MULTIPLE IMPUTATION OF MISSING DATA IN CLINICAL TRIALS

By

Paul Zhang

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Graduate Department of Public Health Sciences
University of Toronto

Multiple Imputation of Missing Data in Clinical Trials

Paul Zhang

Doctor of Philosophy (Biostatistics), 2001

Graduate Department of Public Health Sciences, University of Toronto

Abstract

Missing data or incomplete data are very common in almost every statistical situation. The simple application of complete-data based methods without a careful thinking of the missingness mechanism may result in an invalid conclusion. The statistical inferences on the unknown parameters may be complicated and the power of the inferences may be lowered due to the missing data.

A large number of methods have been developed to handle the missing data. These methods can be categorized as (1) the likelihood-based approach and (2) the multiple imputation approach. The likelihood-based methods utilize the observed data only; all inferences are based on the observed-data likelihood function. The multiple imputation is to impute the missing data by some 'plausible' values multiple times, the analyses results based on each imputed complete data are then combined to make a repeated imputation inference (Rubin, 1987). In order to make a valid repeated imputation inference, an appropriate imputation model needs to be built and the missing data need to be imputed by independent draws from this model.

In this research, we develop new analytical methods for imputing the missing data arising from clinical trials in which the repeated measurement design are commonly employed. A particular covariance structure, the ante-dependence structure, is fully discussed and the methods for imputing the missing data with the ante-dependence covariance structure are the core components of this research. The merits of multiple imputation versus no imputation are investigated by simulations. The performance of the proposed methods and two other developed multiple imputation methods, the predictive model method and the MCMC method, are also compared by simulations. The application of these methods to a clinical trial data is presented.

Acknowledgments

I would like to thank my supervisor, Professor Paul Corey, for his continuous guidance, support and advice during my stay at the University of Toronto and during the completion of this thesis. Without his support, this thesis would not been accomplished.

I would like to thank my thesis committee member, Professor Michael Escobar, for his thorough advice and support during the completion of this thesis. I benefited a lot from a series of valuable discussions with him.

I would like to thank my thesis committee member, Dr. Hau Lei, for his long term encouragement, support and advice during my stay at GlaxoWellcome Inc. and during the completion of this thesis. His timely and detailed comments are very much appreciated.

I would like to thank my family members, Ping, Andrew and Jenny, for their understanding and support. Their support is so important for me during the completion of the thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Missing data or incomplete data are very common in almost every statistical situation, especially in clinical trials and sample surveys. In these two situations, missing data are inevitable no matter how well the clinical trials or the surveys are designed, and no matter how large the effort to collect the complete data. In clinical trials, subjects are screened against the entry criterion, the qualified subjects are then randomized into the treatment period. The same measurements of some aspects are planned to be taken on each subject at baseline and at multiple scheduled post-randomization time points. For various reasons, however, these repeated measurements are not always collected at all time points. In sample surveys, the questions on different aspects are asked, usually at one time point. The results, even when great effort is expended to have every question answered, are usually incomplete. A large number of methods have been developed for making valid and efficient inferences based on the complete data, either in the univariate or multivariate situations. But when data are incomplete, the validity and efficiency of these methods can not be guaranteed. In the last couple of decades, several methods have been developed to make a valid statistical inference when data are incomplete, such as the EM algorithm, random effects model, etc. These likelihood-based methods can make valid inferences on the unknown parameters, although the likelihood function could be complicated by the missing data. One issue pertinent to the missing data is that if a large amount of information is missing, although the likelihood-based methods are valid, the power of the analysis will be low. Another different research area is to fill in the missing data with some 'plausible' values, and then apply those complete-data based methods to make valid and efficient inferences. This method is called the imputation of missing data. It avoids the complexity caused by the missing data and works on the relatively simple complete data situation. Rubin (1976) introduced this concept, he mentioned that when doing imputation, the uncertainty due to filling in the missing data from the imputation model as well as the uncertainty due to selecting the imputation model need to be taken into account. This research is to develop the methods for imputing the missing data coming from clinical trials in which the repeated measurement designs are commonly employed.

In this chapter, some definitions of imputation and some basic results are summarized and presented. These definitions and results are developed by Rubin (1976), Rubin (1987), Little and Rubin (1987). In section 1.1, the definitions of the missingness mechanism are introduced. These definitions are fundamental for the missing data issues. Only when the probabilities of

missingness are clearly defined, can the appropriate methods then be developed. In sections 1.2 and 1.3, the observed-data likelihood function and the observed-data posterior distribution are derived respectively. The inferences on the unknown parameters can be made either from the observed-data likelihood or from the observed-data posterior. In section 1.4, the posterior predictive distribution of the missing data given the observed data is discussed. The missing data are imputed by random draws from this distribution. In section 1.5, the concept of multiple imputation is presented. The rules for repeated imputation inference are presented. The extra uncertainty due to the missing data need to be incorporated into the repeated imputation inferences. Finally, a brief summary of this chapter is presented in section 1.6.

1.1 Missingness Mechanism

Let Y be a $n \times p$ data matrix, $Y = (y_1, y_2, \cdots, y_n)^T$, where $y_i = (y_{i1}, \cdots y_{ip})^T$ is a random sample from a p-dimensional multivariate probability distribution $P(Y \mid \theta)$, which is characterized by the parameters θ. In the following contexts, we refer the rows of Y as the observations, denoted by y_i $(i = 1,2,\cdots,n)$, and the columns of Y as the variables, denoted by Y_j $(j = 1,\cdots, p)$. If data are incomplete, we denote the observed portion and missing portion of Y by Y_{obs} and Y_{mis} respectively, i.e. $Y = (Y_{obs}, Y_{mis})$. Meanwhile we define a $n \times p$ missingness indicator matrix $R = (r_{ij})$ such that

$$r_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is missing} \\ 0 & \text{if } y_{ij} \text{ is observed} \end{cases}$$

Obviously, when data are complete, R is a matrix of 0's with probability 1. When data are incomplete, R will be subject to a probability distribution characterized by the parameters ξ. In this case, the joint probability distribution of these two random components, the response variables and the missingness indicator variables, need to be jointly considered in order to investigate the implications on the inferences about θ in the presence of missing data. Denote this joint distribution by $P(Y, R \mid \theta, \xi)$, then it can be expressed as

$$P(Y, R \mid \theta, \xi) = P(Y \mid \theta) P(R \mid \theta, \xi, Y) \qquad (1.1.1)$$

where $P(Y \mid \theta)$ is the marginal distribution of the response variables, and $P(R \mid \theta, \xi, Y)$ is the conditional distribution of missingness given the response variables.

The two sets of parameters θ and ξ are said to be distinct by Little and Rubin (1987) and Rubin (1987) if (1) From a frequentist perspective, the joint parameter space of (θ,ξ) is the Cartesian cross-product of each individual parameter space for θ and ξ. (2) From a Bayesian perspective, the joint prior distribution of (θ,ξ) can be factored into independent marginal priors for θ and ξ. This assumption is intuitively reasonable in many real situations since knowing θ will provide little information about ξ and vice versa. Under this assumption, (1.1.1) becomes

$$P(Y,R\mid\theta,\xi)=P(Y\mid\theta)P(R\mid\xi,Y) \tag{1.1.2}$$

Note in this probability model, there are two sets of parameters, the parameters pertaining to the response variables and the parameters pertaining to the missingness mechanism. When data are complete, $P(R\mid\xi,Y)=c$, so the inferences about the parameters θ only depends on the probability model for the response variables, i.e. $P(Y\mid\theta)$. When data are incomplete, the correct inferences about θ need to be conducted based on the joint probability model (1.1.2). Hence these inferences depend on how the probability model for the missingness is defined, i.e. how the missingness depends on Y. Rubin (1976), Little and Rubin (1987) categorized the missingness mechanism into the following three categories based on the conditional distribution $P(R\mid\xi,Y)$.

(1) If $P(R\mid\xi,(Y_{obs},Y_{mis}))=P(R\mid\xi)$, the missingness is independent of the responses (observed and missing), then the missingness mechanism is defined as Missing Completely At Random (MCAR).

(2) If $P(R\mid\xi,(Y_{obs},Y_{mis}))=P(R\mid\xi,Y_{obs})$, the missingness is independent of the missing responses given the observed values. In this case the missingness mechanism is defined as Missing At Random (MAR).

(3) If $P(R\mid\xi,(Y_{obs},Y_{mis}))=P(R\mid\xi,Y_{obs},Y_{mis})$, the missingness depends on both observed and missing responses. This missingness mechanism is defined as non-ignorable or informative.

By the above definitions, MCAR is the most restrictive one in which the missing values do not depend on the response variables, neither observed values nor missing values. In this case, the missing values are a simple random sample of all data values, so the distribution of missing values is the same as the distribution of observed values. In contrast to MCAR, MAR is a less restrictive situation in which the missing values depend on the response variables only through the observed values. In this case, the missing values are a simple random sample of all data

values within subgroups defined by the observed values, the distribution of missing values is the same as the distribution of observed values within each subgroup. The missingness mechanism of MCAR and MAR are called *ignorable* missingness (Little and Rubin, 1987; Rubin, 1987), in the sense that the missingness mechanism can be ignored when making statistical inferences on the parameters of interest θ. In this research, we will only study the ignorable missingness mechanism.

1.2 Observed-Data Likelihood Function

When data are incomplete, the full probability model to describe the data should include the sub-model for the data themselves as well as the sub-model for the missingness. The sub-model for the data is the marginal distribution of the response variables $P(Y_{obs}, Y_{mis} \mid \theta)$, and the sub-model for the missingness is the conditional distribution of the indicator variables given the response variables $P(R \mid \xi, Y_{obs}, Y_{mis})$. The full probability model is the joint probability model $P(Y_{obs}, Y_{mis}, R \mid \theta, \xi)$. Since Y_{mis} are unknown, this distribution can not be determined. We therefore investigate the model for the observed-data. By the definition, this model is the marginal distribution of the joint distribution integrated over Y_{mis}, i.e.

$$P(Y_{obs}, R \mid \theta, \xi) = \int P(Y_{obs}, Y_{mis}, R \mid \theta, \xi) dY_{mis} = \int P(R \mid Y_{obs}, Y_{mis}, \xi) P(Y_{obs}, Y_{mis} \mid \theta) dY_{mis} \qquad (1.2.1)$$

Under the MCAR assumption, (1.2.1) becomes

$$P(Y_{obs}, R \mid \theta, \xi) = P(R \mid \xi) \int P(Y_{obs}, Y_{mis} \mid \theta) dY_{mis} = P(R \mid \xi) P(Y_{obs} \mid \theta) \qquad (1.2.2)$$

and under the MAR assumption, (1.2.1) becomes

$$P(Y_{obs}, R \mid \theta, \xi) = P(R \mid \xi, Y_{obs}) \int P(Y_{obs}, Y_{mis} \mid \theta) dY_{mis} = P(R \mid \xi, Y_{obs}) P(Y_{obs} \mid \theta) \qquad (1.2.3)$$

Thus, under the assumptions of either MCAR or MAR, the observed-data distribution (1.2.1) can be factored into two components, one pertains to the parameters of interest θ and another one pertains to the nuisance parameters ξ. When parameters θ and ξ are distinct, the likelihood-based inferences about θ can be conducted based on $P(Y_{obs} \mid \theta)$ alone, without concerning of the missingness mechanism. That is, the missingness mechanism can be completely ignored.

Therefore, if the missingness mechanism is ignorable, the joint observed-data distribution $P(Y_{obs}, R \mid \theta, \xi)$, can be replaced by the factor pertaining to θ, i.e. $P(Y_{obs} \mid \theta)$, for the purpose of inferences on the parameters θ. This function is referred by Little and Rubin (1987) as the observed-data likelihood function ignoring the missingness mechanism.

$$L(\theta \mid Y_{obs}) \propto P(Y_{obs} \mid \theta) \tag{1.2.4}$$

Note that this observed-data likelihood is of the parameters θ which pertains to the model for the complete data $Y = (Y_{obs}, Y_{mis})$, not the parameters for the distribution of Y_{obs} alone. Although in some situations, some parameters of complete-data model may not appear in the model for Y_{obs}. For example, if the values of the last variable Y_p are missing for all subjects, the mean and variance of this variable will not appear in $P(Y_{obs} \mid \theta)$. In this case, the parameters associated with this variable can not be estimated in any way. If the primary interest is on the θ which only pertains to the distribution of Y_{obs}, the above assumptions like MCAR or MAR would not be necessary. For detailed discussion of this point, please see the exchange between Efron (1994) and Rubin (1994). Our ultimate interest is to draw valid inferences on the parameters pertaining to the model for the complete data $Y = (Y_{obs}, Y_{mis})$, rather than the inferences on the parameters for the distribution of Y_{obs} alone. It has been shown that this task can be performed based on the observed-data likelihood function (1.2.4) as long as the missingness mechanism can be ignored.

1.3 Observed-Data Posterior Distribution

In the Bayesian framework, the inferences about the unknown parameters are based on their posterior probability distributions. When data are incomplete, the unknown parameters are (θ, ξ), and the observed portions of the data are Y_{obs} and R. By Bayes's Theorem, the posterior distribution for (θ, ξ) is the product of the probability distribution for the data and a prior probability distribution for the unknown parameters (θ, ξ), i.e.

$$P(\theta, \xi \mid Y_{obs}, Y_{mis}, R) \propto P(Y_{obs}, Y_{mis}, R \mid \theta, \xi) \times \pi(\theta, \xi) \tag{1.3.1}$$

where $\pi(\theta, \xi)$ is a joint prior distribution for θ and ξ. The prior distribution is one's prior knowledge about (θ, ξ), it has nothing to do with the intended data. Therefore, the prior distribution is completely independent of Y_{obs} and Y_{mis}. The observed-data posterior distribution of (θ, ξ), by definition, is the complete-data posterior (1.3.1) integrated over Y_{mis}, i.e.

$$\begin{aligned} P(\theta, \xi \mid Y_{obs}, R) &\propto \int P(Y_{obs}, Y_{mis}, R \mid \theta, \xi) \times \pi(\theta, \xi) dY_{mis} \\ &= \pi(\theta, \xi) \times \int P(Y_{obs}, Y_{mis}, R \mid \theta, \xi) dY_{mis} \end{aligned} \tag{1.3.2}$$

When the missingness mechanism is ignorable, the observed-data posterior distribution of (θ,ξ) can be factored as

$$P(\theta,\xi \mid Y_{obs},R) \propto P(R|\xi)\times P(Y_{obs} \mid \theta)\times\pi(\theta,\xi) \qquad (1.3.3)$$

or

$$P(\theta,\xi \mid Y_{obs},R) \propto P(R \mid \xi,Y_{obs})\times P(Y_{obs} \mid \theta)\times\pi(\theta,\xi) \qquad (1.3.4)$$

The Bayesian inferences on (θ,ξ) can be made simultaneously based on this observed-data posterior distribution. But as for the likelihood-based inferences on the parameters of interest θ, our ultimate goal is to make Bayesian inferences about θ alone. These Bayesian inferences on θ can be made based on the marginal observed-data posterior distribution obtained by integrating either (1.3.3) or (1.3.4) over the nuisance parameters ξ. When θ and ξ are distinct as defined by Little and Rubin (1987) and Rubin (1987), the joint prior distribution can be factored as

$$\pi(\theta,\xi)=\pi_{\theta}(\theta)\times\pi_{\xi}(\xi)$$

the marginal observed-data posterior distribution of θ is

$$P(\theta \mid Y_{obs},R)=\int P(\theta,\xi \mid Y_{obs},R)d\xi \propto L(Y_{obs} \mid \theta)\times\pi_{\theta}(\theta) \qquad (1.3.5)$$

Note that R does not appear on the right-hand side of (1.3.5), so $P(\theta \mid Y_{obs},R)=P(\theta \mid Y_{obs})$. Thus it has been shown that the observed-data posterior distribution of θ is independent of the missingness mechanism,

$$P(\theta \mid Y_{obs}) \propto L(Y_{obs} \mid \theta)\times\pi_{\theta}(\theta) \qquad (1.3.6)$$

Since the observed-data posterior distribution depends on the probability model for the response variables as well as the prior probability model for the parameters of interest θ, it seems that making the inferences based on the observed-data posterior is more difficult than that based on the observed-data likelihood. But in applications, by properly choosing a prior distribution for θ within a natural conjugate family, the observed-data posterior can be greatly simplified when data are complete, see Box and Tiao (1973). In this case, the inferences can be made by either analytical approach or the Monte Carlo approach in which the values of θ are simulated from the observed-data posterior distribution. However, if data are incomplete, the usual natural conjugate prior will not guarantee that the observed-data posterior is tractable and easily to be summarized. In this case, the Bayesian inferences are made by the Monte Carlo

approach or some complicated iterative simulation techniques if the values of θ can not be directly simulated from the observed-data posterior distribution.

1.4 Posterior Predictive Distribution

From the Bayesian point of view, any thing unknown is treated as random. When data are complete, we have two random components, Y and θ, where Y is a random realization from the probability model for the response variables and θ is governed by its posterior distribution given Y. When data are incomplete, we have five random components, Y_{obs}, Y_{mis}, θ, R and ξ. Y_{obs} and R are observable, Y_{mis}, θ and ξ are non-observable. If the missingness mechanism is ignorable, we can safely ignore the two random components R and ξ, leaving three random components for consideration, Y_{obs}, Y_{mis} and θ. Obviously the joint distribution of these three components is very difficult to handle, and is not our primary interest. These three random components are inter-dependent, the full description of one component depends on the specifications of others. Our primary interest is to get the best estimates of θ and/or Y_{mis} based on the direct or indirect evidence provided by Y_{obs}. In doing so we consider some marginal distributions and conditional distributions interesting to us. In section 1.2 we discussed the marginal distribution of Y_{obs}, the observed-data likelihood. That is,

$$L(\theta \mid Y_{obs}) \propto P(Y_{obs} \mid \theta) = \int P(Y_{obs}, Y_{mis} \mid \theta) dY_{mis} \qquad (1.4.1)$$

Any statistical inferences on θ based on this observed-data likelihood or based on the observed-data posterior distribution $L(Y_{obs} \mid \theta) \times \pi_\theta(\theta)$ are valid. These inferences utilize the observed data Y_{obs} only. The likelihood-based inferences are usually done analytically, Bayesian inferences can be done either analytically or by the Monte Carlo method. If a large amount of data is missing, even these inferences are valid, the power of these inferences will be low. Also when data are incomplete, the MLE of θ usually requires an iterative algorithm. A large amount of missing data may raise the concerns on the convergence issues.

Another approach of handling missing data is to impute the missing data Y_{mis}, and then apply those complete-data based methods to the imputed completed-data to make inferences on the parameters θ. The missing data Y_{mis} should be imputed from some probability distributions such that the imputed values preserve the intrinsic correlations between Y_{mis} and Y_{obs}. Also

these probability distributions should engage the same set of parameters θ as the one in the probability model for the response variables, the inferences on θ then can be made from the imputed completed-data. Therefore, the probability model for imputing Y_{mis} should depend on Y_{obs} and θ. Since both Y_{mis} and θ are non-observable, the probability distribution of Y_{mis} can not be determined without specifying the values of θ. To avoid this dependency on the values of θ, the expected probability distribution of Y_{mis} is derived over the parameter space of θ, i.e.

$$P(Y_{mis} \mid Y_{obs}) = \int P(Y_{mis} \mid Y_{obs}, \theta) P(\theta \mid Y_{obs}) d\theta \qquad (1.4.2)$$

where $P(Y_{mis} \mid Y_{obs}, \theta)$ is the *conditional predictive* distribution of Y_{mis} given Y_{obs} and θ, and $P(\theta \mid Y_{obs})$ is the observed-data posterior distribution of θ. This expected probability distribution of Y_{mis}, $P(Y_{mis} \mid Y_{obs})$, is called the *posterior predictive distribution* which is the conditional predictive distribution $P(Y_{mis} \mid Y_{obs}, \theta)$ averaged over the observed-data posterior distribution of θ. The imputations of Y_{mis} by random draws from the posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$ reflects the uncertainty due to the random sampling of Y_{mis} when the parameters θ are given, as well as the extra uncertainty due to the fact that the parameters θ are unknown. Therefore, the posterior predictive distribution of Y_{mis} is the ideal distribution from which the missing values Y_{mis} can be imputed.

1.5 Multiple Imputation

Multiple imputation (MI) is a general approach for handling missing data. This concept was first proposed by Rubin in early 1970's as a possible solution to the problem of survey non-response. He emphasized that the missing data should be handled based on some principled methods, rather than based on some ad hoc methods. Multiple imputation is a principled method and consists of three steps. The first step is to create $m\,(m > 1)$ complete data sets by substituting the missing values m times using m independent draws from an imputation model. The imputation model is constructed to reasonably approximate the true distributional relationship between the missing values and the observed values. In the second step the m imputed complete data sets are analyzed by treating each imputed complete data set as a 'real' complete data set. Standard complete data procedures and software can be utilized directly. In the third step the analysis results based on the m complete data sets are combined by a simple,

appropriate way to obtain the so-called repeated imputation inferences (Rubin, 1987). The variances of combined estimates consist of within imputation and between imputation variances, so the uncertainty in the imputed data is being properly incorporated into the final inference. This method overcomes the drawback of single imputation, which underestimates the standard error of estimate.

The first step, drawing random samples from an imputation model, is the most fundamental part of the multiple imputation. This task involves building an imputation model and drawing random samples from the built imputation model. A multiple imputation is called a *proper* multiple imputation (Rubin, 1987), if (1) the imputation model preserves the correlations between the missing values and the observed values. (2) The missing values are imputed by the *independent* random draws from the imputation model, and (3) the estimation of θ is approximately unbiased over the imputed data distribution. If the imputation model does not preserve the correlations between the missing values and the observed values, the inferences about these correlations based on the imputed complete data will be biased. For example, if the imputation model does not include the variables which will be used in the inferences based on the imputed complete data, then the correlations between these omitted variables and the imputed variables will be biased towards zero. If the multiple imputation are not based on the independent draws from the imputation model, it will result in the between imputation variance being underestimated. For the third requirement, approximately unbiased estimate of θ over the imputation model is to assure that the inferences based on repeated imputations are valid from the frequentists point of view.

The theory for making a repeated or combined imputation inference is derived from a Bayesian model (Rubin, 1987). Let Q be a generic scalar quantity to be estimated, such as treatment effect, odds ratio or regression coefficient, and then the observed-data posterior of Q is given by

$$P(Q \mid Y_{obs}) = \int P(Q \mid Y_{obs}, Y_{mis}) P(Y_{mis} \mid Y_{obs}) dY_{mis} \qquad (1.5.1)$$

That is, the observed-data posterior is the completed-data posterior averaged over the posterior predictive distribution. Hence the moments summaries can be obtained from the observed-data posterior distribution as

$$E(Q \mid Y_{obs}) = E[E(Q \mid Y_{obs}, Y_{mis}) \mid Y_{obs}] = Ave(\hat{Q})$$

$$V(Q \mid Y_{obs}) = E[V(Q \mid Y_{obs}, Y_{mis}) \mid Y_{obs}] + V[E(Q \mid Y_{obs}, Y_{mis}) \mid Y_{obs}] = Ave(\hat{U}) + (1 + m^{-1}) V(\hat{Q})$$

where $\hat{Q}$ and $\hat{U}$ are the estimates of Q and var(Q) based on the imputed complete data respectively, and Ave and V are the average and variance over the repeated imputations.

After the missing data Y_{mis} have been imputed by m sets of independent draws from the imputation model, the repeated imputation inference is performed as follows. Let $\hat{Q} = \hat{Q}(y_{obs}, y_{mis})$ denote the statistic that would be used to estimate Q if complete data are available and let $\hat{U} = \hat{U}(y_{obs}, y_{mis})$ be its squared standard error. It is assumed that with complete data, the hypothesis test and confidence interval are based on the normal approximation

$$(\hat{Q} - Q)/\sqrt{\hat{U}} \sim N(0,1) \tag{1.5.2}$$

We may need to make a transformation to have the estimand meet this approximation. Now based on m independent imputations, we calculate the imputed-data estimates $\hat{Q}^{(k)} = \hat{Q}(y_{obs}, y_{mis}^{(k)})$ along with their estimated variances $\hat{U}^{(k)} = \hat{U}(y_{obs}, y_{mis}^{(k)})$, $k = 1,2,\cdots,m$. The overall estimate of Q is simply the average of these imputed estimates,

$$\overline{Q} = \frac{1}{m}\sum_{k=1}^{m}\hat{Q}^{(k)} \tag{1.5.3}$$

The standard error of $\overline{Q}$ is

$$T = \left\{(1+m^{-1})B + \overline{U}\right\}^{1/2} \tag{1.5.4}$$

where $B = \frac{1}{m-1}\sum_{k=1}^{m}(\hat{Q}^{(k)} - \overline{Q})^2$ is the between imputation variance and $\overline{U} = \frac{1}{m}\sum_{k=1}^{m}\hat{U}^{(k)}$ is the within imputation variance. The hypothesis test and the confidence interval are then based on an approximate Student-t distribution

$$(\overline{Q} - Q)/T \sim t_{\nu} \tag{1.5.5}$$

with the degrees of freedom

$$\nu = (m-1)\left[1 + r^{-1}\right]^2 \tag{1.5.6}$$

where r is the relative increase in variance due to missing data

$$r = (1+m^{-1})B/\overline{U} \tag{1.5.7}$$

If there is no multiple imputation, the between imputation variance B is zero and the within imputation variance $\overline{U}$ is the complete data variance, so the index r is a measurement of increase in variance caused by multiple imputation. It is noticed that the value of r is going down while the value of ν is going up when number of imputation m goes up.

Another important diagnostic index of multiple imputation is the fraction of missing information about Q,

$$\lambda = \frac{r + 2/(v+3)}{r+1} \tag{1.5.8}$$

This index is a measurement of how the missing data influence the uncertainty of estimate of Q. Its value is determined by the proportion of missing data and the performance of multiple imputation. When number of imputation m goes up, the value of λ goes down.

The above repeated imputation inference is valid as long as the imputations $y_{mis}^{(1)}, y_{mis}^{(2)}, \cdots, y_{mis}^{(m)}$ are generated properly. The imputations should, on average, give reasonable predictions for the missing values, and the variability among the imputations must reflect an appropriate degree of uncertainty.

The inference for multidimensional estimands can be generalized from single estimand. For detailed discussions, see Li, Raghunnthan and Rubin (1991), Meng and Rubin (1992), Schafer (1997).

1.6 Summary

In this chapter, we mentioned two approaches for handling missing data. One is the likelihood-based approach and another one is the imputation-based approach. The concept of imputation, specifically the multiple imputation, was introduced by Rubin in the 1970's. The fundamental work on imputation has been well developed by Rubin (1976), Rubin (1987), Little and Rubin (1987). The definitions of the missingness mechanism, the ignorability of missingness are presented in this chapter. The observed-data likelihood function, the observed-data posterior distribution and the posterior predictive distribution are defined and derived. The concept of multiple imputation and the rule for combined multiple imputation inference are presented. One of the most important concepts in multiple imputation is the *proper* multiple imputation. Only when the multiple imputation is proper, will the combined multiple imputation inference be valid.

Chapter 2. Bayesian Statistics

Multiple imputation is to fill in the missing data by some plausible values multiple times. The imputation process involves two sources of uncertainties, one is the uncertainty due to the random sampling of Y_{mis} from a probability distribution and the second is the uncertainty due to the identification of this probability distribution. The inference based on the imputed data should consider these two uncertainties. The posterior predictive distribution of Y_{mis} given Y_{obs}, $P(Y_{mis} \mid Y_{obs})$, is an ideal probability distribution from which the missing values Y_{mis} are drawn. The posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$ is the conditional predictive distribution $P(Y_{mis} \mid Y_{obs}, \theta)$ averaged over the observed-data posterior distribution of unknown parameters θ. Therefore, the identification of $P(Y_{mis} \mid Y_{obs})$ involves the knowledge of the observed-data posterior distribution of θ. All these concepts are within the Bayesian frame. To lay out a foundation for the following chapters, we review some definitions and properties of Bayesian statistics in this chapter. For detailed descriptions of Bayesian statistics, see Box and Tiao (1973), Gelman, Rubin, Carlin and Stern (1995). In section 2.1, Bayes theorem is presented. In section 2.2, different prior distributions are discussed. Jeffreys' rules (Jeffreys, 1961) for choosing the non-informative priors are reviewed. In section 2.3, posterior distributions are reviewed. The robustness of posterior distribution to the choice of prior distribution is discussed. Section 2.4 presents a brief summary.

2.1 Bayes Theorem

Assume that y is a random sample from a probability distribution $P(y \mid \theta)$ which is characterized by the distribution parameter θ. We also assume that the parameter θ itself has a probability distribution $P(\theta)$. Then the joint probability distribution of y and θ is,

$$P(y,\theta) = P(y \mid \theta)P(\theta) = P(\theta \mid y)P(y) \tag{2.1.1}$$

Given the observed data y, the conditional distribution of θ is

$$P(\theta \mid y) = \frac{P(y \mid \theta)P(\theta)}{P(y)} \tag{2.1.2}$$

where the marginal distribution of y, $P(y)$, is the joint distribution $P(y,\theta)$ integrated over the admissible range of θ, i.e.

$$P(y) = \int P(y,\theta)d\theta = \int P(y \mid \theta)P(\theta)d\theta \qquad (2.1.3)$$

This marginal distribution is merely a 'normalizing' constant with respect to θ, it ensures that the conditional distribution $P(\theta \mid y)$ integrates to one over the admissible range of θ. Therefore, (2.1.2) can be expressed as

$$P(\theta \mid y) = cP(y \mid \theta)P(\theta) \qquad (2.1.4)$$

This formula is referred as Bayes' theorem (Bayes, 1763). In this expression, the marginal distribution $P(\theta)$, which tells us what is known about θ before the data are observed, is called the prior distribution of θ or the distribution of θ a priori. Correspondingly, the conditional distribution $P(\theta \mid y)$, which tells us what is known about θ after the data are observed, is called the posterior distribution of θ or the distribution of θ a posteriori. The conditional distribution function $P(y \mid \theta)$ is the distribution we assumed on the response variable Y. If we treat the observed value y as fixed and the parameter θ as varied, then $P(y \mid \theta)$ can be regarded as a function of θ but not of y. When so regarded, this function is called the likelihood function of θ for given y. Thus, Bayes's formula can be written as

$$P(\theta \mid y) = L(\theta \mid y)P(\theta) \qquad (2.1.5)$$

That is, the posterior knowledge of θ can be viewed as the prior knowledge of θ modified through the likelihood function when data are observed.

2.2 Prior Distribution

A prior distribution represents what is known about the unknown parameters θ before the data are available, or it represents a state of ignorance about θ before the data are available. In Bayes theorem, the prior distribution plays an important role. The Bayesian inferences about the unknown parameters are made based on the posterior distribution. The posterior distribution itself is influenced by the prior distribution. In this section, the definitions of different priors and the approach to select the priors are presented. For a detailed description of prior distribution, please see Spiegelhalter, Freedman and Palmer (1994).

Improper prior:

A basic property of a probability density function $P(x)$ is that it integrates to 1 over its admissible range. If a prior distribution $P(\theta)$ is not integrated to 1 over the admissible range of θ, then this prior distribution is called an improper prior.

Sometimes, although the prior is improper, the posterior distribution, which combines the likelihood and the prior, maintains as a proper distribution.

Likelihood dominated prior:

A likelihood dominated prior is one which does not change very much over the region in which the likelihood is appreciable and does not assume large values outside that region (Box and Tiao, 1973, p23).

A prior distribution with these properties is referred as a *locally uniform prior* (Box and Tiao, 1973, p23).

For such a prior distribution, the Bayes' formula can be approximated by substituting a constant for the prior distribution so that

$$P(\theta \mid y) = \frac{L(\theta \mid y)P(\theta)}{\int L(\theta \mid y)P(\theta)d\theta} \approx \frac{L(\theta \mid y)}{\int L(\theta \mid y)d\theta} \tag{2.2.1}$$

That is, for a locally uniform prior, the posterior distribution density function is approximately numerically equal to the standardized likelihood function.

Although there are some cases where the prior distribution density is not dominated by the likelihood function, in most scientific investigations it is reasonable to assume that the prior distributions are dominated by the likelihood functions, see Box and Tiao (1973, p21), Lee (1997, p45).

Reference prior:

A prior distribution which is convenient to use as a standard is called reference prior. A reference prior may or may not be dominated by the likelihood function.

Conjugate prior distributions:

The posterior distribution of θ is given by $P(\theta \mid y) = L(\theta \mid y)P(\theta)$. This distribution may have a quite different functional format from either the likelihood function $L(\theta \mid y)$ or the prior distribution $P(\theta)$. However, it is sometimes desirable for the posterior distribution to have the same functional format as the prior distribution. In this case, this prior is called a conjugate prior.

The formal definition of conjugate prior distributions is given as follows (Lee, 1997, p59):

A class ω of prior distributions is said to form a conjugate family if the posterior distribution is also in the class ω for all y whenever the prior distribution is in ω.

Non-informative prior distribution:

A prior distribution is supposed to represent little a priori knowledge about the parameters before the data are available. Box and Tiao (1973, p25) mentioned that 'knowing little a priori' can only have meaning relative to the information provided by the data. Thus, the main issue is how to select a prior which provides little information relative to what is expected to be provided by the intended data. To say that we know little a priori relative to what the data is going to tell us is equivalent to saying that we are almost equally willing to accept one value of θ as another. The prior distributions with this state of indifference are called non-informative priors with respect to the data (Box and Tiao, 1973, p27). When $P(\theta)$ is locally uniform, it is non-informative in a local area in which the likelihood is appreciable.

Jeffreys' rule for single parameter (Jeffreys, 1961):

The prior distribution for a single parameter θ is approximately non-informative if it is taken proportional to the square root of Fisher's information measure (Fisher, 1922, 1925). That is if

$$P(\theta) \propto I^{1/2}(\theta)$$

where

$$I(\theta) = E_{y|\theta}\left[-\frac{\partial^2 \log L(\theta \mid y)}{\partial\theta\,\partial\theta}\right]$$

then $P(\theta)$ is approximately non-informative.

Example (Box and Tiao, 1973, p43):

Suppose $y^T = (y_1, y_2, \cdots, y_n)$ is a set of n independent random samples from $N_1(\mu,\sigma^2)$.

a) μ unknown, σ known. The loglikelihood function is

$$\ell(\mu \mid \sigma, y) = -\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{2\sigma^2}$$

and

$$\frac{\partial^2 \ell(\mu \mid \sigma, y)}{\partial\mu\,\partial\mu} = c$$

Therefore,

$$P(\mu \mid \sigma) \propto c$$

b) μ known, σ unknown. The loglikelihood function is

$$\ell(\sigma \mid \mu, y) = -n\log(\sigma) - \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma^2}$$

and

$$E\left[-\frac{\partial^2 \ell(\sigma \mid \mu, y)}{\partial\sigma\,\partial\sigma}\right] = \frac{n}{\sigma^2}$$

Therefore,

$$P(\sigma \mid \mu) \propto \frac{1}{\sigma} \qquad \text{or} \qquad P(\log(\sigma) \mid \mu) \propto c$$

In some situations, it is appropriate to assume that the location parameters are independent of the scale parameter. This is because any prior knowledge about the location parameters would usually not be much influenced by the prior knowledge of the scale parameter. Thus under this assumption $P(\mu \mid \sigma) = P(\mu)$. When σ is known, a non-informative prior for μ is obtained by taking $P(\mu \mid \sigma)$ locally uniform. With additional independence assumption this implies that $P(\mu)$ should be uniform. A similar argument leads to take $P(\sigma) \propto 1/\sigma$. Thus

$$P(\mu,\sigma) = P(\mu)P(\sigma) \propto 1/\sigma$$

In this case, the posterior distribution of (μ,σ) is

$$P(\mu,\sigma \mid y) \propto \sigma^{-(n+1)} \exp\left[-\frac{n(\mu - \bar{y})^2}{2\sigma^2} - \frac{(n-1)s^2}{2\sigma^2}\right]$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$.

Jeffreys' rule for multiple parameters (Jeffreys, 1961):

An approximate non-informative prior distribution for a set of parameters is taken to be proportional to the square root of the determinant of the Fisher's information matrix. That is

$$P(\theta) \propto |I(\theta)|^{1/2}$$

Example (Box and Tiao, 1973, p51):

Suppose $y^T = (y_1, y_2, \cdots, y_n)$ is a set of normally and independently distributed random samples having a common variance σ^2, and the expected value of y_i is a linear function of p parameters $\theta = (\theta_1, \theta_2, \cdots, \theta_p)$ such that

$$E(y_i) = \theta_1 x_{i1} + \theta_2 x_{i2} + \cdots + \theta_p x_{ip} \qquad i = 1, 2, \cdots, n$$

where the x's are known constants. Let X be the $n \times p$ matrix of $\{x_{ij}\}$, then the n equations can be written concisely as

$$E(y) = X\theta$$

When both θ and σ are unknown, the likelihood function is

$$L(\theta, \sigma \mid y) \propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}(y - X\theta)^T (y - X\theta)\right\}$$

$$\propto \left(\frac{1}{\sigma}\right)^n \exp\left\{-\frac{(n-p)s^2}{2\sigma^2} - \frac{(\theta - \hat{\theta})^T X^T X (\theta - \hat{\theta})}{2\sigma^2}\right\}$$

where

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

and

$$s^2 = \frac{1}{n-p}(y - X\hat{\theta})^T (y - X\hat{\theta})$$

A non-informative prior in this case is the one for which approximately $\log\sigma$ and $(\theta_1, \theta_2, \cdots, \theta_p)$ are locally uniform. That is, locally,

$$P(\theta, \log\sigma) \propto c$$

or equivalently,

$$P(\theta, \sigma) \propto \sigma^{-1}$$

The posterior distribution of (θ, σ) is

$$P(\theta, \sigma \mid y) \propto \left(\frac{1}{\sigma}\right)^n \left(\frac{1}{\sigma}\right) \exp\left\{-\frac{(n-p)s^2}{2\sigma^2} - \frac{(\theta - \hat{\theta})^T X^T X (\theta - \hat{\theta})}{2\sigma^2}\right\}$$

The choice of a prior does not necessarily represent one's prior knowledge about the parameters of interest, but this prior should at least represent an unprejudiced state about the parameters. In this research, our primary interest is to impute the missing data from the posterior predictive distribution. By doing so, we need to acquire the knowledge about the posterior distribution. For this purpose, the choice of the prior distribution is served as to simplify the posterior distribution. Non-informative priors can simplify the posterior distribution in a broad range of situations, but it is not always the case.

2.3 Posterior Distribution

The posterior distribution of θ given y is $P(\theta \mid y) = L(\theta \mid y)P(\theta)$. This distribution depends on the choice of the prior distribution $P(\theta)$ and the likelihood function of θ given y. The likelihood function $L(\theta \mid y)$ is determined by the probability model for the response variable Y which is relatively easy to verify. But the choice of the prior distribution depends on one's prior knowledge which is relatively hard to verify. Therefore, we would like to see that the posterior distribution is robust to the choice of the prior distribution, and the choice of the prior distribution can simplify the posterior distribution.

In fact, if the prior distribution is dominated by the likelihood, that is, the one that does not change very much over the region in which the likelihood is appreciable and does not assume large values outside that region. Then the posterior distribution is robust to the changes of the prior distribution. This result comes from the following theorem.

Theorem (Lindley, 1965): A random sample $y = (y_1, y_2, \cdots, y_n)$ of size n is taken from $N(\mu, \sigma)$ where σ is known. Suppose that there exist positive constants α, ε, M and c depending on y, such that in the interval I_α defined by

$$I_\alpha = \left\{\mu:\ \bar{y} - \lambda_\alpha \sqrt{\sigma / n} \le \mu \le \bar{y} + \lambda_\alpha \sqrt{\sigma / n}\right\}$$

where

$$\Phi(-\lambda_\alpha) = \alpha / 2$$

the prior distribution density of μ lies between $c(1-\varepsilon)$ and $c(1+\varepsilon)$, and outside I_α the prior density is bounded by Mc, that is $|P(\mu)| \le Mc$. Then inside I_α, the posterior distribution density $P(\mu \mid y)$ satisfies

$$\frac{(1-\varepsilon)}{(1+\varepsilon)(1-\alpha)+M\alpha}(2\pi\sigma/n)^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\bar{y}-\mu)^2/(\sigma/n)\}$$

$$\le\ P(\mu \mid y)\ \le \frac{(1+\varepsilon)}{(1-\varepsilon)(1-\alpha)}(2\pi\sigma/n)^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\bar{y}-\mu)^2/(\sigma/n)\}$$

and outside I_α it satisfies

$$0\ \le\ P(\mu \mid y)\ \le\ \frac{M}{(1-\varepsilon)(1-\alpha)}(2\pi\sigma/n)^{-\frac{1}{2}}\exp\{-\frac{1}{2}\lambda_\alpha^2\}$$

Note that in this theorem the expression $(2\pi\sigma/n)^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\bar{y}-\mu)^2/(\sigma/n)\}$ is the likelihood function, so this theorem tells us that inside the interval I_α, the posterior density has the lower bound

$$\frac{(1-\varepsilon)}{(1+\varepsilon)(1-\alpha)+M\alpha}L(\mu\mid y)$$

and the upper bound

$$\frac{(1+\varepsilon)}{(1-\varepsilon)(1-\alpha)}L(\mu\mid y)$$

respectively. Similar results hold for multiple parameters.

The interpretation of this theorem is as follows: Take $\alpha=0.01$, since $\lambda_{0.01}=2.58$, the likelihood function of μ is appreciable in the interval $I_\alpha=\{\bar{y}-2.58\sqrt{\sigma/n}\,,\,\bar{y}+2.58\sqrt{\sigma/n}\}$. If the prior distribution $P(\mu)$ is bounded by $c(1-\varepsilon)$ and $c(1+\varepsilon)$ inside this interval, and $|P(\mu)|\leq Mc$ outside I_α, then the posterior distribution lies within the multiplies $\frac{(1-\varepsilon)}{(1+\varepsilon)(1-\alpha)+M\alpha}$ and $\frac{(1+\varepsilon)}{(1-\varepsilon)(1-\alpha)}$ of the likelihood function.

Therefore, it has been shown that as long as the values of the prior distribution are limited both within and outside the interval I_α, the posterior distribution will not change much within the interval I_α. Also from section 2.2, we know that if the prior distribution is dominated by the likelihood, the posterior distribution can be simplified as the standardized likelihood.

In many real applications, the prior knowledge or relative ignorance about the parameters can be well represented without prejudice by the likelihood dominated prior distribution. In these situations, the posterior distribution can be simplified. In some other situations choosing a non-informative prior can also lead to the simplification of the posterior distribution.

2.4 Summary

Posterior predictive distribution is an ideal distribution from which the missing data are imputed. In order to obtain the posterior predictive distribution, we need to acquire the knowledge of the posterior distribution. The posterior distribution is a combination of the likelihood function and the prior distribution. The definitions of different priors are presented in this chapter, most of those definitions are from Box and Tiao (1973). According to them, the

likelihood dominated prior is equivalent to the locally uniform prior. The non-informative prior has no strong preference on one value of the parameter than for another value. In many situations, non-informative prior can simplify the posterior distribution, and Jeffreys' rules are convenient to use for choosing the non-informative priors. This research is intended to impute the missing data rather than making inference about the parameters from the posterior distribution. For this purpose, it is convenient for us to choose priors such that the posterior distribution can be simplified, and the values of the parameters can be easily simulated from the posterior distribution, either analytically or by the Monto Carlo method.

Chapter 3. Literature Review

Standard methods have been developed to analyze complete data. When data are incomplete, the application of these complete-data based methods without a careful consideration of the missingness mechanism may result in an invalid conclusion (Rubin, 1987). If the missingness mechanism can be safely ignored in the sense defined by Little and Rubin (1987), Rubin (1987), those likelihood-based methods can be directly applied to the observed-data likelihood functions. But since the data are incomplete, it complicates the likelihood function, some properties associated with those methods, such as integrity, simplicity etc., will no longer hold. Correspondingly, making a statistical inference on the incomplete data may become a very difficult task. To overcome these difficulties, several methods have been developed for handling incomplete data. Basically, there are two approaches to handle missing data. The first one is to estimate the unknown parameters and make inferences on them by utilizing the observed data only. A second approach is to impute the missing data with some plausible values, and then make the estimation and inferences of the unknown parameters based on the imputed complete data. In this chapter, four methods of handling incomplete data are reviewed. In section 3.1 the likelihood-based methods, especially the EM algorithm, are presented. In section 3.2 the propensity score based multiple imputation is discussed. In section 3.3 the predictive model based multiple imputation is described. In section 3.4 a new multiple imputation technique, the Markov Chain Monto Carlo (MCMC), is discussed and a brief summary is given in section 3.5.

3.1 The EM Algorithm

When data are complete, the maximum likelihood estimate (MLE) of the parameters of interest $\theta = (\mu, \Sigma)$ is obtained by maximizing the likelihood function $L(\theta | y)$ with respect to θ. For a detailed description of the ML estimation, see Cox and Hinkley (1974).

When data are incomplete, if the missingness mechanism is ignorable, the observed-data likelihood of $\theta = (\mu, \Sigma)$ is given by

$$L(\theta | Y_{obs}) = \int P(Y_{obs}, Y_{mis} | \theta) dY_{mis} . \qquad (3.1.1)$$

The maximum likelihood estimate (MLE) of θ is obtained by maximizing this observed-data likelihood function with respect to θ. In one formal sense there is no difference between ML estimation for complete data and ML estimation for incomplete data (Little and Rubin, 1987, p88).

The ML estimation procedures are summarized as follows (Little and Rubin, 1987, p128).

If the likelihood function is differentiable and unimodal, the MLE of θ can be obtained by solving the likelihood equation

$$S(\theta \mid Y_{obs}) \equiv \frac{\partial \ell(\theta \mid Y_{obs})}{\partial \theta} = 0 \tag{3.1.2}$$

where $\ell(\theta \mid Y_{obs}) = \ln L(\theta \mid Y_{obs})$ is the observed-data loglikelihood function, and $S(\theta \mid Y_{obs})$ is the observed-data score function. When a closed form solution of (3.1.2) can not be found, iterative methods will be applied. The Newton-Raphson algorithm is one of them: Given an initial estimate of θ, say $\theta^{(0)}$, let $\theta^{(t)}$ be the estimate at the t^{th} iteration, then the $(t+1)^{th}$ iteration is defined by the equation

$$\theta^{(t+1)} = \theta^{(t)} + I^{-1}(\theta^{(t)} \mid Y_{obs}) S(\theta^{(t)} \mid Y_{obs}) \tag{3.1.3}$$

where $I(\theta \mid Y_{obs})$ is the observed information matrix, which is the second order derivatives of the observed-data loglikelihood $\ell(\theta \mid Y_{obs})$ with respect to θ,

$$I(\theta \mid Y_{obs}) = -\frac{\partial^2 \ell(\theta \mid Y_{obs})}{\partial \theta \, \partial \theta}.$$

Under general assumptions on the observed-data likelihood function (3.1.1), the sequence of iterates $\{\theta^{(t)}\}$ converges to a point estimate $\hat{\theta}$, which is the MLE of θ.

A variant of this procedure is the Method of Scoring, where the observed information in (3.1.3) is replaced by the expected information

$$J(\theta) = E\{I(\theta \mid Y_{obs}) \mid \theta\} = -\int \frac{\partial^2 \ell(\theta \mid Y_{obs})}{\partial \theta \, \partial \theta} P(Y_{obs} \mid \theta) dY_{obs}.$$

Both these methods involve calculating the matrix of second order derivatives of the loglikelihood. For complex patterns of incomplete data, the entries in this matrix tend to be complicated functions of θ.

A third alternative algorithm (Berndt et al., 1974) is to replace the observed information in (3.1.3) by the sampling covariance matrix of the score function $S(\theta \mid Y_{obs})$, since this matrix is a consistent estimate of the information matrix in the neighborhood of $\hat{\theta}$. The resulting iterative equation is

$$\theta^{(t+1)} = \theta^{(t)} + \lambda_t Q^{-1}(\theta^{(t)}) S(\theta^{(t)} \mid Y_{obs})$$

where $Q(\theta) = \sum_{i=1}^{n} (\partial \ell_i / \partial \theta)(\partial \ell_i / \partial \theta)^T$, ℓ_i is the observed-data loglikelihood of the i^{th} observation and λ_t is a positive scale designed to ensure the convergence to a local maximum.

These iterative methods will generate valid maximum likelihood estimates of θ as long as the missingness mechanism can be safely ignored. These methods require the calculations or approximations of the second order derivatives of the observed-data loglikelihood function. When data are incomplete, the observed-data likelihood (3.1.1) may be complicated, and the entries in the second order derivative matrix tend to be complicated function of θ. As a result, the calculation or approximation of this matrix may become very complicated (Little and Rubin, 1987, p128).

An alternative approach, which does not require the second order derivatives to be calculated or approximated, is the EM algorithm (Dempster, Laird and Rubin, 1977). This algorithm fills in the missing data Y_{mis} based on the estimate of θ, and then re-estimate θ based on Y_{obs} and filled-in Y_{mis} and iterates until the estimates converge. This intuitive idea of filling in missing data and iterating has appeared in statistical literature as far back as in 1926 by McKendrick, Hartley (1958), Baum et al. (1970), Orchard and Woodbury (1972), Sundberg (1974), Beale and Little (1975). The term EM (*Expectation-Maximization)* was introduced by Dempster, Laird and Rubin (1977). In that paper, they proved that the loglikelihood function $\ell(\theta \mid Y_{obs})$ increases at each iteration.

The theoretical background of the EM algorithm is laid out as the follows:
The probability model for complete data can be factored as

$$P(Y \mid \theta) = P(Y_{obs} \mid \theta) P(Y_{mis} \mid Y_{obs}, \theta) \tag{3.1.4}$$

and the loglikelihood can be expressed as

$$\ell(\theta \mid Y) = \ell(\theta \mid Y_{obs}) + \log P(Y_{mis} \mid Y_{obs}, \theta) \tag{3.1.5}$$

where $\ell(\theta \mid Y) = \log P(Y \mid \theta)$ is the complete data loglikelihood, $\ell(\theta \mid Y_{obs}) = \log P(Y_{obs} \mid \theta)$ is the observed-data loglikelihood. The term $P(Y_{mis} \mid Y_{obs}, \theta)$ is the conditional predictive distribution of Y_{mis} given Y_{obs} and θ. When viewed as a probability distribution of Y_{mis} it summarizes the knowledge about Y_{mis} for any given value of θ; when viewed as a function of θ it conveys the information about θ contained in Y_{mis} beyond that already provided by Y_{obs}.

The complete data loglikelihood $\ell(\theta \mid Y)$ can not be calculated directly because $P(Y_{mis} \mid Y_{obs}, \theta)$ can not be calculated due to Y_{mis} is unknown. Instead the expected loglikelihood over the distribution of missing data Y_{mis} is calculated. Given observed data Y_{obs} and current estimate of θ, say $\theta^{(t)}$,

$$Q(\theta \mid \theta^{(t)}) = \ell(\theta \mid Y_{obs}) + H(\theta \mid \theta^{(t)}) \tag{3.1.6}$$

where

$$Q(\theta \mid \theta^{(t)}) = \int \ell(\theta \mid Y) P(Y_{mis} \mid Y_{obs}, \theta^{(t)}) dY_{mis}$$

and

$$H(\theta \mid \theta^{(t)}) = \int \log P(Y_{mis} \mid Y_{obs}, \theta) P(Y_{mis} \mid Y_{obs}, \theta^{(t)}) dY_{mis}$$

Suppose $\theta^{(t+1)}$ is the value which maximizes $Q(\theta \mid \theta^{(t)})$, i.e. $Q(\theta \mid \theta^{(t)}) \leq Q(\theta^{(t+1)} \mid \theta^{(t)})$, then Dempster, Laird and Rubin (1977) showed that $\theta^{(t+1)}$ is a better estimate than $\theta^{(t)}$ in the sense that its observed-data loglikelihood is at least as high as that of $\theta^{(t)}$,

$$\ell(\theta^{(t)} \mid Y_{obs}) \leq \ell(\theta^{(t+1)} \mid Y_{obs}) \tag{3.1.7}$$

Therefore, the sequence $\{\theta^{(t)}\}$ increases the values of the observed-data loglikelihood at each iteration, and if the observed-data loglikelihood $\ell(\theta \mid Y_{obs})$ is bounded, the sequence $\left\{\ell(\theta^{(t)} \mid Y_{obs})\right\}$ converges to a stationary value of $\ell(\theta \mid Y_{obs})$.

The iteration of EM consists of two distinct steps:

The Expectation or E step: The E step is to find the conditional expectation of the missing data given the observed data and current estimate of θ, and then substitute these expectations for the missing data. Note it is not the individual missing data that is being substituted by its expectation, it is the function of missing data appearing in the complete data loglikelihood that is being substituted. Specifically, The E step is to find the expected value of the complete data loglikelihood over the distribution of missing data Y_{mis}, given observed data Y_{obs} and current estimate of θ, say $\theta^{(t)}$, i.e. $Q(\theta \mid \theta^{(t)}) = \int \ell(\theta \mid Y) P(Y_{mis} \mid Y_{obs}, \theta^{(t)}) dY_{mis}$.

The Maximization or M step: M step is to perform the maximum likelihood estimation of θ as if there were no missing data, i.e. as if the maximization is based on the complete data

loglikelihood. Thus, maximizing $Q(\theta \mid \theta^{(t)})$ uses the identical computational methods as for the maximization of complete data loglikelihood.

Alternatively performing the E and M steps beginning with a starting value $\theta^{(0)}$ generates a sequence of estimates $\{\theta^{(t)}\}$ and a sequence of observed-data loglikelihood values $\left\{\ell(\theta^{(t)} \mid Y_{obs})\right\}$. Dempster, Laird and Rubin (1977) and Wu (1983) provided conditions under which the loglikelihood sequence converges reliably to a stationary value of observed-data loglikelihood. The convergence is linear with rate proportional to the fraction of information about θ in $\ell(\theta \mid Y)$ that is observed. Quite generally, if the sequence $\{\theta^{(t)}\}$ converges, it converges to a local maximum or saddle point of $\ell(\theta \mid Y_{obs})$ (Little and Rubin, 1987, p130).

Two points need to be noted. (1) Since the EM algorithm does not require the knowledge of the second order derivatives of the observed-data loglikelihood, the asymptotic covariance matrix for $\hat{\theta}$ will not be automatically generated as the way in Newton-Raphson algorithm. (2) Although the EM algorithm performs well under general conditions, sometimes it does not converge to a unique global maximum or converges very slow. Therefore, the issues of convergence need to be carefully monitored in real applications.

3.2 Propensity Score Based Multiple Imputation

Lavori et al. (1995) proposed this propensity score based multiple imputation method. Basically, this method is a nonparametric approach in which the missing values are imputed by re-sampling of the observed values.

By defining the missingness indicator variables r_{ij} as

$$r_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is missing} \\ 0 & \text{if } y_{ij} \text{ is observed} \end{cases} \tag{3.2.1}$$

It is assumed that the missingness can be explained by a linear logistic regression model. Suppose that the values for variables $Y_1, \cdots, Y_{j-1}$ are complete, either fully observed or consisting of observed and imputed values, then the probability of observing y_{ij}, i.e. $p_{ij} = \text{Prob}\{r_{ij} = 0\}$, can be modeled by

$$\log(\frac{p_{ij}}{1-p_{ij}}) = \beta_0 + \beta_1 y_{i1} + \cdots + \beta_{j-1} y_{i,j-1} + e_{ij} \tag{3.2.2}$$

The conditional probability of observing y_{ij}, given the previous complete history, is called the propensity score s_{ij} (Lavori et al., 1995), i.e.

$$s_{ij} = \Pr\{r_{ij} = 0 \mid y_{i1}, \cdots, y_{i,j-1}\}$$

After the regression coefficients in (3.2.2) are estimated based on the indicator variable r_{ij} for the response variable Y_j and completed data for variables $Y_1, \cdots, Y_{j-1}$, each observation can be assigned an estimated propensity score.

$$\hat{s}_{ij} = \frac{\exp\{\hat{\beta}_0 + \hat{\beta}_1 y_{i1} + \cdots + \hat{\beta}_{j-1} y_{i,j-1}\}}{1 + \exp\{\hat{\beta}_0 + \hat{\beta}_1 y_{i1} + \cdots + \hat{\beta}_{j-1} y_{i,j-1}\}} \tag{3.2.3}$$

All observations are then stratified into q strata based on the quantiles of assigned propensity scores. Note that not all previous variables $Y_1, \cdots, Y_{j-1}$ are necessarily included in the model, the choice of previous variables should make the distributions of observed values and missing values of Y_j balanced within each stratum.

After each observation is stratified according to its propensity score, a 'donor pool' is created by applying the approximate Bayesian bootstrap (ABB) within each stratum. That is, a random sample is created by random draws with replacement from the observed values of Y_j within each stratum. The ABB method is applied in order to reflect the extra uncertainty about the posterior predictive distribution of missing values of Y_j, given the observed values of Y_j within each stratum. This method is roughly equivalent to choosing the parameters for the conditional posterior predictive distribution $P(Y_{mis} \mid Y_{obs}, \theta)$ from their observed-data posterior distribution $P(\theta \mid Y_{obs})$. Each missing value of Y_j is then imputed by a single random draw from its donor pool. The m sets of proper multiple imputation are obtained by creating m donor pools independently for each individual missing value and then drawing a single random sample from each donor pool. Note that imputing a missing value by a random draw from its stratum rather than from its donor pool would result in an improper multiple imputation in the sense that the between imputation variance is underestimated.

It is important to note that it is the *missingness* being modeled rather than the *missing values* being modeled. The logistic model (3.2.2) explains the relationships between the missingness indicator r_{ij} and $(Y_1, \cdots, Y_{j-1})$, it does not model the relationships between $Y_{mis,j}$ and $(Y_1, \cdots, Y_{j-1})$. This is pointed out by Schafer on his web page regarding the usage of Solas. He

commented that this method is effective for the analysis pertaining to the distribution of the missing values of Y_j, but it is not appropriate in general for the analysis involving the relationships between Y_j and $(Y_1,\cdots,Y_{j-1})$. One hypothetical example in his comment is that if some variables are highly correlated with $Y_{mis.j}$ but unrelated to the missingness indicator r_{ij}, then those variables have no influence in the logistic regression model. Therefore, the imputed values of Y_j bear no relationship to those variables, the estimate of the correlation between Y_j and those variables based on the multiple imputed data sets will be biased toward zero. In this case, propensity score based multiple imputation is unable to preserve important features of the joint distribution of Y_j and $(Y_1,\cdots,Y_{j-1})$.

3.3 Predictive Model Based Multiple imputation

In linear regression models, the values of the dependent variable can be predicted based on the given values of the covariates and the estimated regression coefficients. This is the idea for the predictive model based multiple imputation (Rubin, 1987).

The predictive model for the missing values of variable Y_j is the linear regression of Y_j on the covariates $Y_1,\cdots,Y_{j-1},Y_{j+1},\cdots,Y_p$. The missing values of Y_j are imputed by the predicted values from the linear regression model, given the observed values of $Y_1,\cdots,Y_{j-1},Y_{j+1},\cdots,Y_p$ and the simulated regression parameters which are randomly drawn from their observed-data posterior distributions. In this way, the extra uncertainty due to the fact that the regression parameters can be estimated, but not determined, from the observed values of Y_j and $Y_1,\cdots,Y_{j-1},Y_{j+1},\cdots,Y_p$ is reflected. Using estimated regression parameters rather than the simulated values from their observed-data posterior distribution will result in improper multiple imputation in the sense that the between imputation variance is underestimated (Rubin, 1987, p167).

The approaches to derive the observed-data posterior distributions for the regression coefficients and the residual variance are different depending on whether the values for the covariates are complete or incomplete. Let $Y_{obs.j}$ and $Y_{mis.j}$ be the observed values and missing values for Y_j respectively. For the simplicity, we assume that the first n_j observations of Y_j are observed and the rest of (n_j+1) to n observations are missing. Let X be the data

matrix for $Y_1,\cdots,Y_{j-1},Y_{j+1},\cdots,Y_p$ and augmented ones in the first column to incorporate the intercept. Let X_{obs} and X_{mis} be the rows of X corresponding to $Y_{obs,j}$ and $Y_{mis,j}$ respectively.

a) The values of all covariates are observed.

In this case, the probability model for Y_j given $Y_1,\cdots,Y_{j-1},Y_{j+1},\cdots,Y_p$ is an univariate normal,

$$Y_j \sim N_1(\mu_j,\sigma_j^2) \tag{3.3.1}$$

where $\mu_j = \beta_0 + \beta_1 Y_1 + \cdots + \beta_{j-1}Y_{j-1} + \beta_{j+1}Y_{j+1} + \cdots \beta_p Y_p$. The observed-data likelihood function of the regression parameters $\theta = (\beta_0,\beta_1,\cdots,\beta_{j-1},\beta_{j+1},\cdots\beta_p,\sigma_j^2)$ is

$$L(\mu_j,\sigma_j^2 \mid Y_{obs}) \propto \sigma_j^{-n_j} \exp\left\{-\frac{1}{2\sigma_j^2}\sum_{i=1}^{n_j}(y_{ij} - X_{obs(i)}\beta)^2\right\}$$

The observed-data posterior distribution of θ depends on the observed-data likelihood as well as the prior distribution for θ. If there is no strong preference on the values of θ, a non-informative prior $\pi(\theta) = \sigma_j^{-1}$ is assumed for (β,σ_j). After some manipulation, the observed-data posterior can be expressed as

$$\sigma_j^{-\frac{p}{2}} \exp\left\{-\frac{1}{2\sigma_j^2}(\beta-\hat{\beta})^T(X_{obs}^T X_{obs})(\beta-\hat{\beta})\right\} \times \sigma_j^{-(\frac{n_j-p}{2})-1} \exp\left\{-\frac{1}{2\sigma_j^2}\sum_{i=1}^{n_j}(y_{ij} - X_{obs(i)}\hat{\beta})^2\right\}$$

Which is a normal scaled inverted-chisquare distribution,

$$\beta \mid Y_{obs},\sigma_j^2 \sim N_p(\hat{\beta},\sigma_j^2(X_{obs}^T X_{obs})^T) \tag{3.3.2}$$

$$\sigma_j^2 \mid Y_{obs} \sim \hat{\varepsilon}^T\hat{\varepsilon}\,\chi_{n_j-p}^{-2} \tag{3.3.3}$$

where $\hat{\beta} = (X_{obs}^T X_{obs})^{-1} X_{obs}^T Y_{obs,j}$ is the MLE of β from the observed-data likelihood, and $\hat{\varepsilon} = Y_{obs,j} - X_{obs}\hat{\beta}$ is the residual vector.

The values of (β,σ_j^2) can be simulated from their observed-data posterior distribution (3.3.2)-(3.3.3) as follows: A random sample $\tilde{\sigma}_j^2$ is obtained by letting $\tilde{\sigma}_j^2 = \hat{\varepsilon}^T\hat{\varepsilon}/\sigma^*$, where σ^* is a random draw from $\chi_{n_j-p}^2$, and then a random sample $\tilde{\beta}$ is drawn from (3.3.2) given $\tilde{\sigma}_j^2$. After a random sample $(\tilde{\beta},\tilde{\sigma}_j^2)$ has been drawn from their observed-data posterior

distribution, the missing values of Y_j are imputed by independent random draws from the conditional predictive distribution $P(Y_{mis,j} \mid Y_{obs}, \tilde{\beta}, \tilde{\sigma}_j^2)$, which is an univariate normal distribution $N_1(X_{mis}\tilde{\beta}, \tilde{\sigma}_j^2)$. Each missing value is imputed once.

In order to obtain m sets of proper multiple imputation, m independent random samples are drawn from the observed-data posterior distribution (3.3.2)-(3.3.3), say $(\tilde{\beta}^{(k)}, \tilde{\sigma}_j^{2(k)})$ $k = 1, \cdots, m$. For each simulated regression parameters $(\tilde{\beta}^{(k)}, \tilde{\sigma}_j^{2(k)})$, the missing values of Y_j are imputed once by independent random draws from $N_1(X_{mis}\tilde{\beta}^{(k)}, \tilde{\sigma}_j^{2(k)})$.

b) The values of covariates are not all observed.

If not all values of covariates are observed, either the regression parameters (β, σ_j^2) can not be estimated, so the observed-data posterior can not be obtained, or the conditional predictive distribution $N_1(X_{mis}\tilde{\beta}, \tilde{\sigma}_j^2)$ can not be determined. In this case, the response indicators are introduced. Let c be the number of covariates with incomplete data, and let

$$r_{ik} = \begin{cases} 1 & \text{if } y_{ik} \text{ is observed} \\ 0 & \text{if } y_{ik} \text{ is missing} \end{cases}, \qquad k = 1, \cdots, c$$

then the linear model is (SOLAS v3.0: User Reference, 2001)

$$y_{ij} = \beta_0 + \beta_{01}(1 - r_{i1}) + \cdots + \beta_{0c}(1 - r_{ic}) + \beta_1 r_{i1} y_{i1} + \cdots + \beta_c r_{ic} y_{ic} + \beta_{c+1} y_{i,c+1} + \cdots + \beta_p y_p + \varepsilon_i$$

In this model, if y_{ik} is missing, the term $\beta_k r_{ik} y_{ik}$ is zero and the intercept is adjusted by the term $\beta_{0k}(1 - r_{ik})$. If y_{ik} is observed, then $\beta_k y_{ik}$ appears in the model and there is no adjustment to the intercept. Let X be the adjusted data matrix constructed as (1) the first column of X consists of 1's. (2) The second to $(c+1)^{th}$ columns of X consist of 1's and 0's, such that the i^{th} entry of the t^{th} column equals 0 when the i^{th} data entry of t^{th} covariate Y_t is observed, and equals to 1 when this data entry is missing. (3) The $(c+2)^{th}$ to $(c+p)^{th}$ columns of X consist of the values of $Y_1, \cdots, Y_{j-1}, Y_{j+1}, \cdots, Y_p$, the missing data entries of these variables will be replaced by 0. Let X_{obs} and X_{mis} be the rows of X corresponding to $Y_{obs.j}$ and $Y_{mis.j}$ respectively. The observed-data posterior distribution of $(\beta_0, \beta_{01}, \cdots, \beta_{0c}, \beta_1, \cdots, \beta_{j-1}, \beta_{j+1}, \cdots \beta_p, \sigma_j^2)$ is given by

$$\beta \mid Y_{obs}, \sigma_j^2 \sim N_{c+p}(\hat{\beta}, \sigma_j^2 (X_{obs}^T X_{obs})^T) \tag{3.3.4}$$

$$\sigma_j^2 \mid Y_{obs} \sim \hat{\varepsilon}^T \hat{\varepsilon} \chi_{n_j-c-p}^{-2} \tag{3.3.5}$$

To get m sets of proper multiple imputation for the missing values of Y_j, we take m independent random samples from the observed-data posterior distribution (3.3.4)-(3.3.5), say $(\tilde{\beta}^{(k)}, \tilde{\sigma}_j^{2(k)})$ $k = 1, \cdots, m$. For each $(\tilde{\beta}^{(k)}, \tilde{\sigma}_j^{2(k)})$, the missing values of Y_j are imputed once by independent random draws from $N_1(X_{mis}\tilde{\beta}^{(k)}, \tilde{\sigma}_j^{2(k)})$.

3.4 Markov Chain Monte Carlo Based Multiple imputation

Markov chain Monte Carlo (MCMC) method is Monte Carlo integration using Markov chains. This method has been successfully applied in a large number of statistical situations. Most of these applications are oriented towards Bayesian inferences. For a comprehensive overview including the theory and the application, please see Gilks, Richardson and Spiegelhalter (1996).

Given the observed data y, the posterior distribution of θ is

$$P(\theta \mid y) = \frac{P(y \mid \theta) P(\theta)}{\int P(y \mid \theta) P(\theta) d\theta} \tag{3.4.1}$$

This posterior distribution is the object of all Bayesian inference. The moments, quantiles, highest density regions can all be expressed in terms of posterior expectations. The posterior expectation of a function $f(\theta)$ is

$$E\{f(\theta) \mid y\} = \int f(\theta) P(\theta \mid y) d\theta \tag{3.4.2}$$

The integrations in this expression have been the most difficult part in Bayesian inference, especially in high dimensions.

Monte Carlo integration:

Drawing samples $\{\theta^{(t)} : t = 1, \cdots, n\}$ from $P(\theta \mid y)$ and then approximating the integration $E\{f(\theta) \mid y\}$ by

$$E\{f(\theta) \mid y\} \approx \frac{1}{n} \sum_{t=1}^{n} f(\theta^{(t)}) \tag{3.4.3}$$

In this way, the population mean of $P(\theta \mid y)$ is estimated by a sample mean. If the samples $\{\theta^{(t)}\}$ are independent, the accuracy can be improved by increasing the sample size. However, $\{\theta^{(t)}\}$ need not necessarily be independent. The $\{\theta^{(t)}\}$ can be generated by any process. One way of doing this is through Markov chain which has $P(\theta \mid y)$ as its stationary distribution.

Markov chains:

Suppose we generate a sequence of random variables $\{X_0, X_1, X_2, \cdots\}$ such that at each time $t \geq 0$, the next state X_{t+1} is sampled from a distribution $P(X_{t+1} \mid X_t)$ which depends only on the current state of the chain, X_t. This sequence of random variables is called a Markov chain, and the distribution $P(.|.)$ is called the *transition kernel* of the chain. It is assumed that the chain is time-homogenous, that is, $P(.|.)$ does not depend on t (Gilks, Richardson and Spiegelhalter, 1996, p5).

Subject to regularity conditions, the chain will gradually 'forget' its initial state and will eventually converge to a unique stationary (or invariant) distribution which does not depend on t or X_0. We denote the stationary distribution by $P(.)$. Thus, after a sufficiently long *burn-in* period, say k iterations, the sequence $\{X_t : t = k+1, \cdots, n\}$ will be *dependent* samples from a distribution which is an approximation of $P(.)$.

Therefore, to estimate the expectation $E\{f(\theta) \mid y\}$, we can use the output from a Markov chain $\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(n)}\}$, where the stationary distribution of this chain is the posterior distribution of θ. The burn-in sample are usually discarded for this calculation, giving an estimation

$$E\{f(\theta) \mid y\} = \frac{1}{n-k} \sum_{t=k+1}^{n} f(\theta^{(t)}) \qquad (3.4.4)$$

This is called an *ergodic average*. Convergence to the required expectation is ensured by the ergodic theorem. For more details, please see Roberts (1995) and Tierney (1995).

There is a number of ways to generate Markov chains, but all of them, including the Gibbs sampler (Geman and Geman, 1984), are special cases of the general framework of Metropolis et al. (1953) and Hastings (1970). There is also a number of ways to evaluate the convergence, for a recent review, please see Cowles and Carlin (1996).

As mentioned above, in most cases the MCMC is applied to estimate the posterior expectation. Schafer (1997) applied this method slightly different from this purpose. He applied the MCMC for the purpose of multiple imputation by utilizing the data augmentation algorithm developed by Tanner and Wong (1987). A multiple imputation which can yield valid inference is called a *proper* multiple imputation (Rubin, 1987). In order to apply the MCMC to generate multiple imputation which can yield valid inference, Schafer (1997, p105) gave a slightly different definition of proper multiple imputation: A multiple imputation is called *Bayesianly proper* if the imputations are *independent* drawn from the posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$, under some completed-data model $P(Y_{mis}, Y_{obs} \mid \theta)$ and prior $P(\theta)$.

If the missingness mechanism is ignorable, the response variables Y can be described by the probability model $P(Y \mid \theta)$. If data are incomplete, the analytical form of this model is $P(Y_{obs} \mid \theta)$. The likelihood-based estimation and inference about θ can be made from this function, like the EM algorithm. Since Y_{mis} and θ are unknown, from Bayesian point view, they both are treated as random variables, the joint conditional distribution of Y_{mis} and θ given Y_{obs} is

$$P(Y_{mis}, \theta \mid Y_{obs}) \tag{3.4.5}$$

If this distribution is tractable or can be easily simulated, then Y_{mis} and θ can be simultaneously drawn from this distribution. Unfortunately, this distribution is generally intractable and can not be easily simulated. Therefore, we look at the marginal distribution of (3.4.5)

$$P(Y_{mis} \mid Y_{obs}) = \int P(Y_{mis} \mid Y_{obs}, \theta) P(\theta \mid Y_{obs}) d\theta \tag{3.4.6}$$

This distribution is the posterior predictive distribution of Y_{mis} given Y_{obs}. $P(Y_{mis} \mid Y_{obs}, \theta)$ is the conditional predictive distribution given Y_{obs} and θ, and $P(\theta \mid Y_{obs})$ is the observed-data posterior distribution of θ.

Note that the posterior predictive distribution (3.4.6) is the conditional predictive distribution $P(Y_{mis} \mid Y_{obs}, \theta)$ averaged over the observed-data posterior distribution of θ. The conditional predictive distribution $P(Y_{mis} \mid Y_{obs}, \theta)$ is usually analytically tractable once the probability model for the response variables is specified. For example, if the probability model for the response variables is a multivariate normal, then the conditional predictive distribution of Y_{mis} given Y_{obs} and θ is a multivariate normal or an univariate normal. However, the observed-data

posterior distribution $P(\theta \mid Y_{obs})$ is not always tractable, it could be a very complicated distribution. In this case, the posterior predictive distribution (3.4.6) is not tractable and can not be easily simulated. Even if the observed-data posterior is tractable, the integration in (3.4.6) will not guarantee the posterior predictive distribution tractable. Therefore, imputing Y_{mis} by random draws from the posterior predictive distribution could be a very difficult task.

The multiple imputation by the MCMC is conducted as follows (Schafer, 1997, p72). Augment the missing data Y_{mis} by some assumed values, then θ can be easily simulated from the resulting completed-data posterior distribution $P(\theta \mid Y_{obs}, Y_{mis})$. Let $\theta^{(t)}$ be the current simulated value of θ, then the next iterative sample of Y_{mis}, $Y_{mis}^{(t+1)}$, can be drawn from the conditional predictive distribution of Y_{mis} given Y_{obs} and $\theta^{(t)}$, i.e.

$$Y_{mis}^{(t+1)} \sim P(Y_{mis} \mid Y_{obs}, \theta^{(t)}) \tag{3.4.7}$$

conditioning on $Y_{mis}^{(t+1)}$, the next iterative simulation of θ can be drawn from its completed-data posterior distribution,

$$\theta^{(t+1)} \sim P(\theta \mid Y_{obs}, Y_{mis}^{(t+1)}) \tag{3.4.8}$$

Repeating the random draws of (3.4.7)-(3.4.8) from a starting value of $\theta^{(0)}$ yields a stochastic sequence $\left\{(\theta^{(t)}, Y_{mis}^{(t)}) : t = 1,2,\cdots\right\}$. The stationary distribution of this sequence is the joint distribution of θ and Y_{mis} given Y_{obs}, $P(Y_{mis}, \theta \mid Y_{obs})$. Consequently, the marginal stationary distributions of the subsequence $\{\theta^{(t)} : t = 1,2,\cdots\}$ and $\{Y_{mis}^{(t)} : t = 1,2,\cdots\}$ are the observed-data posterior distribution $P(\theta \mid Y_{obs})$ and the posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$ respectively. When t is sufficiently large, $\theta^{(t)}$ can be viewed as a single simulation from the approximate observed-data posterior distribution $P(\theta \mid Y_{obs})$, and $Y_{mis}^{(t)}$ can be viewed as a single imputation from the approximate posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$. The random draw of (3.4.7) is to impute the missing data Y_{mis}, and the random draw of (3.4.8) is to simulate the unknown parameters θ, therefore (3.4.7) and (3.4.8) are referred as the Imputation or I-step and the Posterior or P-step respectively (Tanner and Wong, 1987). The first use of this algorithm seems to have been made by Li (1988) who presented an argument for convergence and used it to impute the missing data Y_{mis}.

Proper multiple imputation, which can yield valid inference, requires the multiplely imputed Y_{mis} to be independent. Therefore, even after a long period of burn-in, the multiple imputation can not be done by successive iterates from the stationary distribution because the successive iterates tend to be correlated. One way to get proper multiple imputation is to take subsample from the chain. For example, take every k^{th} iterate after the burn-in period, where k is chosen large enough so that the dependence between the imputed values can be neglected. Alternatively, we can generate m independent chains, after the burn-in period, take the final values of each chain as the imputed values of Y_{mis} (Schafer, 1997, p106).

It is pointed out by Schafer (1997, p73) that the imputation by the MCMC bears a strong resemblance to the EM algorithm. In the EM algorithm, the missing sufficient statistics are filled in by their expected values and then the maximization is carried out on the filled in completed-data loglikelihood. The procedure is repeated until the estimates of the parameters converge to the stationary point, which is the MLE of the parameters based on the observed-data loglikehood. In the MCMC, each individual missing value is filled in by random draws and then the current simulated parameters based on the filled in completed-data are updated. the whole process is repeated until the convergence is reached.

The convergence of the MCMC is more complicated than the convergence of the EM algorithm. In EM the estimates sequence $\{\theta^{(t)}\}$ converges to a point estimates $\hat{\theta}$. That is, the convergence is a *point* convergence. When the EM sequence converges, the estimates of θ no longer change from one iteration to the next. In the MCMC, the convergence of the Markov chain $\left\{(\theta^{(t)}, Y_{mis}^{(t)})\right\}$ is a *distribution* convergence. When the MCMC converges, the distribution of (θ, Y_{mis}) no longer change from one iteration to the next, but the random draws from this distribution do continuously change.

The MCMC method avoids the complicated calculation of the observed-data posterior distribution of the unknown parameters θ and the posterior predictive distribution of the missing data Y_{mis}. But as for any iterative methodology, convergence needs to be carefully monitored, especially when there is a large amount of missing data. Other than that, the MCMC is quite successful for multiple imputation and can be used for almost any kind of incomplete data.

3.5 Summary

In this chapter, four methods of handling incomplete data are reviewed. The EM algorithm is a likelihood-based method. The interdependence between the missing data Y_{mis} and the parameters θ is utilized to estimate θ. The most attractive features of the EM relative to other likelihood-based methods are its simplicity and its stability (Schafer, 1997, p51). The propensity score based multiple imputation is a nonparametric method. The propensity score of each observation is estimated by a linear logistic model. Observations with similar estimated propensity scores are grouped, the approximate Bayesian bootstrap is applied within each group to impute the missing data by observed ones. The predictive model based multiple imputation is an univariate approach. Given the values of covariates, the response variable is modeled by a linear regression model. The regression parameters are simulated from their observed-data posterior distribution rather than estimated from the regression model. The missing data are imputed by the predictive values from the regression model given the simulated regression parameters and the values of covariates. The MCMC method for multiple imputation mentioned in section 3.4 is little bit different from the MCMC method for the Bayesian inference. In the following chapters, the MCMC is referred as the MCMC method for multiple imputation. This method is a multivariate approach. In order to have *proper* multiple imputation, (1) a sufficiently large numbers of iterates must be burned so that the marginal stationary distribution of the subsequence $\{Y_{mis}^{(t)} : t = 1,2,\cdots\}$ is the approximate posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$, and (2) imputed values of Y_{mis} must be independently drawn from this approximate posterior predictive distribution (Schafer, 1997).

The following software packages have implemented the multiple imputation methods discussed in this chapter. For a complete and most recent review of the multiple imputation software packages, please see Horton and Lipsitz (2001).

Solas v3.0:	The propensity score method and the predictive model method.
BMDP Professional 2.0:	The predictive model method.
SAS v8.2:	The propensity score method and the predictive model method for monotone missing pattern. The MCMC method for arbitrary missing pattern.
NORM:	The MCMC method. (free at http://www.stat.psu.edu/~jls)

Chapter 4. Multivariate Normal Distribution

The multivariate normal distribution is often a useful and reasonable probability model for continuous response variables, especially for repeated measurement data. In many situations, although the response variables are not normally distributed, the normal distribution still can be used as an approximation to the underlying distribution. The broad application of the multivariate normal distribution is due to the central limit theorem and its good analytical properties. In this chapter, some of these properties, such as the alternative parameterization, the likelihood function and the posterior distribution are reviewed. These properties will be utilized when establishing an imputation model later. For a thorough discussion of the multivariate normal distribution, see Anderson (1984). In section 4.1, the derivation of the likelihood function and the posterior distribution are discussed when data are complete. In section 4.2 the discussion is focused on these properties when data are incomplete. One particular missing pattern is assumed for the incomplete data, that is, the monotone missing pattern. Section 4.3 presents a very useful tool, which can be used to perform the transformation between the normal parameters and the alternative parameters, the sweep operator. A brief summary is given in section 4.4.

4.1 Complete Data

Let Y be a $n \times p$ data matrix, $Y = (y_1, y_2, \cdots, y_n)^T$. where $y_i = (y_{i1}, \cdots, y_{ip})^T$ is a random sample from a p-dimensional multivariate normal distribution,

$$y_i \sim N_p(\mu, \Sigma)$$

with the density function

$$f(y_i \mid \mu, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y_i - \mu)^T \Sigma^{-1} (y_i - \mu)\right\} \qquad (4.1.1)$$

4.1.1 Likelihood Function

When $y_1, y_2, \cdots, y_n$ are iid samples from (4.1.1) and have no missing values, the complete-data likelihood function is given by

$$L(\mu, \Sigma \mid Y) \propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (y_i - \mu)^T \Sigma^{-1} (y_i - \mu)\right\} \qquad (4.1.2)$$

By some algebraic manipulation, the complete-data loglikelihood function of μ and Σ can be written as

$$l(\mu,\Sigma|Y) = -\frac{n}{2}\log|\Sigma| - \frac{n}{2}\mu^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}T_1 - \frac{1}{2}tr\Sigma^{-1}T_2$$

where $$T_1 = \sum_{i=1}^{n} y_i = Y^T 1_n$$

and $$T_2 = \sum_{i=1}^{n} y_i y_i^T = Y^T Y$$

are the sufficient statistics for μ and Σ respectively. The maximum likelihood estimates of μ and Σ can be obtained by equating the sampled values of T_1 and T_2 to their expected values $E(T_1) = n\mu$ and $E(T_2) = n(\Sigma + \mu\mu^T)$, i.e.

$$\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{4.1.3}$$

$$\hat{\Sigma} = S = \frac{1}{n} y^T y - \bar{y}\ \bar{y}^T = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})(y_i-\bar{y})^T \tag{4.1.4}$$

Thus when data are complete, the maximum likelihood estimates of unknown parameters μ and Σ can be obtained by a non-iterative straightforward method. $\hat{\mu}$ is an unbiased estimate of μ and S is a biased estimate of Σ, the unbiased estimate of Σ can be obtained by $n(n-1)^{-1}S$.

4.1.2 Posterior Distribution

The posterior distribution of the parameters depends on the assumption about the response variables as well as the assumption about the prior distribution of the parameters. When no strong preference is available about the parameters $\theta = (\mu,\Sigma)$, a convenient non-informative prior is assumed,

$$\pi(\theta) \propto |\Sigma|^{-(\frac{p+1}{2})} \tag{4.1.5}$$

This prior is the limiting form of the following normal inverted-Wishart distribution as $\tau \to 0$, $m \to -1$ and $\Lambda^{-1} \to 0$ (Schafer, 1997, p154).

$$\mu|\Sigma \sim N(\mu_0, \tau^{-1}\Sigma)$$

$$\Sigma \sim W^{-1}(m,\Lambda)$$

Note that μ does not appear on the right-hand side of (4.1.5). In this case the prior distribution of μ is assumed to be multivariate uniform over the p-dimensional real space.

Under this non-informative prior, the complete-data posterior distribution of $\theta=(\mu,\Sigma)$ is a normal inverted-Wishart distribution with the density function

$$P(\theta\,|\,Y) \propto |\Sigma|^{-(\frac{n+p+1}{2})} \exp\left\{-\frac{n}{2}\mathrm{tr}\Sigma^{-1}S\right\}\times\exp\left\{-\frac{n}{2}(\mu-\bar{y})^T\Sigma^{-1}(\mu-\bar{y})\right\}$$

The marginal complete-data posterior distribution of Σ is a p-dimensional inverted-Wishart distribution with the degrees of freedom $n-1$, the conditional complete-data posterior of μ given Σ is a p-dimensional multivariate normal, i.e.

$$\mu\,|\,\Sigma,Y \sim N_p(\bar{y},n^{-1}\Sigma) \tag{4.1.6}$$

$$\Sigma\,|\,Y \sim W_p^{-1}(n-1,(nS)^{-1}) \tag{4.1.7}$$

Since (4.1.6) and (4.1.7) both have closed forms and have good analytical properties, it has been shown when data are complete, the posterior distribution of the unknown parameters can be easily summarized analytically, or the values of parameters can be easily simulated.

It is noted that the non-informative prior (4.1.5) also arises by applying the Jeffreys invariance principle to $\theta=(\mu,\Sigma)$, see Box and Tiao (1973). In this case, the Bayesian estimates of μ and the maximum likelihood estimates of μ are the same. $E(\mu\,|\,Y)=MLE(\mu)=\bar{y}$ (Schafer, 1997, p155).

4.1.3 Alternative Parameterization

Let the multivariate vector $Y=(Y_1,\cdots Y_p)^T$ be partitioned as $Y=(Z_1,Z_2)^T$, where Z_1 and Z_2 are subvectors of length p_1 and $p_2=p-p_1$ respectively. The marginal distribution of Z_1 and Z_2 are $N(\mu_1,\Sigma_{11})$ and $N(\mu_2,\Sigma_{22})$ respectively, where $\mu=(\mu_1,\mu_2)^T$ and $\Sigma=\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$ are the partition of μ and Σ corresponding to the partition $Y=(Z_1,Z_2)^T$.

The conditional distribution of Z_2 given Z_1 is also a normal distribution with mean $\mu_{2\cdot1}$ and covariance matrix $\Sigma_{22\cdot1}$. The conditional mean vector $\mu_{2\cdot1}$ and the conditional covariance matrix $\Sigma_{22\cdot1}$ are given by

$$\mu_{2\cdot1} = \mu_2 + B_{2\cdot1}(Z_1 - \mu_1) = \alpha_{2\cdot1} + B_{2\cdot1}Z_1$$

$$\Sigma_{22\cdot1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

where $\alpha_{2\cdot1} = \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1$ and $B_{2\cdot1} = \Sigma_{21}\Sigma_{11}^{-1}$ are the vector of intercepts and the matrix of slopes from the linear regression of Z_2 on Z_1, while $\Sigma_{22\cdot1}$ is the residual covariance matrix. Since $P(Z_1, Z_2) = P(Z_1)P(Z_2 \mid Z_1)$, we can always characterize the distribution $P(Z_1, Z_2)$ either by the *normal parameters* $\theta = (\mu, \Sigma)$ or by the *alternative parameters* $\phi = (\phi_1, \phi_2)$, where $\phi_1 = (\mu_1, \Sigma_{11})$ and $\phi_2 = (\alpha_{2\cdot1}, B_{2\cdot1}, \Sigma_{22\cdot1})$. The conversion of ϕ from θ is a one-to-one transformation $\phi = g(\theta)$, and the reverse transformation $\theta = g^{-1}(\phi)$ is given by

$$\mu_2 = \alpha_{2\cdot1} + B_{2\cdot1}\mu_1$$

$$\Sigma_{12} = \Sigma_{11}B_{2\cdot1}^T$$

$$\Sigma_{22} = \Sigma_{22\cdot1} + B_{2\cdot1}\Sigma_{11}B_{2\cdot1}^T$$

The alternative parameters ϕ_1 and ϕ_2 are distinct in the sense that the parameter space of ϕ is the Cartesian product of the individual parameter spaces of ϕ_1 and ϕ_2. That is, any choice of $\alpha_{2\cdot1}$, $B_{2\cdot1}$ and $\Sigma_{22\cdot1} > 0$ (positive definite) will yield a valid normal parameter $\theta = (\mu, \Sigma)$ with $\Sigma > 0$ (Anderson, 1984).

4.2 Incomplete Data

When the data matrix Y is incomplete, we use the notation Y_{obs} and Y_{mis} to represent the observed and missing portion of Y respectively, and use $y_{i(obs)}$ and $y_{i(mis)}$ to denote the observed and missing portion of i^{th} observation respectively. Therefore, $Y = (Y_{obs}, Y_{mis})$ and $y_i = (y_{i(obs)}^T, y_{i(mis)}^T)^T$. When missingness mechanism is ignorable, the observed-data likelihood function is the complete-data likelihood function (4.1.2) integrated over Y_{mis}.

$$L(\mu, \Sigma \mid Y_{obs}) \propto \int |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (y_i - \mu)^T \Sigma^{-1}(y_i - \mu)\right\} dY_{mis}$$

$$= \prod_{i=1}^{n} |\Sigma_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_{i(obs)} - \mu_i)^T \Sigma_i^{-1}(y_{i(obs)} - \mu_i)\right\} \qquad (4.2.1)$$

where μ_i and Σ_i are the marginal mean vector and marginal covariance matrix corresponding to $y_{i(obs)}$ respectively. Generally, this observed-data likelihood function has a complicated functional format, and the maximum likelihood estimation of $\theta = (\mu, \Sigma)$ will require some iterative algorithm. However, if the incomplete data have a special pattern, (4.2.1) can be simplified and the maximum likelihood estimation of $\theta = (\mu, \Sigma)$ can be obtained by a non-iterative method.

Monotone missing pattern is defined as follow. A multivariate observation $y_i = (y_{i1}, y_{i2}, \cdots, y_{ip})^T$ is said to be monotone if whenever y_{ij} is missing, then y_{ik} is also missing for all $k > j$ (Anderson, 1957; Rubin, 1974; Little and Rubin, 1987). Monotone missing pattern is often seen in clinical trials with repeated measurements. A subject may prematurely discontinue the trial at any time during the trial, then all the subsequent measurements on this subject after this time will be missing. Let n_j denote the number of observed values of variable Y_j, then when missing data follow a monotone pattern, we have $n = n_1 \geq n_2 \geq \cdots \geq n_{p-1} \geq n_p$. The data matrix with a monotone missing pattern will look like the following:

		Y_1	Y_2	Y_3	...	Y_p
Rows	1	y_{11}	y_{12}	y_{13}		y_{1p}
	2	y_{21}	y_{22}	y_{23}		y_{2p}
	3	y_{31}	y_{32}	y_{33}		
	4	y_{41}	y_{42}	y_{43}		
	5	y_{51}	y_{52}			
	6	y_{61}	y_{62}			
	.					
	.					
	.					
	$n-1$	$y_{n-1,1}$				
	n	$y_{n,1}$				

4.2.1 Factorization of Likelihood Function

It is well known that a multivariate distribution can always be factored by a series of independent conditional distributions, $P(Y_1, \cdots, Y_p) = P(Y_1)P(Y_2 \mid Y_1) \cdots P(Y_p \mid Y_1, \cdots, Y_{p-1})$. For the multivariate normal distribution, this factorization has more practical meaning since each

factor is an univariate normal distribution. Let $\theta=(\mu,\Sigma)$ be the normal parameters, and $\phi=(\phi_1,\phi_2,\cdots,\phi_p)$ be the alternative parameters, where ϕ_j contains the intercept, the slopes and the residual variance from the linear regression of Y_j on $Y_1,Y_2,\cdots,Y_{j-1}$. That is, ϕ_j contains the parameters from the conditional distribution of Y_j given $Y_1,Y_2,\cdots,Y_{j-1}$,

$$P(Y_1,\cdots,Y_p\mid\theta)=P(Y_1\mid\phi_1)P(Y_2\mid Y_1,\phi_2)\cdots P(Y_p\mid Y_1,\cdots,Y_{p-1},\phi_p) \tag{4.2.2}$$

On the left-hand side of (4.2.2), the multivariate normal distribution is characterized by the normal parameters θ, and on the right-hand side of (4.2.2), it is characterized by the alternative parameters ϕ.

When missing data follow a monotone pattern and the missingness mechanism is ignorable, the likelihood function of θ can be factored into the independent likelihood functions of $\phi_1,\phi_2,\cdots,\phi_p$. The complete-data likelihood function is

$$\begin{aligned}L(\phi\mid y)&=\prod_{i=1}^{n}P(y_{i1},\cdots,y_{ip}\mid\theta)\\&=\prod_{i=1}^{n}\prod_{j=1}^{p}P(y_{ij}\mid y_{i1},\cdots,y_{i,j-1},\phi_j)\\&=\prod_{j=1}^{p}\prod_{i=1}^{n}P(y_{ij}\mid y_{i1},\cdots,y_{i,j-1},\phi_j)\end{aligned} \tag{4.2.3}$$

and the observed-data likelihood function, by definition, is the integral of complete-data likelihood function over Y_{mis}, i.e.

$$L(\phi\mid Y_{obs})=\int L(\phi\mid Y_{obs},Y_{mis})dY_{mis}$$

The inner product in (4.2.3) can be expressed as

$$\prod_{i=1}^{n}P(y_{ij}\mid y_{i1},\cdots,y_{i,j-1},\phi_j)=\prod_{i=1}^{n_j}P(y_{ij}\mid y_{i1},\cdots,y_{i,j-1},\phi_j)\prod_{i=n_j+1}^{n}P(y_{ij}\mid y_{i1},\cdots,y_{i,j-1},\phi_j)$$

The first product of this expression is the observed part of variable Y_j, while the second product is the missing part of Y_j. The integral of the second product over Y_{mis} is unity (Schafer, 1997, p220). Therefore, the observed-data likelihood function in terms of $\phi_1,\phi_2,\cdots,\phi_p$ is

$$L(\phi \mid Y_{obs}) = \prod_{j=1}^{p} L(\phi_j \mid Y_{obs}) \tag{4.2.4}$$

where

$$L(\phi_j \mid Y_{obs}) = \prod_{i=1}^{n_j} P(y_{ij} \mid y_{i1}, \cdots, y_{i,j-1}, \phi_j) \tag{4.2.5}$$

is the likelihood function for the normal linear regression of Y_j on $Y_1, Y_2, \cdots, Y_{j-1}$ based on the rows $1, \cdots, n_j$ of the data matrix. Thus the observed-data likelihood (4.2.4) is factored into a sequence of complete-data regressions over the subsets of the data matrix (Little and Rubin, 1987, p107).

4.2.2 Maximum Likelihood Estimation

The factorization of observed-data likelihood (4.2.4) enables us to make maximum likelihood estimation of ϕ without iteration (Little and Rubin, 1987, p108). Since the parameters $\phi_1, \phi_2, \cdots, \phi_p$ are distinct, maximizing $L(\phi \mid Y_{obs})$ can be realized by maximizing each factor $L(\phi_j \mid Y_{obs})$ separately for $j = 1, \cdots, p$. Let the maximum likelihood estimation of ϕ be $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p)$, then since the maximum likelihood estimation is invariant under the transformation of parameters, the maximum likelihood estimation of θ is $\hat{\theta} = g^{-1}(\hat{\phi})$.

The maximization of $L(\phi_j \mid Y_{obs})$ is accomplished by ordinary least squares regression of Y_j on $Y_1, \cdots, Y_{j-1}$ based on the first n_j rows of the data matrix. Let z_j denote the observed data of variable Y_j, $z_j = (y_{1j}, y_{2j}, \cdots, y_{n_j,j})^T$, and let X_j denote the upper-left $n_j \times (j-1)$ submatrix of Y, augmented by a column of ones for the intercept.

$$X_j = \begin{bmatrix} 1 & y_{11} & y_{12} & \cdots & y_{1,j-1} \\ 1 & y_{21} & y_{22} & \cdots & y_{2,j-1} \\ . & . & . & \cdots & . \\ 1 & y_{n_j,1} & y_{n_j,2} & \cdots & y_{n_j,j-1} \end{bmatrix} \tag{4.2.6}$$

the conditional distribution of z_j given X_j and ϕ_j is

$$z_j \mid X_j, \phi_j \sim N(X_j \beta_j, \delta_j I)$$

where $\phi_j = (\beta_j^T, \delta_j)^T$, β_j are the regression coefficients of Y_j on $Y_1, \cdots, Y_{j-1}$, and δ_j is the residual variance.

Using the properties of normal linear regression model, the maximum likelihood estimation of $\phi_j = (\beta_j^T, \delta_j)^T$ are given by

$$\hat{\beta}_j = (X_j^T X_j)^{-1} X_j^T z_j \tag{4.2.7}$$

$$\hat{\delta}_j = \frac{1}{n_j}(z_j - X_j\hat{\beta}_j)^T(z_j - X_j\hat{\beta}_j) \tag{4.2.8}$$

Calculating (4.2.7) and (4.2.8) for $j = 1, \cdots, p$ separately, we obtain the maximum likelihood estimation of the alternative parameters $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p)$. By performing the reverse transformation, we obtain the maximum likelihood estimation of the normal parameters θ as $\hat{\theta} = g^{-1}(\hat{\phi})$. Note that the maximum likelihood estimation of $\theta = (\mu, \Sigma)$ obtained through (4.2.7) and (4.2.8) are identical to that obtained through (4.1.3) and (4.1.4) when data are complete. Therefore, it has been shown that when data are complete or when incomplete data follow a monotone missing pattern, the maximum likelihood estimation of θ can be obtained by using a non-iterative algorithm.

4.2.3 Observed-data Posterior Distribution

When data are complete, the posterior distribution of θ are derived in (4.1.6) and (4.1.7). This complete-data posterior distribution has a closed form with attractive analytical properties. When data are incomplete, to derive the observed-data posterior is no longer an easy task and the observed-data posterior distribution could be very complicated. Tanner and Wong (1987) developed a method for calculating the observed-data posterior distribution by using data augmentation. However, if the incomplete data have a monotone missing pattern, the observed-data posterior distribution still can be derived in a closed form with good analytical properties.

Schafer (1997, p224) discussed the derivation of the observed-data posterior distribution. When incomplete data follow a monotone missing pattern and the missingness mechanism can be ignored, the observed-data likelihood function of θ can be factored into *independent* observed-data likelihood functions of ϕ. The observed-data posterior distributions of ϕ, therefore, can be derived from these independent observed-data likelihood functions. In order to do so, we need to assume a joint prior distribution for $\phi = (\phi_1, \phi_2, \cdots, \phi_p)$. As in section 4.1, a convenient non-

informative prior distribution of $\theta = (\mu, \Sigma)$ is assumed, $\pi_\theta(\theta) \propto |\Sigma|^{-(p+1)/2}$. The joint prior distribution for $\phi = (\phi_1, \phi_2, \cdots, \phi_p)$, through the transformation $\phi = g(\theta)$, is given by

$$\pi_\phi(\phi) = \pi_\theta(g^{-1}(\phi))\|J\|^{-1} \tag{4.2.9}$$

where J is a Jacobian matrix of the transformation $\phi = g(\theta)$, $\|J\|$ is the absolute value of the determinant of J. Using a property of determinants, $|\Sigma|$ can be partitioned as

$$|\Sigma| = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix} = |\Sigma_{11}||\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|$$

where Σ_{11} and Σ_{22} are any square submatrices of Σ. Notice that $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ is the conditional covariance matrix of the variables corresponding to Σ_{22} given the variables corresponding to Σ_{11}. If we take $\Sigma_{22} = \sigma_{pp}$, the variance of Y_p, the determinant of Σ becomes

$$|\Sigma| = |\Sigma_{11}|\delta_p \tag{4.2.10}$$

where Σ_{11} is the Σ without the last row and last column. Applying (4.2.10) recursively to Σ_{11} leads to

$$|\Sigma| = \prod_{j=1}^{p} \delta_j \tag{4.2.11}$$

where δ_j is the regression residual variance of Y_j on $Y_1, \cdots, Y_{j-1}$, $j = 1,2,\cdots,p$.

In order to find the prior distribution $\pi_\phi(\phi)$, we also need to evaluate $\|J\|$. Suppose that we first transform θ to the intermediate parameter (ξ_{p-1}, ϕ_p), where ξ_{p-1} represents the portions of μ and Σ pertaining to the marginal distribution of $Y_1, \cdots, Y_{p-1}$, and ϕ_p pertaining to the regression of Y_p on $Y_1, \cdots, Y_{p-1}$. Then the determinant of the Jacobian for the transformation of θ to (ξ_{p-1}, ϕ_p) is $|\Sigma_{11}|^{-1}$, where Σ_{11} is the marginal covariance matrix of $Y_1, \cdots, Y_{p-1}$. From (4.2.10), we know that $|\Sigma_{11}| = \delta_1\delta_2\cdots\delta_{p-1}$, so the determinant of the Jacobian for this intermediate transformation is $(\delta_1\delta_2\cdots\delta_{p-1})^{-1}$. If we then transform ξ_{p-1} to (ξ_{p-2}, ϕ_{p-1}), where ξ_{p-2} contains the portions of μ and Σ pertaining to $Y_1, \cdots, Y_{p-2}$, and ϕ_{p-1} pertains to the regression of Y_{p-1} on $Y_1, \cdots, Y_{p-2}$. The determinant of the Jacobian for this transformation

is $(\delta_1\delta_2\cdots\delta_{p-2})^{-1}$. We can repeat this procedure until we have reached the final parameterization $\phi=(\phi_1,\phi_2,\cdots,\phi_p)$, and the determinant of the Jacobian for the transformation $\phi=g(\theta)$ is the product of the determinants for each of the intermediate transformations, i.e.

$$\|J\| = \delta_1^{-(p-1)}\delta_2^{-(p-2)}\cdots\delta_{p-1}^{-1} \tag{4.2.12}$$

Therefore, under the assumption of non-informative prior $\pi_\theta(\theta)\propto|\Sigma|^{-(p+1)/2}$, the non-informative prior for $\phi=g(\theta)$ is given by

$$\pi_\phi(\phi) \propto \prod_{j=1}^{p}\delta_j^{-(\frac{p+1}{2}-p+j)} \tag{4.2.13}$$

The observed-data posterior distribution of ϕ then can be factorized by each individual observed-data posterior for ϕ_j.

$$P(\phi\,|\,Y_{obs}) \propto \prod_{j=1}^{p}L(\phi_j\,|\,Y_{obs})\delta_j^{-(\frac{p+1}{2}-p+j)}$$

$$= \prod_{j=1}^{p}P(\phi_j\,|\,Y_{obs})$$

where the observed-data posterior distribution of ϕ_j is

$$L(\phi_j\,|\,Y_{obs})\delta_j^{-(\frac{p+1}{2}-p+j)} \propto \delta_j^{-(\frac{n_j}{2})}\exp\left\{-\frac{1}{2\delta_j}(z_j-X_j\beta_j)^T(z_j-X_j\beta_j)\right\}\times\delta_j^{-(\frac{p+1}{2}-p+j)}$$

$$=\delta_j^{-\frac{j}{2}}\exp\left\{-\frac{1}{2\delta_j}(\beta_j-\hat{\beta}_j)^T X_j^T X_j(\beta_j-\hat{\beta}_j)\right\}\times\delta_j^{-((n_j-p+j-1)/2)-1}\exp\left\{-\frac{1}{2\delta_j}\hat{\varepsilon}_j{}^T\hat{\varepsilon}_j\right\}$$

which is the product of a conditional multivariate normal density and a marginal scaled inverted-chisquare density,

$$\beta_j\,|\,Y_{obs},\delta_j \sim N_j(\hat{\beta}_j,\delta_j(X_j^TX_j)^{-1}) \tag{4.2.14}$$

$$\delta_j\,|\,Y_{obs} \sim \hat{\varepsilon}_j^T\hat{\varepsilon}_j\chi^{-2}_{n_j-p+j-1} \tag{4.2.15}$$

where $\hat{\beta}_j=(X_j^TX_j)^{-1}X_j^Tz_j$ and $\hat{\varepsilon}=z_j-X_j\hat{\beta}_j$ are the estimated regression coefficients and the estimated residuals based on the submatrix X_j of Y respectively.

It has been shown that when incomplete data follow a monotone missing pattern and the missingness mechanism is ignorable, the observed-data posterior distribution of ϕ can be factored into independent observed-data posterior distributions of $\phi_1,\cdots,\phi_p$. This factorization is named as the *monotone distinct* (Rubin, 1987). Based upon this factorization, Bayesian inferences about ϕ can be conducted as a sequence of independent inferences about ϕ_j from each individual observed-data posterior distribution of ϕ_j, $j=1,\cdots,p$. For example, a simulated value $\tilde{\phi}=(\tilde{\phi}_1,\tilde{\phi}_2,\cdots,\tilde{\phi}_p)$ of ϕ can be obtained by independently drawing ϕ_j from its observed-data posterior distribution $P(\phi_j \mid Y_{obs})$. A simulated value $\tilde{\theta}=(\tilde{\mu},\tilde{\Sigma})$ of θ then can be obtained through the reverse transformation $\tilde{\theta}=g^{-1}(\tilde{\phi})$.

4.3 Sweep Operator

Since we need to frequently conduct the transformation from the normal parameters $\theta=(\mu,\Sigma)$ to the alternative parameters $\phi=(\phi_1,\phi_2,\cdots,\phi_p)$ and vice verse, Little and Rubin (1987) reviewed and applied a very useful tool known as the *sweep* operator, first introduced by Beaton (1964). Suppose that G is a $p\times p$ symmetric matrix with elements g_{ij}. The sweep operator SWP[k] operates on G by generating another $p\times p$ symmetric matrix H, $H=SWP[k]G$, where the elements of H are given by

$$
\begin{aligned}
h_{kk} &= -1/g_{kk} \\
h_{jk} &= h_{kj} = g_{jk}/g_{kk} \qquad \text{for } j\neq k \\
h_{jl} &= h_{lj} = g_{jl}-g_{jk}g_{kl}/g_{kk} \qquad \text{for } j\neq k \text{ and } l\neq k
\end{aligned}
$$

After such sweep operation on G, the matrix H is said to have been *swept on position* k.

Suppose G is partitioned as

$$
G=\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}
$$

where G_{11} is a $p_1\times p_1$ submatrix. After sweeping on positions $1,2,\cdots,p_1$, the result matrix becomes

$$
SWP[1,2,\cdots,p_1]G=\begin{bmatrix} -G_{11}^{-1} & G_{11}^{-1}G_{12} \\ G_{21}G_{11}^{-1} & G_{22}-G_{21}G_{11}^{-1}G_{12} \end{bmatrix}
$$

The notation $SWP[1,2,\cdots,p_1]$ indicates successive sweep operations, i.e.

$$SWP[1,2,\cdots,p_1]G = SWP[p_1]\cdots SWP[2]SWP[1]G$$

After defining the sweep operator, we also need to define a *reverse-sweep* operator. This operator returns a swept matrix to its original form. The reverse-sweep operator RSW[k] operates on G by generating another same dimensional symmetric matrix H, $H = RSW[k]G$, where the elements of H are given by

$$\begin{aligned} h_{kk} &= -1/g_{kk} \\ h_{jk} &= h_{kj} = -g_{jk}/g_{kk} \qquad \text{for } j \neq k \\ h_{jl} &= h_{lj} = g_{jl} - g_{jk}g_{kl}/g_{kk} \qquad \text{for } j \neq k \text{ and } l \neq k \end{aligned}$$

By definition, we have $RSW[k]\,SWP[k]\,G = SWP[k]\,RSW[k]\,G = G$.

Little and Rubin (1987, p112) demonstrated the usefulness of sweep operator for transforming the normal parameters and the alternative parameters.

Transformation between $\theta = (\mu, \Sigma)$ and $\phi = (\phi_1, \phi_2)$

Suppose we have the normal parameters $\theta = (\mu, \Sigma)$, we want to compute the alternative parameters $\phi = (\phi_1, \phi_2)$, where $\phi_1 = (\mu_1, \Sigma_{11})$ and $\phi_2 = (\alpha_{2\cdot1}, B_{2\cdot1}, \Sigma_{22\cdot1})$ are corresponding to the partition $Y = (Z_1, Z_2)^T$. Arranging the normal parameters $\theta = (\mu, \Sigma)$ as a $(p+1)\times(p+1)$ matrix in the following form,

$$\theta = \begin{bmatrix} -1 & \mu^T \\ \mu & \Sigma \end{bmatrix} = \begin{bmatrix} -1 & \mu_1^T & \mu_2^T \\ \mu_1 & \Sigma_{11} & \Sigma_{12} \\ \mu_2 & \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{4.3.1}$$

then sweeping this θ-matrix on the positions $2,3,\cdots,p_1+1$, we have

$$SWP[2,3,\cdots,p_1+1]\theta = \begin{bmatrix} -1-\mu^T\Sigma_{11}^{-1}\mu_1 & \mu_1^T\Sigma_{11}^{-1} & \alpha_{2\cdot1}^T \\ \Sigma_{11}^{-1}\mu_1 & -\Sigma_{11}^{-1} & B_{2\cdot1} \\ \alpha_{2\cdot1}^T & B_{2\cdot1} & \Sigma_{22\cdot1} \end{bmatrix} \tag{4.3.2}$$

Reverse sweeping the upper-left $(p_1+1)\times(p_1+1)$ submatrix of (4.3.2), we have

$$RSW[2,3,\cdots,p_1+1]\begin{bmatrix} -1-\mu^T\Sigma_{11}^{-1}\mu_1 & \mu_1^T\Sigma_{11}^{-1} \\ \Sigma_{11}^{-1}\mu_1 & -\Sigma_{11}^{-1} \end{bmatrix} = \begin{bmatrix} -1 & \mu_1^T \\ \mu_1 & \Sigma_{11} \end{bmatrix} \qquad (4.3.3)$$

Therefore, it has been shown that the alternative parameters $\phi=(\mu_1,\Sigma_{11},\alpha_{2\cdot1},B_{2\cdot1},\Sigma_{22\cdot1})$, expressed in the matrix form as

$$\phi=\begin{bmatrix} -1 & \mu_1^T & \alpha_{2\cdot1}^T \\ \mu_1 & \Sigma_{11} & B_{2\cdot1}^T \\ \alpha_{2\cdot1} & B_{2\cdot1} & \Sigma_{22\cdot1} \end{bmatrix} \qquad (4.3.4)$$

can be computed from the $\theta-$matrix (4.3.1) by first sweeping the full matrix on positions $2,3,\cdots,p_1+1$, and then reverse-sweeping the upper-left $(p_1+1)\times(p_1+1)$ submatrix of result matrix on the same positions. The calculation of $\theta=(\mu,\Sigma)$ from $\phi=(\phi_1,\phi_2)$ can be performed by reversing the above procedures.

Transformation between $\theta=(\mu,\Sigma)$ and $\phi=(\phi_1,\phi_2,\cdots,\phi_p)$

Now suppose we have the alternative parameters $\phi=(\phi_1,\phi_2,\cdots,\phi_p)$, where $\phi_j=(\beta_j^T,\delta_j)^T$ is a $(j+1)\times1$ vector, β_j are the regression coefficients of Y_j on $Y_1,\cdots,Y_{j-1}$, and δ_j is the residual variance, we want to perform the transformation between the normal parameters $\theta=(\mu,\Sigma)$ and the alternative parameters $\phi=(\phi_1,\phi_2,\cdots,\phi_p)$.

First, we create the $\theta-$matrix as in (4.3.1), then we create the ϕ-matrix as the following

$$\phi=\begin{bmatrix} -1 & & & & \\ & \phi_1 & & & \\ & & \phi_2 & & \\ & & & \phi_3 & \\ & & & & \cdots \end{bmatrix} \qquad (4.3.5)$$

Note that the dimension of the $\theta-$matrix and the ϕ-matrix are the same, $(p+1)\times(p+1)$. By sweeping the $\theta-$matrix on the positions $2,3,\cdots,j$, we have a new matrix whose $(j+1)^{th}$ column is ϕ_j. Therefore, if we sweep the full $\theta-$matrix on the positions $2,3,\cdots,p$, then we get a new matrix with ϕ_p appearing in the $(p+1)^{th}$ column. If we reverse sweeping all but the last row and column of the new matrix on the position p, then ϕ_{p-1} appears in the p^{th} column. By

reverse sweeping all but the last two rows and columns on the position $p-1$, we get ϕ_{p-1} appeared in column $p-1$, and so on. Hence it has been shown that the ϕ-matrix can be obtained by sweeping the $\theta-$matrix and then reverse sweeping the result new matrix. The $\theta-$matrix can be obtained through a reversal of the above procedures on the ϕ-matrix.

4.4 Summary

Multivariate normal distributions will be assumed as the underlying probability model for continuous response variables in the following chapters, so this chapter reviewed some its attractive analytical properties. When data are complete, the estimation of $\theta=(\mu,\Sigma)$ can be obtained easily without using iterative algorithm. By choosing a convenient non-informative prior distribution for $\theta=(\mu,\Sigma)$, the complete-data posterior distribution is a normal inverted-Wishart distribution with good analytical properties. The normal parameters can be easily simulated from this distribution. When data are incomplete, the estimation of $\theta=(\mu,\Sigma)$ generally requires some iterative algorithms, the observed-data posterior distribution is usually complicated. However, if the missing pattern is monotone, this estimation can be obtained by non-iterative method. The observed-data posterior distribution of the alternative parameters $\phi=(\phi_1,\phi_2,\cdots,\phi_p)$ are independent normal inverted-chisquare distributions. The alternative parameters can be easily simulated from these distributions independently and converted to the normal parameters. The sweep operator is conveniently used to make the transformation between the normal parameters and the alternative parameters.

Chapter 5. Monotone Incomplete Data

In Chapter 4, some relevant properties of the multivariate normal distribution are reviewed. Some of these properties are applied when making inferences about the unknown parameters $\theta = (\mu, \Sigma)$, either based on the observed-data likelihood function or the observed-data posterior distribution. However, our primary interests focus on imputing the missing data rather than making inferences based on the observed data. In this chapter we propose several imputation methods when incomplete data are from the multivariate normal distribution and have a monotone missing pattern. In section 5.1 we discuss how to build the imputation model and generate proper multiple imputation from the imputation model. We also review the predictive model method which is an univariate approach. In section 5.2 we develop a new multivariate approach for doing the multiple imputation when incomplete data are from a single multivariate normal distribution with an unstructured covariance matrix. In section 5.3 we discuss the same procedure as in section 5.2 except the covariance matrix is assumed to have a special structure, the ante-dependence structure. In section 5.4 we consider a situation where the data are from more than one multivariate normal distributions. The procedures for imputing the missing data are discussed. At the end, a brief summary of this chapter is presented in section 5.5.

5.1 Imputation Model and Proper Multiple Imputation

When data are incomplete, we want to fill in the missing data Y_{mis} with values randomly drawn from a probability distribution. When the missingness mechanism is ignorable, two probability distributions are derived: the observed-data posterior distribution of θ and the posterior predictive distribution of Y_{mis} given Y_{obs}. We refer, hereinafter, the random draws from the observed-data posterior distribution as the simulation of θ and the random draws from the posterior predictive distribution as the imputation of Y_{mis} (Schafer, 1997).

The posterior predictive distribution of Y_{mis} given Y_{obs} is given by

$$P(Y_{mis} \mid Y_{obs}) = \int P(Y_{mis} \mid Y_{obs}, \theta) P(\theta \mid Y_{obs}) d\theta \qquad (5.1.1)$$

Because of the integration in (5.1.1), it is generally difficult to have a closed form for the posterior predictive distribution. Also it is often difficult to draw a sample from this distribution numerically. However, the conditional predictive distribution is rather easy to obtain once the values of the parameters θ are given. For example, when the multivariate normal distribution is

assumed for the response variables, the conditional predictive distribution of Y_{mis} given Y_{obs} and θ is either a multivariate normal or an univariate normal. Thus if the observed-data posterior $P(\theta \mid Y_{obs})$ is either in a closed form or the values of θ can be randomly drawn from it without much difficulty, the Rao-Blackwellized estimates can be used to approximate the posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$, (Gelfand and Smith, 1990)

$$\int P(Y_{mis} \mid Y_{obs}, \theta) P(\theta \mid Y_{obs}) d\theta \approx \frac{1}{k} \sum_{t=1}^{k} P(Y_{mis} \mid Y_{obs}, \tilde{\theta}^{(t)}) \qquad (5.1.2)$$

where $\tilde{\theta}^{(t)}$ $(t = 1, \cdots, k)$ are the random draws from the observed-data posterior distribution $P(\theta \mid Y_{obs})$. When k is large enough, the right-hand side of (5.1.2) will provide a consistent estimate of the posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$.

Note that the right-hand side of (5.1.2) is a mixture distribution of the conditional predictive distributions over the simulated values $\tilde{\theta}^{(t)}$ from its observed-data posterior. A random draw from a mixture distribution of k component distributions consists of two steps, (1) randomly draw a distribution among the k distributions with the equal probability of $1/k$, and (2) randomly draw a sample from the selected distribution (Gilks, Richardson and Spiegelhalter, 1996, p176). Since each component distribution in the right-hand side of (5.1.2) is indexed by $\tilde{\theta}^{(t)}$, the above two-step random sampling from a mixture distribution is equivalent to creating a sampling pool $\tilde{\theta}^{(t)}$ $(t = 1, \cdots, k)$ from the observed-data posterior distribution $P(\theta \mid Y_{obs})$, and then randomly draw a value from this sampling pool. This is actually the same as taking a random draw from $P(\theta \mid Y_{obs})$ directly. Therefore, a single set of imputed values of Y_{mis} is obtained by taking random draws from the conditional predictive distribution $P(Y_{mis} \mid Y_{obs}, \tilde{\theta})$ where $\tilde{\theta}$ is a random draw from the observed-data posterior distribution. In order to generate m sets of proper multiple imputations, m simulated values of θ are independently drawn from the observed-data posterior distribution, say $\tilde{\theta}^{(t)}$ $(t = 1, \cdots, m)$. For each simulated value $\tilde{\theta}^{(t)}$, a single set of imputed values of Y_{mis} is obtained by taking random draws separately from each corresponding conditional posterior predictive distribution $P(Y_{mis} \mid Y_{obs}, \tilde{\theta}^{(t)})$.

Rubin (1987) proposed the predictive model of multiple imputation. When incomplete data have a monotone missing pattern, the observed-data likelihood function of the normal parameters

$\theta = (\mu, \Sigma)$ can be factored into a series of independent observed-data likelihood functions of individual alternative parameters ϕ_j $(j = 1, \cdots, p)$, where $\phi_j = (\beta_j^T, \delta_j)^T$ are the regression coefficients and the residual variance of Y_j on $Y_1, \cdots, Y_{j-1}$, i.e.

$$L(\theta \mid Y_{obs}) = \prod_{i=1}^{n_1} P(y_{i1} \mid \phi_1) \prod_{i=1}^{n_2} P(y_{i2} \mid y_{i1}, \phi_2) \cdots \prod_{i=1}^{n_p} P(y_{ip} \mid y_{i1}, \cdots, y_{i,p-1}, \phi_p)$$

Since $\prod_{i=1}^{n_j} P(y_{ij} \mid y_{i1}, \cdots, y_{i,j-1}, \phi_j)$ is the linear regression of Y_j on $Y_1, \cdots, Y_{j-1}$ based on n_j observed values of $Y_1, \cdots, Y_{j-1}$ and Y_j, this regression model is used to impute the missing values of Y_j. In order to generate proper multiple imputation, the regression parameters $\phi_j = (\beta_j^T, \delta_j)^T$ need to be simulated from their observed-data posterior distribution rather than estimated from the regression model.

To obtain the observed-data posterior distribution of the regression parameters, a prior distribution need to be assumed. In the absence of a strong knowledge about the prior distribution of the regression parameters, a convenient non-informative prior distribution of $\phi = (\phi_1, \cdots, \phi_p)$

$$\pi(\phi) = \delta_1^{-1} \delta_2^{-1} \cdots \delta_p^{-1} \tag{5.1.3}$$

is assumed. This is Jeffreys prior for the factored density (Box and Tiao, 1973). Note that this prior distribution is the product of a uniform prior distribution for $(\beta_1, \beta_2, \cdots, \beta_p)$ over the $p(p+1)/2$ dimensional real space and the non-informative prior distribution for $(\delta_1, \delta_2, \cdots, \delta_p)$. Thus the observed-data posterior distribution of ϕ can be factored into the observed-data posterior distribution of individual parameter ϕ_j,

$$P(\phi \mid Y_{obs}) \propto \prod_{j=1}^{p} P(\phi_j \mid Y_{obs}) \tag{5.1.4}$$

and the individual observed-data posterior distribution of ϕ_j can be expressed in a closed form

$$P(\phi_j \mid Y_{obs}) = L(\phi_j \mid Y_{obs}) \times \delta_j^{-1}$$

$$= \delta_j^{-\frac{n_j}{2}} \exp\left\{ -\frac{1}{2\delta_j} (z_j - X_j \beta_j)^T (z_j - X_j \beta_j) \right\} \times \delta_j^{-1}$$

$$= \delta_j^{-\frac{j}{2}} \exp\left\{-\frac{1}{2\delta_j}(\beta_j - \hat{\beta}_j)^T X_j^T X_j (\beta_j - \hat{\beta}_j)\right\} \times \delta_j^{-(\frac{n_j - j}{2})-1} \exp\left\{-\frac{1}{2\delta_j}\hat{\varepsilon}_j^T \hat{\varepsilon}_j\right\}$$

where $z_j = (y_{1j}, y_{2j}, \cdots, y_{n_j,j})^T$ is the observed values of Y_j, X_j is the upper-left $n_j \times (j-1)$ submatrix of Y augmented by a column of ones as in (4.2.6), $\hat{\beta}_j = (X_j^T X_j)^{-1} X_j^T z_j$ is the estimates of the regression coefficients, and $\hat{\varepsilon}_j = z_j - X_j \hat{\beta}_j$ is a vector of residuals. This observed-data posterior distribution of ϕ_j is the product of a j-dimensional normal distribution given δ_j and a scaled inverted-chisquare distribution with the degrees of freedom $n_j - j$. The marginal observed-data posterior distribution of δ_j and the conditional observed-data posterior distribution of β_j are given by

$$\beta_j \mid Y_{obs}, \delta_j \sim N_j(\hat{\beta}_j, \delta_j (X_j^T X_j)^{-1}) \tag{5.1.5}$$

$$\delta_j \mid Y_{obs} \sim \hat{\varepsilon}_j^T \hat{\varepsilon}_j \chi^{-2}_{n_j - j} \tag{5.1.6}$$

The regression parameters $\phi_j = (\beta_j, \delta_j)$, therefore, can be simulated from this observed-data posterior distribution.

After a simulated value $\tilde{\phi}_j$ has been drawn from this distribution, the i^{th} missing value of Y_j is imputed from the univariate normal distribution $N(\tilde{\beta}_{0j} + \tilde{\beta}_{1j} y_{i1} + \cdots + \tilde{\beta}_{j-1,j} y_{i,j-1}, \tilde{\delta}_j)$. The multiple imputations of Y_j are created by drawing m sets of the regression parameters independently from the observed-data posterior distribution, for each simulated parameters $(\tilde{\beta}_j^{(l)}, \tilde{\delta}_j^{(l)})$, the missing values of Y_j are imputed single time. Once the missing values of $Y_2, \cdots, Y_{j-1}$ have been imputed sequentially, the missing values of Y_j can be imputed given the imputed complete data of $Y_1, Y_2 \cdots, Y_{j-1}$. Continuing this process until the missing values of Y_p are imputed. This is the predictive model based imputation.

5.2 Imputation for Single Population With Unstructured Covariance

Let $Y = (y_1, y_2, \cdots, y_n)^T$ be a $n \times p$ data matrix, and we assume that $y_1, y_2, \cdots, y_n$ are iid samples from $N_p(\mu, \Sigma)$, where $\mu = (\mu_1, \cdots, \mu_p)$ is the mean vector and $\Sigma = (\sigma_{ij})$ is an

unstructured covariance matrix. The dimension of the unknown parameters $\theta = (\mu, \Sigma)$ is $p + p(p+1)/2$.

When incomplete data have a monotone missing pattern, the observed-data likelihood function of the alternative parameters $\phi = (\phi_1, \cdots, \phi_p)$ is

$$L(\phi \mid Y_{obs}) = \prod_{i=1}^{n_1} P(y_{i1} \mid \phi_1) \prod_{i=1}^{n_2} P(y_{i2} \mid y_{i1}, \phi_2) \cdots \prod_{i=1}^{n_p} P(y_{ip} \mid y_{i1}, \cdots, y_{i,p-1}, \phi_p)$$

The alternative parameter $\phi_j = (\beta_j^T, \delta_j)^T$ can be independently estimated from the linear regression model of Y_j on $Y_1, \cdots, Y_{j-1}$ based on n_j observed values of $Y_1, \cdots, Y_{j-1}$ and Y_j. Let the estimated alternative parameters be $\hat{\phi}_j = (\hat{\beta}_j^T, \hat{\delta}_j)^T$ $(j = 1,2,\cdots, p)$, where

$$\hat{\beta}_j = (X_j^T X_j)^{-1} X_j^T z_j \tag{5.2.1}$$

and

$$\hat{\delta}_j = (z_j - X_j \hat{\beta}_j)^2 / n_j \tag{5.2.2}$$

Then the estimated normal parameters $\theta = (\mu, \Sigma)$ can be obtained by the reverse transformation $\hat{\theta} = g^{-1}(\hat{\phi})$. Note that the denominator of $\hat{\delta}_j$ is n_j instead of $n_j - j$, so these estimators are biased. When data are complete, the estimations by this approach are the same as by (4.1.6)-(4.1.7).

To impute the missing data, we need to simulate the parameters from their observed-data posterior distribution rather than estimate them from the regression models. If we assume the same prior as (5.1.3), the joint observed-data posterior distribution of $\phi = (\phi_1, \cdots, \phi_p)$ is the product of the observed-data posterior distributions of individual component ϕ_j. That is, $P(\phi \mid Y_{obs}) \propto \prod_{j=1}^{p} P(\phi_j \mid Y_{obs})$. Thus a Bayesian simulation of ϕ can be obtained by independently simulating individual ϕ_j from its own observed-data posterior distribution which is a normal scaled inverted-chisquare distribution given by (5.15)-(5.1.6). A simulated value $\tilde{\phi}_j = (\tilde{\beta}_j, \tilde{\delta}_j)$ can be obtained by letting $\tilde{\delta}_j = \hat{\varepsilon}_j^T \hat{\varepsilon}_j / \delta_j^*$, where δ_j^* is a random draw from $\chi^2_{n_j - j}$, and $\tilde{\beta}_j$ is a random sample from (5.1.5) given $\tilde{\delta}_j$.

After the values of $\phi = (\phi_1, \cdots, \phi_p)$ have been simulated from the observed-data posterior distribution of ϕ, say $\tilde{\phi} = (\tilde{\phi}_1, \cdots, \tilde{\phi}_p)$, the simulated normal parameters $\theta = (\mu, \Sigma)$ can be obtained from the reverse transformation $\tilde{\theta} = g^{-1}(\tilde{\phi})$. This reverse transformation can be accomplished by applying the sweep operator on the $\tilde{\phi}$-matrix as described in section 4.3. Having obtained the simulated values $\tilde{\theta}$, a single set of imputed values of Y_{mis} is obtained by taking random draws from the conditional predictive distribution $P(Y_{mis} \mid Y_{obs}, \tilde{\theta})$. Since the rows of Y, $y_1, y_2, \cdots, y_n$, are conditionally independent given $\tilde{\theta}$, the imputation is carried out independently on each individual observation $y_i = (y_{i(obs)}^T, y_{i(mis)}^T)^T$. The $y_{i(mis)}$ are imputed by taking a random draw from the conditional predictive distribution $P(y_{i(mis)} \mid y_{i(obs)}, \tilde{\theta})$, which is either a multivariate normal distribution or an univariate normal distribution. In order to generate m sets of proper multiple imputations, m simulated values $\tilde{\theta}^{(t)}$ $(t = 1, \cdots, m)$ are obtained through the reverse transformation $\tilde{\theta}^{(t)} = g^{-1}(\tilde{\phi}^{(t)})$, where $\tilde{\phi}^{(t)}$ are independent samples from the observed-data posterior distribution of ϕ. For each of $\tilde{\theta}^{(t)}$, a single set of imputed values of Y_{mis} is obtained from the corresponding conditional posterior predictive distribution $P(Y_{mis} \mid Y_{obs}, \tilde{\theta}^{(t)})$. Let $Y_{mis}^{(t)}$ be the imputed values of Y_{mis} from $P(Y_{mis} \mid Y_{obs}, \tilde{\theta}^{(t)})$, then $\{Y_{mis}^{(1)}, \cdots, Y_{mis}^{(m)}\}$ consists of a proper multiple imputation of Y_{mis} since each set of $Y_{mis}^{(t)}$ is independently drawn from the approximate posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$.

5.3 Imputation for Single Population With Structured Covariance

In this section, the assumptions about the response variables are the same as those in section 5.2, i.e. $Y = (y_1, y_2, \cdots, y_n)^T$ is a $n \times p$ data matrix, and $y_1, y_2, \cdots, y_n$ are assumed to be iid samples from a multivariate normal distribution $N_p(\mu, \Sigma)$, except here the covariance matrix $\Sigma = (\sigma_{ij})$ is assumed to have an ante-dependence structure of order g. Under this assumption, the dimension of the unknown parameters $\theta = (\mu, \Sigma)$ is $p + (g+1)(2p-g)/2$.

The ante-dependence structure was introduced by Gabriel (1961,1962) to describe the conditional independence among the multiple variables. Since then a number of authors have applied this covariance structure into the analysis of multivariate data and longitudinal data. Kenward (1987) assumed an ante-dependence structure while comparing repeated measurement profiles for monotone samples. Patel (1991) proposed a likelihood-based method to analyze repeated measurement data with ante-dependence covariance structure and monotone missing pattern.

A multivariate $Y=(Y_1,Y_2,\cdots,Y_p)^T$ is said to have an ante-dependence structure of order g if $Y_j \mid Y_{j-1},Y_{j-2},\cdots,Y_{j-g}$ and $Y_1,Y_2,\cdots,Y_{j-g-1}$ are uncorrelated (Gabriel, 1962).

When $Y=(Y_1,Y_2,\cdots,Y_p)^T$ has an ante-dependence structure of order g, the elements of the covariance matrix $\Sigma=(\sigma_{ij})$ satisfy the following condition (SAS/STAT Software: Changes and Enhancements Through Release 6.12, p629)

$$\sigma_{ij}=(\sigma_{ii}\sigma_{jj})^{1/2}\prod_{k=i}^{j-g}\rho_k$$

where ρ_k is the lag-k autocorrelation parameters satisfying $|\rho_k|<1$. For example, when $p=4$ and $g=1$, the covariance matrix $\Sigma=(\sigma_{ij})$ is

$$\Sigma=\begin{bmatrix} \sigma_{11} & (\sigma_{11}\sigma_{22})^{1/2}\rho_1 & (\sigma_{11}\sigma_{33})^{1/2}\rho_1\rho_2 & (\sigma_{11}\sigma_{44})^{1/2}\rho_1\rho_2\rho_3 \\ & \sigma_{22} & (\sigma_{22}\sigma_{33})^{1/2}\rho_2 & (\sigma_{22}\sigma_{44})^{1/2}\rho_2\rho_3 \\ & & \sigma_{33} & (\sigma_{33}\sigma_{44})^{1/2}\rho_3 \\ & & & \sigma_{44} \end{bmatrix}$$

This covariance structure is a flexible model which contains the autoregressive model, complete independence model and unstructured covariance model as special cases. In the ante-dependence model the variances of the variables and the distances between the variables are not required to be equal. When all variances σ_{ii} and all lag-k autocorrelation parameters ρ_k are equal, this model becomes a first order autoregressive model. When the order of ante-dependence is 0, we have complete independence of $Y_1,Y_2,\cdots,Y_p$. When the order of ante-dependence is $p-1$, we have unstructured covariance. Therefore, the ante-dependence structure is a very general covariance structure. This covariance structure is naturally applicable to the repeated measurement data in clinical trials, in which the measurements are ordered in time and the present measurement depends only on the immediate antecedent measurements.

Under the assumption of ante-dependence structure of order g, the conditional distribution of Y_j given $Y_1,\cdots,Y_{j-1}$ depends on the g immediate previous variables only,

$$P(Y_j \mid Y_1,Y_2,\cdots,Y_{j-1},\phi_j) = P(Y_j \mid Y_{j-g},\cdots,Y_{j-1},\phi_j) \tag{5.3.1}$$

That is, the regression of Y_j on $Y_1,\cdots,Y_{j-1}$ has zero regression coefficients corresponding to the predictor variables $Y_1,\cdots,Y_{j-g-1}$ and non-zero regression coefficients corresponding to $Y_{j-g},\cdots,Y_{j-1}$.

When incomplete data follow a monotone missing pattern, the observed-data likelihood function of the alternative parameters $\phi=(\phi_1,\cdots,\phi_p)$ is the product of the observed-data likelihood functions of individual alternative parameter ϕ_j $(j=1,\cdots,p)$,

$$L(\phi \mid Y_{obs}) = \prod_{j=1}^{p} L(\phi_j \mid Y_{obs})$$

where

$$L(\phi_j \mid Y_{obs}) = \prod_{i=1}^{n_j} P(y_{ij} \mid y_{i,1},\cdots,y_{i,j-1},\phi_j) = \prod_{i=1}^{n_j} P(y_{ij} \mid y_{i,j-g},\cdots,y_{i,j-1},\phi_j)$$

is the observed-data likelihood function of ϕ_j. This is a linear regression of Y_j on $Y_{j-g},\cdots,Y_{j-1}$, where g is the given order of the ante-dependence structure, based on n_j observations in which the values of Y_j and $Y_{j-g},\cdots,Y_{j-1}$ are all observed.

The maximum likelihood estimates of the alternative parameters ϕ can be obtained by maximizing $L(\phi_j \mid Y_{obs})$ independently for $j=1,\cdots,p$. The maximization of $L(\phi_j \mid Y_{obs})$ is accomplished by the ordinary least square regression of Y_j on $Y_{j-g},\cdots,Y_{j-1}$ based on the first n_j rows of the data matrix Y. Let $z_j=(y_{1j},y_{2j},\cdots,y_{n_j,j})^T$ denote the observed values for variable Y_j, X_j denote a $n_j\times(g+1)$ submatrix of Y corresponding to $Y_{j-g},\cdots,Y_{j-1}$, and augmented by a column of ones.

$$X_j = \begin{bmatrix} 1 & y_{1,j-g} & y_{1,j-g+1} & \cdots & y_{1,j-1} \\ 1 & y_{2,j-g} & y_{2,j-g+1} & \cdots & y_{2,j-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & y_{n_j,j-g} & y_{n_j,j-g+1} & \cdots & y_{n_j,j-1} \end{bmatrix} \tag{5.3.2}$$

The conditional distribution of z_j given X_j and ϕ_j is

$$z_j \mid X_j, \phi_j \sim N(X_j \beta_j, \delta_j I)$$

where $\phi_j = (\beta_j, \delta_j)$, β_j is a vector of regression coefficients and δ_j is the conditional variance of Y_j given $Y_{j-g}, \cdots, Y_{j-1}$. Using the properties of linear regression model, the maximum likelihood estimators of $\phi_j = (\beta_j, \delta_j)$ is given by

$$\hat{\beta}_j = (X_j^T X_j)^{-1} X_j^T z_j \tag{5.3.3}$$

$$\hat{\delta}_j = \frac{1}{n_j}(z_j - X_j \hat{\beta}_j)^T (z_j - X_j \hat{\beta}_j) \tag{5.3.4}$$

Note that $\hat{\beta}_j = (\hat{\beta}_{j,0}, \hat{\beta}_{j,j-g}, \cdots, \hat{\beta}_{j,j-1})^T$ is a $(g+1)$-element vector consisting of the estimates of the intercept and the slopes of Y_j on $Y_{j-g}, \cdots, Y_{j-1}$. Let $\hat{\beta}_{(j)}$ denote an augmented j-element vector consisting of $\hat{\beta}_j$, and augmented by zeros in the middle positions corresponding to $Y_1, \cdots, Y_{j-g-1}$, that is, $\hat{\beta}_{(j)} = (\hat{\beta}_{j,0}, 0, \cdots, 0, \hat{\beta}_{j,j-g}, \cdots, \hat{\beta}_{j,j-1})^T$. For example, if $j = 4$ and $g = 1$, then $\hat{\beta}_4 = (\hat{\beta}_{4,0}, \hat{\beta}_{4,3})^T$ and $\hat{\beta}_{(4)} = (\hat{\beta}_{4,0}, 0, 0, \hat{\beta}_{4,3})^T$. Let $\hat{\phi}_{(j)} = (\hat{\beta}_{(j)}, \hat{\delta}_j)$, then $\hat{\phi} = (\hat{\phi}_{(1)}, \cdots, \hat{\phi}_{(p)})$ is the maximum likelihood estimates of the alternative parameters $\phi = (\phi_1, \cdots, \phi_p)$ under the assumption of the ante-dependence structure of order g. By applying the sweep operator on the $\hat{\phi}$-matrix, we can obtain the maximum likelihood estimates of the normal parameters $\theta = (\mu, \Sigma)$ under the assumption of the ante-dependence structure of order g. When data are complete, the maximum likelihood estimates of μ, but not of Σ, are the same regardless if the covariance matrix Σ is assumed to be unstructured or ante-dependence structured. When data are incomplete, the estimates of μ and Σ are different under the different assumptions about Σ.

When the same non-informative prior distribution of $\phi = (\phi_1, \cdots, \phi_p)$ is assumed as in (5.1.3), the observed-data posterior distribution of ϕ is the product of the observed-data posterior distributions of individual parameter ϕ_j $(j = 1, \cdots, p)$, and the observed-data posterior distribution of ϕ_j has the same form as those in (5.1.5) and (5.1.6). That is, the joint density function of the observed-data posterior distribution of ϕ_j is the product of a $(g+1)$-

dimensional normal density given δ_j and a scaled inverted-chisquare distribution density with the degrees of freedom $n_j - g - 1$. Therefore, the marginal observed-data posterior distribution of δ_j and the conditional observed-data posterior distribution of β_j are given by

$$\beta_j \mid Y_{obs}, \delta_j \sim N_{g+1}(\hat{\beta}_j, \delta_j (X_j^T X_j)^{-1}) \tag{5.3.5}$$

$$\delta_j \mid Y_{obs} \sim n_j \hat{\delta}_j \chi^{-2}_{n_j - g - 1} \tag{5.3.6}$$

where $\hat{\beta}_j$ and $\hat{\delta}_j$ are obtained from (5.3.3) and (5.3.4).

The simulated values of the alternative parameters $\phi = (\phi_1, \cdots, \phi_p)$ are obtained by independently simulating the individual parameters $\phi_j = (\beta_j, \delta_j)$ from their own observed-data posterior distributions (5.3.5)-(5.3.6). A simulated value of δ_j is obtained by having $\tilde{\delta}_j = n_j \hat{\delta}_j / \delta_j^*$, where δ_j^* is a random draw from $\chi^2_{n_j - g - 1}$, and a random sample $\tilde{\beta}_j = (\tilde{\beta}_{j,0}, \tilde{\beta}_{j,j-g}, \cdots, \tilde{\beta}_{j,j-1})$ is drawn from (5.3.5) given $\tilde{\delta}_j$. Let $\tilde{\beta}_{(j)}$ denote an augmented j-element vector consisting of $\tilde{\beta}_j$ and augmented by zeros in the middle positions corresponding to $Y_1, \cdots, Y_{j-g-1}$. That is, $\tilde{\beta}_{(j)} = (\tilde{\beta}_{j,0}, 0, \cdots, 0, \tilde{\beta}_{j,j-g}, \cdots, \tilde{\beta}_{j,j-1})$. Let $\tilde{\phi}_{(j)} = (\tilde{\beta}_{(j)}, \tilde{\delta}_j)$, then $\tilde{\phi} = (\tilde{\phi}_{(1)}, \cdots, \tilde{\phi}_{(p)})$ is the simulated values from the observed-data posterior distribution of the alternative parameters $\phi = (\phi_1, \cdots, \phi_p)$ under the assumption of the ante-dependence structure of order g. By applying the sweep operator on the $\tilde{\phi}$ -matrix, we can obtain the simulated values of the normal parameters $\theta = (\mu, \Sigma)$ under the assumption of the ante-dependence structure of order g, say $\tilde{\theta} = (\tilde{\mu}, \tilde{\Sigma})$.

After the simulated values of $\theta = (\mu, \Sigma)$ have been obtained under the assumption of ante-dependence structure of order g, the imputation procedures for filling in the missing data Y_{mis} are the same as those when the covariance matrix is unstructured. The missing data Y_{mis} are imputed observation by observation, the $y_{i(mis)}$ $(i = 1, \cdots, n)$ are imputed by random draws from the conditional predictive distribution $P(y_{i(mis)} \mid y_{i(obs)}, \tilde{\theta})$. The covariance matrix of this distribution is the conditional covariance matrix given that the unconditional covariance matrix has an ante-dependence structure of order g, so the imputed values of $y_{i(mis)}$ preserve the

structured correlations between $y_{i(mis)}$ and $y_{i(obs)}$. In order to have m sets of proper multiple imputations. m simulated values $\tilde{\theta}^{(t)}$ $(t=1,\cdots,m)$ are obtained through the reverse transformation $\tilde{\theta}^{(t)} = g^{-1}(\tilde{\phi}^{(t)})$, where $\tilde{\phi}^{(t)} = (\tilde{\phi}_{(1)}^{(t)},\cdots,\tilde{\phi}_{(p)}^{(t)})$ are the independent draws from the observed-data posterior distribution under the assumption of ante-dependence structure of order g. For each of $\tilde{\theta}^{(t)}$, Y_{mis} are imputed from each corresponding conditional posterior predictive distribution $P(Y_{mis} \mid Y_{obs}, \tilde{\theta}^{(t)})$. Let $Y_{mis}^{(t)}$ be the imputed values of Y_{mis} from $P(Y_{mis} \mid Y_{obs}, \tilde{\theta}^{(t)})$, then $\left\{Y_{mis}^{(1)},\cdots,Y_{mis}^{(m)}\right\}$ consists of a proper multiple imputations of Y_{mis} since each $Y_{mis}^{(t)}$ are independently drawn from the approximate posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$.

5.4 Imputation for Multiple Populations With Structured Covariance

In sections 5.2 and 5.3, we discussed the situations where all multivariate random samples are from a single probability population. In this section we consider a probability model where the multivariate random samples are from more than one probability populations. Suppose that we have f design factors with levels $b_1, b_2, \cdots, b_f$ respectively, the total number of probability populations formed by the f factors is $b = b_1 \times b_2 \times \cdots \times b_f$. The design factors are usually treated as fixed, such as the treatment groups and clinical sites in a clinical trial, so they are not subject to a probability distribution. Let $Y = (y_1, y_2, \cdots, y_n)^T$ be a $n \times p$ data matrix for the response variables, $D = (d_1, d_2, \cdots, d_b)$ be a $n \times b$ matrix with d_k being the dummy variable pertinent to the k^{th} population. Then the probability model for the response variables is assumed to be a conditional multivariate normal distribution given the design factors,

$$y_i \mid d_k = 1 \ \sim \ N_p(\mu_k, \Sigma) \qquad (5.4.1)$$

The mean vectors μ_k $(k = 1,2,\cdots,b)$ are allowed to vary from population to population, but the covariance matrix Σ is assumed to be the same across the populations with an ante-dependence structure of order g. Let $\theta = (\theta_1, \theta_2, \cdots, \theta_b)$ be the unknown parameters, where $\theta_k = (\mu_k, \Sigma)$ are the normal parameters of the k^{th} population. Then under this model assumption, the dimension of the unknown parameters is $k \times p + (g+1)(2p-g)/2$.

The matrix D is assumed to be complete and the data matrix Y is assumed to be incomplete. When incomplete data follow a monotone missing pattern, the observed-data likelihood function of $\phi=(\phi_1,\cdots,\phi_p)$, where $\phi_j=(\phi_{j(1)},\cdots,\phi_{j(b)})$ are the alternative parameters of Y_j given $Y_{j-g},\cdots Y_{j-1}$ across the populations, can be factored into a series of independent observed-data likelihood functions of individual alternative parameters, $\phi_{j(k)}$ $(j=1,\cdots,p\ ;\ k=1,\cdots,b)$, i.e.

$$L(\phi\mid Y_{obs})=\prod_{j=1}^{p}\prod_{k=1}^{b}L(\phi_{j(k)}\mid Y_{obs}) \tag{5.4.2}$$

where

$$\prod_{k=1}^{b}L(\phi_{j(k)}\mid Y_{obs})=\prod_{k=1}^{b}\prod_{i=1}^{n_{j(k)}}P(y_{ij}\mid y_{i,j-g},\cdots,y_{i,j-1},\phi_{j(k)})$$

$$=\prod_{i=1}^{n_j}P(y_{ij}\mid y_{i,j-g},\cdots,y_{i,j-1},\phi_j) \tag{5.4.3}$$

where $n_{j(k)}$ is the number of observed values of variable Y_j in the k^{th} population, and $n_j=n_{j(1)}+\cdots+n_{j(b)}$ is the number of observed values of variable Y_j across all populations. Therefore, (5.4.3) is a linear regression of Y_j on $Y_{j-g},\cdots Y_{j-1}$ based on all observed data across the populations.

The maximum likelihood estimates of θ are obtained using the ordinary least squares. Let $z_j=(y_{1j},y_{2j},\cdots,y_{n_j,j})^T$ be the observed values of variable Y_j across all populations, and let X_j be the data matrix containing the first n_j rows of the values for the dummy variables $d_1,d_2,\cdots,d_b$ and the interaction terms of $d_1,d_2,\cdots,d_b$ and $Y_{j-g},\cdots Y_{j-1}$. The matrix X_j has the dimension of $n_j\times(b+b\times g)$. The conditional distribution of z_j given X_j and ϕ_j is

$$z_j\mid X_j,\phi_j\ \sim\ N(X_j\beta_j,\delta_jI) \tag{5.4.4}$$

where $\phi_j=(\phi_{j(1)},\cdots,\phi_{j(b)})$, $\phi_{j(k)}=(\beta_{j(k)}^T,\delta_j)^T$ and $\beta_j=(\beta_{j(1)}^T,\cdots,\beta_{j(b)}^T)^T$. Note in this model, $\beta_{j(k)}$ are the intercept and slopes of Y_j on $Y_{j-g},\cdots,Y_{j-1}$ in the k^{th} population, they

vary from population to population, while δ_j is a constant residual variance across the populations.

The maximum likelihood estimates of β_j and δ_j are given by

$$\hat{\beta}_j = (X_j^T X_j)^{-1} X_j^T z_j$$

$$\hat{\delta}_j = \frac{1}{n_j}(z_j - X_j\hat{\beta}_j)^T(z_j - X_j\hat{\beta}_j)$$

Note that $\hat{\beta}_j = (\hat{\beta}_{j(1)}^T, \cdots, \hat{\beta}_{j(b)}^T)^T$. where $\hat{\beta}_{j(k)} = (\hat{\beta}_{j(k),0}, \hat{\beta}_{j(k),j-g}, \cdots, \hat{\beta}_{j(k),j-1})$ is a $(g+1)\times 1$ vector. We augment this vector by zeros in the middle positions corresponding to $Y_1, \cdots, Y_{j-g-1}$ as described in section 5.3. To keep the simplicity, we use the same notation to denote the augmented $j \times 1$ vector, i.e. $\hat{\beta}_{j(k)} = (\hat{\beta}_{j(k),0}, 0, \cdots, 0, \hat{\beta}_{j(k),j-g}, \cdots, \hat{\beta}_{j(k),j-1})$. The estimated population means $\hat{\mu}_k$ and covariance matrix $\hat{\Sigma}$ then can be obtained from the reverse transformation $\hat{\theta}_k = g^{-1}(\hat{\phi}_{1(k)}, \hat{\phi}_{2(k)}, \cdots, \hat{\phi}_{p(k)})$, where $\hat{\phi}_{j(k)} = (\hat{\beta}_{j(k)}, \hat{\delta}_j)$ $j = 1,2,\cdots, p$. Note that $\hat{\delta}_j$ is a constant across all populations, whereas $\hat{\beta}_{j(k)}$ varies across the populations. When $\hat{\theta}_k = (\hat{\mu}_k, \hat{\Sigma})$ is calculated through the reverse transformation, both $\hat{\mu}_k$ and $\hat{\Sigma}$ vary across the populations. Denoting the estimated covariance matrix of k^{th} population by $\hat{\Sigma}_k$, a common covariance matrix can be estimated by $\hat{\Sigma} = \frac{1}{b}\sum_{k=1}^{b} \hat{\Sigma}_k$. Therefore, it has been shown that $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_b)$, where $\hat{\theta}_k = (\hat{\mu}_k, \hat{\Sigma})$, are the maximum likelihood estimates of $\theta = (\theta_1, \theta_2, \cdots, \theta_b)$ under the assumption of the ante-dependence structure of order g for the common covariance matrix Σ.

For imputation purposes, the values of μ_k and Σ need to be simulated from their observed-data posterior distributions rather than estimated from the observed-data likelihood functions. The observed-data posterior distributions of μ_i and Σ, therefore, need to be investigated. The observed-data likelihood function of $\phi_j = (\phi_{j(1)}, \cdots, \phi_{j(b)})$ is

$$L(\phi_j \mid Y_{obs}) = \prod_{k=1}^{b}\prod_{i=1}^{n_{j(k)}} P(y_{ij} \mid y_{i,j-g}, \cdots, y_{i,j-1}, \phi_{j(k)})$$

$$= \prod_{i=1}^{n_j} P(y_{ij} \mid y_{i,j-g}, \cdots, y_{i,j-1}, \phi_j)$$

$$\propto \delta_j^{-(\frac{n_j}{2})} \exp\left\{-\frac{1}{2\delta_j}(z_j - X_j\beta_j)^T(z_j - X_j\beta_j)\right\}$$

Since the missing pattern is monotone, the observed-data likelihood function of $\phi = (\phi_1, \cdots, \phi_p)$ is the product of the observed-data likelihood function of ϕ_j.

$$L(\phi \mid Y_{obs}) = \prod_{j=1}^{p} L(\phi_j \mid Y_{obs})$$

In the absence of strong preference on the parameters $\phi = (\phi_1, \cdots, \phi_p)$, a non-informative prior distribution

$$\pi(\phi) = \delta_1^{-1}\delta_2^{-1}\cdots\delta_p^{-1}$$

is assumed. This prior distribution is the product of a multivariate uniform distribution for $(\beta_1, \beta_2, \cdots, \beta_p)$, where $\beta_j = (\beta_{j(1)}^T, \cdots, \beta_{j(b)}^T)^T$, over the $b \times p(p+1)/2$ dimensional real space and a non-informative prior distribution for $(\delta_1, \delta_2, \cdots, \delta_p)$. Thus the observed-data posterior distribution of ϕ is given by

$$P(\phi \mid Y_{obs}) = L(\phi \mid Y_{obs}) \times \pi(\phi) = \prod_{j=1}^{p} L(\phi_j \mid Y_{obs}) \times \delta_j^{-1} = \prod_{j=1}^{p} P(\phi_j \mid Y_{obs})$$

This observed-data posterior distribution of ϕ_j can be expressed in a closed form as

$$P(\phi_j \mid Y_{obs}) \propto \delta_j^{-(\frac{n_j}{2})} \exp\left\{-\frac{1}{2\delta_j}(z_j - X_j\beta_j)^T(z_j - X_j\beta_j)\right\} \times \delta_j^{-1}$$

$$= \delta_j^{\frac{b(g+1)}{2}} \exp\left\{-\frac{1}{2\delta_j}(\beta_j - \hat{\beta}_j)^T X_j^T X_j (\beta_j - \hat{\beta}_j)\right\} \times \delta_j^{-(\frac{n_j - b(g+1)}{2})-1} \exp\left\{-\frac{1}{2\delta_j}\hat{\varepsilon}_j{}^T\hat{\varepsilon}_j\right\}$$

This joint observed-data posterior distribution of β_j and δ_j is the product of a $b(g+1)$-dimensional multivariate normal distribution and a scaled inverted-chisquare distribution with the degrees of freedom $n_j - b(g+1)$. The marginal observed-data posterior distribution of δ_j and the conditional observed-data posterior distribution of β_j given δ_j are

$$\beta_j \mid Y_{obs}, \delta_j \sim N_{b(g+1)}(\hat{\beta}_j, \delta_j(X_j^T X_j)^{-1}) \qquad (5.4.5)$$

$$\delta_j \mid Y_{obs} \sim \hat{\varepsilon}_j^T \hat{\varepsilon}_j \chi^{-2}_{n_j - b(g+1)} \tag{5.4.6}$$

The simulated values of $\phi = (\phi_1, \cdots, \phi_p)$ can be obtained by independently drawing the values of $\phi_j = (\phi_{j(1)}, \cdots, \phi_{j(b)})$ from their own observed-data posterior distribution (5.4.5)-(5.4.6). A simulated value of δ_j is obtained by having $\tilde{\delta}_j = \hat{\varepsilon}_j^T \hat{\varepsilon}_j / \delta_j^*$, where δ_j^* is a random draw from $\chi^2_{n_j - b(g+1)}$, and then a random sample $\tilde{\beta}_j = (\tilde{\beta}_{j(1)}^T, \cdots, \tilde{\beta}_{j(b)}^T)^T$, where $\tilde{\beta}_{j(k)} = (\tilde{\beta}_{j(k),0}, \tilde{\beta}_{j(k),j-g}, \cdots, \tilde{\beta}_{j(k),j-1})$, is drawn from (5.4.5) given $\tilde{\delta}_j$. We augment the $(g+1)$-element vector $\tilde{\beta}_{j(k)}$ by zeros in the middle positions corresponding to $Y_1, \cdots, Y_{j-g-1}$ as described in section 5.3. To keep the simplicity, we use the same notation to denote the augmented j-element vector, i.e. $\tilde{\beta}_{j(k)} = (\tilde{\beta}_{j(k),0}, 0, \cdots, 0, \tilde{\beta}_{j(k),j-g}, \cdots, \tilde{\beta}_{j(k),j-1})$. Let $\tilde{\phi}_j = (\tilde{\phi}_{j(1)}, \cdots, \tilde{\phi}_{j(b)})$, where $\tilde{\phi}_{j(k)} = (\tilde{\beta}_{j(k)}, \tilde{\delta}_j)$, then $\tilde{\phi} = (\tilde{\phi}_1, \cdots, \tilde{\phi}_p)$ are the simulated values from the observed-data posterior distribution of the alternative parameters $\phi = (\phi_1, \cdots, \phi_p)$ under the assumption of the ante-dependence structure of order g. The simulated values of the population means and covariance matrix $\theta_k = (\mu_k, \Sigma)$ can be obtained from the reverse transformation $\tilde{\theta}_k = g^{-1}(\tilde{\phi}_{1(k)}, \tilde{\phi}_{2(k)}, \cdots, \tilde{\phi}_{p(k)})$, $k = 1,2,\cdots,b$. Note that $\tilde{\delta}_j$ is simulated from (5.4.6) and it is a constant across all populations. $\tilde{\beta}_j = (\tilde{\beta}_{j(1)}^T, \cdots, \tilde{\beta}_{j(b)}^T)^T$ are simulated from (5.4.5) given $\tilde{\delta}_j$. $\tilde{\beta}_{j(k)}$ varies from population to population. When the simulated values $\tilde{\theta}_k = (\tilde{\mu}_k, \tilde{\Sigma})$ are derived from the reverse transformation, both $\tilde{\mu}_k$ and $\tilde{\Sigma}$ vary across the populations. Denoting the simulated covariance matrix of k^{th} population by $\tilde{\Sigma}_k$, a simulated common covariance matrix can be obtained from

$$\tilde{\Sigma} = \frac{1}{b} \sum_{k=1}^{b} \tilde{\Sigma}_k \tag{5.4.7}$$

Thus it has been shown that $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \cdots, \tilde{\theta}_b)$, where $\tilde{\theta}_k = (\tilde{\mu}_k, \tilde{\Sigma})$, are the simulated values from the observed-data posterior distribution of $\theta = (\theta_1, \theta_2, \cdots, \theta_b)$ under the assumption of the ante-dependence structure of order g for the common covariance matrix Σ.

After the simulated values of $\theta = (\theta_1, \theta_2, \cdots, \theta_b)$ are obtained under the assumption of ante-dependence structure of order g, the imputation of the missing data of Y_{mis} are conducted as follows. The missing data Y_{mis} are imputed observation by observation within each population, the $y_{i(mis)} \mid d_k = 1$ are imputed by random draws from the conditional predictive distribution $P(y_{i(mis)} \mid y_{i(obs)}, \tilde{\mu}_k, \tilde{\Sigma})$. The covariance matrix of this distribution is the simulated common covariance matrix given in (5.4.7). In this way, the imputed values of Y_{mis} preserve the structured correlations between Y_{mis} and Y_{obs} across the populations, and the missing values of $y_{i(mis)}$ are drawn from their pertinent populations. In order to have m sets of proper multiple imputations, m independently simulated values $\tilde{\theta}^{(t)}$ $(t = 1, \cdots, m)$ are obtained from the reverse transformation $\tilde{\theta}^{(t)} = g^{-1}(\tilde{\phi}^{(t)})$, where $\tilde{\phi}^{(t)} = (\tilde{\phi}_1^{(t)}, \cdots, \tilde{\phi}_p^{(t)})$ are the independent draws from the observed-data posterior distribution (5.4.5)-(5.4.6). For each simulated value $\tilde{\theta}^{(t)}$, Y_{mis} are imputed from the corresponding conditional posterior predictive distribution $P(Y_{mis} \mid Y_{obs}, \tilde{\theta}^{(t)})$. Let $Y_{mis}^{(t)}$ be the imputed values of Y_{mis} from $P(Y_{mis} \mid Y_{obs}, \tilde{\theta}^{(t)})$, then $\{Y_{mis}^{(1)}, \cdots, Y_{mis}^{(m)}\}$ consists of a proper multiple imputation of Y_{mis} since each $Y_{mis}^{(t)}$ are independently drawn from the approximate posterior predictive distribution $P(Y_{mis} \mid Y_{obs})$.

5.5 Summary

Multiple imputation includes building an imputation model, imputing the missing data from this imputation model and making combined multiple imputation inference. In this chapter, we discussed how to build the imputation model and generate proper multiple imputation from the imputation model. In section 5.1 we discussed how to generate proper multiple imputation from the conditional predictive distribution. We also reviewed the predictive model method. When incomplete data have a monotone missing pattern, the observed-data posterior distribution of the regression parameters of Y_j on $Y_1, \cdots, Y_{j-1}$ is derived by Rubin (1987). The missing values of Y_j are imputed by an univariate approach. In section 5.2, a multivariate approach of multiple imputation is proposed. When incomplete data are from the multivariate normal distribution with a general covariance structure, the normal parameters are estimated or simulated by the reverse transformation after the alternative parameters are independently estimated or simulated.

The missing data Y_{mis} are imputed from the conditional predictive distribution given Y_{obs} and the simulated normal parameters, which is either a multivariate normal or univariate normal distribution. In section 5.3, method for imputing Y_{mis} is proposed when the covariance matrix has an ante-dependence structure. Ante-dependence structure is a natural interpretation of ordered response variables through successive regression model. This model reduces the number of covariance parameters from $p(p+1)/2$ to $(g+1)(2p-g)/2$, so it results in a more efficient estimation since the smaller the number of unknown parameters, the smaller the variance of the estimation (Altham, 1984). In section 5.4, method for imputing Y_{mis} is proposed when incomplete data are from multiple multivariate normal distributions with varied mean vectors and a common structured covariance matrix. Different mean vectors and common covariance matrix are simulated from the observed-data posterior distribution, the missing data are imputed for given simulated mean vectors and covariance matrix within each population.

Chapter 6. Non-Monotone Incomplete Data

In chapter 5 we proposed methods for imputing the missing data when the incomplete data have a monotone missing pattern. In such a case, the observed-data posterior distribution of the unknown parameters can be analytically derived in a closed form. However, when the incomplete data have a non-monotone missing pattern, the observed-data posterior distribution may not be easily derived in a closed form. Subsequently, simulating the unknown parameters from their observed-data posterior distribution will face a lot challenge. In this chapter we work on methods of imputing the missing data when incomplete data have a non-monotone missing pattern. In section 6.1, we review the implementation of the EM algorithm. When incomplete data are from the multivariate normal distribution, the mean vector and the covariance matrix can be quite easily estimated by the EM algorithm. In section 6.2, we discuss the imputation when incomplete data are from a single multivariate normal distribution with an unstructured covariance matrix. In section 6.3, we discuss the same procedures as in section 6.2 except the covariance matrix is assumed to have a special structure, the ante-dependence structure. In section 6.4, the incomplete data are assumed to be from more than one multivariate normal distributions. The procedures for imputing the missing data are discussed. In section 6.5, a brief summary is presented.

6.1 Implementation of the EM Algorithm

The EM algorithm is a likelihood-based iterative technique for estimating the normal parameters θ when data are incomplete. When data are incomplete, the maximum likelihood estimates of θ are the maximizer of the observed-data loglikelihood function $\ell(\theta \mid Y_{obs})$. This function itself, however, may be very complicated due to the incompleteness of the data. Thus the maximization of this function may become very difficult. The EM algorithm does not maximize the observed-data loglikelihood directly, it fills in the functions of missing data appearing in the complete data loglikelihood by their expected values, and then maximizes the filled in complete-data loglikelihood. Specifically, the functions of missing data appearing in the complete-data loglikelihood are filled in by (Dempster, Laird and Rubin, 1987)

$$Q(\theta \mid \theta^{(t)}) = \int \ell(\theta \mid Y) P(Y_{mis} \mid Y_{obs}, \theta^{(t)}) dY_{mis} \qquad (6.1.1)$$

where $\ell(\theta \mid Y)$ is the complete-data loglikelihood and $P(Y_{mis} \mid Y_{obs}, \theta^{(t)})$ is the conditional predictive distribution of Y_{mis} given Y_{obs} and current estimates $\theta^{(t)}$. The function $Q(\theta \mid \theta^{(t)})$

is equivalent to the complete-data loglikelihood once the functions of missing data have been filled in. The maximization with respect to θ then can be performed on this function. In this section, we review the application of the EM algorithm to estimate the mean vector and the covariance matrix from incomplete multivariate normal data with an arbitrary missing pattern. These works are due to Orchard and Woodbury (1972) and Beale and Little (1975).

Let $Y=(y_1, y_2, \cdots, y_n)^T$ be the data matrix, where $y_i=(y_{i1}, \cdots, y_{ip})^T$ is the i^{th} observation. We assume $y_i \sim N_p(\mu, \Sigma)$, where $\mu=(\mu_1, \cdots, \mu_p)^T$ is the mean vector and $\Sigma=(\sigma_{ij})$ is an unstructured covariance matrix. To implement the EM algorithm, we need to assume that the missingness mechanism is ignorable.

The E step of the EM algorithm is to fill in the functions of missing data appearing in the complete-data loglikelihood by their expected values. In multivariate normal situation, these functions are the sufficient statistics

$$\sum_{i=1}^{n} y_{ij} \text{ and } \sum_{i=1}^{n} y_{ij} y_{ik}, \qquad j,k=1,\cdots,p$$

The expected values of these sufficient statistics, given current estimates $\theta^{(t)}$, are

$$E\left(\sum_{i=1}^{n} y_{ij} \mid y_{obs}, \theta^{(t)}\right)=\sum_{i=1}^{n} y_{ij}^{(t)}, \qquad j=1,\cdots,p \tag{6.1.2}$$

$$E\left(\sum_{i=1}^{n} y_{ij} y_{ik} \mid y_{obs}, \theta^{(t)}\right)=\sum_{i=1}^{n} (y_{ij}^{(t)} y_{ik}^{(t)} + c_{jki}^{(t)}), \qquad j,k=1,\cdots,p \tag{6.1.3}$$

where

$$y_{ij}^{(t)}=\begin{cases} y_{ij} & \text{if } y_{ij} \text{ is observed} \\ E(y_{ij} \mid y_{i(obs)}, \theta^{(t)}) & \text{if } y_{ij} \text{ is missing} \end{cases}$$

and

$$c_{jki}^{(t)}=\begin{cases} 0 & \text{if } y_{ij} \text{ or } y_{ik} \text{ are observed} \\ Cov(y_{ij}, y_{ik} \mid y_{i(obs)}, \theta^{(t)}) & \text{if } y_{ij} \text{ and } y_{ik} \text{ are missing} \end{cases}$$

The missing values $y_{i(mis)}$ are replaced by the conditional means of $y_{i(mis)}$ given the current estimates $\theta^{(t)}$ and $y_{i(obs)}$. The expected sufficient statistics are filled in by these conditional means and nonzero conditional covariance of the missing values. The conditional means of $y_{i(mis)}$ are the multivariate linear regression of $y_{i(mis)}$ on $y_{i(obs)}$. The regression parameters

are obtained from the given normal parameters $\theta^{(t)}$, not obtained from the estimation based on the data. They can be obtained by performing the sweep operator on the $\theta^{(t)}$-matrix.

The M step of the EM algorithm is straightforward. The new estimates $\theta^{(t+1)}$ are calculated from the filled-in complete data sufficient statistics,

$$\mu_j^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} y_{ij}^{(t)}, \qquad j = 1, \cdots, p \tag{6.1.4}$$

$$\sigma_{jk}^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\left[(y_{ij}^{(t)} - \mu_j^{(t+1)})(y_{ik}^{(t)} - \mu_k^{(t+1)}) + c_{jki}^{(t)}\right], \qquad j,k = 1, \cdots, p. \tag{6.1.5}$$

The convergence of the sequence $\{\theta^{(t)}\}$ is discussed by Dempster, Laird and Rubin (1977), and Wu (1983). Under general conditions, the estimates $\{\theta^{(t)}\}$ will converge reliably to its stationary point $\hat{\theta}$, which is the maximum likelihood estimates of θ from the observed-data loglikelihood function $\ell(\theta \mid Y_{obs})$. This convergence can be easily evaluated by the values of the observed-data loglikelihood function $\{\ell(\theta^{(t)} \mid Y_{obs})\}$.

The remaining issue to implement the EM algorithm is to choose the initial values of θ for the first iteration. Several choices of initial values are described by Little and Rubin (1987). Unless the fractions of missing information for some components of θ are very high, the choice of initial values is usually not crucial. For simplicity, we use the sample means and variances as the initial values for the means and variances and set all correlations equal to zero.

6.2 Imputation for Single Population With Unstructured Covariance

Suppose that we have a random sample of size n from a multivariate normal distribution $N_p(\mu, \Sigma)$ with the mean vector μ and the covariance matrix Σ assumed to be the same for all samples. That is, all samples are from a single multivariate normal distribution. Furthermore, the covariance matrix $\Sigma = (\sigma_{ij})$ is assumed to be unstructured. When the data are incomplete, and the missing pattern is non-monotone, we use $Y_{obs.j}$ and n_j to denote the observed values and the number of observed values of variable Y_j $(j = 1,2,\cdots, p)$ respectively, and use $y_{i(obs)}$ and $y_{i(mis)}$ to denote the observed portion and missing portion of the i^{th} observation respectively.

If the missingness mechanism is ignorable, any inference on θ based on the observed-data likelihood function is valid. Due to its simplicity and stability, we use the EM algorithm

reviewed in section 6.1 to estimate the normal parameters $\theta = (\mu, \Sigma)$. Denoting the maximum likelihood estimates of θ by $\hat{\theta} = (\hat{\mu}, \hat{\Sigma})$. By the parameter transformation, we obtain the maximum likelihood estimates of the alternative parameters $\hat{\phi} = g(\hat{\theta})$. where $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p)$ and $\hat{\phi}_j = (\hat{\beta}_j, \hat{\delta}_j)$. Note that the EM estimates $\hat{\theta} = (\hat{\mu}, \hat{\Sigma})$ are based on all observed data, that is, $\hat{\theta} = (\hat{\mu}, \hat{\Sigma})$ are obtained from the joint observed-data likelihood function

$$L(\mu, \Sigma \mid Y_{obs}) = \prod_{i=1}^{n} L(\mu_i, \Sigma_i \mid y_{i(obs)}) \tag{6.2.1}$$

where

$$L(\mu_i, \Sigma_i \mid y_{i(obs)}) \propto |\Sigma_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_{i(obs)} - \mu_i)^T \Sigma_i^{-1} (y_{i(obs)} - \mu_i)\right\} \tag{6.2.2}$$

is the observed-data likelihood function for the i^{th} observation. The μ_i and Σ_i are the marginal mean vector and marginal covariance matrix corresponding to $y_{i(obs)}$ respectively.

When incomplete data have a non-monotone missing pattern, the successive regression method described in section 5.1 can not be applied directly since the conditional distribution of Y_j given $Y_1, \cdots, Y_{j-1}$ can not be determined due to the missing values of $Y_1, \cdots, Y_{j-1}$. Therefore, we impute the missing data sequentially. First, impute the missing data of variable Y_1, and then impute the missing data of Y_2 conditioned on the imputed complete data of Y_1. Continuing this process until the missing data of Y_p are imputed conditioned on the imputed complete data of $Y_1, Y_2, \cdots, Y_{p-1}$. To do this, we first look at the marginal observed-data likelihood function and the observed-data posterior distribution of the alternative parameters $\phi_1 = (\mu_1, \delta_1)$. For simplicity, we assume that the first n_1 values of Y_1 are observed and the remaining of $m_1 = n - n_1$ values are missing. The marginal observed-data likelihood function is

$$L(\mu_1, \delta_1 \mid Y_{obs,1}) \propto \delta_1^{-\frac{n_1}{2}} \exp\left\{-\frac{1}{2\delta_1}\sum_{i=1}^{n_1}(y_{i1} - \mu_1)^2\right\} \tag{6.2.3}$$

The maximum likelihood estimate of μ_1 based on this marginal observed-data likelihood is $\frac{1}{n_1}\sum_{i=1}^{n_1} y_{i1}$.

If a non-informative prior $\pi(\phi)=\delta_1^{-1}\delta_2^{-1}\cdots\delta_p^{-1}$ is assumed for the alternative parameters $\phi=(\phi_1,\cdots,\phi_p)$, the Bayesian inference on ϕ_1 is independent of the inferences on $\phi_2,\cdots,\phi_p$. Hence the marginal observed-data posterior distribution of $\phi_1=(\mu_1,\delta_1)$ can be expressed as

$$P(\mu_1,\delta_1\mid Y_{obs,1}) \propto \delta_1^{-\frac{n_1}{2}}\exp\left\{-\frac{1}{2\delta_1}\sum_{i=1}^{n_1}(y_{i1}-\mu_1)^2\right\}\times\delta_1^{-1} \qquad (6.2.4)$$

This marginal observed-data posterior distribution density function can be decomposed as

$$P(\mu_1,\delta_1\mid Y_{obs,1}) \propto \delta_1^{-\frac{n_1}{2}}\exp\left\{-\frac{1}{2\delta_1}\sum_{i=1}^{n_1}(y_{i1}-\mu_1)^2\right\}\times\delta_1^{-1}$$

$$=\delta_1^{-\frac{n_1}{2}-1}\exp\left\{-\frac{n_1}{2\delta_1}(\mu_1-\hat{\mu}_1^*)^2\right\}\exp\left\{-\frac{1}{2\delta_1}\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1^*)^2\right\}\exp\left\{-\frac{(\mu_1-\hat{\mu}_1^*)}{\delta_1}\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1^*)\right\}$$

where $\hat{\mu}_1^*$ is a given value within the admissible range of μ_1. The term $\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1^*)$ is not necessarily equal to zero. However, if $\hat{\mu}_1^*$ is the maximum likelihood estimate of μ_1 based on the marginal observed-data likelihood function (6.2.3), $\hat{\mu}_1^*=\frac{1}{n_1}\sum_{i=1}^{n_1}y_{i1}$, then $\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1^*)=0$. In this case, the marginal observed-data posterior distribution (6.2.4) becomes

$$P(\mu_1,\delta_1\mid Y_{obs,1}) \propto \delta_1^{-\frac{1}{2}}\exp\left\{-\frac{n_1}{2\delta_1}(\mu_1-\hat{\mu}_1^*)^2\right\}\times\delta_1^{-(\frac{n_1-1}{2})-1}\exp\left\{-\frac{1}{2\delta_1}\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1^*)^2\right\}$$

which is the product of a conditional normal density and a marginal scaled inverted-chisquare density with the degrees of freedom n_1-1, i.e.

$$\mu_1\mid Y_{obs,1},\delta_1 \sim N_1(\hat{\mu}_1^*,\frac{1}{n_1}\delta_1) \qquad (6.2.5)$$

$$\delta_1\mid Y_{obs,1} \sim \sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1^*)^2\chi^{-2}_{n_1-1} \qquad (6.2.6)$$

Thus the exact observed-data posterior distribution of $\phi_1=(\mu_1,\delta_1)$ is a normal scaled inverted-chisquare distribution. Note that $\hat{\mu}_1^*$ is the maximizer of the marginal observed-data likelihood (6.2.3), it utilizes the observed values of Y_1 only. The correlations between Y_1 and $Y_2,\cdots,Y_p$

does not influence the estimation of μ_1 and δ_1, the observed values of $Y_2,\cdots,Y_p$ do not make any contribution to this estimation.

In order to incorporate the influence of the correlations between Y_1 and $Y_2,\cdots,Y_p$ on the estimation of μ_1 and δ_1, we look for some estimators which can utilize the observed values of Y_1 as well as the correlations between Y_1 and $Y_2,\cdots,Y_p$. Suppose that $\hat{\mu}_1$ is the maximum likelihood estimate of μ_1 by the EM algorithm based on the joint observed-data likelihood function (6.2.1), then this estimation of μ_1 utilizes all observed values through the correlations between Y_1 and $Y_2,\cdots,Y_p$. These two estimates $\hat{\mu}_1$ and $\hat{\mu}_1^*$ are both unbiased, and if there is no missing value for Y_1, they will be identical. Thus $\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1)=0$, and the observed-data posterior distribution of (μ_1,δ_1) can be approximated by

$$P(\mu_1,\delta_1 \mid y_{obs}) \approx \delta_1^{-\frac{1}{2}}\exp\left\{-\frac{n_1}{2\delta_1}(\mu_1-\hat{\mu}_1)^2\right\}\times\delta_1^{-(\frac{n_1-1}{2})-1}\exp\left\{-\frac{1}{2\delta_1}\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1)^2\right\}$$

i.e.

$$\mu_1 \mid Y_{obs,1},\delta_1 \approx N_1(\hat{\mu}_1,\frac{1}{n_1}\delta_1) \tag{6.2.7}$$

$$\delta_1 \mid Y_{obs,1} \approx \sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1)^2\chi_{n_1-1}^{-2} \tag{6.2.8}$$

Note that the distributions (6.2.5)-(6.2.6) and (6.2.7)-(6.2.8) have the same shape, they only differ at the locations $\hat{\mu}_1^*$ and $\hat{\mu}_1$, and differ at the scales $\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1^*)^2$ and $\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1)^2$. Since $E(\hat{\mu}_1^*)=E(\hat{\mu}_1)=\mu_1$, these two distributions approximately center at the same location. But since $\sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1^*)^2 \le \sum_{i=1}^{n_1}(y_{i1}-\hat{\mu}_1)^2$, the approximate observed-data posterior distribution (6.2.7)-(6.2.8) always has a larger variation than the exact observed-data posterior distribution. This is the penalty we will pay by simulating $\phi_1=(\mu_1,\delta_1)$ from the approximate observed-data posterior distribution rather than simulating it from the exact posterior distribution.

The missing values of Y_1 are imputed in the following way. A simulated value $\tilde{\delta}_1$ is obtained by letting $\tilde{\delta}_1 = \sum_{i=1}^{n_1}(y_{i1} - \hat{\mu}_1)^2 / \delta^*$, where δ^* is a random draw from $\chi^2_{n_1-1}$, and then a simulated value $\tilde{\mu}_1$ is randomly drawn from (6.2.7) given $\tilde{\delta}_1$. After simulated values $(\tilde{\mu}_1, \tilde{\delta}_1)$ have been obtained from their approximate observed-data posterior distribution, a single set of m_1 missing values of Y_1 is imputed by m_1 iid random draws from the conditional predictive distribution $P(Y_{mis,1} \mid Y_{obs}, \tilde{\mu}_1, \tilde{\delta}_1)$, which is $N_1(\tilde{\mu}_1, \tilde{\delta}_1)$. In order to obtain m sets of proper multiple imputations, m simulated values $(\tilde{\mu}_1^{(t)}, \tilde{\delta}_1^{(t)})$ $t = 1, \cdots, m$ are independently drawn from (6.2.7)-(6.2.8). Corresponding to each simulated values $(\tilde{\mu}_1^{(t)}, \tilde{\delta}_1^{(t)})$, m_1 iid random draws are generated from $N_1(\tilde{\mu}_1^{(t)}, \tilde{\delta}_1^{(t)})$, hence $(Y_{mis,1}^{(1)}, Y_{mis,1}^{(2)}, \cdots, Y_{mis,1}^{(m)})$ consists of a proper multiple imputations of $Y_{mis,1}$.

Once the m sets of imputed values of Y_1 have been generated, imputing the missing values of Y_j $(j = 2,3,\cdots, p)$ can be conducted conditioned on the imputed complete data of $Y_1, \cdots, Y_{j-1}$. Suppose that the missing values of $Y_1, \cdots, Y_{j-1}$ have been imputed m times, let $X_j^{(t)}$ be the data matrix consisting of the observed and t^{th} imputed values of $Y_1, \cdots, Y_{j-1}$. The rows of $X_j^{(t)}$ are corresponding to the observed values of Y_j, and the first column is augmented by 1 to incorporate the intercept term. Therefore, $X_j^{(t)}$ is a $n_j \times j$ data matrix. The conditional observed-data likelihood function of Y_j given $Y_1, \cdots, Y_{j-1}$, in terms of $\phi_j = (\beta_j, \delta_j)$, where $\beta_j = (\beta_{j,0}, \beta_{j,1}, \cdots, \beta_{j,j-1})$, can be expressed as

$$L(\beta_j, \delta_j \mid Y_{obs}) \propto \delta_j^{-\frac{n_j}{2}} \exp\left\{-\frac{1}{2\delta_j}(Y_{obs,j} - X_j^{(t)}\beta_j)^T (Y_{obs,j} - X_j^{(t)}\beta_j)\right\} \qquad (6.2.9)$$

and the conditional observed-data posterior distribution of $\phi_j = (\beta_j, \delta_j)$ is given by

$$L(\beta_j, \delta_j \mid Y_{obs}) \propto \delta_j^{-\frac{n_j}{2}} \exp\left\{-\frac{1}{2\delta_j}(Y_{obs,j} - X_j^{(t)}\beta_j)^T (Y_{obs,j} - X_j^{(t)}\beta_j)\right\} \times \delta_j^{-1} \qquad (6.2.10)$$

Now suppose we have obtained the maximum likelihood estimates $\hat{\theta} = (\hat{\mu}, \hat{\Sigma})$ by applying the EM algorithm to the joint observed-data likelihood function (6.2.1). By performing the sweep

operator on these estimates, we obtain the maximum likelihood estimates of the alternative parameters $\hat{\phi}_j = (\hat{\beta}_j, \hat{\delta}_j)$. Then the observed-data posterior distribution (6.2.10) can be approximated by

$$P(\phi_j \mid Y_{obs}) \propto \delta_j^{-\frac{j}{2}} \exp\left\{-\frac{1}{2\delta_j}(\beta_j - \hat{\beta}_j)^T X_j^{(t)T} X_j^{(t)} (\beta_j - \hat{\beta}_j)\right\} \times \delta_j^{-(\frac{n_j - j}{2})-1} \exp\left\{-\frac{1}{2\delta_j}\hat{\varepsilon}_j{}^T \hat{\varepsilon}_j\right\}$$

where $\hat{\beta}_j = (\hat{\beta}_{j,0}, \hat{\beta}_{j,1}, \cdots, \hat{\beta}_{j,j-1})$ and $\hat{\varepsilon}_j = Y_{obs,j} - X_j^{(t)} \hat{\beta}_j$. Thus the observed-data posterior distribution of $\phi_j = (\beta_j, \delta_j)$ is an approximate normal scaled inverted-chisquare distribution,

$$\beta_j \mid Y_{obs}, \delta_j \;=\; N_j(\hat{\beta}_j, \delta_j (X_j^{(t)T} X_j^{(t)})^{-1}) \tag{6.2.11}$$

$$\delta_j \mid Y_{obs} \;=\; (\mathrm{Y}_{\mathrm{obs,j}} - X_j^{(t)} \hat{\beta}_j)^T (\mathrm{Y}_{\mathrm{obs,j}} - X_j^{(t)} \hat{\beta}_j) \chi^{-2}_{n_j - j} \tag{6.2.12}$$

The values of the parameters $\phi_j = (\beta_j, \delta_j)$ can be simulated from this approximate observed-data posterior distribution. Let $\tilde{\delta}_j = (Y_{obs,j} - X_j^{(t)} \hat{\beta}_j)^T (Y_{obs,j} - X_j^{(t)} \hat{\beta}_j) / \delta^*$, where δ^* is a random draw from $\chi^2_{n_j - j}$, then β_j can be simulated from $N_j(\hat{\beta}_j, \tilde{\delta}_j (X_j^{(t)T} X_j^{(t)})^{-1})$ given $\tilde{\delta}_j$. Since the observations of Y_j are conditionally independent of each other given $\phi_j = (\beta_j, \delta_j)$ and $Y_1, \cdots, Y_{j-1}$, once the parameters $\phi_j = (\beta_j, \delta_j)$ have been simulated, the missing values of Y_j are then imputed observation by observation. Let $X_{-j}^{(t)}$ be the data matrix consisting of the observed and the t^{th} imputed values of $Y_1, \cdots, Y_{j-1}$ and augmented by 1 in the first column, the rows of $X_{-j}^{(t)}$ are corresponding to the missing values of Y_j. Then the missing values $Y_{mis,j}$ are imputed by a single random draw from their conditional predictive distribution $P(Y_{mis,j} \mid X_{-j}^{(t)}, \tilde{\beta}_j, \tilde{\delta}_j) = N_{m_j}(X_{-j}^{(t)} \tilde{\beta}_j \;,\; \tilde{\delta}_j I_{m_j})$, where $m_j = n - n_j$ and I_{m_j} is a $m_j \times m_j$ unity matrix. The m sets of proper multiple imputations are generated by taking a single random draw from the approximate observed-data posterior distribution (6.2.11)-(6.2.12) corresponding to each $X_j^{(t)}$ and $\hat{\beta}_j$, say $(\tilde{\beta}_j^{(t)}, \tilde{\delta}_j^{(t)})$ $t = 1, \cdots, m$, and then for each random draw $(\tilde{\beta}_j^{(t)}, \tilde{\delta}_j^{(t)})$, impute $Y_{mis,j}$ by taking a single random draw from each corresponding conditional predictive distribution of $Y_{mis,j}$, i.e. $N_{m_j}(X_{-j}^{(t)} \tilde{\beta}_j^{(t)} \;,\; \tilde{\delta}_j^{(t)} I_{m_j})$. Repeat this procedure

until all the missing values $Y_{mis,2},\cdots,Y_{mis,p}$ are imputed. $\left\{Y_{mis,1}^{(t)},Y_{mis,2}^{(t)},\cdots,Y_{mis,p}^{(t)}\right\}t=1,2,\cdots,m$ consist of a proper multiple imputation of Y_{mis}.

Note that if $(\hat{\beta}_j,\hat{\delta}_j)$ is the maximum likelihood estimates of (β_j,δ_j) from the conditional observed-data likelihood function (6.2.9) rather than from the joint observed-data likelihood function (6.2.1), the observed-data posterior distribution (6.2.11)-(6.2.12) will become an exact normal scaled inverted-chisquare distribution. In this case, the posterior distribution does not utilize the observed values of $Y_{j+1},\cdots,Y_p$. When the missing is non-monotone, these observed values will influence the estimation of (β_j,δ_j) through the correlations between $Y_1,\cdots,Y_j$ and $Y_{j+1},\cdots,Y_p$. If the missing is monotone, these observed values of $Y_{j+1},\cdots,Y_p$ will not make any contribution to the estimation of (β_j,δ_j) even though $Y_{j+1},\cdots,Y_p$ are correlated to $Y_1,\cdots,Y_j$, because the EM algorithm fills in the missing sufficient statistics by their expected values given the observed ones. Therefore, in the case of monotone missing, the estimates of (β_j,δ_j) using the EM algorithm from the joint observed-data likelihood function (6.2.1) and the estimates using the ordinary least squares from the conditional observed-data likelihood function (6.2.9) are identical. The observed-data posterior distribution is an exact normal scaled inverted-chisquare distribution centered at $\hat{\beta}_j$ and the scale of $(Y_{obs,j}-X_j^{(t)}\hat{\beta}_j)^T(Y_{obs,j}-X_j^{(t)}\hat{\beta}_j)$. Thus the method proposed in this section can be viewed as a generalization of the method proposed in the section 5.1.

6.3 Imputation for Single Population With Structured Covariance

In this section we have the same assumptions about the data as those in section 6.2. That is, the non-monotone incomplete multivariate data are assumed to be iid from a multivariate normal distribution $N_p(\mu,\Sigma)$ with a common mean vector μ and covariance matrix Σ. Except in this section we further assume that the covariance matrix Σ has an ante-dependence structure of order g. Under this assumption, the dimension of the unknown parameters $\theta=(\mu,\Sigma)$ is $p+(g+1)(2p-g)/2$.

The EM algorithm reviewed in section 6.1 is used to estimate $\theta=(\mu,\Sigma)$ under the general covariance structure. In order to obtain the estimated alternative parameters $\phi=(\phi_1,\phi_2,\cdots,\phi_p)$ under the ante-dependence structure, we use the transformation formula derived by Macchiavelli

and Arnold (1994). Let $\hat{\theta} = (\hat{\mu}, \hat{\Sigma})$, where $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \cdots, \hat{\mu}_p)$ and $\hat{\Sigma} = (\hat{\sigma}_{ij})$, be the EM estimates based on the joint observed-data likelihood function (6.2.1) without assumption of the ante-dependence structure. Then the maximum likelihood estimates of the alternative parameters $\phi = (\phi_1, \phi_2, \cdots, \phi_p)$ under the assumption of ante-dependence structure of order g can be obtained from the following transformation:

$$\hat{\phi}_1 = (\hat{\mu}_1, \hat{\delta}_1), \text{ where } \hat{\delta}_1 = \hat{\sigma}_{11} \tag{6.3.2}$$

and

$$\hat{\phi}_j = ((\hat{\beta}_{j,0}, \hat{\beta}_{j,j-g}, \cdots, \hat{\beta}_{j,j-1}), \hat{\delta}_j), \; j = 2,3,\cdots, p$$

where

$$\hat{\delta}_j = \hat{\Sigma}_{jj} - \hat{\Sigma}_{j,(j-g):(j-1)} \hat{\Sigma}^{-1}_{(j-g):(j-1),(j-g):(j-1)} \hat{\Sigma}_{(j-g):(j-1),j} \tag{6.3.3}$$

$$(\hat{\beta}_{j,j-g}, \cdots, \hat{\beta}_{j,j-1}) = \hat{\Sigma}_{(j-g):(j-1),j} \hat{\Sigma}^{-1}_{(j-g):(j-1),(j-g):(j-1)} \tag{6.3.4}$$

$$\hat{\beta}_{j,0} = \hat{\mu}_j - (\hat{\beta}_{j,j-g} \hat{\mu}_{j-g} + \cdots + \hat{\beta}_{j,j-1} \hat{\mu}_{j-1}) \tag{6.3.5}$$

Once the maximum likelihood estimates of these alternative parameters have been obtained, the observed-data posterior distribution of the alternative parameters can be approximated. The observed-data posterior distribution of $\phi_1 = (\mu_1, \delta_1)$ can be approximated by

$$\mu_1 \mid Y_{obs,1}, \delta_1 \;\; \approx \;\; N_1(\hat{\mu}_1, \frac{1}{n_1} \delta_1) \tag{6.3.6}$$

$$\delta_1 \mid Y_{obs,1} \;\; \approx \;\; \sum_{i=1}^{n_1} (y_{i1} - \hat{\mu}_1)^2 \chi^{-2}_{n_1 - 1} \tag{6.3.7}$$

A single simulated value $\tilde{\phi}_1 = (\tilde{\mu}_1, \tilde{\delta}_1)$ is randomly drawn from (6.3.6) and (6.3.7), and the m_1 missing values of Y_1 are imputed by m_1 iid random draws from the conditional predictive distribution $P(Y_{mis,1} \mid Y_{obs,1}, \tilde{\mu}_1, \tilde{\delta}_1)$, which is $N_1(\tilde{\mu}_1, \tilde{\delta}_1)$. The m sets of proper multiple imputations of $Y_{mis,1}$ are generated by independently drawing m simulated values $(\tilde{\mu}_1^{(t)}, \tilde{\delta}_1^{(t)})$ $t = 1, \cdots, m$ from the approximate observed-data posterior distribution (6.3.6) and (6.3.7). and then for each simulated value $(\tilde{\mu}_1^{(t)}, \tilde{\delta}_1^{(t)})$, taking m_1 iid random draws from $N_1(\tilde{\mu}_1^{(t)}, \tilde{\delta}_1^{(t)})$.

Suppose that the missing values of $Y_{j-g},\cdots,Y_{j-1}$ have been imputed m times. the missing values of Y_j $(j=2,3,\cdots,p)$ are then imputed conditioned on the imputed complete data of $Y_{j-g},\cdots,Y_{j-1}$. Let $X_j^{(t)}$ be the data matrix consisting of the observed and the t^{th} imputed values of $Y_{j-g},\cdots,Y_{j-1}$, the rows of $X_j^{(t)}$ are corresponding to the observed values of Y_j, and the first column is augmented by 1 to incorporate the intercept term. The dimension of the data matrix $X_j^{(t)}$ is $n_j\times(g+1)$. The conditional observed-data likelihood function of Y_j given $Y_{j-g},\cdots,Y_{j-1}$, in terms of $\phi_j=(\beta_j,\delta_j)$, where $\beta_j=(\beta_{j,0},\beta_{j,j-g},\cdots,\beta_{j,j-1})$, can be expressed as

$$L(\beta_j,\delta_j\mid Y_{obs}) \propto \delta_j^{-\frac{n_j}{2}}\exp\left\{-\frac{1}{2\delta_j}(Y_{obs,j}-X_j^{(t)}\beta_j)^T(Y_{obs,j}-X_j^{(t)}\beta_j)\right\} \qquad (6.3.8)$$

and the conditional observed-data posterior distribution of $\phi_j=(\beta_j,\delta_j)$, given the non-informative prior $\pi(\phi)=\delta_1^{-1}\delta_2^{-1}\cdots\delta_p^{-1}$, is given by

$$L(\beta_j,\delta_j\mid Y_{obs}) \propto \delta_j^{-\frac{n_j}{2}}\exp\left\{-\frac{1}{2\delta_j}(Y_{obs,j}-X_j^{(t)}\beta_j)^T(Y_{obs,j}-X_j^{(t)}\beta_j)\right\}\times\delta_j^{-1} \qquad (6.3.9)$$

This observed-data posterior distribution can be approximated by

$$\delta_j^{-\frac{g+1}{2}}\exp\left\{-\frac{1}{2\delta_j}(\beta_j-\hat{\beta}_j)^T X_j^{(t)T}X_j^{(t)}(\beta_j-\hat{\beta}_j)\right\}\times\delta_j^{-(\frac{n_j-g-1}{2})-1}\exp\left\{-\frac{1}{2\delta_j}\hat{\varepsilon}_j{}^T\hat{\varepsilon}_j\right\} \qquad (6.3.10)$$

where $\hat{\beta}_j=(\hat{\beta}_{j,0},\hat{\beta}_{j,j-g},\cdots,\hat{\beta}_{j,j-1})$ is the maximum likelihood estimates of $\beta_j=(\beta_{j,0},\beta_{j,j-g},\cdots,\beta_{j,j-1})$ obtained from (6.3.4)-(6.3.5), and $\hat{\varepsilon}_j=Y_{obs,j}-X_j^{(t)}\hat{\beta}_j$. The distribution (6.3.10) is a normal scaled inverted-chisquare distribution,

$$\beta_j\mid Y_{obs},\delta_j \;\approx\; N_{g+1}(\hat{\beta}_j,\delta_j(X_j^{(t)T}X_j^{(t)})^{-1}) \qquad (6.3.11)$$

$$\delta_j\mid Y_{obs} \;\approx\; (Y_{obs,j}-X_j^{(t)}\hat{\beta}_j)^T(Y_{obs,j}-X_j^{(t)}\hat{\beta}_j)\chi^{-2}_{n_j-g-1} \qquad (6.3.12)$$

After the values of $\phi_j=(\beta_j,\delta_j)$ have been simulated from this approximate observed-data posterior distribution, the missing values of Y_j can be imputed from the conditional predictive distribution of $Y_{mis,j}$ given the imputed complete data of $Y_{j-g},\cdots,Y_{j-1}$ and $\tilde{\phi}_j=(\tilde{\beta}_j,\tilde{\delta}_j)$. Let $X_{-j}^{(t)}$ be the data matrix consisting of the observed and the t^{th} imputed values of $Y_{j-g},\cdots,Y_{j-1}$

and augmented by 1 in the first column, the rows of $X_{-j}^{(t)}$ are corresponding to the missing values of Y_j. Then the missing values $Y_{mis,j}$ are imputed by a single random draw from $P(Y_{mis,j} \mid X_{-j}^{(t)}, \tilde{\beta}_j, \tilde{\delta}_j) = N_{m_j}(X_{-j}^{(t)}\tilde{\beta}_j \ , \ \tilde{\delta}_j I_{m_j})$, where $m_j = n - n_j$ and I_{m_j} is a $m_j \times m_j$ unity matrix. The m sets of proper multiple imputations are generated by taking a single random draw of β_j and δ_j from the approximate observed-data posterior distribution (6.3.11)-(6.3.12) corresponding to each $X_j^{(t)}$, say $(\tilde{\beta}_j^{(t)}, \tilde{\delta}_j^{(t)})$ $t = 1, \cdots, m$, and then for each random draw $(\tilde{\beta}_j^{(t)}, \tilde{\delta}_j^{(t)})$, impute $Y_{mis,j}$ by taking a single random draw from each corresponding conditional predictive distribution of $Y_{mis,j}$, i.e. $N_{m_j}(X_{-j}^{(t)}\tilde{\beta}_j^{(t)}, \tilde{\delta}_j^{(t)} I_{m_j})$. Repeat this procedure until all the missing values of $Y_{mis,2}, \cdots, Y_{mis,p}$ are imputed. $\left\{Y_{mis,1}^{(t)}, Y_{mis,2}^{(t)}, \cdots, Y_{mis,p}^{(t)}\right\}$ $t = 1,2,\cdots,m$ consist of a proper multiple imputation of Y_{mis}.

6.4 Imputation for Multiple Populations With Structured Covariance

In sections 6.2 and 6.3, we discussed the situations where incomplete multivariate data are from a single probability population. In this section we consider a probability model where incomplete multivariate data are from multiple probability populations which are indexed by the design factors. Suppose we have f design factors with levels $b_1, b_2, \cdots, b_f$ respectively, then the total number of probability populations formed by the f design factors is $b = b_1 \times b_2 \times \cdots \times b_f$. The design factors are usually treated as fixed, such as the treatment groups and the clinical sites in a clinical trial, so they are not subject to a probability distribution. Let $Y = (y_{11}, \cdots, y_{1N_1}, y_{21}, \cdots, y_{2N_2}, \cdots\cdots, y_{b1}, \cdots, y_{bN_b})^T$ be a $N \times p$ data matrix of the response variables, where $y_{ki} = (y_{ki1}, y_{ki2}, \cdots, y_{kip})$ is the i^{th} random sample from the k^{th} population, N_k is the number of samples from the k^{th} population, and $N = N_1 + \cdots + N_b$ is the total number of samples. We denote the number of observed values of Y_j in the k^{th} population by n_{kj}. It is assumed that these samples are from multivariate normal distributions

$$y_{ki} \ \sim \ N_p(\mu_k, \Sigma) \qquad (6.4.1)$$

The mean vectors μ_k $(k = 1,2,\cdots,b)$ are allowed to vary from population to population, but the covariance matrix Σ is assumed to be the same across the populations. Furthermore, the common covariance matrix Σ is assumed to have an ante-dependence structure of order g. Let $\theta = (\theta_1, \theta_2, \cdots, \theta_b)$ be the unknown parameters, where $\theta_k = (\mu_k, \Sigma)$ are the normal parameters pertinent to the k^{th} population. Then under this model assumption, the dimension of the unknown parameters is $k \times p + (g+1)(2p-g)/2$.

When incomplete data have a non-monotone missing pattern, the observed-data likelihood function of $\phi = (\phi_1, \cdots, \phi_p)$, where $\phi_j = (\phi_{1j}, \cdots, \phi_{bj})$ are the alternative parameters of Y_j given $Y_{j-g}, \cdots Y_{j-1}$ across the populations, can not be factored into a series of independent observed-data likelihood functions of each individual alternative parameter, ϕ_{kj}. In this case, we attempt to estimate the normal parameters $\theta = (\theta_1, \theta_2, \cdots, \theta_b)$ by the EM algorithm, and then get the estimates of the alternative parameters $\phi = (\phi_1, \cdots, \phi_p)$ by the parameter transformation.

The joint observed-data likelihood function of $\theta = (\theta_1, \theta_2, \cdots, \theta_b)$ is given by

$$L(\theta \mid Y_{obs}) = \prod_{k=1}^{b} L(\mu_k, \Sigma \mid Y_{obs}) = \prod_{k=1}^{b} \prod_{i=1}^{N_k} L(\mu_{ki}, \Sigma_{ki} \mid y_{ki(obs)}) \tag{6.4.2}$$

where

$$L(\mu_{ki}, \Sigma_{ki} \mid y_{ki(obs)}) \propto |\Sigma_{ki}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_{ki(obs)} - \mu_{ki})^T \Sigma_{ki}^{-1} (y_{ki(obs)} - \mu_{ki})\right\} \tag{6.4.3}$$

is the observed-data likelihood function of the i^{th} observation in the k^{th} population, μ_{ki} and Σ_{ki} are the marginal mean vector and marginal covariance matrix corresponding to $y_{ki(obs)}$ respectively.

The EM algorithm is applied to the joint observed-data likelihood function (6.4.2) to estimate the mean vectors and the common covariance matrix. The E-step of the EM algorithm is applied to fill in the missing sufficient statistics without any alteration. The missing values $y_{ki(mis)}$ and the non-zero conditional covariance of the missing values are replaced by the conditional means of $y_{ki(mis)}$ and the conditional covariance given the current estimates $\theta^{(t)}$ and $y_{ki(obs)}$. However, the M-step needs to be altered to estimate the different mean vectors $\mu_k = (\mu_{k1}, \mu_{k2}, \cdots, \mu_{kp})$

for the k^{th} population. The new estimates $\mu_k^{(t+1)}$ are calculated from the filled-in complete data sufficient statistics,

$$\mu_{kj}^{(t+1)} = \frac{1}{N_k}\sum_{i=1}^{N_k} y_{kij}^{(t)}, \qquad j = 1,\cdots,p \tag{6.4.4}$$

The common covariance matrix $\Sigma = (\sigma_{jt})$ is estimated without assumption of ante-dependence structure as described in section 6.1, i.e. the elements σ_{jt} of Σ are calculated by

$$\sigma_{jt}^{(t+1)} = \frac{1}{n}\sum_{k=1}^{b}\sum_{i=1}^{N_k}\left[(y_{kij}^{(t)} - \mu_{kj}^{(t+1)})(y_{kit}^{(t)} - \mu_{kt}^{(t+1)}) + c_{jti}^{(t)}\right], \quad j,t = 1,\cdots,p \tag{6.4.5}$$

where $y_{kij}^{(t)}$, $y_{kit}^{(t)}$ and $c_{jti}^{(t)}$ are given in (6.1.3) and (6.1.4) respectively.

After the maximum likelihood estimates of the normal parameters $\theta_k = (\mu_k, \Sigma)$ $k = 1,2,\cdots,b$ have been obtained under the general covariance structure, the maximum likelihood estimates of the alternative parameters $\phi = (\phi_1,\cdots,\phi_p)$, under the assumption of ante-dependence structure of order g, can be obtained by the following parameter transformation formula which is modified from the one given by Macchiavelli and Arnold(1994).

$$\hat{\phi}_{k1} = (\hat{\mu}_{k,1}, \hat{\delta}_1), \text{ where } \hat{\delta}_1 = \hat{\sigma}_{11} \tag{6.4.6}$$

and

$$\hat{\phi}_{kj} = ((\hat{\beta}_{kj,0}, \hat{\beta}_{kj,j-g},\cdots,\hat{\beta}_{kj,j-1}), \hat{\delta}_j), \; j = 2,3,\cdots,p \text{ and } k = 1,2,\cdots,b$$

where

$$\hat{\delta}_j = \hat{\Sigma}_{jj} - \hat{\Sigma}_{j,(j-g):(j-1)}\hat{\Sigma}^{-1}_{(j-g):(j-1),(j-g):(j-1)}\hat{\Sigma}_{(j-g):(j-1),j} \tag{6.4.7}$$

$$(\hat{\beta}_{kj,j-g},\cdots,\hat{\beta}_{kj,j-1}) = \hat{\Sigma}_{(j-g):(j-1),j}\hat{\Sigma}^{-1}_{(j-g):(j-1),(j-g):(j-1)} \tag{6.4.8}$$

$$\hat{\beta}_{kj,0} = \hat{\mu}_{kj} - (\hat{\beta}_{kj,j-g}\hat{\mu}_{k,j-g} + \cdots + \hat{\beta}_{kj,j-1}\hat{\mu}_{k,j-1}) \tag{6.4.9}$$

Note that $\hat{\phi}_j = (\hat{\phi}_{1j},\cdots,\hat{\phi}_{bj})$. The regression coefficients $\hat{\beta}_{kj} = (\hat{\beta}_{kj,0}, \hat{\beta}_{kj,j-g},\cdots,\hat{\beta}_{kj,j-1})$ vary across the populations and the residual variance δ_j is a constant across all the populations.

If the non-informative prior distribution $\pi(\phi) = \delta_1^{-1}\delta_2^{-1}\cdots\delta_p^{-1}$ is assumed for the alternative parameters $\phi = (\phi_1,\cdots,\phi_p)$, then the observed-data posterior distribution of ϕ_{k1} $(k = 1,\cdots,b)$, where $\phi_{k1} = (\mu_{k,1}, \delta_1)$, can be approximated by

$$\mu_{k,1} \mid Y_{obs,1}, \delta_1 \approx N_1(\hat{\mu}_{k,1}, \frac{1}{n_1}\delta_1) \tag{6.4.10}$$

$$\delta_1 \mid Y_{obs,1} \approx \sum_{k=1}^{b}\sum_{i=1}^{n_{k1}}(y_{ki1} - \hat{\mu}_{k,1})^2 \chi^{-2}_{n_1-1} \tag{6.4.11}$$

where $n_1 = n_{11} + n_{21} + \cdots + n_{b1}$ is the total number of observed values of Y_1 across all the populations.

A single simulated value $\tilde{\phi}_1 = (\tilde{\phi}_{11}, \tilde{\phi}_{21}, \cdots, \tilde{\phi}_{b1})$ is randomly drawn from (6.4.10) and (6.4.11), and the missing values of Y_1 are imputed by random draws from the conditional predictive distributions $P(Y_{mis,1} \mid Y_{obs,1}, \tilde{\mu}_{k,1}, \tilde{\delta}_1)$, which are $N_1(\tilde{\mu}_{k,1}, \tilde{\delta}_1)$ $k = 1,2,\cdots,b$. That is, $Y_{mis,1}$ are imputed from the different populations. For given k^{th} population, which is indexed by $\tilde{\phi}_{k1} = (\tilde{\mu}_{k,1}, \tilde{\delta}_1)$, a single set of imputed values of $y_{ki(mis)}$ is obtained by taking iid random draws from the corresponding conditional predictive distribution, that is $N_1(\tilde{\mu}_{k,1}, \tilde{\delta}_1)$. The m sets of proper multiple imputations of $Y_{mis,1}$ are generated by independently drawing m sets of simulated values $(\tilde{\mu}_{k,1}^{(t)}, \tilde{\delta}_1^{(t)})$ $k = 1,\cdots,b$, $t = 1,\cdots,m$ from the approximate observed-data posterior distribution (6.4.10) and (6.4.11). Then for each set of simulated values $(\tilde{\mu}_{k,1}^{(t)}, \tilde{\delta}_1^{(t)})$ $k = 1,2,\cdots,b$, a single set of imputed values of $y_{ki(mis)}$ is obtained by taking iid random draws from the corresponding conditional predictive distributions $N_1(\tilde{\mu}_{k,1}^{(t)}, \tilde{\delta}_1^{(t)})$.

Suppose that the missing values of $Y_{j-g}, \cdots, Y_{j-1}$ have been imputed m times, the imputed values of $Y_{mis,j}$ $(j = 2,3,\cdots,p)$ are generated conditioned on the imputed complete data of $Y_{j-g}, \cdots, Y_{j-1}$. Let $X_{kj}^{(t)}$ be the data matrix consisting of the observed and the t^{th} imputed values of $Y_{j-g}, \cdots, Y_{j-1}$ in the k^{th} population, the rows of $X_{kj}^{(t)}$ are corresponding to the observed values of Y_j in the k^{th} population, the first column of $X_{kj}^{(t)}$ is augmented by 1 to incorporate the intercept term. Therefore, $X_{kj}^{(t)}$ is a $n_{kj} \times (g+1)$ data matrix. The conditional observed-data likelihood function of Y_j given $Y_{j-g}, \cdots, Y_{j-1}$, in terms of $\phi_j = (\phi_{1j}, \cdots, \phi_{bj})$, where $\phi_{kj} = (\beta_{kj}, \delta_j)$, and $\beta_{kj} = (\beta_{kj,0}, \beta_{kj,j-g}, \cdots, \beta_{kj,j-1})$, can be expressed as

$$L(\phi_j \mid Y_{obs}) \propto \prod_{k=1}^{b} \delta_j^{-\frac{n_{kj}}{2}} \exp\left\{-\frac{1}{2\delta_j}(Y_{obs.j} - X_{kj}^{(t)}\beta_{kj})^T (Y_{obs.j} - X_{kj}^{(t)}\beta_{kj})\right\} \qquad (6.4.12)$$

and the conditional observed-data posterior distribution of $\phi_j = (\phi_{1j}, \cdots, \phi_{bj})$ is given by

$$P(\phi_j \mid Y_{obs}) \propto \prod_{k=1}^{b} P(\phi_{kj} \mid Y_{obs})$$

$$= \prod_{k=1}^{b} \delta_j^{-\frac{n_{kj}}{2}} \exp\left\{-\frac{1}{2\delta_j}(Y_{obs.j} - X_{kj}^{(t)}\beta_{kj})^T (Y_{obs.j} - X_{kj}^{(t)}\beta_{kj})\right\} \times \delta_j^{-1} \qquad (6.4.13)$$

The observed-data posterior distribution $P(\phi_{kj} \mid Y_{obs})$ of $\phi_{kj} = (\beta_{kj}, \delta_j)$ can be approximated by

$$\delta_j^{-\frac{g+1}{2}-(\frac{n_j-g-1}{2})-1} \exp\left\{-\frac{1}{2\delta_j}(\beta_{kj} - \hat{\beta}_{kj})^T X_{kj}^{(t)T} X_{kj}^{(t)} (\beta_{kj} - \hat{\beta}_{kj}) - \frac{1}{2\delta_j}\hat{\varepsilon}_j^T \hat{\varepsilon}_j\right\} \qquad (6.4.14)$$

where $\hat{\beta}_{kj} = (\hat{\beta}_{kj.0}, \hat{\beta}_{kj.j-g}, \cdots, \hat{\beta}_{kj.j-1})$ is the maximum likelihood estimates of β_{kj} obtained from (6.4.8)-(6.4.9) and $\hat{\varepsilon}_j^T \hat{\varepsilon}_j = \sum_{k=1}^{b} (Y_{obs.j} - X_{kj}^{(t)}\hat{\beta}_{kj})^T (Y_{obs.j} - X_{kj}^{(t)}\hat{\beta}_{kj})$. This distribution is an approximate normal scaled inverted-chisquare distribution.

$$\beta_{kj} \mid Y_{obs}, \delta_j \approx N_{g+1}(\hat{\beta}_{kj}, \delta_j (X_{kj}^{(t)T} X_{kj}^{(t)})^{-1}) \qquad (6.4.15)$$

$$\delta_j \mid Y_{obs} \approx \sum_{k=1}^{b} (Y_{obs,j} - X_{kj}^{(t)}\hat{\beta}_{kj})^T (Y_{obs,j} - X_{kj}^{(t)}\hat{\beta}_{kj}) \chi_{n_j-g-1}^{-2} \qquad (6.4.16)$$

where $n_j = n_{1j} + n_{2j} + \cdots + n_{bj}$ is the total number of observed values of Y_j across all the populations.

After the values of $\phi_{kj} = (\beta_{kj}, \delta_j)$ $k = 1,2,\cdots,b$ have been simulated from the approximate observed-data posterior distributions (6.4.15)-(6.4.16), the missing values of Y_j can be imputed from the independent conditional predictive distribution of $Y_{mis.j}$ given the imputed complete data of $Y_{j-g}, \cdots, Y_{j-1}$ and $\hat{\phi}_{kj} = (\hat{\beta}_{kj}, \hat{\delta}_j)$. Let $X_{-kj}^{(t)}$ be the data matrix consisting of the observed and the t^{th} imputed values of $Y_{j-g}, \cdots, Y_{j-1}$ in the k^{th} population and augmented by 1 in the first column, the rows of $X_{-kj}^{(t)}$ correspond to the missing values of Y_j in the k^{th}

population. Then the missing values $Y_{mis,j}$ in the k^{th} population are imputed by a single random draw from the conditional predictive distribution $N_{m_{kj}}(X_{-kj}^{(t)}\tilde{\beta}_{kj}, \tilde{\delta}_j I_{m_{kj}})$. where $m_{kj} = N_k - n_{kj}$ and $I_{m_{kj}}$ is a $m_{kj} \times m_{kj}$ unity matrix. The m sets of proper multiple imputations are generated by taking a single random draw of β_{kj} and δ_j from the approximate observed-data posterior distribution (6.4.15)-(6.4.16) corresponding to each $X_{kj}^{(t)}$, say $(\tilde{\beta}_{kj}^{(t)}, \tilde{\delta}_j^{(t)})$ $t = 1, \cdots, m$, and then for each random draw $(\tilde{\beta}_{kj}^{(t)}, \tilde{\delta}_j^{(t)})$, impute $Y_{mis,j}$ by independently taking a single random draw from each corresponding conditional predictive distribution of $Y_{mis,j}$, i.e. $N_{m_{kj}}(X_{-kj}^{(t)}\tilde{\beta}_{kj}^{(t)}, \tilde{\delta}_j^{(t)} I_{m_j})$. Repeat this procedure until all the missing values $Y_{mis,2}, \cdots, Y_{mis,p}$ are imputed. $\left\{Y_{mis,1}^{(t)}, Y_{mis,2}^{(t)}, \cdots, Y_{mis,p}^{(t)}\right\}$ $t = 1,2, \cdots, m$ consist of a proper multiple imputation of Y_{mis}.

6.5 Summary

When incomplete data have a non-monotone missing pattern, the observed-data posterior distribution of the alternative parameters can not be derived in a closed form. A non-analytical method, the MCMC method, can be used to simulate the unknown parameters as well as to impute the missing data. In this chapter, we proposed analytical methods for simulating the parameters and imputing the missing data. By these methods, non-monotone missing data are imputed sequentially. In section 6.1, the EM algorithm for estimating the normal parameters under the general covariance structure was reviewed. The EM algorithm utilizes all observed data, the correlations between the response variables are incorporated into the estimation. In section 6.2, a new method is proposed to impute the missing data. After the normal parameters have been estimated using the EM algorithm, the alternative parameters are estimated by applying the sweep operator on the estimated normal parameters, then the alternative parameters are simulated from the approximate observed-data posterior distribution given the estimated values. Missing values of Y_j are imputed after the missing values of $Y_1, \cdots, Y_{j-1}$ have been imputed. In section 6.3, non-monotone incomplete data are assumed from the multivariate normal distribution with an ante-dependence covariance structure. The normal parameters are still estimated using the EM algorithm under the general covariance structure, but the alternative parameters are estimated under the ante-dependence structure by transforming the estimated

normal parameters. The alternative parameters, under the ante-dependence structure, are simulated from the approximate observed-data posterior distribution given the estimated values. The imputing procedures are the same as for the incomplete data with unstructured covariance. In section 6.4, method for imputing the missing data is proposed when incomplete data are from multiple multivariate normal distributions with varied mean vectors and a common covariance matrix. Varied mean vectors and common covariance matrix are estimated using the modified EM algorithm, the alternative parameters are estimated under the ante-dependence structure by transforming the estimated normal parameters. After the alternative parameters are simulated from the approximated observe-data posterior distribution, missing data are imputed from each individual population.

Chapter 7. Simulation Studies

In chapter 5 and chapter 6, we proposed methods for imputing the missing data when incomplete data follow a monotone and non-monotone missing pattern respectively. The purpose of imputing missing data is to retrieve the missing information from an appropriate imputation model. The retrieved information ideally should be a true reflection of the underlying probability model for the response variables. In this chapter, we conduct some simulation studies to investigate the properties of proposed methods. As a comparison, two other established methods are also investigated, the predictive model method and the MCMC method. All simulations are done by Splus functions. The Splus functions for proposed methods and for the predictive model method are written by author. The Splus functions for the predictive model method are verified with SOLAS v3.0. All these Splus functions are listed in the Appendix 2. The Splus functions for the EM algorithm and the MCMC method are written by Schafer (free at www.stat.psu.edu/~jls/).

In section 7.1, simulated complete data are generated from a multivariate normal distribution with a structured covariance matrix and an unstructured covariance matrix separately, then proposed estimation methods in sections 5.2 and 5.3 are applied to these simulated complete data to estimate the normal parameters. In section 7.2, the monotone incomplete data are created in the sense that the missing data are missing at random (MAR). The missing data are imputed by proposed imputation methods, the predictive model method and the MCMC method. Based on the imputed complete data, the normal parameters are estimated, the mean square errors of these estimates are also calculated. In section 7.3, monotone incomplete data from multiple normal populations are created and then proposed method is applied to impute the missing data. The mean vectors of the multiple populations are estimated, the standard errors of these estimates and the mean square errors are estimated from the imputed complete data. In section 7.4, non-monotone incomplete data are created from a single normal population with structured and unstructured covariance matrix. The EM algorithm is applied to estimate the normal parameters and the missing data are imputed by proposed imputation methods of chapter 6. The predictive model method and the MCMC method are also applied to impute the missing data. Several properties are compared based on the imputed complete data. In section 7.5, non-monotone incomplete data are created from multiple normal populations with different mean vectors and a common structured covariance matrix. The proposed method is applied to impute

the missing data and the population means and their standard errors are estimated from the imputed complete data. A brief summary is given in section 7.6.

7.1 Complete Data From a Single Normal Population

In sections 5.2 and 5.3, the joint observed-data likelihood function is factorized by the independent observed-data likelihood function of individual alternative parameter. The alternative parameters are estimated through linear regression and the estimated normal parameters are obtained by reverse parameter transformation. In this section, we study the properties of these methods when data are complete.

The complete data are generated by randomly drawing samples $y_i = (y_{i1}, y_{i2}, y_{i3})$ from a 3-dimensional multivariate normal distribution $N_3(\mu, \Sigma)$, where

$$\mu^T = (0,0,0) \text{ and } \Sigma = \begin{bmatrix} 1.0000 & 0.1930 & 0.0109 \\ 0.1930 & 2.0000 & 0.1130 \\ 0.0109 & 0.1130 & 3.0000 \end{bmatrix} \quad (7.1.1)$$

are the mean vector and the covariance matrix. The covariance matrix Σ has an ante-dependence structure of order 1. For each simulated complete data, the methods described in section 5.2 and section 5.3 are applied to estimate the alternative parameters ϕ, and then the estimated normal parameters are obtained by the reverse parameter transformation $\hat{\theta} = g^{-1}(\hat{\phi})$. Repeating this procedure 1000 times, we obtain a sampling distribution of the estimated normal parameters. Table 7.1.1 presents the mean, standard deviation (SD) and the mean square error (MSE) of the sampling distribution for each of the normal parameters.

Table 7.1.1. Summary Statistics of Estimated Mean Parameters and Covariance Parameters: Complete Data With **Structured** Covariance (n=30)

		μ_1=0	μ_2=0	μ_3=0	σ_{11}= 1.0000	σ_{12}= 0.1930	σ_{13}= 0.0109	σ_{22}= 2.0000	σ_{23}= 0.1130	σ_{33}= 3.0000
Unstructured Method	Mean	0.0051	-0.0026	0.0251	0.9399	0.1654	**-0.0058**	1.9159	0.1199	2.9104
	SD	0.1765	0.2511	0.3140	0.2450	0.2566	**0.3029**	0.5254	0.4204	0.7583
	MSE	0.0311	0.0630	0.0991	0.0636	0.0665	**0.0919**	0.2829	0.1766	0.5824
Structured Method	Mean	0.0051	-0.0026	0.0251	0.9399	0.1654	**0.0125**	1.9159	0.1199	2.9104
	SD	0.1765	0.2511	0.3140	0.2450	0.2566	**0.0682**	0.5254	0.4204	0.7583
	MSE	0.0311	0.0630	0.0991	0.0636	0.0665	**0.0046**	0.2829	0.1766	0.5824

The whole process is performed again with the complete data being simulated from an unstructured 3-dimensional multivariate normal distribution $N_3(\mu,\Sigma)$, where

$$\mu^T = (0,0,0) \text{ and } \Sigma = \begin{bmatrix} 1.0 & 0.5 & 0.9 \\ 0.5 & 2.0 & 0.3 \\ 0.9 & 0.3 & 3.0 \end{bmatrix} \quad (7.1.2)$$

The simulation results (based on 1000 times of simulations) are presented in Table 7.1.2.

Table 7.1.2. Summary Statistics of Estimated Mean Parameters and Covariance Parameters: Complete Data With **Unstructured** Covariance (n =30)

		μ_1=0	μ_2=0	μ_3=0	σ_{11}= 1.0	σ_{12}= 0.5	σ_{13}= 0.9	σ_{22}= 2.0	σ_{23}= 0.3	σ_{33}= 3.0
Unstructured Method	Mean	0.0029	-0.0055	0.0036	0.9720	0.4868	**0.8798**	1.9562	0.2934	2.9147
	SD	0.1805	0.2459	0.3083	0.2589	0.2793	**0.3577**	0.5355	0.4516	0.7528
	MSE	0.0325	0.0604	0.0949	0.0677	0.0781	**0.1282**	0.2884	0.2038	0.5734
Structured Method	Mean	0.0029	-0.0055	0.0036	0.9720	0.4868	**0.0997**	1.9562	0.2934	2.9147
	SD	0.1805	0.2459	0.3083	0.2589	0.2793	**0.1504**	0.5355	0.4516	0.7528
	MSE	0.0325	0.0604	0.0949	0.0677	0.0781	**0.6629**	0.2884	0.2038	0.5734

When data are complete with the unstructured covariance, the estimated normal parameters by unstructured regression method proposed in section 5.2 are identical to those estimated by non-iterative maximum likelihood method (4.1.3) and (4.1.4). When data are three dimensional, the impact of the ante-dependence of order 1 and unstructured covariance on the estimation of the normal parameters is reflected from the estimation of σ_{13}. From Tables 7.1.1 and 7.1.2, it can be seen that the mean vector is well estimated by both structured and unstructured regression methods. The structured method has much better estimation of σ_{13} than that of unstructured method when data have a structured covariance. But when data have an unstructured covariance, the unstructured method has much better estimation of σ_{13} than that of structured method.

7.2 Monotone Incomplete Data From a Single Normal Population

When data are incomplete, as long as the missingness mechanism is ignorable, any likelihood-based inference is valid. The inferences based on the multiple imputation are also valid (Little and Rubin, 1987, Rubin, 1987). In this section we investigate some aspects of non-imputation vs imputation by conducting some simulations. First, we simulate a complete data

from a 3-dimensional multivariate normal distribution with structured and unstructured covariance matrix given in (7.1.1) and (7.1.2) respectively. Then the monotone incomplete data are created from the simulated complete data in the following way: The values of Y_2 and Y_3 are set to be missing if the corresponding values of Y_1 are less than –0.1, and the values of Y_3 are set to be missing if the values of Y_2 are less than –0.1, i.e.

$$y_{i2} = y_{i3} = \text{missing} \qquad \text{if } y_{i1} < -0.1 \tag{7.2.1}$$
$$y_{i3} = \text{missing} \qquad \text{if } y_{i2} < -0.1$$

In this way, the missing data are created such that the missingness depends on the observed data, but does not depend on the missing data. That is, the missingness mechanism is missing at random (MAR). The missing pattern of this incomplete data is monotone. Since $P(Y_1 < -0.1 \mid \sigma_1^2 = 1) = 0.4601$ and $P(Y_2 < -0.1 \mid \sigma_2^2 = 2) = 0.4718$, the average proportion of missing data of Y_2 is about 46% and the average proportion of missing data of Y_3 is about 70% for structured covariance. The corresponding proportions for unstructured covariance are 46% and 66% respectively.

The simulations are conducted by a split-plot design. For each of the covariance structures, 1000 simulated monotone incomplete datasets are created, the simulated dataset (block) are nested within the covariance structure. For each of the blocks, 4 methods of handling incomplete data are applied to estimate the normal parameters $\theta = (\mu, \Sigma)$ and the mean square errors of these estimates $MSE = (\hat{\theta} - \theta)^2$. For the mean vector $\mu = (\mu_1, \mu_2, \mu_3)$, the t values, where $t_j = \hat{\mu}_j / se(\hat{\mu}_j)$, are calculated. The mean square errors of these t values are also calculated as $MSE = (t_j - 0)^2$.

The first method, the estimation method described in section 5.2, is used to estimate the normal parameters without imputation. The standard errors of the estimated mean parameters are $se(\hat{\mu}_j) = \sqrt{\hat{\sigma}_{jj} / n_j}$, where n_j is the number of observed values of Y_j. The other three methods, the proposed method, the predictive model method and the MCMC method, are applied to generate imputed datasets. The proposed method and the predictive model method assume the non-informative prior (5.1.3) and the MCMC method assumes the non-informative prior (4.1.5) for the multivariate normal distribution. Each simulated incomplete dataset is imputed multiple times ($m = 5$) by each of these methods. The normal parameters are then estimated based on

each imputed complete dataset, the standard errors of the estimated mean parameters are $se(\hat{\mu}_j) = \sqrt{\hat{\sigma}_{jj}/n}$, where n is the sample size. The combined estimates of the normal parameters are the average of estimated values based on the five imputed complete data. The combined standard errors of the estimated mean parameters are obtained by applying Rubin's rule (1.5.3)-(1.5.4). These results are presented in Table 7.2.1 and Table 7.2.2.

Table 7.2.1. Estimated Normal Parameters and Calculated t Values: Monotone Incomplete Data With **Structured** Covariance (n=30)

	μ_1=0				μ_2=0				μ_3=0			
	$\bar{\hat{\mu}}_1$	MSE_μ	$\bar{t}_1$	MSE_t	$\bar{\hat{\mu}}_2$	MSE_μ	$\bar{t}_2$	MSE_t	$\bar{\hat{\mu}}_3$	MSE_μ	$\bar{t}_3$	MSE_t
No Imputation	0.004	0.029	0.023	1.022	0.013	0.337	0.093	2.344	-0.076	1.567	-0.048	2.724
Proposed Method	0.004	0.029	0.023	0.988	0.015	0.404	0.050	1.183	-0.087	2.355	-0.016	1.047
Predictive Model	0.004	0.029	0.023	0.988	0.002	0.393	0.039	1.093	-0.217	45.84	0.017	1.060
MCMC	0.004	0.029	0.023	0.988	0.015	0.417	0.031	1.206	-0.022	3.543	0.030	1.640
	σ_{11}=1.000		σ_{12}=0.193		σ_{13}=0.011		σ_{22}=2.000		σ_{23}=0.113		σ_{33}=3.000	
	$\bar{\hat{\sigma}}_{11}$	MSE	$\bar{\hat{\sigma}}_{12}$	MSE	$\bar{\hat{\sigma}}_{13}$	MSE	$\bar{\hat{\sigma}}_{22}$	MSE	$\bar{\hat{\sigma}}_{23}$	MSE	$\bar{\hat{\sigma}}_{33}$	MSE
No Imputation	0.957	0.064	0.189	0.348	0.035	0.245	2.141	0.856	0.205	3.182	3.722	14.30
Proposed Method	0.990	0.067	0.193	0.430	0.016	0.547	2.878	2.623	0.342	9.367	9.448	261.9
Predictive Model	0.990	0.067	0.207	0.437	0.256	48.19	2.882	2.471	0.515	103.3	373.5	8E7
MCMC	0.990	0.067	0.193	0.440	-0.005	2.318	2.891	2.766	0.181	10.15	9.709	209.1

Table 7.2.2. Estimated Normal Parameters and Calculated t Values: Monotone Incomplete Data With **Unstructured** Covariance (n=30)

	μ_1=0				μ_2=0				μ_3=0			
	$\bar{\hat{\mu}}_1$	MSE_μ	$\bar{t}_1$	MSE_t	$\bar{\hat{\mu}}_2$	MSE_μ	$\bar{t}_2$	MSE_t	$\bar{\hat{\mu}}_3$	MSE_μ	$\bar{t}_3$	MSE_t
No Imputation	-0.003	0.035	-0.072	1.187	-0.010	0.323	0.140	2.439	-0.108	1.770	0.090	3.590
Proposed Method	-0.003	0.035	-0.015	1.148	-0.015	0.392	0.030	1.350	-0.063	2.401	0.024	1.210
Predictive Model	-0.003	0.035	-0.015	1.148	-0.018	0.395	0.025	1.191	-0.161	4.030	0.068	1.218
MCMC	-0.003	0.035	-0.015	1.148	-0.020	0.421	0.041	1.259	-0.117	2.333	-0.071	1.836
	σ_{11}=1.0		σ_{12}=0.5		σ_{13}=0.9		σ_{22}=2.0		σ_{23}=0.3		σ_{33}=3.0	
	$\bar{\hat{\sigma}}_{11}$	MSE	$\bar{\hat{\sigma}}_{12}$	MSE	$\bar{\hat{\sigma}}_{13}$	MSE	$\bar{\hat{\sigma}}_{22}$	MSE	$\bar{\hat{\sigma}}_{23}$	MSE	$\bar{\hat{\sigma}}_{33}$	MSE
No Imputation	0.958	0.061	0.492	0.355	0.922	1.100	2.110	1.122	0.410	2.782	4.267	13.70

Proposed Method	0.991	0.064	0.514	0.443	0.936	1.533	2.782	2.762	0.351	7.018	10.98	679.9
Predictive Model	0.991	0.064	0.516	0.460	0.996	2.261	2.825	3.113	0.389	7.611	23.91	1E5
MCMC	0.991	0.064	0.517	0.482	0.959	1.552	2.821	3.331	0.455	6.135	6.931	57.76

From the above tables, it is noticed that the mean square errors of the estimated normal parameters are smaller for non-imputation method than for those with multiple imputations. But the mean square errors of the calculated t values are larger for non-imputation method than for those with multiple imputations.

For the mean parameters $\mu = (\mu_1, \mu_2, \mu_3)$, the inference can be made based on the calculated t values, where $t_j = \hat{\mu}_j / se(\hat{\mu}_j)$. These t values have the Student-t distribution as their approximate reference distribution. When missing data are not imputed, the degrees of freedom of these distributions are $n_j - 1$, where n_j is the number of observed values of Y_j. When missing data are imputed multiple times, the degrees of freedom of these distributions are given by (1.5.6) (Rubin, 1987). We calculate 95% confidence interval for each mean parameter in each simulation run, and then the empirical coverage rates of these intervals are computed as the proportion of the intervals which include the true values of the parameters. These results are presented in Table 7.2.3 and Table 7.2.4.

Table 7.2.3. Confidence Interval and Coverage Rate of Mean Parameters: Monotone Incomplete Data With **Structured** Covariance (n=30)

	μ_1=0		μ_2=0		μ_3=0	
	95% CI	Rate	95% CI	Rate	95% CI	Rate
No Imputation	-0.3577 , 0.3666	94.9%	-0.7547 , 0.7821	84.9%	-1.5366 , 1.3841	84.6%
Proposed Method	-0.3485 , 0.3575	94.6%	-1.5526 , 1.5838	94.3%	-3.9532 , 3.7774	96.6%
Predictive Model	-0.3485 , 0.3575	94.6%	-1.5990 , 1.6042	96.6%	-7.9525 , 7.5183	95.6%
MCMC	-0.3485 , 0.3575	94.6%	-1.5703 , 1.6018	95.0%	-3.6973 , 3.6531	92.6%

Table 7.2.4. Confidence Interval and Coverage Rate of Mean Parameters: Monotone Incomplete Data With **Unstructured** Covariance (n=30)

	μ_1=0		μ_2=0		μ_3=0	
	95% CI	Rate	95% CI	Rate	95% CI	Rate
No Imputation	-0.3659 , 0.3593	94.4%	-0.7760 , 0.7543	83.1%	-1.5625 , 1.3458	78.8%
Proposed Method	-0.3568 , 0.3501	93.6%	-1.4838 , 1.4522	93.9%	-4.3381 , 4.2104	94.9%

Predictive Model	-0.3568 , 0.3501	93.6%	-1.5559 , 1.5188	94.4%	-4.8409 , 4.5178	95.8%
MCMC	-0.3568 , 0.3501	93.6%	-1.5382 , 1.4980	94.3%	-3.1181 , 2.8823	91.9%

Since the calculated t values have smaller mean $\bar{t}$ (smaller bias) and smaller mean square errors for the multiple imputation methods than those without multiple imputation, the multiple imputation provides better approximation to the Student-t distribution. Therefore, the inference based on the Student-t distribution can be improved by multiple imputation. Tables 7.2.3 and 7.2.4 showed that the empirical coverage rates of the multiple imputation are close to the nominal level while the empirical coverage rates of the non-imputation are far less than the nominal level.

When missing data are imputed multiple times, the combined standard errors of the estimations will be increased compared to those without multiple imputation. This is because the combined standard errors incorporate the within and the between the imputation variation. In order to compare the performance of different multiple imputation methods, two diagnostic indexes, *relative increase in variance due to missing data*, denoted by r in (1.5.7), and *fraction of missing information about* θ, denoted by λ in (1.5.8), are calculated for each multiple imputation method. The method with the smaller values of these two indexes is preferred.

Table 7.2.5. Relative Increase in Variance and Fraction of Missing Information: Monotone Incomplete Data With **Structured** Covariance (n=30)

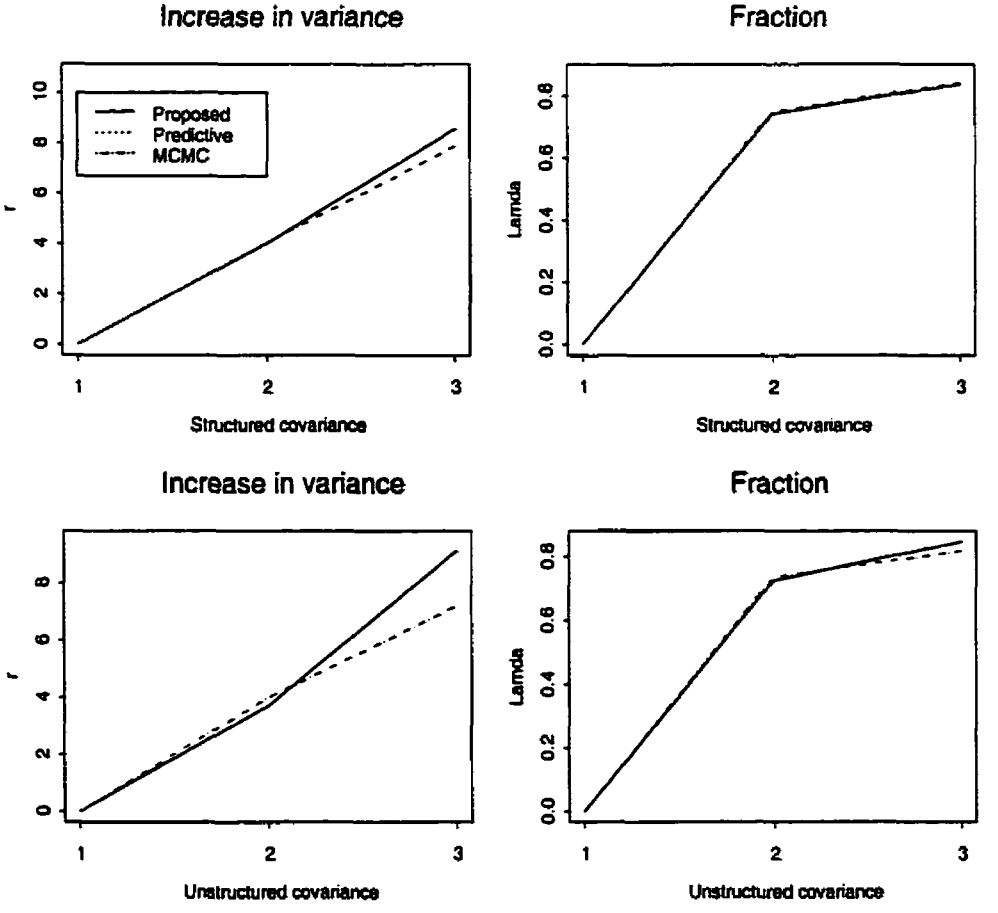
	μ_1=0		μ_2=0		μ_3=0	
	r	λ	r	λ	r	λ
Proposed Method	0.0000	0.0000	3.9978	0.7433	8.5529	0.8389
Predictive Model	0.0000	0.0000	4.1314	0.7585	10.696	0.8680
MCMC	0.0000	0.0000	4.0129	0.7498	7.8629	0.8446

Table 7.2.6. Relative Increase in Variance and Fraction of Missing Information: Monotone Incomplete Data With **Unstructured** Covariance (n=30)

	μ_1=0		μ_2=0		μ_3=0	
	r	λ	r	λ	r	λ
Proposed Method	0.0000	0.0000	3.7059	0.7251	9.1225	0.8455
Predictive Model	0.0000	0.0000	4.0373	0.7429	9.4328	0.8440
MCMC	0.0000	0.0000	3.9891	0.7353	7.1868	0.8170

When incomplete data are monotone, the proposed method and the MCMC method are indistinguishable, they both perform better than the predictive model method in terms of the relative increase in variance and the fraction of missing information.

Figure 7.2.1. Relative Increase in Variance and Fraction of Missing Information - Monotone Incomplete Data



These simulation results are also analyzed by an ANOVA model. The mean square errors of the mean parameters, mean square errors of the calculated t values and the relative increase in variance and the fraction of missing information are analyzed. The SAS procedure PROC MIXED is employed to analyze the split-plot design data in which the simulated dataset (block) are treated as random effects nested within the covariance structure. The SAS codes are

PROC MIXED;

CLASS STRUCTUR METHOD BLOCK;

```
MODEL Y = STRUCTUR METHOD STRUCUTR*METHOD / DDFM=SATTERTH;
RANDOM BLOCK(STRUCTUR);
LSMEANS STRUCTURE*METHOD / SLICE=STRUCTURE DIFF=ALL;
RUN;
```

Since the proposed methods and the predictive model method take different approaches under the different covariance structure, and one of the primary interest is to compare the performance of different multiple imputation methods within each covariance structure, we report the ANOVA results under each covariance structure. For detailed SAS output, see Appendix 3.

Table 7.2.7. Least Square Means and Overall Comparison of Mean Parameters

	Covariance	LSM_1	LSM_2	LSM_3	LSM_4	Overall
$MSE_{\mu 2}$	Structured	0.3378	0.4042	0.3937	0.4177	p=0.0001
	Unstructured	0.3239	0.3923	0.3955	0.4211	p=0.0001
MSE_{t2}	Structured	2.3440	1.1831	1.0931	1.2063	p=0.0001
	Unstructured	2.4393	1.3509	1.1911	1.2590	p=0.0001
$MSE_{\mu 3}$	Structured	1.5678	2.3557	45.844	3.5437	p=0.0287
	Unstructured	1.7703	2.4016	4.0303	2.3334	p=0.9994
MSE_{t3}	Structured	2.7247	1.0478	1.0602	1.6402	p=0.0001
	Unstructured	3.5900	1.2103	1.2180	1.8365	p=0.0001

The mean square errors of calculated t values, where $t = \hat{\mu} / se(\hat{\mu})$, can be used to evaluate different methods since it incorporates the estimated values and their standard errors. From this table, it is noticed that the mean square errors of the multiple imputation methods are significantly smaller than that of the non-imputation method.

Table 7.2.8. Least Square Means and Pairwise Comparison of Relative Increase in Variance and Fraction of Missing Information

	Covariance	LSM_2	LSM_3	LSM_4	Overall	Pairwise Comparison	
						M2 vs M3	M2 vs M4
r_2	Structured	3.9978	4.1314	4.0129	p=0.4926	p=0.2773	p=0.9022
	Unstructured	3.7059	4.0373	3.9891	p=0.0144	p=0.0071	p=0.0213
λ_2	Structured	0.7433	0.7585	0.7498	p=0.0640	p=0.0195	p=0.3166

	Unstructured	0.7251	0.7429	0.7353	p=0.0227	p=0.0061	p=0.1176
r_3	Structured	8.5529	10.696	7.8629	p=0.0001	P=0.0001	p=0.0387
	Unstructured	9.1225	9.4328	7.1868	p=0.0001	p=0.3524	p=0.0001
λ_3	Structured	0.8389	0.8680	0.8446	p=0.0001	p=0.0001	p=0.3015
	Unstructured	0.8455	0.8440	0.8170	p=0.0001	p=0.7829	p=0.0001

The performance of the different multiple imputation methods depends on the proportion of missing data. When missing data are moderate, the proportion of 40%~50%, the proposed method is significantly better than the predictive model method in terms of r and λ, but it is not significantly different from the MCMC method. When missing data are severe, the proportion of 60%~70%, the proposed method is significantly better than the predictive model method when data have the ante-dependence covariance structure. When data have unstructured covariance, the MCMC method has the best performance while the proposed method and the predictive model method are not significantly different.

7.3 Monotone Incomplete Data From Multiple Normal Populations

In section 5.4, a method for imputing the missing data when incomplete data are from multivariate normal distributions with a common covariance matrix and difference mean vectors is proposed. In this section, simulations are conducted to investigate the properties of this method. Multivariate samples $y_i = (y_{i1}, y_{i2}, y_{i3})$ are generated from three multivariate normal distributions $N_3(\mu_k, \Sigma)$ $k = 1, 2, 3$, where

$$\mu_1^T = (0,0,0) \quad \mu_2^T = (2,2,2) \quad \mu_3^T = (4,4,4)$$

and

$$\Sigma = \begin{bmatrix} 1.0000 & 0.1930 & 0.0109 \\ 0.1930 & 2.0000 & 0.1130 \\ 0.0109 & 0.1130 & 3.0000 \end{bmatrix}$$

are the mean vectors and the common covariance matrix. The common covariance matrix Σ has an ante-dependence structure of order 1. A random sample of size 20 is randomly drawn from each distribution and incomplete data are created in the following way. The values of Y_2 and Y_3 are set to be missing if the corresponding values of Y_1 are less than $\mu_{k1} - 0.5$, and the values of Y_3 are set to be missing if the values of Y_2 are less than $\mu_{k2} - 0.5$, i.e.

$$y_{ki2} = y_{ki3} = \text{missing} \qquad \text{if } y_{ki1} < \mu_{k1} - 0.5, \qquad k = 1,2,3$$

$$y_{ki3} = \text{missing} \qquad \text{if } y_{ki2} < \mu_{k2} - 0.5, \quad k = 1,2,3$$

In this way the incomplete data from each population has a monotone missing pattern and the overall missing pattern of all sample is also monotone. Since $P(Y_{k1} < \mu_{k1} - 0.5) = 0.30$, the average proportion of missing data of Y_2 is about 30% and the average proportion of missing data of Y_3 is about 45%.

For each simulated incomplete data, the estimation method of section 5.4 is applied to estimate the $\hat{\mu}_{kj}$ and $MSE = (\hat{\mu}_{kj} - \mu_{kj})^2$. The standard error of the estimated value $\hat{\mu}_{kj}$, $se(\hat{\mu}_{kj}) = \sqrt{\hat{\sigma}_{jj} / m_{kj}}$ where m_{kj} is the number of non-missing data of Y_j in the k^{th} sample, is calculated and 95% confidence interval of $\hat{\mu}_{kj}$ is derived based on the approximate Student-t distribution. The degrees of freedom of this distribution is $\sum_{k=1}^{3} m_{kj} - 1$. Then the missing data $y_{i(mis)}$ are imputed $m = 5$ times from the conditional predictive distribution $N(y_{i(mis)} \mid y_{i(obs)}, \tilde{\mu}_k, \tilde{\Sigma})$, where $\tilde{\theta} = (\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3, \tilde{\Sigma})$ are obtained by the reverse transformation $\tilde{\theta} = g^{-1}(\tilde{\phi})$ and $\tilde{\phi}$ are the random draws from the observed-data posterior distribution (5.4.5)-(5.4.6). For each imputed complete data, $\hat{\mu}_{kj}$ and $se(\hat{\mu}_{kj})$ are calculated, where $se(\hat{\mu}_{kj}) = \sqrt{\hat{\sigma}_{jj} / n_{kj}}$ and n_{kj} is the sample size of Y_j in the k^{th} sample. Then the combined estimations and the standard errors are computed by Rubin's rule. The confidence interval is computed from the approximated Student-t distribution with the degrees of freedom given by (7.2.4). Table 7.3.1 presents these estimations and the empirical coverage rate based on 1000 simulations.

Table 7.3.1. Summary Statistics of Estimated Mean Parameters and Their Standard Errors: Monotone Incomplete Data With Structured Covariance (n =20 for each sub-sample)

	No Imputation				Proposed Method			
	$\bar{\hat{\mu}}$	MSE	95% CI	Rate	$\bar{\hat{\mu}}$	MSE	95% CI	Rate
μ_{11}=0	-0.0042	0.0459	-0.4401 , 0.4317	93.5%	-0.0042	0.0459	-0.4423 , 0.4339	93.5%
μ_{12}=0	-0.0802	0.2228	-0.8739 , 0.7133	95.0%	-0.0769	0.3712	-1.4945 , 1.3405	98.6%
μ_{13}=0	0.0293	0.7063	-1.2336 , 1.2922	84.1%	0.0176	1.2288	-3.8234 , 3.8588	98.4%

μ_{21}=2	1.9767	0.0538	1.5407 , 2.4127	94.5%	1.9767	0.0538	1.5386 , 2.4148	94.5%
μ_{22}=2	2.1096	0.3019	1.3108 , 2.9084	85.0%	2.1182	0.8230	0.6905 , 3.5458	94.1%
μ_{23}=2	1.9681	1.0028	0.7068 , 3.2294	80.2%	1.7309	6.4286	-2.3655 , 5.8273	93.3%
μ_{31}=4	4.0046	0.0316	3.5686 , 4.4406	98.0%	4.0046	0.0316	3.5665 , 4.4427	98.0%
μ_{32}=4	3.9785	0.2513	3.1832 , 4.7739	88.1%	3.8749	1.2865	2.4975 , 5.2522	95.6%
μ_{33}=4	4.0018	0.8857	2.7389 , 5.2648	82.4%	4.4109	12.264	-0.5443 , 9.3663	94.3%

Similar to the results observed in section 7.2, multiple imputation yields larger mean square error compared to no imputation, but improves the confidence interval coverage.

7.4 Non-monotone Incomplete Data From A Single Normal Population

When incomplete data have a non-monotone missing pattern, the normal parameters can not be estimated by the regression method. They can be estimated by the EM algorithm under the assumption of unstructured covariance matrix. The alternative parameters under the assumption of ante-dependence can be estimated by the parameter transformation described in (6.3.2)-(6.3.5) and can be simulated from the approximate observed-data posterior distributions (6.3.11)-(6.3.12). Then the missing data are imputed sequentially. The missing values of Y_1 are imputed first, then the missing values of Y_2 are imputed given the observed and imputed values of Y_1. This process is continued until the missing values of Y_p are imputed.

In this section we simulate non-monotone multivariate incomplete data. First, complete data are simulated from a 3-dimensional multivariate normal distribution with structured and unstructured covariance matrix given in (7.1.1) and (7.1.2) respectively. Second, the non-monotone incomplete data are created from the simulated complete data in the following way. The values of Y_3 are set to be missing if the values of Y_2 are less than –0.1, the values of Y_2 are set to be missing if the corresponding values of Y_1 are less than –0.1, and one third of values of Y_1 are randomly set to be missing. That is,

$$y_{i3} = \text{missing} \qquad \text{if } y_{i2} < -0.1 \tag{7.4.1}$$

$$y_{i2} = \text{missing} \qquad \text{if } y_{i1} < -0.1$$

$$1/3 \text{ of } y_{i1} = \text{missing randomly}$$

In this way, the missing data are created such that the missingness depends on the observed data, but does not depend on the missing data. That is, the missingness mechanism is missing at random (MAR). The missingness pattern of this incomplete data is non-monotone. The average proportions of missing data of Y_2 and Y_3 are about 46% and 47% respectively, and the proportion of missing data of Y_1 is about 33%.

The simulations are conducted by a split-plot design. For each of the covariance structures, 1000 simulated non-monotone incomplete datasets are created. Four methods of handling incomplete data are then applied to each simulated incomplete data to estimate the normal parameters and the mean square errors of the estimations $MSE = (\hat{\theta} - \theta)^2$. For mean vector $\mu = (\mu_1, \mu_2, \mu_3)$, the t values, where $t_j = \hat{\mu}_j / se(\hat{\mu}_j)$, are calculated. The mean square errors of these t values are also calculated as $MSE = (t_j - 0)^2$.

The first method, the EM algorithm, is used to estimate $\theta = (\mu, \Sigma)$ without imputation. The EM algorithm reviewed in section 6.1 is applicable for unstructured covariance. The standard errors of the estimated mean vector are $se(\hat{\mu}_j) = \sqrt{\hat{\sigma}_{jj} / n_j}$, where n_j is the number of observed values of Y_j. The proposed method, the predictive model method and the MCMC method are applied to impute the missing data $m = 5$ times, the normal parameters are then estimated based on each imputed complete data. The standard errors of the estimated mean vector are $se(\hat{\mu}_j) = \sqrt{\hat{\sigma}_{jj} / n}$, where n is the sample size. The combined estimates of the normal parameters are the average of the estimated values based on five imputed complete data. The combined standard errors of the estimated mean vector are obtained by applying Rubin's rule (1.5.3)-(1.5.4). The proposed method and the predictive model method assume the non-informative prior (5.1.3) and the MCMC method assumes the non-informative prior (4.1.5). These results are presented in Table 7.4.1 and Table 7.4.2.

Table 7.4.1. Estimated Normal Parameters and Calculated t Values: Non-monotone Incomplete Data With **Structured** Covariance (n=30)

	μ_1=0				μ_2=0				μ_3=0			
	$\bar{\hat{\mu}}_1$	MSE_μ	$\bar{t}_1$	MSE_t	$\bar{\hat{\mu}}_2$	MSE_μ	$\bar{t}_2$	MSE_t	$\bar{\hat{\mu}}_3$	MSE_μ	$\bar{t}_3$	MSE_t
EM	0.005	0.049	0.023	1.154	0.092	0.266	0.277	2.092	0.069	0.383	0.179	1.802

Proposed Method	0.001	0.049	0.009	1.053	0.084	0.287	0.210	1.468	0.071	0.411	0.129	1.207
Predictive Model	-0.001	0.049	-0.002	1.065	0.113	0.197	0.252	1.003	0.080	0.269	0.145	0.921
MCMC	0.021	0.050	0.089	1.083	0.128	0.293	0.257	1.451	0.075	0.431	0.153	1.370
	σ_{11}=1.000		σ_{12}=0.193		σ_{13}=0.011		σ_{22}=2.000		σ_{23}=0.113		σ_{33}=3.000	
	$\bar{\hat{\sigma}}_{11}$	*MSE*	$\bar{\hat{\sigma}}_{12}$	*MSE*	$\bar{\hat{\sigma}}_{13}$	*MSE*	$\bar{\hat{\sigma}}_{22}$	*MSE*	$\bar{\hat{\sigma}}_{23}$	*MSE*	$\bar{\hat{\sigma}}_{33}$	*MSE*
EM	0.972	0.107	0.100	0.415	0.001	0.472	2.045	0.687	0.043	1.613	3.265	2.840
Proposed Method	1.036	0.129	0.093	0.333	0.012	0.139	2.697	2.093	0.032	1.629	4.583	11.03
Predictive Model	1.033	0.127	0.055	0.178	0.019	0.199	2.409	1.118	0.001	0.509	4.083	5.353
MCMC	1.071	0.144	0.082	0.456	-0.001	0.539	2.511	1.580	0.006	2.093	4.160	6.724

Table 7.4.2. Estimated Normal Parameters and Calculated t Values: Non-monotone Incomplete Data With **Unstructured** Covariance (n=30)

	μ_1=0				μ_2=0				μ_3=0			
	$\bar{\hat{\mu}}_1$	MSE_μ	$\bar{t}_1$	MSE_t	$\bar{\hat{\mu}}_2$	MSE_μ	$\bar{t}_2$	MSE_t	$\bar{\hat{\mu}}_3$	MSE_μ	$\bar{t}_3$	MSE_t
EM	0.012	0.051	0.056	1.226	0.168	0.263	0.601	2.395	-0.028	0.295	-0.035	1.546
Proposed Method	-0.001	0.053	-0.001	1.159	0.167	0.271	0.464	1.472	-0.010	0.373	0.016	0.912
Predictive Model	-0.003	0.053	-0.021	1.151	0.271	0.248	0.681	1.476	0.026	0.263	0.068	0.941
MCMC	0.030	0.052	0.143	1.174	0.197	0.312	0.488	1.545	-0.003	0.335	0.000	1.268
	σ_{11}=1.0		σ_{12}=0.5		σ_{13}=0.9		σ_{22}=2.0		σ_{23}=0.3		σ_{33}=3.0	
	$\bar{\hat{\sigma}}_{11}$	*MSE*	$\bar{\hat{\sigma}}_{12}$	*MSE*	$\bar{\hat{\sigma}}_{13}$	*MSE*	$\bar{\hat{\sigma}}_{22}$	*MSE*	$\bar{\hat{\sigma}}_{23}$	*MSE*	$\bar{\hat{\sigma}}_{33}$	*MSE*
EM	0.970	0.109	0.355	0.359	0.845	0.417	1.895	0.658	0.394	1.055	3.086	1.731
Proposed Method	1.037	0.130	0.299	0.313	0.719	0.413	2.565	1.914	0.363	1.502	5.600	23.47
Predictive Model	1.035	0.129	0.175	0.253	0.576	0.346	2.218	0.945	0.277	0.539	3.975	4.837
MCMC	1.059	0.137	0.347	0.432	0.888	0.466	2.358	1.498	0.290	1.465	3.812	3.917

Since the proportion of missing values of Y_j is smaller than that of section 7.2, the mean square error of $\hat{\sigma}_{33}$ is much smaller than that in Tables 7.2.1 and 7.2.2. The mean square errors of the calculated t values are smaller for the multiple imputation methods than for those without multiple imputation.

The 95% confidence intervals of the mean parameters $\mu=(\mu_1,\mu_2,\mu_3)$ are calculated based on the approximate Student-t distributions. When missing data are not imputed, the degrees of freedom of these distributions are n_j-1, where n_j is the number of observed values of Y_j.

When missing data are multiply imputed, the degrees of freedom of these distributions are given by (1.5.6) (Rubin, 1987). These results are presented in Table 7.4.3 and Table 7.4.4.

Table 7.4.3. Confidence Interval and Coverage Rate of Mean Parameters: Non-monotone Incomplete Data With **Structured** Covariance (n =30)

	μ_1=0		μ_2=0		μ_3=0	
	95% CI	Rate	95% CI	Rate	95% CI	Rate
EM	-0.4492 , 0.4608	95.4%	-0.65830 , 0.8429	86.4%	-0.8900 , 1.0282	90.1%
Proposed Method	-0.4685 , 0.4705	95.6%	-0.9967 , 1.1663	92.0%	-1.2910 , 1.4330	94.3%
Predictive Model	-0.4696 , 0.4677	94.9%	-1.0268 , 1.2543	95.7%	-1.3078 , 1.4690	96.6%
MCMC	-0.4492 , 0.4913	93.9%	-0.9955 , 1.2521	92.1%	-1.2348 , 1.3859	92.5%

Table 7.4.4. Confidence Interval and Coverage Rate of Mean Parameters: Non-monotone Incomplete Data With **Unstructured** Covariance (n =30)

	μ_1=0		μ_2=0		μ_3=0	
	95% CI	Rate	495% CI	Rate	95% CI	Rate
EM	-0.4423 , 0.4666	93.4%	-0.5572 , 0.8935	84.2%	-0.9684 , 0.9117	91.6%
Proposed Method	-0.4721 , 0.4705	93.2%	-0.8928 , 1.2269	92.3%	-1.5393 , 1.5193	95.9%
Predictive Model	-0.4742 , 0.4674	93.4%	-0.8142 , 1.3577	92.4%	-1.3095 , 1.3630	95.6%
MCMC	-0.4216 , 0.4829	93.5%	-0.9029 , 1.2986	91.7%	-1.1729 , 1.1663	92.9%

When the proportion of missing is around 33%, the empirical coverage rates of the EM algorithm are similar to those of multiple imputation. When the missing proportion is around 46%, the empirical coverage rates of the EM algorithm are similar to those of Tables 7.2.3 and 7.2.4 which are generated by the proposed regression method. These non-imputation coverage rates are smaller than those of multiple imputation. It is noticed that the EM coverage rates for μ_3 are higher than those by the regression method. Overall, the multiple imputation improves the coverage rates.

The *relative increase in variance due to missing data* and the *fraction of missing information about the parameter* are computed and presented in Table 7.4.5 and Table 7.4.6. These two indexes are used to evaluate the performance of different multiple imputation methods.

Table 7.4.5. Relative Increase in Variance and Fraction of Missing Information: Non-monotone Incomplete Data With **Structured** Covariance (n=30)

	μ_1=0		μ_2=0		μ_3=0	
	r	λ	r	λ	r	λ
Proposed Method	0.4799	0.3216	1.6530	0.5736	1.5732	0.5592
Predictive Model	0.4794	0.3202	2.1429	0.6371	1.8458	0.5961
MCMC	0.4476	0.2998	1.9987	0.6098	1.6182	0.5594

Table 7.4.6. Relative Increase in Variance and Fraction of Missing Information: Non-monotone Incomplete Data With **Unstructured** Covariance (n=30)

	μ_1=0		μ_2=0		μ_3=0	
	r	λ	r	λ	r	λ
Proposed Method	0.4727	0.3159	1.6880	0.5800	1.6119	0.5619
Predictive Model	0.4706	0.3146	2.0783	0.6271	1.6925	0.5711
MCMC	0.3547	0.2557	2.0451	0.6144	1.3110	0.5074

Compared with the results in section 7.2, it is noticed that the proposed method and the MCMC method are indistinguishable when incomplete data are non-monotone, they both perform better than the predictive model method in terms of the relative increase in variance and the fraction of missing information.

These simulation results are also analyzed by the same ANOVA model as for the monotone simulation results. The mean square errors of the mean parameters, mean square errors of the calculated t values and the relative increase in variance and the fraction of missing information are analyzed.

Since the proposed methods and the predictive model method take different approaches under the different assumptions about the covariance matrix, and one of the primary interest is to compare the performance of different multiple imputation methods within each covariance structure, we report the ANOVA results under each covariance structure in Tables 7.4.7 and 7.4.8. For detailed SAS output, please see Appendix 3.

Figure 7.4.1. Relative Increase in Variance and Fraction of Missing Information – Non-monotone Incomplete Data

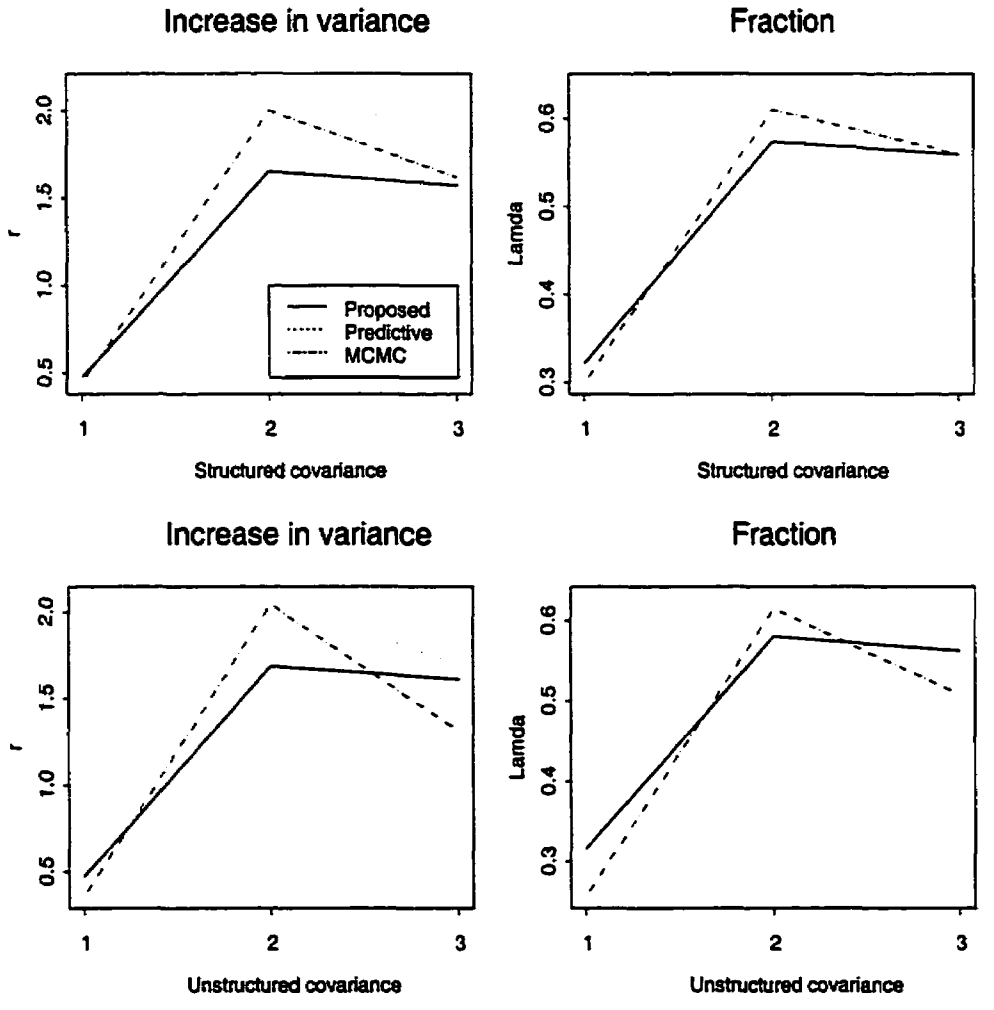


Table 7.4.7. Least Square Means and Overall Comparison of Mean Parameters

	Covariance	LSM_1	LSM_2	LSM_3	LSM_4	Overall
$MSE_{\mu 1}$	Structured	0.0490	0.0493	0.0492	0.0501	p=0.8109
	Unstructured	0.0515	0.0533	0.0536	0.0521	p=0.2318
MSE_{t1}	Structured	1.1546	1.0537	1.0651	1.0837	p=0.0035
	Unstructured	1.2260	1.1591	1.1510	1.1740	p=0.0569
$MSE_{\mu 2}$	Structured	0.2666	0.2877	0.1975	0.2930	p=0.0001

	Unstructured	0.2638	0.2714	0.2483	0.3120	p=0.0001
MSE_{t2}	Structured	2.0928	1.4681	1.0036	1.4513	p=0.0001
	Unstructured	2.3959	1.4729	1.4760	1.5451	p=0.0001
$MSE_{\mu 3}$	Structured	0.3831	0.4115	0.2692	0.4312	p=0.0001
	Unstructured	0.2959	0.3738	0.2630	0.3359	p=0.0001
MSE_{t3}	Structured	1.8025	1.2073	0.9210	1.3704	p=0.0001
	Unstructured	1.5462	0.9124	0.9418	1.2689	p=0.0001

Table 7.4.8. Least Square Means and Pairwise Comparison of Relative Increase in Variance and Fraction of Missing Information

	Covariance	LSM_2	LSM_3	LSM_4	Overall	Pairwise Comparison	
						M2 vs M3	M2 vs M4
r_1	Structured	0.4799	0.4794	0.4476	p=0.0422	p=0.9756	p=0.0282
	Unstructured	0.4727	0.4706	0.3547	p=0.0001	p=0.8866	p=0.0001
λ_1	Structured	0.3216	0.3202	0.2998	p=0.0015	p=0.8300	p=0.0013
	Unstructured	0.3159	0.3146	0.2557	p=0.0001	p=0.8384	p=0.0001
r_2	Structured	1.6530	2.1429	1.9987	p=0.0001	p=0.0001	p=0.0001
	Unstructured	1.6880	2.0783	2.0451	p=0.0001	p=0.0001	p=0.0001
λ_2	Structured	0.5736	0.6371	0.6098	p=0.0001	p=0.0001	p=0.0001
	Unstructured	0.5800	0.6271	0.6144	p=0.0001	p=0.0001	p=0.0001
r_3	Structured	1.5732	1.8458	1.6182	p=0.0001	p=0.0001	p=0.4613
	Unstructured	1.6119	1.6925	1.3110	p=0.0001	p=0.1873	p=0.0001
λ_3	Structured	0.5592	0.5961	0.5594	p=0.0001	p=0.0001	p=0.9812
	Unstructured	0.5619	0.5711	0.5074	p=0.0001	p=0.2567	p=0.0001

The mean square errors of t_2 and t_3 of the multiple imputation methods are significantly smaller than those of the EM methods. When data have unstructured covariance, the mean square error of t_1 of the EM method is smaller than those of multiple imputation, but not significant. Among the multiple imputation methods, overall the proposed method and the MCMC method are indistinguishable in terms of the relative increase in variance and the fraction of missing information. They both have better performance than the predictive model method.

7.5 Non-Monotone Incomplete Data From Multiple Normal Populations

In this section, we simulate non-monotone incomplete data from three multivariate normal distributions $N_3(\mu_k, \Sigma)$ $k = 1, 2, 3$, where

$$\mu_1^T = (0,0,0) \quad \mu_2^T = (2,2,2) \quad \mu_3^T = (4,4,4)$$

and

$$\Sigma = \begin{bmatrix} 1.0000 & 0.1930 & 0.0109 \\ 0.1930 & 2.0000 & 0.1130 \\ 0.0109 & 0.1130 & 3.0000 \end{bmatrix}$$

are the mean vectors and the common covariance matrix. The common covariance matrix Σ has an ante-dependence structure of order 1. A random sample of size 20 is randomly drawn from each distribution and the incomplete data are created in the following way. The values of Y_3 are set to be missing if the corresponding values of Y_2 are less than $\mu_{k2} - 0.5$, the values of Y_2 are set to be missing if the values of Y_1 are less than $\mu_{k1} - 0.5$, and one third of values of Y_1 are randomly set to be missing, i.e.

$$y_{ki3} = \text{missing} \qquad \text{if } y_{ki2} < \mu_{k2} - 0.5, \qquad k = 1,2,3$$

$$y_{ki2} = \text{missing} \qquad \text{if } y_{ki1} < \mu_{k1} - 0.5, \qquad k = 1,2,3$$

$$1/3 \text{ of } y_{ki1} = \text{missing randomly}$$

In this way the incomplete data from each population has a non-monotone missingness pattern and the overall missingness pattern of all sample is also non-monotone. Since $P(Y_{k1} < \mu_{k1} - 0.5) = 0.30$ and $P(Y_{k2} < \mu_{k2} - 0.5) = 0.30$, the average proportion of missing data of Y_2 and Y_3 is about 30% and the proportion of missing data of Y_1 is about 33%.

For each simulated incomplete data, the EM algorithm is applied to estimate the $\hat{\mu}_{kj}$ and $MSE = (\hat{\mu}_{kj} - \mu_{kj})^2$. The standard error $se(\hat{\mu}_{kj}) = \sqrt{\hat{\sigma}_{jj} / m_{kj}}$, where m_{kj} is the number of non-missing data of Y_j in the k^{th} sample, is calculated and 95% confidence interval of $\hat{\mu}_{kj}$ is derived based on the approximate Student-t distribution with the degrees of freedom $\sum_{k=1}^{3} m_{kj} - 1$. Second, the missing data $y_{i(mis)}$ are imputed m=5 times by the method described in section 6.4. For each imputed complete data, $\hat{\mu}_{kj}$ and $se(\hat{\mu}_{kj})$ are calculated, where $se(\hat{\mu}_{kj}) = \sqrt{\hat{\sigma}_{jj} / n_{kj}}$ and n_{kj} is the sample size of Y_j in the k^{th} sample. Then the combined estimations and the combined standard errors are computed by Rubin's rule. The

confidence interval of μ_{kj} is computed from the approximated Student-t distribution with the degrees of freedom given by (7.2.4). Table 7.5.1 presents these estimations and the empirical coverage rate based on 1000 simulations.

Table 7.5.1. Summary Statistics of Estimated Mean Parameters and Their Standard Errors: Non-monotone Incomplete Data With Structured Covariance (n =20 for each sub-sample)

	EM				Proposed Method			
	$\bar{\hat{\mu}}$	MSE	95% CI	Rate	$\bar{\hat{\mu}}$	MSE	95% CI	Rate
μ_{11}=0	0.0223	0.0906	-0.5176 , 0.5624	93.3%	0.0179	0.0867	-0.5501 , 0.5861	93.7%
μ_{12}=0	0.0559	0.2453	-0.7184 , 0.8302	89.8%	0.0610	0.2123	-1.0119 , 1.1340	96.6%
μ_{13}=0	0.0204	0.3997	-0.9897 , 1.0305	89.8%	0.0251	0.3638	-1.4907 , 1.5409	96.2%
μ_{21}=2	2.0100	0.0859	1.4705 , 2.5495	93.8%	2.0091	0.0821	1.4422 , 2.5760	95.1%
μ_{22}=2	2.0705	0.2317	1.2937 , 2.8473	89.0%	2.0643	0.1904	1.0116 , 3.1170	95.8%
μ_{23}=2	2.0116	0.4437	1.0010 , 3.0222	88.4%	2.0253	0.3928	0.5305 , 3.5202	95.8%
μ_{31}=4	4.0002	0.0864	3.4598 , 4.5405	93.6%	3.9968	0.0839	3.4315 , 4.5621	93.5%
μ_{32}=4	4.0538	0.2181	3.2778 , 4.8299	90.8%	4.0535	0.1917	3.0057 , 5.1013	96.4%
μ_{33}=4	4.0021	0.4842	2.9922 , 5.0121	86.1%	4.0139	0.4047	2.4682 , 5.5595	96.3%

It is noted that all mean square errors of the estimated mean parameters are smaller for the proposed method than for the no imputation. The empirical coverage rates are also improved by the multiple imputation.

7.6 Summary

In chapter 5 and chapter 6, we proposed methods for estimating and simulating the normal parameters, we also proposed the methods for imputing incomplete data. In section 7.1, the simulation results showed that when data are complete, the covariance structures are well preserved by the proposed estimation methods. In section 7.2, the simulation results showed that when incomplete data are monotone, the multiple imputation methods yield larger mean square error for the estimated normal parameters than those without imputation, but the mean square errors of the derived t values are smaller than those without imputation. Since the inferences are based on the derived t values which have an approximate Student-t distribution, the simulation

results showed that the empirical coverage rates are significantly improved by the multiple imputation. The similar results are observed in section 7.4 for non-monotone incomplete data. Sections 7.3 and 7.5 presented the simulation results when monotone and non-monotone incomplete data are from multiple multivariate normal distributions. In these situations, the proposed method improves the empirical coverage rates for the mean parameters. Overall, the approximate Student-t distribution derived from the multiple imputation has better properties than that derived without imputation, so the inferences based on the approximate Student-t distribution are improved by the multiple imputation.

Among the three multiple imputation methods, the proposed method and the predictive model method impute the missing data from the conditional predictive distribution analytically. The MCMC method imputes the missing data from the posterior predictive distribution by the Monte Carlo approach. The simulation results have shown that, in terms of the *relative increase in variance* and the *fraction of missing information*, the predictive model method has largest values of these two indexes and the proposed method has values indistinguishable from those of the MCMC method. Therefore, the proposed method and the MCMC method can be substitutes for each other. Since the proposed method employs an analytical approach, it has a computational advantage over the MCMC method. In order to have m independent imputations, the proposed method needs to draw m times from the posterior distribution and the conditional predictive distribution, while the MCMC method needs to draw $m \times k$ times from the posterior distribution and the posterior predictive distribution, where k is the number of iterations for the burn-in period.

Chapter 8. Application

New imputation methods for monotone and non-monotone incomplete data are proposed in chapter 5 and chapter 6. In chapter 7, simulations are conducted to evaluate the performance of these proposed methods and to compare with two other imputation methods. The simulation results confirmed that these multiple imputation methods improve the performance of statistical inference, and the proposed methods preserve the covariance structure between the observed data and the imputed data. The ultimate goal of the multiple imputation is to make valid inferences about the unknown parameters while preserving the inherent data structure when data are incomplete. In this chapter, we apply the proposed method and the predictive model method and the MCMC method to a real data arising from a clinical trial. Section 8.1 presents a brief description of the clinical trial and the observed data. In section 8.2, the missingness mechanism is discussed and the covariance structure of the data is explored. The order of the ante-dependence structure is identified. In section 8.3, An ANCOCA model is used to analyze the data with and without imputing the missing data. When missing data are imputed multiple times, Rubin's rule for repeated imputation inference is applied to make an inference on the treatment effect. A summary is included in section 8.4. As a conclusion of this research, some issues and future research directions are discussed in section 8.5.

8.1 A Clinical Trial of Treatment on PMDD

Premenstrual Dysphoric Disorder (PMDD) is a clinical diagnosis characterized by debilitating mood and behavioral changes and somatic complaints during the luteal phase of a menstrual cycle. The luteal phase of a menstrual cycle is conventionally defined as the 7-days prior to the menstruation. PMDD is characterized by its cyclical nature, symptoms occur in the luteal phase of the menstrual cycle and disappear after the onset of the menstruation. Physiologically, PMDD is thought to be a consequence of complex interactions between ovarian steroid production, endogenous opioid peptides, central neurotransmitters, prostaglandins, and peripheral autonomic and endocrine systems. However, no specific deficiency or abnormality has been identified to explain the cause of PMDD.

A clinical trial was conducted to investigate the efficacy of a combination of progestogen and estrogen in the treatment of PMDD. This was a double-blind, placebo-controlled, parallel design study. Subjects were screened against the entry criteria, the eligible subjects were then entered into a run-in period which consists of two menstrual cycles. At the end of the run-in

period, subjects were checked again for inclusion/exclusion criteria, eligible subjects were randomized into a 3-cycle treatment period. During the run-in and treatment period, subjects were required to call a toll-free number to report their daily PMDD symptoms using an automated voice system. The PMDD symptoms were assessed by a questionnaire called COPE (Calendar of Premenstrual Event), which consists of 22 items that range from 0 (absent) to 3 (severe). The luteal total COPE score was calculated for each run-in cycles and the treatment cycles as the sum of the daily total COPE score during the luteal phase. This score ranges from 0(best) to 462(worst). The average of the two run-in cycles scores was used as the baseline. Table 8.1.1 presents 83 subjects' luteal total scores, treatment groups were P=Placebo and T=Active treatment.

Table 8.1.1. Luteal Total COPE Scores

OBS	Subject	TRT	BASELINE	CYCLE1	CYCLE2	CYCLE3
1	3620058	P	347.0	151.0	275.0	256.0
2	3623002	T	218.0	38.0	62.0	222.0
3	3612110	T	219.0	138.0	167.0	220.0
4	3602025	P	207.0	161.0	127.0	183.0
5	3620074	P	228.0	180.0	251.0	182.0
6	3623041	P	146.0	13.0	108.0	169.0
7	3607026	P	277.0	121.0	118.0	167.0
8	3612075	P	233.0	85.0	111.0	152.0
9	3619038	P	138.0	64.0	138.0	142.0
10	3613024	T	133.0	220.0	201.0	134.0
11	3612044	T	279.0	22.0	71.0	122.0
12	3614002	T	82.0	91.0	140.0	116.0
13	3607115	T	151.0	113.0	92.0	116.0
14	3613008	P	125.0	113.0	172.0	114.0
15	3611005	T	228.0	91.0	102.0	112.5
16	3607046	T	200.0	76.0	115.0	107.0
17	3619021	T	193.0	59.0	73.0	103.0
18	3605015	T	243.0	69.0	106.0	101.0
19	3602024	T	100.0	116.0	81.5	85.0
20	3608020	T	128.0	14.0	2.0	85.0
21	3614025	P	222.0	172.0	104.0	83.0
22	3619058	P	128.0	40.0	63.0	82.0
23	3619041	P	87.0	40.0	54.0	82.0
24	3619048	P	126.0	19.0	139.0	78.0
25	3612043	P	234.0	64.0	65.0	74.0
26	3625004	P	104.0	62.0	94.0	73.0
27	3623036	T	136.0	48.0	46.0	73.0
28	3602015	P	114.0	34.0	81.5	66.5
29	3607062	T	252.0	41.0	146.5	64.0
30	3619094	P	160.0	84.0	74.0	64.0
31	3607101	P	42.0	53.0	35.0	58.0
32	3621007	P	154.0	58.0	41.0	50.5
33	3604055	T	99.0	32.0	32.0	49.0
34	3623020	P	297.0	129.5	58.0	42.5
35	3604044	P	248.0	193.0	149.0	42.0
36	3613002	T	123.0	96.0	80.0	42.0

37	3608021	P	235.0	55.0	108.0	41.0
38	3613001	P	152.0	128.0	158.0	35.0
39	3611003	P	235.0	130.0	22.0	31.0
40	3612039	T	227.0	294.0	86.0	28.0
41	3601038	T	93.0	10.0	74.0	27.0
42	3605013	P	159.0	61.0	19.0	17.0
43	3608019	T	25.0	23.0	43.0	12.0
44	3602028	T	350.0	18.0	19.5	12.0
45	3620055	P	134.0	47.0	117.0	11.5
46	3624007	T	190.0	1.0	27.5	10.0
47	3608003	P	98.0	47.0	62.0	9.0
48	3604026	P	104.5	30.0	16.0	9.0
49	3611007	P	108.0	15.0	7.0	9.0
50	3619027	T	314.0	140.0	174.0	
51	3604006	T	175.0	61.0	170.0	
52	3623005	T	82.0	88.0	121.0	
53	3604046	P	232.0	87.0	113.0	
54	3608034	P	72.0	64.5	92.5	
55	3624014	T	315.0	254.0	62.0	
56	3614023	T	158.0		55.0	
57	3618023	T			40.0	
58	3607023	T	234.0	140.0	39.0	
59	3611010	P	119.0	0.0	32.0	
60	3612109	T	1841.0	53.0	6.0	
61	3606013	T	99.5	54.0	4.0	
62	3615026	T	268.0	197.0		
63	3607009	T	164.0	119.0		
64	3617001	T	123.0	117.0		
65	3619028	P	150.0	102.0		
66	3620066	P	127.0	82.0		
67	3620072	T		46.0		
68	3620051	T	171.0	21.5		
69	3606041	T	146.0	19.0		
70	3603006	T	205.0	12.0		
71	3602013	T	236.0			
72	3611001	T	74.0			
73	3619075	T	227.0			
74	3619081	T	126.0			
75	3619095	T	168.0			
76	3623017	T	288.0			
77	3601052	P	54.0			
78	3602018	P	83.0			
79	3602027	P	264.0			
80	3606032	P	345.0			
81	3607034	P	203.0			
82	3607111	P	145.0			
83	3615035	P	323.0			

It is noticed that this incomplete data have a non-monotone missing pattern. 30% of placebo subjects have at least one missing value, 52% of active treatment subjects have at least one missing value. The overall missing rate 4is 41%.

8.2 Preliminary Analysis

In order to correctly analyze incomplete data, we need to investigate the missingness mechanism so appropriate methods can be applied. The missing data in this trial are caused by multiple reasons, some are pertinent to the design and some are pertinent to the disease itself. (1) Subjects lost to follow-up due to moving away. In this case, the missing is MCAR. (2) Subjects did not record their daily symptoms due to bored of using automated voice system. By the design, subjects were asked to call a voice system every day to answer more than 25 questions. In the late stage of the trial, subjects got tired and did not actively record their daily symptoms as they did in the early stage of the trial. This kind of missing is MCAR. (3) Subjects withdrew due to the side effects. If safety is not considered in the imputation model, the missing is MCAR, otherwise it is MAR. (4) Subjects did not record the first menses day information due to the cyclic characteristic of the disease (most of symptoms are disappeared on the first day of menses). In this case, the missing is MAR. (5) Subjects stopped recording their symptoms when they felt better because of the treatment. This missing is informative (non-ignorable). The majority reasons of missing data were (2) and (4). Although some missing data were caused by non-ignorable reason, the bias caused by ignoring the missingness mechanism is negligible if (i) the non-ignorable missing portion is small and (ii) the subjects who had non-ignorable missing values are also included in the imputation model. Therefore, it is reasonable to assume the missingness mechanism is ignorable. In this case, the likelihood-based methods and the multiple imputation methods discussed in this research can be applied to make a valid inference on the treatment effect.

Next, the covariance structure conveyed by the data is explored. The order of the ante-dependence structure is estimated by using penalized likelihood criteria, such as Akaike's Information Criterion (*AIC*) and Schwarz's Bayesian Information Criterion (*BIC*). For a detailed description of how to estimate the order of the ante-dependence structure, please see Appendix 1. To estimate the order of the ante-dependence structure, the *AIC* and *BIC* are calculated for each admissible value of the order g. For this given dataset, the admissible values of g are 0, 1, 2, and 3, and the number of unknown parameters under the assumption of ante-dependence of order g is $4+(g+1)(2*4-1)/2$. After the *AIC* and *BIC* are calculated, the value of g with the lowest *AIC* or *BIC* is the estimated order of the ante-dependence structure. Table 8.2.1 gives the results of all values of *AIC* and *BIC*.

Table 8.2.1. The *AIC* and *BIC* under the assumption of ante-dependence

Order of ante-dependence	AIC	BIC
0	2436.199	2455.550
1	**2428.420**	**2455.027**
2	2431.793	2463.237
3	2435.131	2468.999

Both *AIC* and *BIC* results suggest that these non-monotone incomplete data have an order 1 of the ante-dependence covariance structure.

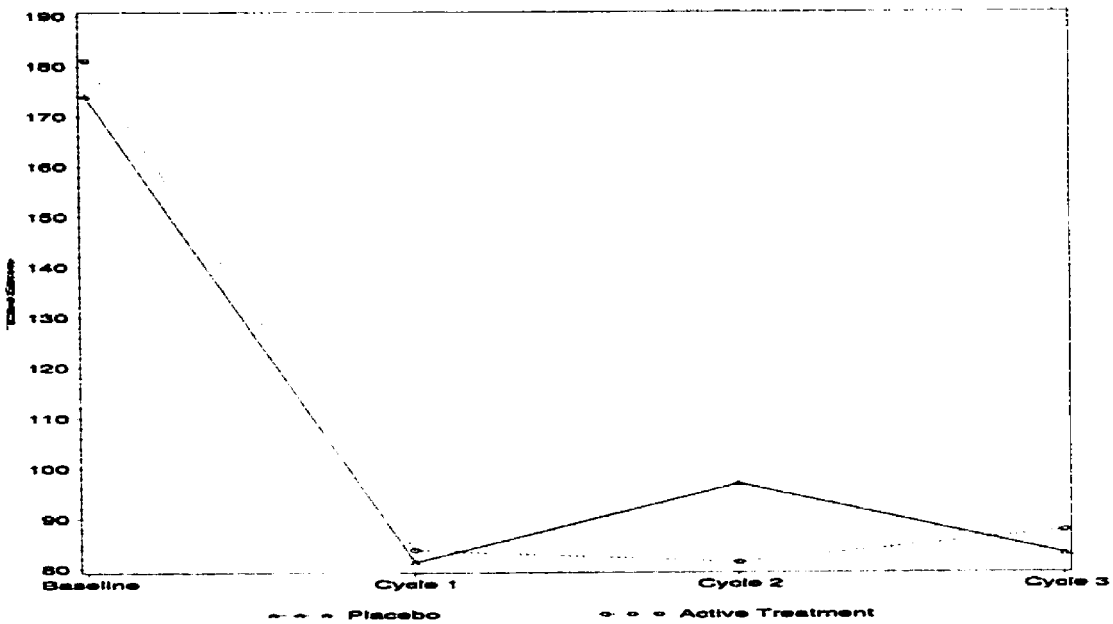
Before we do any modeling, we calculate the raw summary statistics based on the observed data. The mean vectors of each treatment group are calculated based on the scores at each cycle and the common covariance matrix is calculated based all observed data.

$$\hat{\mu}_P = (173.98, 81.36, 96.90, 82.96) \qquad \hat{\mu}_T = (181.13, 83.75, 81.26, 87.64)$$

$$\hat{\Sigma} = \begin{bmatrix} 5849 & 1891 & 1253 & 1541 \\ 1891 & 3860 & 1740 & 1074 \\ 1253 & 1740 & 3448 & 2294 \\ 1541 & 1074 & 2294 & 3815 \end{bmatrix}$$

The raw mean vector of each treatment group are graphically presented in figure 8.2.1.

Figure 8.2.1. Repeated Measurement Data Profiles By Treatment Group



It is noticed from the graph that once the subjects were randomized to take the treatment, the PMDD symptoms were immediately improved for both active treatment group and placebo group subjects. This is because the psychological symptoms of the PMDD are dominating the physical symptoms.

8.3 Analysis of Data

The effect of active treatment on the PMDD versus the placebo is evaluated by an ANCOVA model. First, a weighted ANCOVA is used to analyze the data without imputing the missing data. The average of the post-baseline data is the response variable, the baseline score is the covariate and the treatment group is a factor. The number of non-missing post-baseline score is used as the weight in the model. Table 8.3.1 gives the estimated treatment difference (Placebo – Active treatment), the standard error of this estimate, the calculated t value, the degrees of freedom (v) of the approximate Student-t distribution and the p value.

Table 8.3.1. Weighted Analysis of Covariance of the Average of Post-baseline Scores

$\hat{\Delta}$	$se(\hat{\Delta})$	t	v	p
4.8882	11.0282	0.4432	65	0.6591

Next, the missing data are imputed by three different methods, the proposed method of section 6.4, the MCMC method and the predictive model method. The proposed method is applicable for non-monotone incomplete data with the ante-dependence covariance structure. The MCMC method and the predictive model method are applicable for non-monotone incomplete data with unstructured covariance matrix.

When the proposed method is applied, the EM algorithm described in section 6.4 is first used to estimate the mean vectors and the common covariance matrix,

$$\hat{\mu}_P = (173.98, 84.30, 99.45, 83.57) \quad \hat{\mu}_T = (180.74, 82.88, 80.87, 85.26)$$

$$\hat{\Sigma} = \begin{bmatrix} 5612 & 2057 & 902 & 623 \\ 2057 & 3846 & 1687 & 1166 \\ 902 & 1687 & 3275 & 2263 \\ 623 & 1166 & 2263 & 3753 \end{bmatrix}$$

The missing data are then imputed sequentially from the approximate conditional predictive distribution.

For the predictive model method, the missing data are imputed from a linear regression model which includes the treatment as a factor and the previous variables as covariates. For the MCMC method, the missing data are imputed from the posterior predictive distribution within each treatment separately.

The missing data are imputed $m = 5$ times by each method. For each imputed complete data, a simple ANCOVA model is applied to obtain the estimate of the treatment difference, the standard errors of these estimates, the calculated t value and the p value. The degrees of freedom of the approximate Student-t distributions are all the same for the imputed complete data, ν=80. Table 8.3.2 presents these results.

Table 8.3.2. Analysis Results At Each Imputation.

	Proposed Method				MCMC Method				Predictive Method			
m	$\hat{\Delta}$	$se(\hat{\Delta})$	t	p	$\hat{\Delta}$	$se(\hat{\Delta})$	t	p	$\hat{\Delta}$	$se(\hat{\Delta})$	t	p
1	6.4541	8.5714	0.7529	0.4537	3.1923	9.4487	0.3378	0.7364	5.9507	9.3626	0.6355	0.5269
2	1.5086	8.8088	0.1712	0.8645	10.566	9.2746	1.1392	0.2580	-1.9748	9.5625	-0.2065	0.8369
3	9.1071	9.1620	0.9940	0.3232	4.8054	9.2848	0.5175	0.6062	7.9111	10.605	0.7459	0.4579
4	11.629	9.5836	1.2134	0.2285	10.848	9.0939	1.1929	0.2364	8.0952	9.5657	0.8462	0.3999
5	8.8247	9.2003	0.9591	0.3404	7.5672	8.8805	0.8521	0.3967	9.0832	9.8795	0.9193	0.3607

The combined estimate of the treatment difference and the combined inference is obtained by applying Rubin's rule for repeated imputation inference. The combined estimate of the treatment difference is the average of the estimates at each imputation, i.e.

$$\bar{\Delta} = \frac{1}{5}\sum_{m=1}^{5}\hat{\Delta}^{(m)}$$

The standard error of the combined estimate is given by

$$se(\bar{\Delta}) = \left((1+1/5)B + U\right)^{1/2}$$

where

$$B = \frac{1}{5-1}\sum_{m=1}^{5}(\hat{\Delta}^{(m)} - \bar{\Delta})^2$$

is the between imputation variance and

$$U = \frac{1}{5}\sum_{m=1}^{5} se(\hat{\Delta}^{(m)})^2$$

is the within imputation variance.

The hypothesis of $H_0 : \Delta = 0$ versus $H_1 : \Delta \neq 0$ is tested based on the test statistics

$$t = \frac{\overline{\Delta}}{se(\overline{\Delta})}$$

which follows an approximate Student-t distribution with the degrees of freedom

$$v = (5-1)\left[1 + \frac{U}{(1+1/5)B}\right]^2$$

Table 8.3.3 presents the final inference results after applying Rubin's rule and the inference result without any imputation.

Table 8.3.3. Final inference results

	$\overline{\Delta}$	$se(\overline{\Delta})$	v	t	$p-value$
No Imputation	4.8882	11.0282	65	0.4432	0.6591
Proposed Method	7.5047	9.9906	130	0.7511	0.4539
MCMC Method	7.3958	9.9260	201	0.7450	0.4570
Predictive M	5.8130	10.9740	98	0.5297	0.5975

From this example, it can be seen that multiple imputations improved the power of the inference. The improvement comes from two sources, the size of the treatment effect is enlarged and the standard error of the estimation of the treatment effect is lowered.

8.4 Summary

Due to uncontrollable reasons, incomplete data almost always appear in every clinical trial. When data are incomplete, there are two issues that need to be considered when making a statistical inference. One is the validity of the inference and another is the performance of the inference. When the missingness mechanism is ignorable, all likelihood-based methods and the multiple imputation methods mentioned in this research will generate valid inferences. Due to the incompleteness of the data, we definitely lose some information. The likelihood-based methods do not intend to recover the lost information, all inferences are made based on observed information. By the contrast, the multiple imputation methods try to recover the lost information and then make inferences based on the observed and recovered information. Due to the

uncertainties associated with the recovering process, multiple imputation has a larger variability than no-imputation. But since the repeated imputation inferences are made based on the approximate Student-t distribution, and the approximate Student-t distributions of the multiple imputation has better properties than those of no imputation, the statistical inferences are improved by the multiple imputation. In chapter 7, the simulation results showed that the empirical coverage rates are significantly improved to close to or above the nominal level by the multiple imputation, and in this chapter, a numerical example has shown that the power of the hypothesis test is improved by the multiple imputation.

Proper multiple imputation requires an appropriate imputation model and independent draws from this model. In this research, new analytical imputation methods are proposed. When incomplete data are monotone, the proposed methods are based on a multivariate approach. When incomplete data are non-monotone, the proposed methods are based on an univariate approach, in which the missing data are imputed sequentially. Another analytical imputation method, the predictive model method, is based on the univariate approach. One good thing of the analytical methods is that the missing data can be easily imputed from the conditional predictive distribution independently. The MCMC method is based on the Monte Carlo method. In order to have independent imputations, a Markov chain of $k \times m$ iterations or m independent chains with k iterations of each need to be generated, where k is the number of iterations of the burn-in period. From the simulation results, the proposed method and the MCMC method have better performance than the predictive model method in terms of the increase variance due to the multiple imputation and the fraction of missing information, and the proposed method and the MCMC method have indistinguishable performance. But in terms of the empirical coverage rate, the proposed method has a slightly better performance than the MCMC method.

One purpose of the multiple imputation is to retrieve the missing information while preserving the covariance structure between the observed data and the imputed data. Therefore, how to preserve the covariance structure is an important consideration in this research. In chapter 7, when the covariance structure is the ante-dependence of order 1, the impacts of the proposed method for the structured data are reflected from the estimations of μ_3 and σ_{13}. From Table 7.2.1 and Table 7.4.1, it is observed that the proposed method either has a smaller bias, reflected by $\bar{\hat{\mu}}_3$, or smaller mean square errors than those of the predictive model method and the MCMC method. So the ante-dependence covariance structure is well preserved by the proposed method.

8.1 Discussion

This research discussed the methods of multiple imputation when incomplete data have an ante-dependence covariance structure. Under this assumption, when the order of the ante-dependence is given, the proposed method is applied to generate the imputed complete data. In a real situation, however, the order of the ante-dependence structure is unknown and needs to be estimated. We proposed in the appendix 1 that the penalized likelihood methods, *AIC* or *BIC*, be used to estimate the order of the ante-dependence. Kenward (1987) proposed a likelihood ratio test of the ante-dependence order H_0: order $= g_1$ versus H_1: order $= g_2 > g_1$. For large data sets these likelihood methods are quite reliable. But for small sample sizes these likelihood methods may not perform well. In this case it is recommended that the proposed unstructured method be used.

Rubin (1987) derived the degrees of freedom (1.5.6) for the repeated imputation inference. This formula is based on the assumption that the degrees of freedom are infinite when data are complete and the number of imputations is finite. When the degrees of freedom of the complete data is small and the proportion of missing data is modest, the degrees of freedom defined by formula (1.5.6) can be much larger than the degrees of freedom for the complete data. In this case, equation (1.5.6) is inappropriate. Barnard and Rubin (1999) proposed a modified formula for calculating the degrees of freedom. The modified degrees of freedom has the following properties for fixed m and an estimated fraction of missing information, (1) modified degrees of freedom monotonically increase to complete data degrees of freedom, (2) modified degrees of freedom are always less than or equal to complete data degrees of freedom, and (3) modified degrees of freedom equal complete data degrees of freedom when the latter is infinite. To describe the modified degrees of freedom, we denote the original degrees of freedom by

$$v_m = (m-1)(1+r_m^{-1})^2$$

where

$$r_m = (1+m^{-1})B/\bar{U}$$

the modified degrees of freedom $\tilde{v}_m$ then can be expressed by

$$\tilde{v}_m = \left(\frac{1}{v_m} + \frac{1}{v_{obs}}\right)^{-1}$$

where

$$v_{obs} = \left(\frac{v_{com} + 1}{v_{com} + 3} \right) v_{com} (1 - r_m)$$

is the observed-data degrees of freedom and v_{com} is the complete-data degrees of freedom. This formula for the modified degrees of freedom will result in a larger percentile t-value that will result in a wider confidence interval and a larger p-value.

In longitudinal studies, random effects models are often utilized. More generally, the mixed effects model is defined by $Y = X\beta + Z\gamma + e$ where $X\beta$ is the fixed effect component, $Z\gamma$ is the random effect component and e is the error term. It is usually assumed that $\gamma \sim N(0, G)$. $e \sim N(0, R)$ and γ and e are independent. Under this model, the response variable $Y \sim N(X\beta, ZGZ'+R)$. Little (1995) discussed the dependence of the missingness on the random effects, he referred this dependence on the random effect as the dependence of the future values of the response variable. If the missingness depends on the random effects, the missingness mechanism is non-ignorable. In this case, all multiple imputation methods based on the ignorability are not appropriate. If the missingness is independent of the random effects, the missingness mechanism is ignorable. In this case, either likelihood-based methods or multiple imputation methods can be applied to make valid inferences. The proposed multiple imputation method with an ante-dependence covariance structure can be applied as long as the within subject covariance matrix $\Sigma = ZGZ'+R$ has an ante-dependence structure. If $\Sigma = ZGZ'+R$ does not have an ante-dependence structure, as a further research topic, the EM algorithm may be modified to estimate the normal parameters $\mu = X\beta$ and $\Sigma = ZGZ'+R$, and then transformed to the alternative parameters. The alternative parameters are then simulated from the approximate observed-data posterior distribution given the estimated ones. The missing values are imputed from the conditional predictive distribution given the observed values and simulated parameters.

Throughout the whole research, the response variables are assumed to be multivariate normal and no specific assumptions were applied to the covariates. When continuous covariates X have no missing values the response variables are assumed to follow the conditional multivariate normal given X. If continuous covariates have missing values it is reasonable to assume that

(X,Y) are jointly multivariate normal. In either case the proposed method can be applied directly. When covariates are categorical variables, the multiple imputation of response variables can be divided into two categories. (1) If the covariates have no missing values, the proposed method for multiple populations can be applied directly where each population is indexed by the combination of the covariates. (2) If the covariates have missing values, we still can assume that (X,Y) has a joint multivariate normal as long as these covariates are unimodal and symmetric. See Schafer (1997, section 6.3) for a discussion of this topic and the simulation results of Rubin and Schenker (1986). In some situations data manipulations are required in order to have these categorical variables unimodal and symmetric. In this case, we still can apply the proposed method. Note that when categorical covariates are approximated by the multivariate normal in the imputation model, they are still treated as the categorical variables in the analysis model.

If it is not plausible to assume a joint multivariate normal for (X,Y), the general location model (Olkin and Tate, 1961) can be used as the imputation model. Schafer (1997, chapter 9) described the MCMC imputation for the general location model. The general location model can be described by

$$P(Y,(n_1,\cdots,n_s)\,|\,\mu,\Sigma,(\pi_1,\cdots,\pi_s)) = P((n_1,\cdots,n_s)\,|\,(\pi_1,\cdots,\pi_s))P(Y\,|\,(n_1,\cdots,n_s),\mu,\Sigma)$$

where $P((n_1,\cdots,n_s)\,|\,(\pi_1,\cdots,\pi_s))$ is a marginal multinomial distribution for the number of observations n_i in cell i, the cells are constituted by the combination of the covariates. $P(Y\,|\,(n_1,\cdots,n_s),\mu,\Sigma)$ is the conditional multivariate normal distribution for the response. As a further research topic, if we could find a prior distribution such that the multinomial parameter $(\pi_1,\cdots,\pi_s)$ can be simulated from its posterior distribution, then the proposed method can be extended to this situation.

This research assumes the missingness mechanism is ignorable. In real situations, there may be many causes for missing data and the missingness pattern could be a mixture of ignorable and non-ignorable missing values. As long as a large portion of missing data are ignorable and the non-ignorable portion of missing data are also included in the imputation model, the bias caused by treating all missing data as ignorable is negligible. Diggle and Kenward (1994) proposed a method for identifying the single reason of missingness. Denote

$p_j(y_1,\cdots,y_{j-1},y_j^*;\beta)$ be the probability of missing at visit j given the previous history $(y_1,\cdots,y_{j-1})$ and y_j^* where

$$y_j^* = \begin{cases} y_j & \text{if } y_j \text{ is observed} \\ 0 & \text{if } y_j \text{ is missing} \end{cases}$$

then the missingness is modeled by a logistic regression

$$\text{logit}\{p_j(y_1,\cdots,y_{j-1},y_j^*;\beta)\} = \beta_0 + \beta_1 y_1 + \cdots + \beta_{j-1} y_{j-1} + \beta_j y_j^*$$

(1) if $\beta_1 = \beta_2 = \cdots = \beta_{j-1} = \beta_j = 0$, then the missingness mechanism is MCAR.

(2) if $\beta_j = 0$, then the missingness mechanism is MAR.

(3) if $\beta_j \neq 0$, then the missingness mechanism is non-ignorable.

Rubin (1987) and Schafer (1997) have shown that a small number of multiple imputations can provide estimates of standard errors that are almost fully efficient. Schafer (1999) suggested that no more than 10 imputations are usually required. However, in the situations where the proportion of missing data is large, the inference based on the observed-data likelihood may be unstable and the inference based on the multiple imputation may have quite large variation. In this case, Horton and Lipsitz (2001) suggested a close investigation should be conducted. An increase in the number of imputations may stabilize the variation.

Bibliography

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In second international symposium on information theory, Akademiai Kiado, Budapest, 267-281.

Altham, P.M.E. (1984) Improving the precision of estimation by fitting a model. Journal of Royal Statistical Society, Series B, **46**, 118-119.

Anderson, T.W. (1957) Maximum likelihood estimates for a multivariate distribution when some observations are missing. Journal of the American Statistical Association, **57**, 200-203.

Anderson, T.W. (1984) An introduction to multivariate statistical analysis (Second edition). J. Wiley & Sons, New York.

Barnard, J. and Rubin, D. B. (1999) Small-sample degrees of freedom with multiple imputation. Biometrika, **86**, 948-955.

Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chain. Ann. Math. Statist., **41**, 164-171.

Bayes, T.R. (1763) An essay towards solving a problem in the doctrine of chances. Phil. Trans. Roy. Soc. London, **53**, 370-418.

Beale, E.M.L. and Little, R.J.A. (1975) Missing values in multivariate analysis. Journal of the Royal Statistical Society Series B, **37**, 129-145.

Beaton, A.E. (1964) The use of special matrix operations in statistical calculus. Research Bulletin RB-64-51, Educational Testing Service, Princeton, NJ.

Berndt, E.B., Hall, B., Hall, R., and Hausman, J.A. (1974) Estimation and inference in nonlinear structural models. Ann. Econ. Sco. Meas. **3**, 653-665.

Box, G.E.P. and Tiao, G.C. (1973) Bayesian inference in statistical analysis. J. Wiley and Sons, New York.

Cowles, M.K. and Carlin, B.P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of American Statistical Association, **91**, 883-904.

Cox, D.R. and Hinkley, D.V. (1974) Theoretical Statistics. New York: Wiley.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, Series B, **39**, 1-38.

Diggle, P. and Kenward, M.G. (1994) Informative drop-out in longitudinal analysis (with discussion). Applied Statistics, **43**, 49-93.

Efron, B. (1994) Missing data, imputation, and the bootstrap (with discussion). Journal of the American Statistical Association, **89**, 463-479.

Ezzati-Rice, T.M., Johnson, W., Khare, M., Little, R.J.A., Rubin, D.B. & Schafer, J.L. (1995) A simulation study to evaluate the performance of model-based multiple imputations in NCHS health examination surveys. In proceeding of the annual research conference, 257-266, Bureau of the Census, Washington, D.C.

Fisher, R.A. (1922) On the mathematical foundations of theoretical statistics. Phil. Trans. Roy. Soc. London Series A, **222**, 309-368.

Fisher, R.A. (1925) Theory of statistical information. Proc. Cambridge Philos. Soc., **22**, 700-725.

Gabriel, K.R. (1961) The model of ante-dependence for data of biological growth. Bulletin Institut International Statistique (Paris), **39**, 253-264.

Gabriel, K.R. (1962) Ante-dependence analysis of an ordered set of variables. Ann. Math. Statistics, **33**, 201-212.

Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. Journal of the Aemrican Statistical Association, **85**, 398-409.

Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the Aemrican Statistical Association, **85**, 972-985.

Gelman, A., Rubin, D.B., Carlin, J. and Stern, H. (1995) Bayesian Data Analysis. Chapman & Hall, London.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machines Intelligence, **6**, 721-741.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996) Markov chain Monte Carlo in practice. Chapman & Hall, London.

Hartley, H.O. (1958) Maximum likelihood estimation from incomplete data. Biometrics, **14**,174-194.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika, **57**, 97-109.

Horton, N.J. and Lipsitz, S.R. (2001) Multiple imputation in practice: Comparison of software package for regression models with missing variables. The American Statistician, **55**, 244-254.

Jeffreys, H.S. (1961) Theory of probability. Oxford: University Press.

Kennedy, W.J. and Gentle, J.E. (1980) Statistical Computing. Marcel Dekker, New York.

Kenward, M.G. (1987) A method for comparing profiles of repeated measurements. Journal of Royal Statistical Society, Series C, **36**, 296-308.

Kleijnan, J.P.C. (1974) Statistical Techniques in Simulation - Part I. Marcel Dekker, New York.

Lavori, P.W., Dawson, R. and Shera, D. (1995) A Multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine*, **14**, 1913-1925.

Lee, P.M. (1997) Baysian statistics: An introduction. Second edition. Edward Arnold, London.

Li, K. H. (1988) Imputation using Markov chains. Journal of Statistical Computation and Simulation, **30**, 57-79.

Li, K.H., Raghunathan, T.E. anad Rubin, R.B. (1991) Large-sample significance levels from multiply-imputed data using moment-based statistics and an F reference distribution. Journal of the American Statistical Association, **86**, 1065-1073.

Li, K.H., Meng, X.L., Raghunathan, T.E. anad Rubin, R.B. (1991) Significance levels from repeated p-values with multiply-imputed data. Statistica Sinica, **1**, 65-92.

Lindley, D.V. (1965) Introduction to probability and statistics from a Bayesian viewpoint. (2 vols-Part I: Probability and Part II: Inference), Cambridge: University Press.

Little, R.J.A. and Rubin, D.B. (1987) Statistical analysis with missing data. New York: John Wiley.

Little, R.J.A. (1994) A class of pattern-mixture models for normal missing data. Biometrika, **81**, 471-483.

Little, R. J. A. (1995) Modeling the drop-out mechanism in repeated-measures studies. Journal of American Statistical Association, **90**, 1112-1121.

Macchiavelli, R.E. and Arnold, S.F. (1994) Variable order ante-dependence models. Commun. Statist. – Theory Meth., **23(9)**, 2683-2299.

McKendrick, A.G. (1926) Application of mathematics to medical problems. Proc. Edinburgh Math. Soc. **44**, 98-130.

Meng, X.L. and Rubin, D.B. (1992) Performing likelihood ration tests with multiply-imputed data sets. Biometrika, **79**,103-111.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) Equations of state calculations by fast computing machines. The Journal of Chemical Physica. **21**, 1087-1092.

Olkin, I. and Tate, R.F. (1961) Multivariate correlation models with mixed discrete and continuous variables. Annals of Mathematical Statistics, **32**, 488-465.

Orchard, T. and Woodbury, M.A. (1972) A missing information principle: Theory and applications. Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, **1**, 697-715.

Patel, H.I. (1991) Analysis of incomplete data from a clinical trial with repeated measurements. Biometrika, **78**, 609-619.

Roberts, G.O. (1995) Markov chain concepts related to sampling algorithms. In Markov Chain Monte Carlo in Practice (eds Gilks, Richardson and Spiegelhalter), 45-57. London: Chapman & Hall.

Rubin, D.B. (1976) Inference and missing data. Biometrika, **63**, 581-592.

Rubin, D.B. and Schenker, N. (1986) Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. Journal of the American Statistical Association, **81**, 366-374.

Rubin, D.B. (1987) Multiple imputation for non-response in surveys. New York: John Wiley.

Rubin, D.B. and Shenker, N. (1991) Multiple imputation in health-care database: An overview and some applications. Statistics in Medicine, **10**, 585-598.

Rubin, D.B. (1994) Comments on "Missing data, imputation, and the bootstrap" by B. Efron. Journal of the American Statistical Association, **89**, 475-478.

Rubin, D.B. (1996) Multiple imputation after 18+ years. Journal of American Statistical Association, **91**, 473-489.

SAS/STAT Software: Changes and Enhancements Through Release 6.12 (1997). SAS Institute, North Carolina.

Schafer, J.L. (1997) Analysis of incomplete multivariate data. Chapman & Hall, London.

Schafer, J.L. (1999) Multiple imputation: A primer. Statistical Methods in Medical Research, **8**, 3-15.

Schafer, J.L. (Submitted) Imputation of missing covariates under a multivariate linear mixed model. Biometrics.

Schwarz, G. (1978) Estimating the dimension of a model. Annals Statistics, **6**, 461-464.

Solas 3.0. User Reference (1999) Statistical Solutions Ltd, Ireland.

Spiegelhalter, D.J., Freedman, L.S. and Palmer, M.B. (1994) Bayesian Approaches to random trials. Journal of Royal Statistical Society A, **157**, 357-387.

Sundberg, R. (1974) Maximum likelihood theory for incomplete data from an exponential family. Scand. J. Statist. **1**, 49-58.

Tanner, M.A. and Wong, W.H. (1987) The calculation of posterior distributions by data augmentation (with discussion). Journal of American Statistical Association, **82**, 528-550.

Tanner, M.A. (1993) Tools for Statistical Inference. Methods for the Exploration of Posterior Distributions and Likelihood Functions (Second edition). Springer-Verlag, New York.

Tierney, L. (1995) Introduction to general state-space Markov chain theory. In Markov Chain Monte Carlo in Practice (eds Gilks, Richardson and Spiegelhalter), 59-74. London: Chapman & Hall.

Wu, C.F.J. (1983) On the convergence properties of the EM algorithm. Annals of Statistics, **11**, 95-103.

Appendix 1: Estimate the Order of the Ante-dependence Structure

In real situations, the order of an ante-dependence structure is unknown and needs to be estimated from the data. When a p-dimensional multivariate sample is assumed to have an ante-dependence structure of order $g\,(g=0,1,\cdots,p-1)$, the total number of unknown parameters is $p+(g+1)(2p-g)/2$. Order 0 indicates a complete independence covariance structure and order $p-1$ indicates an unstructured covariance. When the model under the null hypothesis is nested within the model under the alternative hypothesis, the likelihood ratio test can be applied to test the plausibility of the model under the null hypothesis. However, the models with different orders of ante-dependence structure are not nested within each other. The penalized likelihood criteria, like Akaike's Information Criterion (*AIC*) and Schwarz's Bayesian Information Criterion (*BIC*), are used to make the model selection. These penalized likelihoods are essentially the log likelihood values penalized for the number of parameters estimated. The *BIC* imposes heavier penalty than that by *AIC*. The penalized likelihood is given by

$$PL(k) = -2\log L(Y;\hat{\theta}_k) + d_k c(n) \qquad \text{(A.1)}$$

where k indexes the models under the consideration $(k=0,1,\cdots,p-1)$, $L(Y;\hat{\theta}_k)$ is the maximized likelihood for the k^{th} model, d_k is the number of unknown parameters in the k^{th} model. The quantity $c(n)$ is a penalty index, $c(n)=2$ for *AIC* and $c(n)=\log(n)$ for *BIC*.

The penalized likelihood criterion is to minimize (A.1). That is, the model with the lowest value of $PL(k)$ is most desirable and the corresponding k is the most desirable order of the ante-dependence structure supported by the data.

Under the multivariate normal assumption, the unknown parameters are $\theta=(\mu,\Sigma)$. When data are complete or incomplete with a monotone missing pattern, these parameters can be estimated directly by the successive regression method described in section 5.3 under the assumption of an ante-dependence structure. When data are incomplete with a non-monotone missing pattern, $\theta=(\mu,\Sigma)$ are estimated by the EM algorithm under the assumption of an unstructured covariance matrix. Then by using the parameter transformation described in either section 6.3 or section 6.4, we obtain the maximum likelihood estimates of the alternative parameters under the assumption of ante-dependence. Taking the reverse parameter transformation, we obtain the estimates of the normal parameter $\theta=(\mu,\Sigma)$ under the ante-dependence assumption.

Once the maximum likelihood estimates of $\theta = (\mu, \Sigma)$ are obtained, the maximized observed-data likelihood is

$$L(\hat{\mu}, \hat{\Sigma} \mid Y_{obs}) = \prod_{i=1}^{n} L(\hat{\mu}_i, \hat{\Sigma}_i \mid y_{i(obs)}) \tag{A.2}$$

where

$$L(\hat{\mu}_i, \hat{\Sigma}_i \mid y_{i(obs)}) \propto |\hat{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_{i(obs)} - \hat{\mu}_i)^T \hat{\Sigma}_i^{-1}(y_{i(obs)} - \hat{\mu}_i)\right\} \tag{A.3}$$

is the maximized observed-data likelihood for the i^{th} observation, and $\hat{\mu}_i$ and $\hat{\Sigma}_i$ are the estimated marginal mean vector and estimated marginal covariance matrix corresponding to $y_{i(obs)}$ respectively. Hence the penalized likelihood becomes

$$PL(k) = \sum_{i=1}^{n} \left[\log(|\hat{\Sigma}_i|) + (y_{i(obs)} - \hat{\mu}_i)^T \hat{\Sigma}_i^{-1}(y_{i(obs)} - \hat{\mu}_i)\right] + d_k c(n)$$

For each admissible value of k ($k = 0,1,\cdots,p-1$), the corresponding penalized likelihood values are calculated and the order g of the ante-dependence is selected if its penalized likelihood value is the lowest, i.e.

$$PL(g) \le PL(k) \qquad k = 0,1,\cdots,p-1$$

Appendix 2: Splus Functions for Simulations

There are 6 primary Splus functions written by author to perform the simulations.

MSIMU1 – Monotone incomplete data with *structured* covariance.

MSIMU2 – Monotone incomplete data with *unstructured* covariance.

MSIMU3 – Monotone incomplete data with *structured* covariance from 3 multivariate normal.

NSIMU1 – Non-monotone incomplete data with *structured* covariance.

NSIMU2 – Non-monotone incomplete data with *unstructured* covariance.

NSIMU3 – Monotone incomplete data with *unstructured* covariance from 3 multivariate normal.

There are serveral sub-functions to be called by the above primary functions. These sub-functions are available upon requted.

MONOFIT1 – Estimate & simulate the normal parameters for monotone incomplete data (structured covariance).

MONOFIT2 – Estimate & simulate the normal parameters for monotone incomplete data (unstructured covariance).

MONOFIT3 – Estimate & simulate the normal parameters for monotone incomplete data from 3 multivariate normal populations (structured covariance).

MONOMI – Impute the missing data for monotone incomplete data.

PRED1 – Predictive model method of imputation (monotone).

PRED2 – Predictive model method of imputation (non-monotone).

SEQU1 – Sequentially impute the non-monotone data with structured covariance.

SEQU2 – Sequentially impute the non-monotone data with unstructured covariance.

```
MSIMU1:
function(m){
# Monotone Simulation 1 (single population with ante-dependence of order 1)
# 1. Generate complete data from 3-d mutlvariate normal distribution with structured covariance
# 2. Call MK1 to generate monotone incomplete
# 3. Call MONOFIT1 to get estimated & simulated normal parameters (structured covariance)
# 4. Call MONOMI to simulate the missing data
# 5. Call PRED1 to get simulate the missing data by predictive model method
# 6. Call DA.NORM and IMP.NORM to get MCMC simulated data
# 7. Call MI.INFERENCE to get combined estimates, std.err and confidence interval
        n <- 30                         # sample size of multivariate normal
        library(norm)                   # in order to use Schafer's program
        rngseed(1234567)
        R1 <- matrix(0, m, 12)          # Estimated mean, covariance and t (no imputation)
        MSE1 <- matrix(0, m, 12)        # Mean square error of estimates and t
        R2 <- matrix(0, m, 18)          # Proposed method (structured)
        MSE2 <- matrix(0, m, 12)
        R3 <- matrix(0, m, 18)          # Predictive model method
        MSE3 <- matrix(0, m, 12)
        R4 <- matrix(0, m, 18)          # MCMC method
        MSE4 <- matrix(0, m, 12)
        SR1 <- matrix(0, m, 9)          # 95% confidence limits and indicator (no imputation)
        SR2 <- matrix(0, m, 9)
        SR3 <- matrix(0, m, 9)
        SR4 <- matrix(0, m, 9)
        cov1 <- rbind(c(1, 0.193, 0.0109), c(0.193, 2, 0.113), c(0.0109, 0.113, 3))
        xx <- c(0, 0, 0, 0, 0, 0, 1, 0.193, 0.0109, 2, 0.113, 3)
        for(i in 1:m) {
                y <- rmvnorm(n, mean = rep(0, 3), cov = cov1)
        #------ Create monotone incomplete data (at least 3 obs with all data observed) --------
                y1 <- mk1(y)                    # generate monotone incomplete data
                n2 <- nrow(y1[!is.na(y1[, 2]), ])   # number of non-missing of Y[2]
                n3 <- nrow(y1[!is.na(y1[, 3]), ])
                y1.p <- prelim.norm(y1)
        #------ Estimate mean and its standard error by proposed method (no imputation) ------
                para <- monofit1(y1)$Est        # estimated normal parameter matrix
                ser <- c(sqrt(para[2, 2]/n), sqrt(para[3, 3]/n2), sqrt(para[4, 4]/n3))
```

```
        R1[i, c(1, 3, 5, 7, 8, 9, 10, 11, 12)] <- extract(para)     # est. mean & covariance
        R1[i, c(2, 4, 6)] <- c(para[1, 2]/ser[1], para[1, 3]/ser[2], para[1, 4]/ser[3])   # t values
        MSE1[i, ] <- (R1[i, ] - xx)^2        # mean square error
        SR1[i, c(1, 2, 3)] <- CI(R1[i, 1], ser[1], n - 1, 0)
        SR1[i, c(4, 5, 6)] <- CI(R1[i, 3], ser[2], n2 - 1, 0)
        SR1[i, c(7, 8, 9)] <- CI(R1[i, 5], ser[3], n3 - 1, 0)
#------ Proposed method (multiple imputation, m=5) ----------------
        temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
        temp2 <- matrix(0, 5, 3)    # standard error of each imputation
        for(k in 1:5) {
        para <- monofit1(y1)$Sim            # simulated normal parameters
        yt <- monomi(y1, para)              # imputed complete data
        temp1[k, c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
        vt <- var(yt, na.method = "available")
        temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
        temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
        temp1[k, 9] <- vt[3, 3]
        temp2[k, ] <- c(sqrt(vt[1, 1]/n), sqrt(vt[2, 2]/n), sqrt(vt[3, 3]/n))
        }
      est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
      std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
      result <- mi.inference(est, std.err, confidence = 0.95)
      R2[i, c(1, 3, 5)] <- result$est                    # combined estimation of mean
      R2[i, c(2, 4, 6)] <- result$est/result$std.err     # computed t values
      R2[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)  # estimated of covariance
      R2[i, c(13, 14, 15)] <- result$r                   # relative increase in variance due to missing data
      R2[i, c(16, 17, 18)] <- result$fminf               # fraction of missing information about parameter
      MSE2[i, ] <- (R2[i, c(1:12)] - xx)^2
      SR2[i, c(1, 4, 7)] <- result$lower
      SR2[i, c(2, 5, 8)] <- result$upper
      if(SR2[i, 1] <= 0 & 0 <= SR2[i, 2]) {SR2[i, 3] <- 1}
      if(SR2[i, 4] <= 0 & 0 <= SR2[i, 5]) {SR2[i, 6] <- 1}
      if(SR2[i, 7] <= 0 & 0 <= SR2[i, 8]) {SR2[i, 9] <- 1}
#------ Predictive model method (multiple imputation, m=5) ----------------
      temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
      temp2 <- matrix(0, 5, 3)    # standard error of each imputation
      for(k in 1:5) {
```

```
        yt <- pred1(y1)    # imputed complete data
        temp1[k, c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
        vt <- var(yt, na.method = "available")
        temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
        temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
        temp1[k, 9] <- vt[3, 3]
        temp2[k, ] <- c(sqrt(vt[1, 1]/n), sqrt(vt[2, 2]/n), sqrt(vt[3, 3]/n))
        }
    est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
    std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
    result <- mi.inference(est, std.err, confidence = 0.95)
    R3[i, c(1, 3, 5)] <- result$est                    # combined estimation of mean
    R3[i, c(2, 4, 6)] <- result$est/result$std.err     # computed t values
    R3[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)      # estimated of covariance
    R3[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
    R3[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
    MSE3[i, ] <- (R3[i, c(1:12)] - xx)^2
    SR3[i, c(1, 4, 7)] <- result$lower
    SR3[i, c(2, 5, 8)] <- result$upper
    if(SR3[i, 1] <= 0 & 0 <= SR3[i, 2]) {SR3[i, 3] <- 1}
    if(SR3[i, 4] <= 0 & 0 <= SR3[i, 5]) {SR3[i, 6] <- 1}
    if(SR3[i, 7] <= 0 & 0 <= SR3[i, 8]) {SR3[i, 9] <- 1}
#------ MCMC method (multiple imputation, m=5) ----------
    temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
    temp2 <- matrix(0, 5, 3)    # standard error of each imputation
    para <- em.norm(y1.p, showits = F) # initial values of normal parameters
    for(k in 1:5) {
        para <- da.norm(y1.p, para, steps = 50, showits = F)   # simulated parameters
        yt <- imp.norm(y1.p, para, y1)        # imputed complete data
        temp1[k, c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
        vt <- var(yt, na.method = "available")
        temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
        temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
        temp1[k, 9] <- vt[3, 3]
        temp2[k, ] <- c(sqrt(vt[1, 1]/n), sqrt(vt[2, 2]/n), sqrt(vt[3, 3]/n))
        }
    est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
```

```
        std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
        result <- mi.inference(est, std.err, confidence = 0.95)
        R4[i, c(1, 3, 5)] <- result$est                    # combined estimation of mean
        R4[i, c(2, 4, 6)] <- result$est/result$std.err     # computed t values
        R4[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)      # estimated of covariance
        R4[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
        R4[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
        MSE4[i, ] <- (R4[i, c(1:12)] - xx)^2
        SR4[i, c(1, 4, 7)] <- result$lower
        SR4[i, c(2, 5, 8)] <- result$upper
        if(SR4[i, 1] <= 0 & 0 <= SR4[i, 2]) {SR4[i, 3] <- 1}
        if(SR4[i, 4] <= 0 & 0 <= SR4[i, 5]) {SR4[i, 6] <- 1}
        if(SR4[i, 7] <= 0 & 0 <= SR4[i, 8]) {SR4[i, 9] <- 1}
        }
        R1 <- cbind(R1, MSE1)
        R2 <- cbind(R2, MSE2)
        R3 <- cbind(R3, MSE3)
        R4 <- cbind(R4, MSE4)
return(m1=apply(R1,2,mean),m2=apply(R2,2,mean),m3=apply(R3,2,mean),m4=apply(R4,2,mean))
}
```

```
MSIMU2
function(m){
# Monotone Simulation 1 (single population with unstructured covariance)
# 1. Generate complete data from 3-d mutlvariate normal distribution with unstructured covariance
# 2. Call MK1 to generate monotone incomplete
# 3. Call MONOFIT2 to get estimated & simulated normal parameters (unstructured covariance)
# 4. Call MONOMI to simulate the missing data
# 5. Call PRED1 to get simulate the missing data by predictive model method
# 6. Call DA.NORM and IMP.NORM to get MCMC simulated data
# 7. Call MI.INFERENCE to get combined estimates, std.err and confidence interval
        n <- 30                 # sample size of multivariate normal
        library(norm)           # in order to use Schafer's program
        rngseed(1234567)
        R1 <- matrix(0, m, 12)  # Estimated mean, covariance and t (no imputation)
        MSE1 <- matrix(0, m, 12) # Mean square error of estimates and t
        R2 <- matrix(0, m, 18)  # Proposed method (structured)
        MSE2 <- matrix(0, m, 12)
        R3 <- matrix(0, m, 18)  # Predictive model method
        MSE3 <- matrix(0, m, 12)
        R4 <- matrix(0, m, 18)  # MCMC method
        MSE4 <- matrix(0, m, 12)
        SR1 <- matrix(0, m, 9)  # 95% confidence limits and indicator (no imputation)
        SR2 <- matrix(0, m, 9)
        SR3 <- matrix(0, m, 9)
        SR4 <- matrix(0, m, 9)
        cov1 <- rbind(c(1, 0.5, 0.9), c(0.5, 2, 0.3), c(0.9, 0.3, 3))
        xx <- c(0, 0, 0, 0, 0, 0, 1, 0.5, 0.9, 2, 0.3, 3)
        for(i in 1:m) {
                y <- rmvnorm(n, mean = rep(0, 3), cov = cov1)
        #------ Create monotone incomplete data (at least 3 obs with all data observed) ----
                y1 <- mk1(y)    # generate monotone incomplete data
                n2 <- nrow(y1[!is.na(y1[, 2]), ])   # number of non-missing of Y[2]
                n3 <- nrow(y1[!is.na(y1[, 3]), ])
                y1.p <- prelim.norm(y1)
        #------ Estimate mean and its standard error by proposed method (no imputation) ---
                para <- monofit2(y1)$Est            # estimated normal parameter matrix
                ser <- c(sqrt(para[2, 2]/n), sqrt(para[3, 3]/n2), sqrt(para[4, 4]/n3))
```

```
        R1[i, c(1, 3, 5, 7, 8, 9, 10, 11, 12)] <- extract(para)    # estimates mean & covariance
        R1[i, c(2, 4, 6)] <- c(para[1, 2]/ser[1], para[1, 3]/ser[2], para[1, 4]/ser[3])  # t values
        MSE1[i, ] <- (R1[i, ] - xx)^2        # mean square error
        SR1[i, c(1, 2, 3)] <- CI(R1[i, 1], ser[1], n - 1, 0)
        SR1[i, c(4, 5, 6)] <- CI(R1[i, 3], ser[2], n2 - 1, 0)
        SR1[i, c(7, 8, 9)] <- CI(R1[i, 5], ser[3], n3 - 1, 0)
#------ Proposed method (multiple imputation, m=5) ----------------
        temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
        temp2 <- matrix(0, 5, 3)    # standard error of each imputation
        for(k in 1:5) {
        para <- monofit2(y1)$Sim # simulated normal parameters
        yt <- monomi(y1, para)     # imputed complete data
        temp1[k,c(1:3)] <- c(mean(yt[,1],na.rm=T), mean(yt[,2], na.rm=T),mean(yt[,3],na.rm=T))
        vt <- var(yt, na.method = "available")
        temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
        temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
        temp1[k, 9] <- vt[3, 3]
        temp2[k, ] <- c(sqrt(vt[1, 1]/n), sqrt(vt[2, 2]/n), sqrt(vt[3, 3]/n))
        }
    est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
    std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
    result <- mi.inference(est, std.err, confidence = 0.95)
    R2[i, c(1, 3, 5)] <- result$est          # combined estimation of mean
    R2[i, c(2, 4, 6)] <- result$est/result$std.err   # computed t values
    R2[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)       # estimated of covariance
    R2[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
    R2[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
    MSE2[i, ] <- (R2[i, c(1:12)] - xx)^2
    SR2[i, c(1, 4, 7)] <- result$lower
    SR2[i, c(2, 5, 8)] <- result$upper
    if(SR2[i, 1] <= 0 & 0 <= SR2[i, 2]) {SR2[i, 3] <- 1}
    if(SR2[i, 4] <= 0 & 0 <= SR2[i, 5]) {SR2[i, 6] <- 1}
    if(SR2[i, 7] <= 0 & 0 <= SR2[i, 8]) {SR2[i, 9] <- 1}
#------ Predictive model method (multiple imputation, m=5) ----------------
    temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
    temp2 <- matrix(0, 5, 3)    # standard error of each imputation
    for(k in 1:5) {
```

```
            yt <- pred1(y1)    # imputed complete data
            temp1[k,c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
            vt <- var(yt, na.method = "available")
            temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
            temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
            temp1[k, 9] <- vt[3, 3]
            temp2[k, ] <- c(sqrt(vt[1, 1]/n), sqrt(vt[2, 2]/n), sqrt(vt[3, 3]/n))
            }
        est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
        std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
        result <- mi.inference(est, std.err, confidence = 0.95)
        R3[i, c(1, 3, 5)] <- result$est          # combined estimation of mean
        R3[i, c(2, 4, 6)] <- result$est/result$std.err   # computed t values
        R3[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)       # estimated of covariance
        R3[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
        R3[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
        MSE3[i, ] <- (R3[i, c(1:12)] - xx)^2
        SR3[i, c(1, 4, 7)] <- result$lower
        SR3[i, c(2, 5, 8)] <- result$upper
        if(SR3[i, 1] <= 0 & 0 <= SR3[i, 2]) {SR3[i, 3] <- 1}
        if(SR3[i, 4] <= 0 & 0 <= SR3[i, 5]) {SR3[i, 6] <- 1}
        if(SR3[i, 7] <= 0 & 0 <= SR3[i, 8]) {SR3[i, 9] <- 1}
#------ MCMC method (multiple imputation, m=5) ----------
        temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
        temp2 <- matrix(0, 5, 3)    # standard error of each imputation
        para <- em.norm(y1.p, showits = F) # initial values of normal parameters
        for(k in 1:5) {
            para <- da.norm(y1.p, para, steps = 50, showits = F)   # simulated parameters
            yt <- imp.norm(y1.p, para, y1)       # imputed complete data
            temp1[k,c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
            vt <- var(yt, na.method = "available")
            temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
            temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
            temp1[k, 9] <- vt[3, 3]
            temp2[k, ] <- c(sqrt(vt[1, 1]/n), sqrt(vt[2, 2]/n), sqrt(vt[3, 3]/n))
            }
        est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
```

```
        std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
        result <- mi.inference(est, std.err, confidence = 0.95)
        R4[i, c(1, 3, 5)] <- result$est          # combined estimation of mean
        R4[i, c(2, 4, 6)] <- result$est/result$std.err   # computed t values
        R4[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)       # estimated of covariance
        R4[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
        R4[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
        MSE4[i, ] <- (R4[i, c(1:12)] - xx)^2
        SR4[i, c(1, 4, 7)] <- result$lower
        SR4[i, c(2, 5, 8)] <- result$upper
        if(SR4[i, 1] <= 0 & 0 <= SR4[i, 2]) {SR4[i, 3] <- 1}
        if(SR4[i, 4] <= 0 & 0 <= SR4[i, 5]) {SR4[i, 6] <- 1}
        if(SR4[i, 7] <= 0 & 0 <= SR4[i, 8]) {SR4[i, 9] <- 1}
        }
        R1 <- cbind(R1, MSE1)
        R2 <- cbind(R2, MSE2)
        R3 <- cbind(R3, MSE3)
        R4 <- cbind(R4, MSE4)
return(m1=apply(R1,2,mean),m2=apply(R2,2,mean),m3=apply(R3,2,mean),m4 =apply(R4,2,mean))
}
```

```
MSIMU3
function(m){
# Monotone Simulation 3 (multiple population with ante-dependence of order 1)
# 1. Generate complete data from 3-d mutlvariate normal distribution with structured covariance
# 2. Call MK2 to generate monotone incomplete
# 3. Call MONOFIT3 to get estimated & simulated normal parameters (structured covariance)
# 4. Call MONOMI to simulate the missing data within each population
# 5. Call CI to get confidence interval for non-imputation estimates
        n <- 20
        R1 <- matrix(0, m, 9)       # estimated mean vector (no imputation)
        MSE1 <- matrix(0, m, 9)
        SR1 <- matrix(0, m, 27)
        R2 <- matrix(0, m, 9)       # estimated mean vector (multiple imputation)
        MSE2 <- matrix(0, m, 9)
        SR2 <- matrix(0, m, 27)
        A1 <- matrix(-1, 4, 4)
        A2 <- matrix(-1, 4, 4)
        A3 <- matrix(-1, 4, 4)
        xx <- c(0, 0, 0, 2, 2, 2, 4, 4, 4)      # true population mean
        for(i in 1:m) {
        y <- mk2(n)
        nm <- c(nrow(y[!is.na(y[, 2]), ]), nrow(y[!is.na(y[, 3]), ]), nrow(y[!is.na(y[, 4]), ]))
        y1 <- y[y[, 1] == 1, c(2, 3, 4)]
        y2 <- y[y[, 1] == 2, c(2, 3, 4)]
        y3 <- y[y[, 1] == 3, c(2, 3, 4)]
#------ estimate mean and covariance from monotone incomplete data (no imputation) ---
        result<-monofit3
        para <- result$Est
        m1 <- para[1, c(2, 3, 4), 1]# estimated mean vector 1
        m2 <- para[1, c(2, 3, 4), 2]
        m3 <- para[1, c(2, 3, 4), 3]
        R1[i, ] <- c(m1, m2, m3)  # simulated mean vectors of 3 populations
        MSE1[i, ] <- (R1[i, ] - xx)^2
        aev <- (para[c(2,3,4),c(2,3,4),1]+para[c(2,3,4),c(2,3,4),2]+para[c(2,3,4),c(2,3,4),3])/3 # est
        common covariance
        n2 <- nrow(y1[!is.na(y1[, 2]), ])    # number of non-missing of Y[2] in pop 1
        n3 <- nrow(y1[!is.na(y1[, 3]), ])
```

```
        se1 <- c(sqrt(aev[1, 1]/n), sqrt(aev[2, 2]/n2), sqrt(aev[3, 3]/n3))
        SR1[i, c(1, 2, 3)] <- CI(R1[i, 1], se1[1], nm[1] - 1, 0)
        SR1[i, c(4, 5, 6)] <- CI(R1[i, 2], se1[2], nm[2] - 1, 0)
        SR1[i, c(7, 8, 9)] <- CI(R1[i, 3], se1[3], nm[3] - 1, 0)
        n2 <- nrow(y2[!is.na(y2[, 2]), ])     # number of non-missing of Y[2] in pop 2
        n3 <- nrow(y2[!is.na(y2[, 3]), ])
        se2 <- c(sqrt(aev[1, 1]/n), sqrt(aev[2, 2]/n2), sqrt(aev[3, 3]/n3))
        SR1[i, c(10, 11, 12)] <- CI(R1[i, 4], se2[1], nm[1] - 1, 2)
        SR1[i, c(13, 14, 15)] <- CI(R1[i, 5], se2[2], nm[2] - 1, 2)
        SR1[i, c(16, 17, 18)] <- CI(R1[i, 6], se2[3], nm[3] - 1, 2)
        n2 <- nrow(y3[!is.na(y3[, 2]), ])     # number of non-missing of Y[2]
        n3 <- nrow(y3[!is.na(y3[, 3]), ])
        se3 <- c(sqrt(aev[1, 1]/n), sqrt(aev[2, 2]/n2), sqrt(aev[3, 3]/n3))
        SR1[i, c(19, 20, 21)] <- CI(R1[i, 7], se3[1], nm[1] - 1, 4)
        SR1[i, c(22, 23, 24)] <- CI(R1[i, 8], se3[2], nm[2] - 1, 4)
        SR1[i, c(25, 26, 27)] <- CI(R1[i, 9], se3[3], nm[3] - 1, 4)
#------ estimate mean and covariance from monotone incomplete data (multiple imputation) ----
        para <- result$Sim
        m1 <- para[1, c(2, 3, 4), 1]          # simulated mean vector 1
        m2 <- para[1, c(2, 3, 4), 2]
        m3 <- para[1, c(2, 3, 4), 3]
        aev <- (para[c(2, 3, 4), c(2, 3, 4), 1] + para[c(2, 3, 4), c(2, 3, 4), 2] + para[c(2, 3, 4), c(
                        2, 3, 4), 3])/3       # simulated common covariance
        A1[c(2, 3, 4), c(2, 3, 4)] <- aev
        A1[1, c(2, 3, 4)] <- m1
        A1[c(2, 3, 4), 1] <- m1
        A2[c(2, 3, 4), c(2, 3, 4)] <- aev
        A2[1, c(2, 3, 4)] <- m2
        A2[c(2, 3, 4), 1] <- m2
        A3[c(2, 3, 4), c(2, 3, 4)] <- aev
        A3[1, c(2, 3, 4)] <- m3
        A3[c(2, 3, 4), 1] <- m3
        temp1 <- matrix(0, 5, 9)   # estimated mean parameters
        temp2 <- matrix(0, 5, 9)   # standard errors of the estimates
        for(k in 1:5) {yt1 <- sequ1(y1, A1)
                       yt2 <- sequ1(y2, A2)
                       yt3 <- sequ1(y3, A3)
```

```
            m1 <- c(mean(yt1[,1],na.rm=T),mean(yt1[,2],na.rm=T),mean(yt1[,3],na.rm=T))
            m2 <- c(mean(yt2[,1],na.rm=T),mean(yt2[,2],na.rm=T),mean(yt2[,3],na.rm=T))
            m3 <- c(mean(yt3[,1],na.rm=T),mean(yt3[,2],na.rm=T),mean(yt3[,3],na.rm=T))
            aev <- (var(yt1,na.method="available")+var(yt2,na.method="available")+
                  var(yt3, na.method = "available"))/3
            se1 <- c(sqrt(aev[1, 1]/n), sqrt(aev[2, 2]/n), sqrt(aev[3, 3]/n))
            temp1[k, ] <- c(m1, m2, m3)
            temp2[k, ] <- c(se1, se1, se1)
            }
            est <- apply(temp1, 2, mean)        # combined estimate mean
            B <- apply(temp1, 2, var)           # between imputation variance
            U <- apply(temp2^2, 2, mean)        # within imputation variance
            V <- 4 * (1 + U/(1.2 * B))^2        # degrees of freedom
            se <- sqrt(1.2 * B + U)     # combined std error of combined estimate mean
            R2[i, ] <- est
            MSE2[i, ] <- (R2[i, ] - xx)^2
            V[c(1, 4, 7)] <- 0
            for(j in 1:9) {SR2[i,c(3*(j-1)+1,3*(j-1)+2,3*(j-1)+3)]<-CI(est[j],se[j],V[j],xx[j]) }
        }
return(m1 = apply(R1, 2, mean), mse1 = apply(MSE1, 2, mean), CI1 = apply(SR1, 2, mean),
      m2 = apply(R2, 2, mean), mse2 = apply(MSE2, 2, mean), CI2 = apply(SR2, 2, mean))
}
```

```
NSIMU1
function(m){
# Non-monotone Simulation 1 (single population with ante-dependence of order 1)
# 1. Generate complete data from 3-d mutlvariate normal distribution with structured covariance
# 2. Call MK3 to generate non-monotone incomplete data
# 3. Call EM.NORM to get estimated normal parameters
# 4. Call SEQU1 to impute the missing data sequentially
# 5. Call PRED2 to impute the missing data by predictive model method
# 6. Call DA.NORM and IMP.NORM to get MCMC imputed data
# 7. Call MI.INFERENCE to get combined estimates, std.err and confidence interval
        n <- 30          # sample size of multivariate normal
        library(norm)    # in order to use Schafer's program
        rngseed(1234567)
        R1 <- matrix(0, m, 12)    # Estimated mean, covariance and t (no imputation)
        MSE1 <- matrix(0, m, 12) # Mean square error of estimates and t
        R2 <- matrix(0, m, 18)    # Proposed method (structured)
        MSE2 <- matrix(0, m, 12)
        R3 <- matrix(0, m, 18)    # Predictive model method
        MSE3 <- matrix(0, m, 12)
        R4 <- matrix(0, m, 18)    # MCMC method
        MSE4 <- matrix(0, m, 12)
        SR1 <- matrix(0, m, 9)    # 95% confidence limits and indicator (no imputation)
        SR2 <- matrix(0, m, 9)
        SR3 <- matrix(0, m, 9)
        SR4 <- matrix(0, m, 9)
        para0 <- matrix(-1, 4, 4)
        cov1 <- rbind(c(1, 0.193, 0.0109), c(0.193, 2, 0.113), c(0.0109, 0.113, 3))
        xx <- c(0, 0, 0, 0, 0, 0, 1, 0.193, 0.0109, 2, 0.113, 3)
        for(i in 1:m) {
        y <- rmvnorm(n, mean = rep(0, 3), cov = cov1)
#------ Create monotone incomplete data (at least 3 obs with all data observed) ----
        y <- mk3(y)       # generate non-monotone incomplete data
        y1 <- y[!is.na(y[, 1]) | !is.na(y[, 2]) | !is.na(y[, 3]), ]  # at least one Y1,Y2,Y3 observed
        n0 <- nrow(y1)   # actual sample size
        n1 <- nrow(y1[!is.na(y1[, 1]), ])    # number of non-missing of Y[1]
        n2 <- nrow(y1[!is.na(y1[, 2]), ])    # number of non-missing of Y[2]
        n3 <- nrow(y1[!is.na(y1[, 3]), ])
```

```
        y1.p <- prelim.norm(y1)
#------ Estimate mean and its standard error by proposed method (no imputation) ---
        para <- em.norm(y1.p, showits = F)
        para0[1, c(2, 3, 4)] <- getparam.norm(y1.p, para)$mu
        para0[c(2, 3, 4), 1] <- getparam.norm(y1.p, para)$mu
        para0[c(2, 3, 4), c(2, 3, 4)] <- getparam.norm(y1.p, para)$sigma
        ser <- c(sqrt(para0[2, 2]/n1), sqrt(para0[3, 3]/n2), sqrt(para0[4, 4]/n3))
        R1[i, c(1, 3, 5, 7, 8, 9, 10, 11, 12)] <- extract(para0)  # estimates mean & covariance
        R1[i, c(2, 4, 6)] <- c(para0[1, 2]/ser[1], para0[1, 3]/ser[2], para0[1, 4]/ser[3])    # t values
        MSE1[i, ] <- (R1[i, ] - xx)^2        # mean square error
        SR1[i, c(1, 2, 3)] <- CI(R1[i, 1], ser[1], n1 - 1)
        SR1[i, c(4, 5, 6)] <- CI(R1[i, 3], ser[2], n2 - 1)
        SR1[i, c(7, 8, 9)] <- CI(R1[i, 5], ser[3], n3 - 1)
#------- Proposed method (multiple imputation, m=5) ----------------
        temp1 <- matrix(0, 5, 9)   # estimated mean of each imputation
        temp2 <- matrix(0, 5, 3)   # standard error of each imputation
        for(k in 1:5) {
                yt <- sequ1(y1, para0)      # imputed complete data
                temp1[k, c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
                vt <- var(yt, na.method = "available")
                temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
                temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
                temp1[k, 9] <- vt[3, 3]
                temp2[k, ] <- c(sqrt(vt[1, 1]/n0), sqrt(vt[2, 2]/n0), sqrt(vt[3, 3]/n0))
                }
        est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
        std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
        result <- mi.inference(est, std.err, confidence = 0.95)
        R2[i, c(1, 3, 5)] <- result$est        # combined estimation of mean
        R2[i, c(2, 4, 6)] <- result$est/result$std.err   # computed t values
        R2[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)      # estimated of covariance
        R2[i, c(13, 14, 15)] <- result$r      # relative increase in variance due to missing data
        R2[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
        MSE2[i, ] <- (R2[i, c(1:12)] - xx)^2
        SR2[i, c(1, 4, 7)] <- result$lower
        SR2[i, c(2, 5, 8)] <- result$upper
        if(SR2[i, 1] <= 0 & 0 <= SR2[i, 2]) {SR2[i, 3] <- 1}
```

```
        if(SR2[i, 4] <= 0 & 0 <= SR2[i, 5]) {SR2[i, 6] <- 1}
        if(SR2[i, 7] <= 0 & 0 <= SR2[i, 8]) {SR2[i, 9] <- 1}
#------ Predictive model method (multiple imputation, m=5) ----------------
        temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
        temp2 <- matrix(0, 5, 3)    # standard error of each imputation
        for(k in 1:5) {
                yt <- pred2(y1)    # imputed complete data
                temp1[k, c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
                vt <- var(yt, na.method = "available")
                temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
                temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
                temp1[k, 9] <- vt[3, 3]
                temp2[k, ] <- c(sqrt(vt[1, 1]/n0), sqrt(vt[2, 2]/n0), sqrt(vt[3, 3]/n0))
                }
        est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
        std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
        result <- mi.inference(est, std.err, confidence = 0.95)
        R3[i, c(1, 3, 5)] <- result$est          # combined estimation of mean
        R3[i, c(2, 4, 6)] <- result$est/result$std.err    # computed t values
        R3[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)       # estimated of covariance
        R3[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
        R3[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
        MSE3[i, ] <- (R3[i, c(1:12)] - xx)^2
        SR3[i, c(1, 4, 7)] <- result$lower
        SR3[i, c(2, 5, 8)] <- result$upper
        if(SR3[i, 1] <= 0 & 0 <= SR3[i, 2]) {SR3[i, 3] <- 1}
        if(SR3[i, 4] <= 0 & 0 <= SR3[i, 5]) {SR3[i, 6] <- 1}
        if(SR3[i, 7] <= 0 & 0 <= SR3[i, 8]) {SR3[i, 9] <- 1}
#------ MCMC method (multiple imputation, m=5) ----------------
        temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
        temp2 <- matrix(0, 5, 3)    # standard error of each imputation
        for(k in 1:5) {
                para <- da.norm(y1.p, para, steps = 50, showits = F)   # simulated parameters
                yt <- imp.norm(y1.p, para, y1)       # imputed complete data
                temp1[k, c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
                vt <- var(yt, na.method = "available")
                temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
```

```
            temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
            temp1[k, 9] <- vt[3, 3]
            temp2[k, ] <- c(sqrt(vt[1, 1]/n0), sqrt(vt[2, 2]/n0), sqrt(vt[3, 3]/n0))
            }
        est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
        std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
        result <- mi.inference(est, std.err, confidence = 0.95)
        R4[i, c(1, 3, 5)] <- result$est          # combined estimation of mean
        R4[i, c(2, 4, 6)] <- result$est/result$std.err   # computed t values
        R4[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)       # estimated of covariance
        R4[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
        R4[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
        MSE4[i, ] <- (R4[i, c(1:12)] - xx)^2
        SR4[i, c(1, 4, 7)] <- result$lower
        SR4[i, c(2, 5, 8)] <- result$upper
        if(SR4[i, 1] <= 0 & 0 <= SR4[i, 2]) {SR4[i, 3] <- 1}
        if(SR4[i, 4] <= 0 & 0 <= SR4[i, 5]) {SR4[i, 6] <- 1}
        if(SR4[i, 7] <= 0 & 0 <= SR4[i, 8]) {SR4[i, 9] <- 1}
        }
        R1 <- cbind(R1, MSE1)
        R2 <- cbind(R2, MSE2)
        R3 <- cbind(R3, MSE3)
        R4 <- cbind(R4, MSE4)
return(m1=apply(R1,2,mean),m2=apply(R2,2,mean),m3=apply(R3,2,mean),m4=apply(R4,2,mean))
}
```

```
NSIMU2
function(m){
# Non-monotone Simulation 2 (single population with unstructured covariance)
# 1. Generate complete data from 3-d mutlvariate normal distribution with structured covariance
# 2. Call MK3 to generate non-monotone incomplete data
# 3. Call EM.NORM to get estimated normal parameters
# 4. Call SEQU2 to impute the missing data sequentially (unstructured)
# 5. Call PRED2 to impute the missing data by predictive model method
# 6. Call DA.NORM and IMP.NORM to get MCMC imputed data
# 7. Call MI.INFERENCE to get combined estimates, std.err and confidence interval
        n <- 30  # sample size of multivariate normal
        library(norm)      # in order to use Schafer's program
        rngseed(1234567)
        R1 <- matrix(0, m, 12)      # Estimated mean, covariance and t (no imputation)
        MSE1 <- matrix(0, m, 12) # Mean square error of estimates and t
        R2 <- matrix(0, m, 18)      # Proposed method (structured)
        MSE2 <- matrix(0, m, 12)
        R3 <- matrix(0, m, 18)      # Predictive model method
        MSE3 <- matrix(0, m, 12)
        R4 <- matrix(0, m, 18)      # MCMC method
        MSE4 <- matrix(0, m, 12)
        SR1 <- matrix(0, m, 9)      # 95% confidence limits and indicator (no imputation)
        SR2 <- matrix(0, m, 9)
        SR3 <- matrix(0, m, 9)
        SR4 <- matrix(0, m, 9)
        para0 <- matrix(-1, 4, 4)
        cov1 <- rbind(c(1, 0.5, 0.9), c(0.5, 2, 0.3), c(0.9, 0.3, 3))
        xx <- c(0, 0, 0, 0, 0, 0, 1, 0.5, 0.9, 2, 0.3, 3)
        for(i in 1:m) {
                y <- rmvnorm(n, mean = rep(0, 3), cov = cov1)
#------ Create monotone incomplete data (at least 3 obs with all data observed) ----
        y <- mk3(y)        # generate non-monotone incomplete data
        y1 <- y[!is.na(y[, 1]) | !is.na(y[, 2]) | !is.na(y[, 3]), ]  # at least one Y1,Y2,Y3 observed
        n0 <- nrow(y1)   # actual sample size
        n1 <- nrow(y1[!is.na(y1[, 1]), ])    # number of non-missing of Y[1]
        n2 <- nrow(y1[!is.na(y1[, 2]), ])    # number of non-missing of Y[2]
        n3 <- nrow(y1[!is.na(y1[, 3]), ])
```

```
        y1.p <- prelim.norm(y1)
#------ Estimate mean and its standard error by proposed method (no imputation) ----
        para <- em.norm(y1.p, showits = F)
        para0[1, c(2, 3, 4)] <- getparam.norm(y1.p, para)$mu
        para0[c(2, 3, 4), 1] <- getparam.norm(y1.p, para)$mu
        para0[c(2, 3, 4), c(2, 3, 4)] <- getparam.norm(y1.p, para)$sigma
        ser <- c(sqrt(para0[2, 2]/n1), sqrt(para0[3, 3]/n2), sqrt(para0[4, 4]/n3))
        R1[i, c(1, 3, 5, 7, 8, 9, 10, 11, 12)] <- extract(para0)  # estimates mean & covariance
        R1[i, c(2, 4, 6)] <- c(para0[1, 2]/ser[1], para0[1, 3]/ser[2], para0[1, 4]/ser[3])    # t values
        MSE1[i, ] <- (R1[i, ] - xx)^2        # mean square error
        SR1[i, c(1, 2, 3)] <- CI(R1[i, 1], ser[1], n1 - 1, 0)
        SR1[i, c(4, 5, 6)] <- CI(R1[i, 3], ser[2], n2 - 1, 0)
        SR1[i, c(7, 8, 9)] <- CI(R1[i, 5], ser[3], n3 - 1, 0)
#------ Proposed method (multiple imputation, m=5) ----------------
        temp1 <- matrix(0, 5, 9)   # estimated mean of each imputation
        temp2 <- matrix(0, 5, 3)   # standard error of each imputation
        for(k in 1:5) {
                yt <- sequ2(y1, para0)      # imputed complete data
                temp1[k, c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
                vt <- var(yt, na.method = "available")
                temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
                temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
                temp1[k, 9] <- vt[3, 3]
                temp2[k, ] <- c(sqrt(vt[1, 1]/n0), sqrt(vt[2, 2]/n0), sqrt(vt[3, 3]/n0))
                }
        est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
        std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
        result <- mi.inference(est, std.err, confidence = 0.95)
        R2[i, c(1, 3, 5)] <- result$est        # combined estimation of mean
        R2[i, c(2, 4, 6)] <- result$est/result$std.err   # computed t values
        R2[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)      # estimated of covariance
        R2[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
        R2[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
        MSE2[i, ] <- (R2[i, c(1:12)] - xx)^2
        SR2[i, c(1, 4, 7)] <- result$lower
        SR2[i, c(2, 5, 8)] <- result$upper
        if(SR2[i, 1] <= 0 & 0 <= SR2[i, 2]) {SR2[i, 3] <- 1}
```

```
        if(SR2[i, 4] <= 0 & 0 <= SR2[i, 5]) {SR2[i, 6] <- 1}
        if(SR2[i, 7] <= 0 & 0 <= SR2[i, 8]) {SR2[i, 9] <- 1}
#------ Predictive model method (multiple imputation, m=5) ----------------
        temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
        temp2 <- matrix(0, 5, 3)    # standard error of each imputation
        for(k in 1:5) {
                yt <- pred2(y1)    # imputed complete data
                temp1[k, c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
                vt <- var(yt, na.method = "available")
                temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
                temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
                temp1[k, 9] <- vt[3, 3]
                temp2[k, ] <- c(sqrt(vt[1, 1]/n0), sqrt(vt[2, 2]/n0), sqrt(vt[3, 3]/n0))
                }
        est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)], temp1[5,c(1:3)])
        std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
        result <- mi.inference(est, std.err, confidence = 0.95)
        R3[i, c(1, 3, 5)] <- result$est          # combined estimation of mean
        R3[i, c(2, 4, 6)] <- result$est/result$std.err    # computed t values
        R3[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)       # estimated of covariance
        R3[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
        R3[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
        MSE3[i, ] <- (R3[i, c(1:12)] - xx)^2
        SR3[i, c(1, 4, 7)] <- result$lower
        SR3[i, c(2, 5, 8)] <- result$upper
        if(SR3[i, 1] <= 0 & 0 <= SR3[i, 2]) {SR3[i, 3] <- 1}
        if(SR3[i, 4] <= 0 & 0 <= SR3[i, 5]) {SR3[i, 6] <- 1}
        if(SR3[i, 7] <= 0 & 0 <= SR3[i, 8]) {SR3[i, 9] <- 1}
#------ MCMC method (multiple imputation, m=5) ----------------
        temp1 <- matrix(0, 5, 9)    # estimated mean of each imputation
        temp2 <- matrix(0, 5, 3)    # standard error of each imputation
        for(k in 1:5) {
                para <- da.norm(y1.p, para, steps = 50, showits = F)   # simulated parameters
                yt <- imp.norm(y1.p, para, y1)       # imputed complete data
                temp1[k, c(1:3)] <- c(mean(yt[,1],na.rm=T),mean(yt[,2],na.rm=T),mean(yt[,3],na.rm=T))
                vt <- var(yt, na.method = "available")
                temp1[k, c(4, 5, 6)] <- vt[1, c(1, 2, 3)]
```

```
            temp1[k, c(7, 8)] <- vt[2, c(2, 3)]
            temp1[k, 9] <- vt[3, 3]
            temp2[k, ] <- c(sqrt(vt[1, 1]/n0), sqrt(vt[2, 2]/n0), sqrt(vt[3, 3]/n0))
            }
        est <- list(temp1[1,c(1:3)],temp1[2,c(1:3)],temp1[3,c(1:3)],temp1[4,c(1:3)],temp1[5,c(1:3)])
        std.err <- list(temp2[1, ], temp2[2, ], temp2[3, ], temp2[4, ], temp2[5, ])
        result <- mi.inference(est, std.err, confidence = 0.95)
        R4[i, c(1, 3, 5)] <- result$est          # combined estimation of mean
        R4[i, c(2, 4, 6)] <- result$est/result$std.err   # computed t values
        R4[i, c(7:12)] <- apply(temp1[, c(4:9)], 2, mean)       # estimated of covariance
        R4[i, c(13, 14, 15)] <- result$r       # relative increase in variance due to missing data
        R4[i, c(16, 17, 18)] <- result$fminf # fraction of missing information about parameter
        MSE4[i, ] <- (R4[i, c(1:12)] - xx)^2
        SR4[i, c(1, 4, 7)] <- result$lower
        SR4[i, c(2, 5, 8)] <- result$upper
        if(SR4[i, 1] <= 0 & 0 <= SR4[i, 2]) {SR4[i, 3] <- 1}
        if(SR4[i, 4] <= 0 & 0 <= SR4[i, 5]) {SR4[i, 6] <- 1}
        if(SR4[i, 7] <= 0 & 0 <= SR4[i, 8]) {SR4[i, 9] <- 1}
        }
        R1 <- cbind(R1, MSE1)
        R2 <- cbind(R2, MSE2)
        R3 <- cbind(R3, MSE3)
        R4 <- cbind(R4, MSE4)
return(m1=apply(R1,2,mean),m2=apply(R2,2,mean),m3=apply(R3,2,mean),m4=apply(R4,2,mean))
}
```

```
NSIMU3
function(m){
# Non-Monotone Simulation 3 (multiple population with ante-dependence of order 1)
# 1. Generate complete data from 3-d mutlvariate normal distribution with structured covariance
# 2. Call MK4 to generate non-monotone incomplete data
# 3. Call EM.NORM to get estimated normal parameters (unstructured covariance)
# 4. Call SEQU1 to impute the missing data within each population
# 5. Call CI to get confidence interval
        n <- 20
        library(norm)      # in order to use Schafer's program
        rngseed(1234567)
        R1 <- matrix(0, m, 9)       # estimated mean vector (no imputation)
        MSE1 <- matrix(0, m, 9)
        SR1 <- matrix(0, m, 27)     # confidence interval (no imputation)
        R2 <- matrix(0, m, 9)       # estimated mean vector (multiple imputation)
        MSE2 <- matrix(0, m, 9)
        SR2 <- matrix(0, m, 27)
        A1 <- matrix(-1, 4, 4)
        xx <- c(0, 0, 0, 2, 2, 2, 4, 4, 4)       # true population mean
        for(i in 1:m) {
        yy <- mk4(n)
        y <- yy[!is.na(yy[, 2]) | !is.na(yy[, 3]) | !is.na(yy[, 4]), ]          # at least variable observed
        nm <- c(nrow(y[!is.na(y[, 2]), ]), nrow(y[!is.na(y[, 3]), ]), nrow(y[!is.na(y[, 4]), ]))
        y1 <- y[y[, 1] == 1, c(2, 3, 4)]
        y2 <- y[y[, 1] == 2, c(2, 3, 4)]
        y3 <- y[y[, 1] == 3, c(2, 3, 4)]
        y1.p <- prelim.norm(y1)
        y2.p <- prelim.norm(y2)
        y3.p <- prelim.norm(y3)
#------ Estimate mean and its standard error by EM algorithm (no imputation) ---
        y1.para <- getparam.norm(y1.p, em.norm(y1.p, showits = F))
        y2.para <- getparam.norm(y2.p, em.norm(y2.p, showits = F))
        y3.para <- getparam.norm(y3.p, em.norm(y3.p, showits = F))
        m1 <- y1.para$mu
        m2 <- y2.para$mu
        m3 <- y3.para$mu
        aev <- (y1.para$sigma + y2.para$sigma + y3.para$sigma)/3
```

```
        R1[i, ] <- c(m1, m2, m3)  # estimated mean vectors of 3 populations
        MSE1[i, ] <- (R1[i, ] - xx)^2
        n1 <- nrow(y1[!is.na(y1[, 1]), ])     # number of non-missing of Y[1] in pop 1
        n2 <- nrow(y1[!is.na(y1[, 2]), ])
        n3 <- nrow(y1[!is.na(y1[, 3]), ])
        se1 <- c(sqrt(aev[1, 1]/n1), sqrt(aev[2, 2]/n2), sqrt(aev[3, 3]/n3))
        SR1[i, c(1, 2, 3)] <- CI(R1[i, 1], se1[1], nm[1] - 1, 0)
        SR1[i, c(4, 5, 6)] <- CI(R1[i, 2], se1[2], nm[2] - 1, 0)
        SR1[i, c(7, 8, 9)] <- CI(R1[i, 3], se1[3], nm[3] - 1, 0)
        n1 <- nrow(y2[!is.na(y2[, 1]), ])     # number of non-missing of Y[1] in pop 2
        n2 <- nrow(y2[!is.na(y2[, 2]), ])
        n3 <- nrow(y2[!is.na(y2[, 3]), ])
        se2 <- c(sqrt(aev[1, 1]/n1), sqrt(aev[2, 2]/n2), sqrt(aev[3, 3]/n3))
        SR1[i, c(10, 11, 12)] <- CI(R1[i, 4], se2[1], nm[1] - 1, 2)
        SR1[i, c(13, 14, 15)] <- CI(R1[i, 5], se2[2], nm[2] - 1, 2)
        SR1[i, c(16, 17, 18)] <- CI(R1[i, 6], se2[3], nm[3] - 1, 2)
        n1 <- nrow(y3[!is.na(y3[, 1]), ])     # number of non-missing of Y[1] in pop 3
        n2 <- nrow(y3[!is.na(y3[, 2]), ])
        n3 <- nrow(y3[!is.na(y3[, 3]), ])
        se3 <- c(sqrt(aev[1, 1]/n1), sqrt(aev[2, 2]/n2), sqrt(aev[3, 3]/n3))
        SR1[i, c(19, 20, 21)] <- CI(R1[i, 7], se3[1], nm[1] - 1, 4)
        SR1[i, c(22, 23, 24)] <- CI(R1[i, 8], se3[2], nm[2] - 1, 4)
        SR1[i, c(25, 26, 27)] <- CI(R1[i, 9], se3[3], nm[3] - 1, 4)
#--------- Estimate mean and covariance (multiple imputation) -------
        A1[c(2, 3, 4), c(2, 3, 4)] <- aev
        A1[1, c(2, 3, 4)] <- m1
        A1[c(2, 3, 4), 1] <- m1
        A2 <- A1
        A2[1, c(2, 3, 4)] <- m2
        A2[c(2, 3, 4), 1] <- m2
        A3 <- A1
        A3[1, c(2, 3, 4)] <- m3
        A3[c(2, 3, 4), 1] <- m3
        temp1 <- matrix(0, 5, 9)
        temp2 <- matrix(0, 5, 9)
        for(k in 1:5) {
                yt1 <- sequ1(y1, A1)
```

```
            yt2 <- sequ1(y2, A2)
            yt3 <- sequ1(y3, A3)
            m1 <- c(mean(yt1[,1],na.rm=T),mean(yt1[,2],na.rm=T),mean(yt1[,3],na.rm=T))
            m2 <- c(mean(yt2[,1],na.rm=T),mean(yt2[,2],na.rm=T),mean(yt2[,3],na.rm=T))
            m3 <- c(mean(yt3[,1],na.rm=T),mean(yt3[,2],na.rm=T),mean(yt3[,3],na.rm=T))
            aev <- (var(yt1, na.method = "available") + var(yt2, na.method = "available") + var(
                        yt3, na.method = "available"))/3
            se1 <- c(sqrt(aev[1, 1]/n), sqrt(aev[2, 2]/n), sqrt(aev[3, 3]/n))
            temp1[k, ] <- c(m1, m2, m3)
            temp2[k, ] <- c(se1, se1, se1)
            }
        est <- apply(temp1, 2, mean)        # combined estimate mean
        B <- apply(temp1, 2, var)  # between imputation variance
        U <- apply(temp2^2, 2, mean)        # within imputation variance
        V <- 4 * (1 + U/(1.2 * B))^2        # degrees of freedom
        se <- sqrt(1.2 * B + U)     # combined std error of combined estimate mean
        R2[i, ] <- est
        MSE2[i, ] <- (R2[i, ] - xx)^2
        V[c(1, 4, 7)] <- 0
        for(j in 1:9) {SR2[i,c(3*(j-1)+1,3*(j-1)+2,3*(j-1)+3)] <- CI(est[j],se[j], V[j],xx[j])}
        }
return(m1 = apply(R1 , 2 , mean), mse1 = apply(MSE1, 2, mean), CI1 = apply(SR1, 2, mean),
       m2 = apply(R2, 2, mean), mse2 = apply(MSE2, 2, mean), CI2 = apply(SR2, 2, mean))
}
```

Appendix 3: SAS Output of the ANOVA

The MIXED Procedure
Class Level Information

Class	Levels	Values
STRUCTUR	2	1 2
BLOCK	1000	1 . . . 1000
METHOD	4	1 2 3 4

Model Fitting Information for **MSEU2**
Monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.37133359
Residual	0.13714296

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	0.03	0.8568
METHOD	3	5994	21.28	0.0001
STRUCTUR*METHOD	3	5994	0.30	0.8287

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	0.33787805	0.02254942	3074	14.98	0.0001
STRUCTUR*METHOD	1	2	0.40420324	0.02254942	3074	17.93	0.0001
STRUCTUR*METHOD	1	3	0.39373878	0.02254942	3074	17.46	0.0001
STRUCTUR*METHOD	1	4	0.41774837	0.02254942	3074	18.53	0.0001
STRUCTUR*METHOD	2	1	0.32393495	0.02251942	3074	14.37	0.0001
STRUCTUR*METHOD	2	2	0.39238073	0.02254942	3074	17.40	0.0001
STRUCTUR*METHOD	2	3	0.39558066	0.02254942	3074	17.54	0.0001
STRUCTUR*METHOD	2	4	0.42110553	0.02254942	3074	18.67	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	-0.06632520	0.01656158	5994	-4.00	0.0001
STRUCTUR*METHOD	1	1	1	3	-0.05586073	0.01656158	5994	-3.37	0.0007
STRUCTUR*METHOD	1	1	1	4	-0.07987032	0.01656158	5994	-4.82	0.0001
STRUCTUR*METHOD	1	1	2	1	0.01394310	0.03188970	3074	0.44	0.6620
STRUCTUR*METHOD	1	1	2	2	-0.05450268	0.03188970	3074	-1.71	0.0875
STRUCTUR*METHOD	1	1	2	3	-0.05770261	0.03188970	3074	-1.81	0.0705
STRUCTUR*METHOD	1	1	2	4	-0.08322748	0.03188970	3074	-2.61	0.0091
STRUCTUR*METHOD	1	2	1	3	0.01046447	0.01656158	5994	0.63	0.5275
STRUCTUR*METHOD	1	2	1	4	-0.01354513	0.01656158	5994	-0.82	0.4135
STRUCTUR*METHOD	1	2	2	1	0.08026829	0.03188970	3074	2.52	0.0119
STRUCTUR*METHOD	1	2	2	2	0.01182251	0.03188970	3074	0.37	0.7109
STRUCTUR*METHOD	1	2	2	3	0.00862258	0.03188970	3074	0.27	0.7869
STRUCTUR*METHOD	1	2	2	4	-0.01690229	0.03188970	3074	-0.53	0.5961
STRUCTUR*METHOD	1	3	1	4	-0.02400960	0.01656158	5994	-1.45	0.1472
STRUCTUR*METHOD	1	3	2	1	0.06980383	0.03188970	3074	2.19	0.0287
STRUCTUR*METHOD	1	3	2	2	0.00135805	0.03188970	3074	0.04	0.9660
STRUCTUR*METHOD	1	3	2	3	-0.00184188	0.03188970	3074	-0.06	0.9539
STRUCTUR*METHOD	1	3	2	4	-0.02736675	0.03188970	3074	-0.86	0.3909
STRUCTUR*METHOD	1	4	2	1	0.09381342	0.03188970	3074	2.94	0.0033
STRUCTUR*METHOD	1	4	2	2	0.02536764	0.03188970	3074	0.80	0.4264
STRUCTUR*METHOD	1	4	2	3	0.02216771	0.03188970	3074	0.70	0.4870
STRUCTUR*METHOD	1	4	2	4	-0.00335716	0.03188970	3074	-0.11	0.9162
STRUCTUR*METHOD	2	1	2	2	-0.06844578	0.01656158	5994	-4.13	0.0001
STRUCTUR*METHOD	2	1	2	3	-0.07164571	0.01656158	5994	-4.33	0.0001
STRUCTUR*METHOD	2	1	2	4	-0.09717058	0.01656158	5994	-5.87	0.0001
STRUCTUR*METHOD	2	2	2	3	-0.00319993	0.01656158	5994	-0.19	0.8468
STRUCTUR*METHOD	2	2	2	4	-0.02872480	0.01656158	5994	-1.73	0.0829
STRUCTUR*METHOD	2	3	2	4	-0.02552487	0.01656158	5994	-1.54	0.1233

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	8.97	0.0001
STRUCTUR*METHOD	2	3	5994	12.61	0.0001

Model Fitting Information for **MSET2**
Monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	4.71049066
Residual	2.68542736

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	0.99	0.3187
METHOD	3	5994	260.33	0.0001
STRUCTUR*METHOD	3	5994	0.42	0.7369

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	2.34405725	0.08599952	3605	27.26	0.0001
STRUCTUR*METHOD	1	2	1.18313974	0.08599952	3605	13.76	0.0001
STRUCTUR*METHOD	1	3	1.09311550	0.08599952	3605	12.71	0.0001
STRUCTUR*METHOD	1	4	1.20632144	0.08599952	3605	14.03	0.0001
STRUCTUR*METHOD	2	1	2.43938176	0.08599952	3605	28.37	0.0001
STRUCTUR*METHOD	2	2	1.35093756	0.08599952	3605	15.71	0.0001
STRUCTUR*METHOD	2	3	1.19119682	0.08599952	3605	13.85	0.0001
STRUCTUR*METHOD	2	4	1.25905605	0.08599952	3605	14.64	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	1.16091750	0.07328612	5994	15.84	0.0001
STRUCTUR*METHOD	1	1	1	3	1.25094174	0.07328612	5994	17.07	0.0001
STRUCTUR*METHOD	1	1	1	4	1.13773581	0.07328612	5994	15.52	0.0001
STRUCTUR*METHOD	1	1	2	1	-0.09532451	0.12162169	3605	-0.78	0.4332
STRUCTUR*METHOD	1	1	2	2	0.99311969	0.12162169	3605	8.17	0.0001
STRUCTUR*METHOD	1	1	2	3	1.15286043	0.12162169	3605	9.48	0.0001
STRUCTUR*METHOD	1	1	2	4	1.08500120	0.12162169	3605	8.92	0.0001
STRUCTUR*METHOD	1	2	1	3	0.09002424	0.07328612	5994	1.23	0.2193
STRUCTUR*METHOD	1	2	1	4	-0.02318170	0.07328612	5994	-0.32	0.7518
STRUCTUR*METHOD	1	2	2	1	-1.25624202	0.12162169	3605	-10.33	0.0001
STRUCTUR*METHOD	1	2	2	2	-0.16779781	0.12162169	3605	-1.38	0.1678
STRUCTUR*METHOD	1	2	2	3	-0.00805707	0.12162169	3605	-0.07	0.9472
STRUCTUR*METHOD	1	2	2	4	-0.07591631	0.12162169	3605	-0.62	0.5325
STRUCTUR*METHOD	1	3	1	4	-0.11320594	0.07328612	5994	-1.54	0.1225
STRUCTUR*METHOD	1	3	2	1	-1.34626625	0.12162169	3605	-11.07	0.0001
STRUCTUR*METHOD	1	3	2	2	-0.25782205	0.12162169	3605	-2.12	0.0341
STRUCTUR*METHOD	1	3	2	3	-0.09808131	0.12162169	3605	-0.81	0.4200
STRUCTUR*METHOD	1	3	2	4	-0.16594054	0.12162169	3605	-1.36	0.1725
STRUCTUR*METHOD	1	4	2	1	-1.23306032	0.12162169	3605	-10.14	0.0001
STRUCTUR*METHOD	1	4	2	2	-0.14461612	0.12162169	3605	-1.19	0.2345
STRUCTUR*METHOD	1	4	2	3	0.01512462	0.12162169	3605	0.12	0.9010
STRUCTUR*METHOD	1	4	2	4	-0.05273461	0.12162169	3605	-0.43	0.6646
STRUCTUR*METHOD	2	1	2	2	1.08844420	0.07328612	5994	14.85	0.0001
STRUCTUR*METHOD	2	1	2	3	1.24818494	0.07328612	5994	17.03	0.0001
STRUCTUR*METHOD	2	1	2	4	1.18032571	0.07328612	5994	16.11	0.0001
STRUCTUR*METHOD	2	2	2	3	0.15974074	0.07328612	5994	2.18	0.0293
STRUCTUR*METHOD	2	2	2	4	0.09188151	0.07328612	5994	1.25	0.2100
STRUCTUR*METHOD	2	3	2	4	-0.06785923	0.07328612	5994	-0.93	0.3545

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	131.22	0.0001
STRUCTUR*METHOD	2	3	5994	129.54	0.0001

Model Fitting Information for **MSEU3**
Monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	80.95396059
Residual	155972.60700

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	1.46	0.2265
METHOD	3	5994	1.64	0.1774
STRUCTUR*METHOD	3	5994	1.38	0.2465

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1.56787976	12.49213997	7992	0.13	0.9001
STRUCTUR*METHOD	1	2	2.35574890	12.49213997	7992	0.19	0.8504
STRUCTUR*METHOD	1	3	45.84495184	12.49213997	7992	3.67	0.0002
STRUCTUR*METHOD	1	4	3.54372251	12.49213997	7992	0.28	0.7767
STRUCTUR*METHOD	2	1	1.77038636	12.49213997	7992	0.14	0.8873
STRUCTUR*METHOD	2	2	2.40169284	12.49213997	7992	0.19	0.8475
STRUCTUR*METHOD	2	3	4.03030896	12.49213997	7992	0.32	0.7470
STRUCTUR*METHOD	2	4	2.33340259	12.49213997	7992	0.19	0.8518

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	-0.78786915	17.66197084	5994	-0.04	0.9644
STRUCTUR*METHOD	1	1	1	3	-44.27707208	17.66197084	5994	-2.51	0.0122
STRUCTUR*METHOD	1	1	1	4	-1.97584276	17.66197084	5994	-0.11	0.9109
STRUCTUR*METHOD	1	1	2	1	-0.20250660	17.66655376	7992	-0.01	0.9909
STRUCTUR*METHOD	1	1	2	2	-0.83381308	17.66655376	7992	-0.05	0.9624
STRUCTUR*METHOD	1	1	2	3	-2.46242920	17.66655376	7992	-0.14	0.8892
STRUCTUR*METHOD	1	1	2	4	-0.76552284	17.66655376	7992	-0.04	0.9654
STRUCTUR*METHOD	1	2	1	3	-43.48920293	17.66197084	5994	-2.46	0.0138
STRUCTUR*METHOD	1	2	1	4	-1.18797361	17.66197084	5994	-0.07	0.9464
STRUCTUR*METHOD	1	2	2	1	0.58536255	17.66655376	7992	0.03	0.9736
STRUCTUR*METHOD	1	2	2	2	-0.04594394	17.66655376	7992	-0.00	0.9979
STRUCTUR*METHOD	1	2	2	3	-1.67456006	17.66655376	7992	-0.09	0.9245
STRUCTUR*METHOD	1	2	2	4	0.02234631	17.66655376	7992	0.00	0.9990
STRUCTUR*METHOD	1	3	1	4	42.30122932	17.66197084	5994	2.40	0.0166
STRUCTUR*METHOD	1	3	2	1	44.07456548	17.66655376	7992	2.49	0.0126
STRUCTUR*METHOD	1	3	2	2	43.44325900	17.66655376	7992	2.46	0.0140
STRUCTUR*METHOD	1	3	2	3	41.81464288	17.66655376	7992	2.37	0.0180
STRUCTUR*METHOD	1	3	2	4	43.51154924	17.66655376	7992	2.46	0.0138
STRUCTUR*METHOD	1	4	2	1	1.77333616	17.66655376	7992	0.10	0.9200
STRUCTUR*METHOD	1	4	2	2	1.14202967	17.66655376	7992	0.06	0.9485
STRUCTUR*METHOD	1	4	2	3	-0.48658645	17.66655376	7992	-0.03	0.9780
STRUCTUR*METHOD	1	4	2	4	1.21031992	17.66655376	7992	0.07	0.9454
STRUCTUR*METHOD	2	1	2	2	-0.63130648	17.66197084	5994	-0.04	0.9715
STRUCTUR*METHOD	2	1	2	3	-2.25992260	17.66197084	5994	-0.13	0.8982
STRUCTUR*METHOD	2	1	2	4	-0.56301623	17.66197084	5994	-0.03	0.9746
STRUCTUR*METHOD	2	2	2	3	-1.62861612	17.66197084	5994	-0.09	0.9265
STRUCTUR*METHOD	2	2	2	4	0.06829025	17.66197084	5994	0.00	0.9969
STRUCTUR*METHOD	2	3	2	4	1.69690637	17.66197084	5994	0.10	0.9235

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	3.02	0.0287
STRUCTUR*METHOD	2	3	5994	0.01	0.9994

Model Fitting Information for **MSRT3**
Monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	4.54684290
Residual	7.44932875

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	9.31	0.0023
METHOD	3	5994	244.57	0.0001
STRUCTUR*METHOD	3	5994	8.08	0.0001

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	2.72476584	0.10952704	5585	24.88	0.0001
STRUCTUR*METHOD	1	2	1.04784074	0.10952704	5585	9.57	0.0001
STRUCTUR*METHOD	1	3	1.06026541	0.10952704	5585	9.68	0.0001
STRUCTUR*METHOD	1	4	1.64024228	0.10952704	5585	14.98	0.0001
STRUCTUR*METHOD	2	1	3.59002148	0.10952704	5585	32.78	0.0001
STRUCTUR*METHOD	2	2	1.21037544	0.10952704	5585	11.05	0.0001
STRUCTUR*METHOD	2	3	1.21808079	0.10952704	5585	11.12	0.0001
STRUCTUR*METHOD	2	4	1.83655494	0.10952704	5585	16.77	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	1.67692510	0.12206006	5994	13.74	0.0001
STRUCTUR*METHOD	1	1	1	3	1.66450043	0.12206006	5994	13.64	0.0001
STRUCTUR*METHOD	1	1	1	4	1.08452356	0.12206006	5994	8.89	0.0001
STRUCTUR*METHOD	1	1	2	1	-0.86525564	0.15489462	5585	-5.59	0.0001
STRUCTUR*METHOD	1	1	2	2	1.51439040	0.15489462	5585	9.78	0.0001
STRUCTUR*METHOD	1	1	2	3	1.50668505	0.15489462	5585	9.73	0.0001
STRUCTUR*METHOD	1	1	2	4	0.88821090	0.15489462	5585	5.73	0.0001
STRUCTUR*METHOD	1	2	1	3	-0.01242467	0.12206006	5994	-0.10	0.9189
STRUCTUR*METHOD	1	2	1	4	-0.59240154	0.12206006	5994	-4.85	0.0001
STRUCTUR*METHOD	1	2	2	1	-2.54218074	0.15489462	5585	-16.41	0.0001
STRUCTUR*METHOD	1	2	2	2	-0.16253470	0.15489462	5585	-1.05	0.2941
STRUCTUR*METHOD	1	2	2	3	-0.17024005	0.15489462	5585	-1.10	0.2718
STRUCTUR*METHOD	1	2	2	4	-0.78871420	0.15489462	5585	-5.09	0.0001
STRUCTUR*METHOD	1	3	1	4	-0.57997687	0.12206006	5994	-4.75	0.0001
STRUCTUR*METHOD	1	3	2	1	-2.52975608	0.15489462	5585	-16.33	0.0001
STRUCTUR*METHOD	1	3	2	2	-0.15011003	0.15489462	5585	-0.97	0.3325
STRUCTUR*METHOD	1	3	2	3	-0.15781538	0.15489462	5585	-1.02	0.3083
STRUCTUR*METHOD	1	3	2	4	-0.77628953	0.15489462	5585	-5.01	0.0001
STRUCTUR*METHOD	1	4	2	1	-1.94977920	0.15489462	5585	-12.59	0.0001
STRUCTUR*METHOD	1	4	2	2	0.42986684	0.15489462	5585	2.78	0.0055
STRUCTUR*METHOD	1	4	2	3	0.42216149	0.15489462	5585	2.73	0.0064
STRUCTUR*METHOD	1	4	2	4	-0.19631266	0.15489462	5585	-1.27	0.2051
STRUCTUR*METHOD	2	1	2	2	2.37964604	0.12206006	5994	19.50	0.0001
STRUCTUR*METHOD	2	1	2	3	2.37194069	0.12206006	5994	19.43	0.0001
STRUCTUR*METHOD	2	1	2	4	1.75346654	0.12206006	5994	14.37	0.0001
STRUCTUR*METHOD	2	2	2	3	-0.00770535	0.12206006	5994	-0.06	0.9497
STRUCTUR*METHOD	2	2	2	4	-0.62617950	0.12206006	5994	-5.13	0.0001
STRUCTUR*METHOD	2	3	2	4	-0.61847415	0.12206006	5994	-5.07	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	83.30	0.0001
STRUCTUR*METHOD	2	3	5994	169.35	0.0001

Model Fitting Information for **R2**
Monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	5.60747109
Residual	7.55707834

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	1.15	0.2841
METHOD	2	3996	3.67	0.0255
STRUCTUR*METHOD	2	3996	1.28	0.2785

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	3.99784762	0.11473687	4398	34.84	0.0001
STRUCTUR*METHOD	1	3	4.13143880	0.11473687	4398	36.01	0.0001
STRUCTUR*METHOD	1	4	4.01294873	0.11473687	4398	34.98	0.0001
STRUCTUR*METHOD	2	2	3.70595373	0.11473687	4398	32.30	0.0001
STRUCTUR*METHOD	2	3	4.03734146	0.11473687	4398	35.19	0.0001
STRUCTUR*METHOD	2	4	3.98917720	0.11473687	4398	34.77	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	-0.13359118	0.12293965	3996	-1.09	0.2773
STRUCTUR*METHOD	1	2	1	4	-0.01510111	0.12293965	3996	-0.12	0.9022
STRUCTUR*METHOD	1	2	2	2	0.29189389	0.16226244	4398	1.80	0.0721
STRUCTUR*METHOD	1	2	2	3	-0.03949384	0.16226244	4398	-0.24	0.8077
STRUCTUR*METHOD	1	2	2	4	0.00867042	0.16226244	4398	0.05	0.9574
STRUCTUR*METHOD	1	3	1	4	0.11849007	0.12293965	3996	0.96	0.3352
STRUCTUR*METHOD	1	3	2	2	0.42548507	0.16226244	4398	2.62	0.0088
STRUCTUR*METHOD	1	3	2	3	0.09409734	0.16226244	4398	0.58	0.5620
STRUCTUR*METHOD	1	3	2	4	0.14226161	0.16226244	4398	0.88	0.3807
STRUCTUR*METHOD	1	4	2	2	0.30699500	0.16226244	4398	1.89	0.0586
STRUCTUR*METHOD	1	4	2	3	-0.02439273	0.16226244	4398	-0.15	0.8805
STRUCTUR*METHOD	1	4	2	4	0.02377153	0.16226244	4398	0.15	0.8835
STRUCTUR*METHOD	2	2	2	3	-0.33138773	0.12293965	3996	-2.70	0.0071
STRUCTUR*METHOD	2	2	2	4	-0.28322346	0.12293965	3996	-2.30	0.0213
STRUCTUR*METHOD	2	3	2	4	0.04816427	0.12293965	3996	0.39	0.6952

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	0.71	0.4926
STRUCTUR*METHOD	2	2	3996	4.24	0.0144

Model Fitting Information for **R3**
Monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	24.84707608
Residual	55.66852644

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	2.40	0.1213
METHOD	2	3996	57.96	0.0001
STRUCTUR*METHOD	2	3996	7.87	0.0004

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	8.55298623	0.28375271	5035	30.14	0.0001
STRUCTUR*METHOD	1	3	10.69641611	0.28375271	5035	37.70	0.0001
STRUCTUR*METHOD	1	4	7.86291260	0.28375271	5035	27.71	0.0001
STRUCTUR*METHOD	2	2	9.12258267	0.28375271	5035	32.15	0.0001
STRUCTUR*METHOD	2	3	9.43289128	0.28375271	5035	33.24	0.0001
STRUCTUR*METHOD	2	4	7.18680580	0.28375271	5035	25.33	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	-2.14342988	0.33367207	3996	-6.42	0.0001
STRUCTUR*METHOD	1	2	1	4	0.69007363	0.33367207	3996	2.07	0.0387
STRUCTUR*METHOD	1	2	2	2	-0.56959644	0.40128694	5035	-1.42	0.1558
STRUCTUR*METHOD	1	2	2	3	-0.87990504	0.40128694	5035	-2.19	0.0284
STRUCTUR*METHOD	1	2	2	4	1.36618044	0.40128694	5035	3.40	0.0007
STRUCTUR*METHOD	1	3	1	4	2.83350351	0.33367207	3996	8.49	0.0001
STRUCTUR*METHOD	1	3	2	2	1.57383344	0.40128694	5035	3.92	0.0001
STRUCTUR*METHOD	1	3	2	3	1.26352483	0.40128694	5035	3.15	0.0016
STRUCTUR*METHOD	1	3	2	4	3.50961031	0.40128694	5035	8.75	0.0001
STRUCTUR*METHOD	1	4	2	2	-1.25967006	0.40128694	5035	-3.14	0.0017
STRUCTUR*METHOD	1	4	2	3	-1.56997867	0.40128694	5035	-3.91	0.0001
STRUCTUR*METHOD	1	4	2	4	0.67610681	0.40128694	5035	1.68	0.0921
STRUCTUR*METHOD	2	2	2	3	-0.31030861	0.33367207	3996	-0.93	0.3524
STRUCTUR*METHOD	2	2	2	4	1.93577687	0.33367207	3996	5.80	0.0001
STRUCTUR*METHOD	2	3	2	4	2.24608548	0.33367207	3996	6.73	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	39.22	0.0001
STRUCTUR*METHOD	2	2	3996	26.61	0.0001

Model Fitting Information for **LAMDA2**
Monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.01097957
Residual	0.02114635

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	7.23	0.0072
METHOD	2	3996	6.46	0.0016
STRUCTUR*METHOD	2	3996	0.08	0.9189

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	0.74337344	0.00566797	4859	131.15	0.0001
STRUCTUR*METHOD	1	3	0.75857629	0.00566797	4859	133.84	0.0001
STRUCTUR*METHOD	1	4	0.74988668	0.00566797	4859	132.30	0.0001
STRUCTUR*METHOD	2	2	0.72512562	0.00566797	4859	127.93	0.0001
STRUCTUR*METHOD	2	3	0.74297292	0.00566797	4859	131.08	0.0001
STRUCTUR*METHOD	2	4	0.73530407	0.00566797	4859	129.73	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	-0.01520285	0.00650328	3996	-2.34	0.0195
STRUCTUR*METHOD	1	2	1	4	-0.00651324	0.00650328	3996	-1.00	0.3166
STRUCTUR*METHOD	1	2	2	2	0.01824782	0.00801573	4859	2.28	0.0229
STRUCTUR*METHOD	1	2	2	3	0.00040052	0.00801573	4859	0.05	0.9602
STRUCTUR*METHOD	1	2	2	4	0.00806937	0.00801573	4859	1.01	0.3141
STRUCTUR*METHOD	1	3	1	4	0.00868961	0.00650328	3996	1.34	0.1816
STRUCTUR*METHOD	1	3	2	2	0.03345067	0.00801573	4859	4.17	0.0001
STRUCTUR*METHOD	1	3	2	3	0.01560337	0.00801573	4859	1.95	0.0516
STRUCTUR*METHOD	1	3	2	4	0.02327222	0.00801573	4859	2.90	0.0037
STRUCTUR*METHOD	1	4	2	2	0.02476106	0.00801573	4859	3.09	0.0020
STRUCTUR*METHOD	1	4	2	3	0.00691376	0.00801573	4859	0.86	0.3884
STRUCTUR*METHOD	1	4	2	4	0.01458261	0.00801573	4859	1.82	0.0689
STRUCTUR*METHOD	2	2	2	3	-0.01784730	0.00650328	3996	-2.74	0.0061
STRUCTUR*METHOD	2	2	2	4	-0.01017845	0.00650328	3996	-1.57	0.1176
STRUCTUR*METHOD	2	3	2	4	0.00766885	0.00650328	3996	1.18	0.2384

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	2.75	0.0640
STRUCTUR*METHOD	2	2	3996	3.79	0.0227

Model Fitting Information for **LAMDA3**
Monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.00556782
Residual	0.01494803

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	10.66	0.0011
METHOD	2	3996	21.33	0.0001
STRUCTUR*METHOD	2	3996	11.80	0.0001

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	0.83899257	0.00452944	5224	185.23	0.0001
STRUCTUR*METHOD	1	3	0.86804572	0.00452944	5224	191.65	0.0001
STRUCTUR*METHOD	1	4	0.84464277	0.00452944	5224	186.48	0.0001
STRUCTUR*METHOD	2	2	0.84557704	0.00452944	5224	186.68	0.0001
STRUCTUR*METHOD	2	3	0.84407060	0.00452944	5224	186.35	0.0001
STRUCTUR*METHOD	2	4	0.81703621	0.00452944	5224	180.38	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	-0.02905315	0.00546773	3996	-5.31	0.0001
STRUCTUR*METHOD	1	2	1	4	-0.00565021	0.00546773	3996	-1.03	0.3015
STRUCTUR*METHOD	1	2	2	2	-0.00658448	0.00640560	5224	-1.03	0.3040
STRUCTUR*METHOD	1	2	2	3	-0.00507803	0.00640560	5224	-0.79	0.4280
STRUCTUR*METHOD	1	2	2	4	0.02195636	0.00640560	5224	3.43	0.0006
STRUCTUR*METHOD	1	3	1	4	0.02340294	0.00546773	3996	4.28	0.0001
STRUCTUR*METHOD	1	3	2	2	0.02246867	0.00640560	5224	3.51	0.0005
STRUCTUR*METHOD	1	3	2	3	0.02397512	0.00640560	5224	3.74	0.0002
STRUCTUR*METHOD	1	3	2	4	0.05100951	0.00640560	5224	7.96	0.0001
STRUCTUR*METHOD	1	4	2	2	-0.00093427	0.00640560	5224	-0.15	0.8840
STRUCTUR*METHOD	1	4	2	3	0.00057218	0.00640560	5224	0.09	0.9288
STRUCTUR*METHOD	1	4	2	4	0.02760656	0.00640560	5224	4.31	0.0001
STRUCTUR*METHOD	2	2	2	3	0.00150645	0.00546773	3996	0.28	0.7829
STRUCTUR*METHOD	2	2	2	4	0.02854083	0.00546773	3996	5.22	0.0001
STRUCTUR*METHOD	2	3	2	4	0.02703439	0.00546773	3996	4.94	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	15.87	0.0001
STRUCTUR*METHOD	2	2	3996	17.26	0.0001

The MIXED Procedure

Class Level Information

Class	Levels	Values
STRUCTUR	2	1 2
BLOCK	1000	1 . . . 1000
METHOD	4	1 2 3 4

Model Fitting Information for **MSEU1**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.00469544
Residual	0.00067472

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	1.10	0.2954
METHOD	3	5994	0.74	0.5283
STRUCTUR*METHOD	3	5994	1.01	0.3866

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	0.04909428	0.00231736	2427	21.19	0.001
STRUCTUR*METHOD	1	2	0.04930636	0.00231736	2427	21.28	0.001
STRUCTUR*METHOD	1	3	0.04920117	0.00231736	2427	21.23	0.001
STRUCTUR*METHOD	1	4	0.05011381	0.00231736	2427	21.63	0.001
STRUCTUR*METHOD	2	1	0.05156959	0.00231736	2427	22.25	0.001
STRUCTUR*METHOD	2	2	0.05339275	0.00231736	2427	23.04	0.001
STRUCTUR*METHOD	2	3	0.05363150	0.00231736	2427	23.14	0.001
STRUCTUR*METHOD	2	4	0.05218002	0.00231736	2427	22.52	0.001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	-0.00021208	0.00116165	5994	-0.18	0.8551
STRUCTUR*METHOD	1	1	1	3	-0.00010690	0.00116165	5994	-0.09	0.9267
STRUCTUR*METHOD	1	1	1	4	-0.00101953	0.00116165	5994	-0.88	0.3802
STRUCTUR*METHOD	1	1	2	1	-0.00247531	0.00327724	2427	-0.76	0.4501
STRUCTUR*METHOD	1	1	2	2	-0.00429847	0.00327724	2427	-1.31	0.1898
STRUCTUR*METHOD	1	1	2	3	-0.00453722	0.00327724	2427	-1.38	0.1663
STRUCTUR*METHOD	1	1	2	4	-0.00308574	0.00327724	2427	-0.94	0.3465
STRUCTUR*METHOD	1	2	1	3	0.00010519	0.00116165	5994	0.09	0.9279
STRUCTUR*METHOD	1	2	1	4	-0.00080745	0.00116165	5994	-0.70	0.4870
STRUCTUR*METHOD	1	2	2	1	-0.00226323	0.00327724	2427	-0.69	0.4899
STRUCTUR*METHOD	1	2	2	2	-0.00408639	0.00327724	2427	-1.25	0.2126
STRUCTUR*METHOD	1	2	2	3	-0.00432514	0.00327724	2427	-1.32	0.1870
STRUCTUR*METHOD	1	2	2	4	-0.00287366	0.00327724	2427	-0.88	0.3807
STRUCTUR*METHOD	1	3	1	4	-0.00091264	0.00116165	5994	-0.79	0.4321
STRUCTUR*METHOD	1	3	2	1	-0.00236842	0.00327724	2427	-0.72	0.4699
STRUCTUR*METHOD	1	3	2	2	-0.00419158	0.00327724	2427	-1.28	0.2010
STRUCTUR*METHOD	1	3	2	3	-0.00443033	0.00327724	2427	-1.35	0.1766
STRUCTUR*METHOD	1	3	2	4	-0.00297885	0.00327724	2427	-0.91	0.3635
STRUCTUR*METHOD	1	4	2	1	-0.00145578	0.00327724	2427	-0.44	0.6569
STRUCTUR*METHOD	1	4	2	2	-0.00327894	0.00327724	2427	-1.00	0.3172
STRUCTUR*METHOD	1	4	2	3	-0.00351769	0.00327724	2427	-1.07	0.2832
STRUCTUR*METHOD	1	4	2	4	-0.00206621	0.00327724	2427	-0.63	0.5284
STRUCTUR*METHOD	2	1	2	2	-0.00182316	0.00116165	5994	-1.57	0.1166
STRUCTUR*METHOD	2	1	2	3	-0.00206191	0.00116165	5994	-1.77	0.0760
STRUCTUR*METHOD	2	1	2	4	-0.00061043	0.00116165	5994	-0.53	0.5993
STRUCTUR*METHOD	2	2	2	3	-0.00023875	0.00116165	5994	-0.21	0.8372
STRUCTUR*METHOD	2	2	2	4	0.00121273	0.00116165	5994	1.04	0.2965
STRUCTUR*METHOD	2	3	2	4	0.00145148	0.00116165	5994	1.25	0.2115

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	0.32	0.8109
STRUCTUR*METHOD	2	3	5994	1.43	0.2318

Model Fitting Information for **MSET1**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	2.94778129
Residual	0.45193552

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	1.27	0.2596
METHOD	3	5994	6.83	0.0001
STRUCTUR*METHOD	3	5994	0.22	0.8853

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1.15464364	0.05830709	2455	19.80	0.001
STRUCTUR*METHOD	1	2	1.05379155	0.05830709	2455	18.07	0.001
STRUCTUR*METHOD	1	3	1.06515843	0.05830709	2455	18.27	0.001
STRUCTUR*METHOD	1	4	1.08376283	0.05830709	2455	18.59	0.001
STRUCTUR*METHOD	2	1	1.22603878	0.05830709	2455	21.03	0.001
STRUCTUR*METHOD	2	2	1.15913481	0.05830709	2455	19.88	0.001
STRUCTUR*METHOD	2	3	1.15108090	0.05830709	2455	19.74	0.001
STRUCTUR*METHOD	2	4	1.17402813	0.05830709	2455	20.14	0.001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	0.10085209	0.03006445	5994	3.35	0.0008
STRUCTUR*METHOD	1	1	1	3	0.08948521	0.03006445	5994	2.98	0.0029
STRUCTUR*METHOD	1	1	1	4	0.07088081	0.03006445	5994	2.36	0.0184
STRUCTUR*METHOD	1	1	2	1	-0.07139514	0.08245868	2455	-0.87	0.3867
STRUCTUR*METHOD	1	1	2	2	-0.00449116	0.08245868	2455	-0.05	0.9566
STRUCTUR*METHOD	1	1	2	3	0.00356274	0.08245868	2455	0.04	0.9655
STRUCTUR*METHOD	1	1	2	4	-0.01938449	0.08245868	2455	-0.24	0.8142
STRUCTUR*METHOD	1	2	1	3	-0.01136687	0.03006445	5994	-0.38	0.7054
STRUCTUR*METHOD	1	2	1	4	-0.02997128	0.03006445	5994	-1.00	0.3189
STRUCTUR*METHOD	1	2	2	1	-0.17224723	0.08245868	2455	-2.09	0.0368
STRUCTUR*METHOD	1	2	2	2	-0.10534325	0.08245868	2455	-1.28	0.2015
STRUCTUR*METHOD	1	2	2	3	-0.09728935	0.08245868	2455	-1.18	0.2382
STRUCTUR*METHOD	1	2	2	4	-0.12023658	0.08245868	2455	-1.46	0.1449
STRUCTUR*METHOD	1	3	1	4	-0.01860441	0.03006445	5994	-0.62	0.5361
STRUCTUR*METHOD	1	3	2	1	-0.16088035	0.08245868	2455	-1.95	0.0512
STRUCTUR*METHOD	1	3	2	2	-0.09397638	0.08245868	2455	-1.14	0.2545
STRUCTUR*METHOD	1	3	2	3	-0.08592247	0.08245868	2455	-1.04	0.2975
STRUCTUR*METHOD	1	3	2	4	-0.10886971	0.08245868	2455	-1.32	0.1869
STRUCTUR*METHOD	1	4	2	1	-0.14227595	0.08245868	2455	-1.73	0.0846
STRUCTUR*METHOD	1	4	2	2	-0.07537197	0.08245868	2455	-0.91	0.3608
STRUCTUR*METHOD	1	4	2	3	-0.06731806	0.08245868	2455	-0.82	0.4144
STRUCTUR*METHOD	1	4	2	4	-0.09026530	0.08245868	2455	-1.09	0.2738
STRUCTUR*METHOD	2	1	2	2	0.06690398	0.03006445	5994	2.23	0.0261
STRUCTUR*METHOD	2	1	2	3	0.07495788	0.03006445	5994	2.49	0.0127
STRUCTUR*METHOD	2	1	2	4	0.05201065	0.03006445	5994	1.73	0.0837
STRUCTUR*METHOD	2	2	2	3	0.00805391	0.03006445	5994	0.27	0.7880
STRUCTUR*METHOD	2	2	2	4	-0.01489333	0.03006445	5994	-0.50	0.6204
STRUCTUR*METHOD	2	3	2	4	-0.02294724	0.03006445	5994	-0.76	0.4453

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	4.53	0.0035
STRUCTUR*METHOD	2	3	5994	2.51	0.0569

Model Fitting Information for **MSEU2**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.12669085
Residual	0.06566207

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	0.56	0.4545
METHOD	3	5994	34.17	0.0001
STRUCTUR*METHOD	3	5994	6.53	0.0002

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	0.26669219	0.01386914	3473	19.23	0.001
STRUCTUR*METHOD	1	2	0.28779294	0.01386914	3473	20.75	0.001
STRUCTUR*METHOD	1	3	0.19757429	0.01386914	3473	14.25	0.001
STRUCTUR*METHOD	1	4	0.29300746	0.01386914	3473	21.13	0.001
STRUCTUR*METHOD	2	1	0.26386277	0.01386914	3473	19.03	0.001
STRUCTUR*METHOD	2	2	0.27143729	0.01386914	3473	19.57	0.001
STRUCTUR*METHOD	2	3	0.24830786	0.01386914	3473	17.90	0.001
STRUCTUR*METHOD	2	4	0.31208120	0.01386914	3473	22.50	0.001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	-0.02110074	0.01145967	5994	-1.84	0.0656
STRUCTUR*METHOD	1	1	1	3	0.06911790	0.01145967	5994	6.03	0.0001
STRUCTUR*METHOD	1	1	1	4	-0.02631527	0.01145967	5994	-2.30	0.0217
STRUCTUR*METHOD	1	1	2	1	0.00282942	0.01961392	3473	0.14	0.8853
STRUCTUR*METHOD	1	1	2	2	-0.00474510	0.01961392	3473	-0.24	0.8089
STRUCTUR*METHOD	1	1	2	3	0.01838433	0.01961392	3473	0.94	0.3487
STRUCTUR*METHOD	1	1	2	4	-0.04538901	0.01961392	3473	-2.31	0.0207
STRUCTUR*METHOD	1	2	1	3	0.09021864	0.01145967	5994	7.87	0.0001
STRUCTUR*METHOD	1	2	1	4	-0.00521452	0.01145967	5994	-0.46	0.6491
STRUCTUR*METHOD	1	2	2	1	0.02393017	0.01961392	3473	1.22	0.2225
STRUCTUR*METHOD	1	2	2	2	0.01635565	0.01961392	3473	0.83	0.4044
STRUCTUR*METHOD	1	2	2	3	0.03948507	0.01961392	3473	2.01	0.0442
STRUCTUR*METHOD	1	2	2	4	-0.02428827	0.01961392	3473	-1.24	0.2157
STRUCTUR*METHOD	1	3	1	4	-0.09543316	0.01145967	5994	-8.33	0.0001
STRUCTUR*METHOD	1	3	2	1	-0.06628847	0.01961392	3473	-3.38	0.0007
STRUCTUR*METHOD	1	3	2	2	-0.07386299	0.01961392	3473	-3.77	0.0002
STRUCTUR*METHOD	1	3	2	3	-0.05073357	0.01961392	3473	-2.59	0.0097
STRUCTUR*METHOD	1	3	2	4	-0.11450691	0.01961392	3473	-5.84	0.0001
STRUCTUR*METHOD	1	4	2	1	0.02914469	0.01961392	3473	1.49	0.1374
STRUCTUR*METHOD	1	4	2	2	0.02157017	0.01961392	3473	1.10	0.2715
STRUCTUR*METHOD	1	4	2	3	0.04469959	0.01961392	3473	2.28	0.0227
STRUCTUR*METHOD	1	4	2	4	-0.01907375	0.01961392	3473	-0.97	0.3309
STRUCTUR*METHOD	2	1	2	2	-0.00757452	0.01145967	5994	-0.66	0.5087
STRUCTUR*METHOD	2	1	2	3	0.01555490	0.01145967	5994	1.36	0.1747
STRUCTUR*METHOD	2	1	2	4	-0.04821844	0.01145967	5994	-4.21	0.0001
STRUCTUR*METHOD	2	2	2	3	0.02312942	0.01145967	5994	2.02	0.0436
STRUCTUR*METHOD	2	2	2	4	-0.04064391	0.01145967	5994	-3.55	0.0004
STRUCTUR*METHOD	2	3	2	4	-0.06377334	0.01145967	5994	-5.57	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	29.43	0.0001
STRUCTUR*METHOD	2	3	5994	11.27	0.0001

Model Fitting Information for **MSET2**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	4.33061676
Residual	2.37160038

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	4.85	0.0278
METHOD	3	5994	160.63	0.0001
STRUCTUR*METHOD	3	5994	9.34	0.0001

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	2.09289750	0.08186707	3548	25.56	0.001
STRUCTUR*METHOD	1	2	1.46812084	0.08186707	3548	17.93	0.001
STRUCTUR*METHOD	1	3	1.00361162	0.08186707	3548	12.26	0.001
STRUCTUR*METHOD	1	4	1.45130727	0.08186707	3548	17.73	0.001
STRUCTUR*METHOD	2	1	2.39596067	0.08186707	3548	29.27	0.001
STRUCTUR*METHOD	2	2	1.47290422	0.08186707	3548	17.99	0.001
STRUCTUR*METHOD	2	3	1.47604413	0.08186707	3548	18.03	0.001
STRUCTUR*METHOD	2	4	1.54518390	0.08186707	3548	18.87	0.001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	0.62477666	0.06887090	5994	9.07	0.0001
STRUCTUR*METHOD	1	1	1	3	1.08928588	0.06887090	5994	15.82	0.0001
STRUCTUR*METHOD	1	1	1	4	0.64159023	0.06887090	5994	9.32	0.0001
STRUCTUR*METHOD	1	1	2	1	-0.30306317	0.11577752	3548	-2.62	0.0089
STRUCTUR*METHOD	1	1	2	2	0.61999328	0.11577752	3548	5.36	0.0001
STRUCTUR*METHOD	1	1	2	3	0.61685337	0.11577752	3548	5.33	0.0001
STRUCTUR*METHOD	1	1	2	4	0.54771360	0.11577752	3548	4.73	0.0001
STRUCTUR*METHOD	1	2	1	3	0.46450922	0.06887090	5994	6.74	0.0001
STRUCTUR*METHOD	1	2	1	4	0.01681357	0.06887090	5994	0.24	0.8071
STRUCTUR*METHOD	1	2	2	1	-0.92783983	0.11577752	3548	-8.01	0.0001
STRUCTUR*METHOD	1	2	2	2	-0.00478338	0.11577752	3548	-0.04	0.9670
STRUCTUR*METHOD	1	2	2	3	-0.00792328	0.11577752	3548	-0.07	0.9454
STRUCTUR*METHOD	1	2	2	4	-0.07706306	0.11577752	3548	-0.67	0.5057
STRUCTUR*METHOD	1	3	1	4	-0.44769565	0.06887090	5994	-6.50	0.0001
STRUCTUR*METHOD	1	3	2	1	-1.39234904	0.11577752	3548	-12.03	0.0001
STRUCTUR*METHOD	1	3	2	2	-0.46929260	0.11577752	3548	-4.05	0.0001
STRUCTUR*METHOD	1	3	2	3	-0.47243250	0.11577752	3548	-4.08	0.0001
STRUCTUR*METHOD	1	3	2	4	-0.54157228	0.11577752	3548	-4.68	0.0001
STRUCTUR*METHOD	1	4	2	1	-0.94465339	0.11577752	3548	-8.16	0.0001
STRUCTUR*METHOD	1	4	2	2	-0.02159695	0.11577752	3548	-0.19	0.8520
STRUCTUR*METHOD	1	4	2	3	-0.02473685	0.11577752	3548	-0.21	0.8308
STRUCTUR*METHOD	1	4	2	4	-0.09387663	0.11577752	3548	-0.81	0.4175
STRUCTUR*METHOD	2	1	2	2	0.92305644	0.06887090	5994	13.40	0.0001
STRUCTUR*METHOD	2	1	2	3	0.91991654	0.06887090	5994	13.36	0.0001
STRUCTUR*METHOD	2	1	2	4	0.85077677	0.06887090	5994	12.35	0.0001
STRUCTUR*METHOD	2	2	2	3	-0.00313990	0.06887090	5994	-0.05	0.9636
STRUCTUR*METHOD	2	2	2	4	-0.07227968	0.06887090	5994	-1.05	0.2940
STRUCTUR*METHOD	2	3	2	4	-0.06913977	0.06887090	5994	-1.00	0.3155

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	84.51	0.0001
STRUCTUR*METHOD	2	3	5994	85.46	0.0001

Model Fitting Information for MSEU3
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.23859769
Residual	0.10947434

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	6.02	0.0142
METHOD	3	5994	60.99	0.0001
STRUCTUR*METHOD	3	5994	8.11	0.0001

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	0.38314228	0.01865669	3317	20.54	0.001
STRUCTUR*METHOD	1	2	0.41153018	0.01865669	3317	22.06	0.001
STRUCTUR*METHOD	1	3	0.26921544	0.01865669	3317	14.43	0.001
STRUCTUR*METHOD	1	4	0.43126416	0.01865669	3317	23.12	0.001
STRUCTUR*METHOD	2	1	0.29594670	0.01865669	3317	15.86	0.001
STRUCTUR*METHOD	2	2	0.37381449	0.01865669	3317	20.04	0.001
STRUCTUR*METHOD	2	3	0.26302175	0.01865669	3317	14.10	0.001
STRUCTUR*METHOD	2	4	0.33598087	0.01865669	3317	18.01	0.001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	-0.02838790	0.01479691	5994	-1.92	0.0551
STRUCTUR*METHOD	1	1	1	3	0.11392684	0.01479691	5994	7.70	0.0001
STRUCTUR*METHOD	1	1	1	4	-0.04812187	0.01479691	5994	-3.25	0.0012
STRUCTUR*METHOD	1	1	2	1	0.08719558	0.02638454	3317	3.30	0.0010
STRUCTUR*METHOD	1	1	2	2	0.00932780	0.02638454	3317	0.35	0.7237
STRUCTUR*METHOD	1	1	2	3	0.12012053	0.02638454	3317	4.55	0.0001
STRUCTUR*METHOD	1	1	2	4	0.04716142	0.02638454	3317	1.79	0.0740
STRUCTUR*METHOD	1	2	1	3	0.14231474	0.01479691	5994	9.62	0.0001
STRUCTUR*METHOD	1	2	1	4	-0.01973398	0.01479691	5994	-1.33	0.1824
STRUCTUR*METHOD	1	2	2	1	0.11558348	0.02638454	3317	4.38	0.0001
STRUCTUR*METHOD	1	2	2	2	0.03771569	0.02638454	3317	1.43	0.1530
STRUCTUR*METHOD	1	2	2	3	0.14850843	0.02638454	3317	5.63	0.0001
STRUCTUR*METHOD	1	2	2	4	0.07554931	0.02638454	3317	2.86	0.0042
STRUCTUR*METHOD	1	3	1	4	-0.16204872	0.01479691	5994	-10.95	0.0001
STRUCTUR*METHOD	1	3	2	1	-0.02673126	0.02638454	3317	-1.01	0.3111
STRUCTUR*METHOD	1	3	2	2	-0.10459904	0.02638454	3317	-3.96	0.0001
STRUCTUR*METHOD	1	3	2	3	*0.00619369*	*0.02638454*	3317	*0.23*	*0.8144*
STRUCTUR*METHOD	1	3	2	4	-0.06676542	0.02638454	3317	-2.53	0.0114
STRUCTUR*METHOD	1	4	2	1	0.13531746	0.02638454	3317	5.13	0.0001
STRUCTUR*METHOD	1	4	2	2	0.05744967	0.02638454	3317	2.18	0.0295
STRUCTUR*METHOD	1	4	2	3	0.16824241	0.02638454	3317	6.38	0.0001
STRUCTUR*METHOD	1	4	2	4	0.09528329	0.02638454	3317	3.61	0.0003
STRUCTUR*METHOD	2	1	2	2	-0.07786779	0.01479691	5994	-5.26	0.0001
STRUCTUR*METHOD	2	1	2	3	0.03292495	0.01479691	5994	2.23	0.0261
STRUCTUR*METHOD	2	1	2	4	-0.04003417	0.01479691	5994	-2.71	0.0068
STRUCTUR*METHOD	2	2	2	3	0.11079274	0.01479691	5994	7.49	0.0001
STRUCTUR*METHOD	2	2	2	4	0.03783362	0.01479691	5994	2.56	0.0106
STRUCTUR*METHOD	2	3	2	4	-0.07295912	0.01479691	5994	-4.93	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	47.96	0.0001
STRUCTUR*METHOD	2	3	5994	21.15	0.0001

Model Fitting Information for **MSET3**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	3.00926310
Residual	1.46971999

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	3.69	0.0548
METHOD	3	5994	146.29	0.0001
STRUCTUR*METHOD	3	5994	7.21	0.0001

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1.80254535	0.06692521	3395	26.93	0.001
STRUCTUR*METHOD	1	2	1.20732102	0.06692521	3395	18.04	0.001
STRUCTUR*METHOD	1	3	0.92103534	0.06692521	3395	13.76	0.001
STRUCTUR*METHOD	1	4	1.37040873	0.06692521	3395	20.48	0.001
STRUCTUR*METHOD	2	1	1.54623043	0.06692521	3395	23.10	0.001
STRUCTUR*METHOD	2	2	0.91249882	0.06692521	3395	13.63	0.001
STRUCTUR*METHOD	2	3	0.94188226	0.06692521	3395	14.07	0.001
STRUCTUR*METHOD	2	4	1.26898011	0.06692521	3395	18.96	0.001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	1	1	2	0.59522434	0.05421660	5994	10.98	0.0001
STRUCTUR*METHOD	1	1	1	3	0.88151002	0.05421660	5994	16.26	0.0001
STRUCTUR*METHOD	1	1	1	4	0.43213663	0.05421660	5994	7.97	0.0001
STRUCTUR*METHOD	1	1	2	1	0.25631492	0.09464653	3395	2.71	0.0068
STRUCTUR*METHOD	1	1	2	2	0.89004653	0.09464653	3395	9.40	0.0001
STRUCTUR*METHOD	1	1	2	3	0.86066309	0.09464653	3395	9.09	0.0001
STRUCTUR*METHOD	1	1	2	4	0.53356524	0.09464653	3395	5.64	0.0001
STRUCTUR*METHOD	1	2	1	3	0.28628568	0.05421660	5994	5.28	0.0001
STRUCTUR*METHOD	1	2	1	4	-0.16308771	0.05421660	5994	-3.01	0.0026
STRUCTUR*METHOD	1	2	2	1	-0.33890941	0.09464653	3395	-3.58	0.0003
STRUCTUR*METHOD	1	2	2	2	0.29482220	0.09464653	3395	3.11	0.0019
STRUCTUR*METHOD	1	2	2	3	0.26543876	0.09464653	3395	2.80	0.0051
STRUCTUR*METHOD	1	2	2	4	-0.06165909	0.09464653	3395	-0.65	0.5148
STRUCTUR*METHOD	1	3	1	4	-0.44937339	0.05421660	5994	-8.29	0.0001
STRUCTUR*METHOD	1	3	2	1	-0.62519510	0.09464653	3395	-6.61	0.0001
STRUCTUR*METHOD	1	3	2	2	0.00853652	0.09464653	3395	0.09	0.9281
STRUCTUR*METHOD	1	3	2	3	-0.02084693	0.09464653	3395	-0.22	0.8257
STRUCTUR*METHOD	1	3	2	4	-0.34794477	0.09464653	3395	-3.68	0.0002
STRUCTUR*METHOD	1	4	2	1	-0.17582171	0.09464653	3395	-1.86	0.0633
STRUCTUR*METHOD	1	4	2	2	0.45790991	0.09464653	3395	4.84	0.0001
STRUCTUR*METHOD	1	4	2	3	0.42852646	0.09464653	3395	4.53	0.0001
STRUCTUR*METHOD	1	4	2	4	0.10142862	0.09464653	3395	1.07	0.2840
STRUCTUR*METHOD	2	1	2	2	0.63373161	0.05421660	5994	11.69	0.0001
STRUCTUR*METHOD	2	1	2	3	0.60434817	0.05421660	5994	11.15	0.0001
STRUCTUR*METHOD	2	1	2	4	0.27725032	0.05421660	5994	5.11	0.0001
STRUCTUR*METHOD	2	2	2	3	-0.02938344	0.05421660	5994	-0.54	0.5879
STRUCTUR*METHOD	2	2	2	4	-0.35648129	0.05421660	5994	-6.58	0.0001
STRUCTUR*METHOD	2	3	2	4	-0.32709785	0.05421660	5994	-6.03	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	3	5994	92.34	0.0001
STRUCTUR*METHOD	2	3	5994	61.16	0.0001

Model Fitting Information for **R1**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.00768948
Residual	0.10818418

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	15.06	0.0001
METHOD	2	3996	34.19	0.0001
STRUCTUR*METHOD	2	3996	11.09	0.0001

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	0.47993272	0.01076446	5942	44.58	0.001
STRUCTUR*METHOD	1	3	0.47948213	0.01076446	5942	44.54	0.001
STRUCTUR*METHOD	1	4	0.44764539	0.01076446	5942	41.59	0.001
STRUCTUR*METHOD	2	2	0.47273496	0.01076446	5942	43.92	0.001
STRUCTUR*METHOD	2	3	0.47063786	0.01076446	5942	43.72	0.001
STRUCTUR*METHOD	2	4	0.35478690	0.01076446	5942	32.96	0.001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	0.00045060	0.01470946	3996	0.03	0.9756
STRUCTUR*METHOD	1	2	1	4	0.03228734	0.01470946	3996	2.20	0.0282
STRUCTUR*METHOD	1	2	2	2	0.00719776	0.01522325	5942	0.47	0.6364
STRUCTUR*METHOD	1	2	2	3	0.00929486	0.01522325	5942	0.61	0.5415
STRUCTUR*METHOD	1	2	2	4	0.12514583	0.01522325	5942	8.22	0.0001
STRUCTUR*METHOD	1	3	1	4	0.03183674	0.01470946	3996	2.16	0.0305
STRUCTUR*METHOD	1	3	2	2	0.00674717	0.01522325	5942	0.44	0.6576
STRUCTUR*METHOD	1	3	2	3	0.00884426	0.01522325	5942	0.58	0.5613
STRUCTUR*METHOD	1	3	2	4	0.12469523	0.01522325	5942	8.19	0.0001
STRUCTUR*METHOD	1	4	2	2	-0.02508957	0.01522325	5942	-1.65	0.0994
STRUCTUR*METHOD	1	4	2	3	-0.02299248	0.01522325	5942	-1.51	0.1310
STRUCTUR*METHOD	1	4	2	4	0.09285849	0.01522325	5942	6.10	0.0001
STRUCTUR*METHOD	2	2	2	3	0.00209710	0.01470946	3996	0.14	0.8866
STRUCTUR*METHOD	2	2	2	4	0.11794806	0.01470946	3996	8.02	0.0001
STRUCTUR*METHOD	2	3	2	4	0.11585097	0.01470946	3996	7.88	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	3.17	0.0422
STRUCTUR*METHOD	2	2	3996	42.12	0.0001

Model Fitting Information for **R2**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.72795246
Residual	2.36774418

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	0.01	0.9193
METHOD	2	3996	45.76	0.0001
STRUCTUR*METHOD	2	3996	0.79	0.4551

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1.65308265	0.05563898	5397	29.71	0.001
STRUCTUR*METHOD	1	3	2.14299883	0.05563898	5397	38.52	0.001
STRUCTUR*METHOD	1	4	1.99879083	0.05563898	5397	35.92	0.001
STRUCTUR*METHOD	2	2	1.68808514	0.05563898	5397	30.34	0.001
STRUCTUR*METHOD	2	3	2.07837988	0.05563898	5397	37.35	0.001
STRUCTUR*METHOD	2	4	2.04515669	0.05563898	5397	36.76	0.001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	-0.48991619	0.06881488	3996	-7.12	0.0001
STRUCTUR*METHOD	1	2	1	4	-0.34570819	0.06881488	3996	-5.02	0.0001
STRUCTUR*METHOD	1	2	2	2	-0.03500249	0.07868541	5397	-0.44	0.6565
STRUCTUR*METHOD	1	2	2	3	-0.42529723	0.07868541	5397	-5.41	0.0001
STRUCTUR*METHOD	1	2	2	4	-0.39207404	0.07868541	5397	-4.98	0.0001
STRUCTUR*METHOD	1	3	1	4	0.14420800	0.06881488	3996	2.10	0.0362
STRUCTUR*METHOD	1	3	2	2	0.45491370	0.07868541	5397	5.78	0.0001
STRUCTUR*METHOD	1	3	2	3	0.06461896	0.07868541	5397	0.82	0.4116
STRUCTUR*METHOD	1	3	2	4	0.09784214	0.07868541	5397	1.24	0.2138
STRUCTUR*METHOD	1	4	2	2	0.31070570	0.07868541	5397	3.95	0.0001
STRUCTUR*METHOD	1	4	2	3	-0.07958904	0.07868541	5397	-1.01	0.3118
STRUCTUR*METHOD	1	4	2	4	-0.04636586	0.07868541	5397	-0.59	0.5557
STRUCTUR*METHOD	2	2	2	3	-0.39029474	0.06881488	3996	-5.67	0.0001
STRUCTUR*METHOD	2	2	2	4	-0.35707155	0.06881488	3996	-5.19	0.0001
STRUCTUR*METHOD	2	3	2	4	0.03322319	0.06881488	3996	0.48	0.6293

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	26.77	0.0001
STRUCTUR*METHOD	2	2	3996	19.78	0.0001

Model Fitting Information for **R3**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.63076270
Residual	1.86464877

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	7.90	0.0050
METHOD	2	3996	25.08	0.0001
STRUCTUR*METHOD	2	3996	8.06	0.0003

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1.57325757	0.04995409	5315	31.49	0.0001
STRUCTUR*METHOD	1	3	1.84587605	0.04995409	5315	36.95	0.0001
STRUCTUR*METHOD	1	4	1.61825107	0.04995409	5315	32.39	0.0001
STRUCTUR*METHOD	2	2	1.61196970	0.04995409	5315	32.27	0.0001
STRUCTUR*METHOD	2	3	1.69250914	0.04995409	5315	33.88	0.0001
STRUCTUR*METHOD	2	4	1.31100552	0.04995409	5315	26.24	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	-0.27261848	0.06106797	3996	-4.46	0.0001
STRUCTUR*METHOD	1	2	1	4	-0.04499351	0.06106797	3996	-0.74	0.4613
STRUCTUR*METHOD	1	2	2	2	-0.03871213	0.07064576	5315	-0.55	0.5837
STRUCTUR*METHOD	1	2	2	3	-0.11925157	0.07064576	5315	-1.69	0.0915
STRUCTUR*METHOD	1	2	2	4	0.26225205	0.07064576	5315	3.71	0.0002
STRUCTUR*METHOD	1	3	1	4	0.22762498	0.06106797	3996	3.73	0.0002
STRUCTUR*METHOD	1	3	2	2	0.23390635	0.07064576	5315	3.31	0.0009
STRUCTUR*METHOD	1	3	2	3	0.15336691	0.07064576	5315	2.17	0.0300
STRUCTUR*METHOD	1	3	2	4	0.53487053	0.07064576	5315	7.57	0.0001
STRUCTUR*METHOD	1	4	2	2	0.00628137	0.07064576	5315	0.09	0.9292
STRUCTUR*METHOD	1	4	2	3	-0.07425806	0.07064576	5315	-1.05	0.2932
STRUCTUR*METHOD	1	4	2	4	0.30724555	0.07064576	5315	4.35	0.0001
STRUCTUR*METHOD	2	2	2	3	-0.08053944	0.06106797	3996	-1.32	0.1873
STRUCTUR*METHOD	2	2	2	4	0.30096418	0.06106797	3996	4.93	0.0001
STRUCTUR*METHOD	2	3	2	4	0.38150362	0.06106797	3996	6.25	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	11.46	0.0001
STRUCTUR*METHOD	2	2	3996	21.69	0.0001

Model Fitting Information for **LAMDA1**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.00175964
Residual	0.02302720

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	18.07	0.0001
METHOD	2	3996	47.21	0.0001
STRUCTUR*METHOD	2	3996	10.71	0.0001

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	0.32167287	0.00497864	5934	64.61	0.0001
STRUCTUR*METHOD	1	3	0.32021573	0.00497864	5934	64.32	0.0001
STRUCTUR*METHOD	1	4	0.29981324	0.00497864	5934	60.22	0.0001
STRUCTUR*METHOD	2	2	0.31599003	0.00497864	5934	63.47	0.0001
STRUCTUR*METHOD	2	3	0.31460560	0.00497864	5934	63.19	0.0001
STRUCTUR*METHOD	2	4	0.25570751	0.00497864	5934	51.36	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	0.00145715	0.00678634	3996	0.21	0.8300
STRUCTUR*METHOD	1	2	1	4	0.02185964	0.00678634	3996	3.22	0.0013
STRUCTUR*METHOD	1	2	2	2	0.00568285	0.00704086	5934	0.81	0.4196
STRUCTUR*METHOD	1	2	2	3	0.00706727	0.00704086	5934	1.00	0.3155
STRUCTUR*METHOD	1	2	2	4	0.06596537	0.00704086	5934	9.37	0.0001
STRUCTUR*METHOD	1	3	1	4	0.02040249	0.00678634	3996	3.01	0.0027
STRUCTUR*METHOD	1	3	2	2	0.00422570	0.00704086	5934	0.60	0.5484
STRUCTUR*METHOD	1	3	2	3	0.00561013	0.00704086	5934	0.80	0.4256
STRUCTUR*METHOD	1	3	2	4	0.06450822	0.00704086	5934	9.16	0.0001
STRUCTUR*METHOD	1	4	2	2	-0.01617679	0.00704086	5934	-2.30	0.0216
STRUCTUR*METHOD	1	4	2	3	-0.01479237	0.00704086	5934	-2.10	0.0357
STRUCTUR*METHOD	1	4	2	4	0.04410573	0.00704086	5934	6.26	0.0001
STRUCTUR*METHOD	2	2	2	3	0.00138443	0.00678634	3996	0.20	0.8384
STRUCTUR*METHOD	2	2	2	4	0.06028252	0.00678634	3996	8.88	0.0001
STRUCTUR*METHOD	2	3	2	4	0.05889809	0.00678634	3996	8.68	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	6.49	0.0015
STRUCTUR*METHOD	2	2	3996	51.42	0.0001

Model Fitting Information for **LAMDA2**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.00879223
Residual	0.03320431

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	0.00	0.9571
METHOD	2	3996	47.21	0.0001
STRUCTUR*METHOD	2	3996	1.22	0.2942

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	0.57368128	0.00648047	5511	88.52	0.0001
STRUCTUR*METHOD	1	3	0.63718674	0.00648047	5511	98.32	0.0001
STRUCTUR*METHOD	1	4	0.60981426	0.00648047	5511	94.10	0.0001
STRUCTUR*METHOD	2	2	0.58007787	0.00648047	5511	89.51	0.0001
STRUCTUR*METHOD	2	3	0.62716516	0.00648047	5511	96.78	0.0001
STRUCTUR*METHOD	2	4	0.61445739	0.00648047	5511	94.82	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	-0.06350545	0.00814915	3996	-7.79	0.0001
STRUCTUR*METHOD	1	2	1	4	-0.03613298	0.00814915	3996	-4.43	0.0001
STRUCTUR*METHOD	1	2	2	2	-0.00639658	0.00916477	5511	-0.70	0.4852
STRUCTUR*METHOD	1	2	2	3	-0.05348387	0.00916477	5511	-5.84	0.0001
STRUCTUR*METHOD	1	2	2	4	-0.04077611	0.00916477	5511	-4.45	0.0001
STRUCTUR*METHOD	1	3	1	4	0.02737248	0.00814915	3996	3.36	0.0008
STRUCTUR*METHOD	1	3	2	2	0.05710887	0.00916477	5511	6.23	0.0001
STRUCTUR*METHOD	1	3	2	3	0.01002158	0.00916477	5511	1.09	0.2742
STRUCTUR*METHOD	1	3	2	4	0.02272934	0.00916477	5511	2.48	0.0132
STRUCTUR*METHOD	1	4	2	2	0.02973639	0.00916477	5511	3.24	0.0012
STRUCTUR*METHOD	1	4	2	3	-0.01735089	0.00916477	5511	-1.89	0.0584
STRUCTUR*METHOD	1	4	2	4	-0.00464313	0.00916477	5511	-0.51	0.6124
STRUCTUR*METHOD	2	2	2	3	-0.04708729	0.00814915	3996	-5.78	0.0001
STRUCTUR*METHOD	2	2	2	4	-0.03437953	0.00814915	3996	-4.22	0.0001
STRUCTUR*METHOD	2	3	2	4	0.01270776	0.00814915	3996	1.56	0.1190

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	30.56	0.0001
STRUCTUR*METHOD	2	2	3996	17.87	0.0001

Model Fitting Information for **LAMDA3**
Non-monotone Incomplete Data

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
BLOCK(STRUCTUR)	0.01163899
Residual	0.03294194

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
STRUCTUR	1	1998	13.56	0.0002
METHOD	2	3996	38.37	0.0001
STRUCTUR*METHOD	2	3996	11.37	0.0001

Least Squares Means

Effect	STRUCTUR	METHOD	LSMEAN	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	0.55925098	0.00667690	5275	83.76	0.0001
STRUCTUR*METHOD	1	3	0.59614566	0.00667690	5275	89.28	0.0001
STRUCTUR*METHOD	1	4	0.55944200	0.00667690	5275	83.79	0.0001
STRUCTUR*METHOD	2	2	0.56195579	0.00667690	5275	84.16	0.0001
STRUCTUR*METHOD	2	3	0.57116258	0.00667690	5275	85.54	0.0001
STRUCTUR*METHOD	2	4	0.50741529	0.00667690	5275	76.00	0.0001

Differences of Least Squares Means

Effect	STRUCTUR	METHOD	_STRUCTU	_METHOD	Difference	Std Error	DF	t	Pr > |t|
STRUCTUR*METHOD	1	2	1	3	-0.03689468	0.00811689	3996	-4.55	0.0001
STRUCTUR*METHOD	1	2	1	4	-0.00019102	0.00811689	3996	-0.02	0.9812
STRUCTUR*METHOD	1	2	2	2	-0.00270481	0.00944256	5275	-0.29	0.7745
STRUCTUR*METHOD	1	2	2	3	-0.01191160	0.00944256	5275	-1.26	0.2072
STRUCTUR*METHOD	1	2	2	4	0.05183569	0.00944256	5275	5.49	0.0001
STRUCTUR*METHOD	1	3	1	4	0.03670366	0.00811689	3996	4.52	0.0001
STRUCTUR*METHOD	1	3	2	2	0.03418987	0.00944256	5275	3.62	0.0003
STRUCTUR*METHOD	1	3	2	3	0.02498308	0.00944256	5275	2.65	0.0082
STRUCTUR*METHOD	1	3	2	4	0.08873037	0.00944256	5275	9.40	0.0001
STRUCTUR*METHOD	1	4	2	2	-0.00251379	0.00944256	5275	-0.27	0.7901
STRUCTUR*METHOD	1	4	2	3	-0.01172058	0.00944256	5275	-1.24	0.2146
STRUCTUR*METHOD	1	4	2	4	0.05202671	0.00944256	5275	5.51	0.0001
STRUCTUR*METHOD	2	2	2	3	-0.00920679	0.00811689	3996	-1.13	0.2567
STRUCTUR*METHOD	2	2	2	4	0.05454050	0.00811689	3996	6.72	0.0001
STRUCTUR*METHOD	2	3	2	4	0.06374729	0.00811689	3996	7.85	0.0001

Tests of Effect Slices

Effect	STRUCTUR	NDF	DDF	F	Pr > F
STRUCTUR*METHOD	1	2	3996	13.70	0.0001
STRUCTUR*METHOD	2	2	3996	36.04	0.0001