA.

Zouaoui, F. and J. R. Wilson. 2003. Accounting for parameter uncertainty in simulation input modeling. *IIE Transactions*, *35*, 781 – 792.

Zouaoui, F. and J. R. Wilson. 2004. Accounting for input-model and input-parameter uncertainties in simulation. *IIE Transactions*, *36*, 1135 – 1151.

Xu, S. H. 1999. Structural analysis of a queueing system with multi-classes of correlated arrivals and blocking. *Operations Research*, *47*, 264 – 276.