BAYESIAN MODEL POOLING FOR THE ANALYSIS OF GENERALIZED LINEAR MODELS WITH MISSING-NOT-AT-RANDOM COVARIATES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SEZGIN ÇIFTÇI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
STATISTICS

SEPTEMBER 2020

Approval of the thesis:

BAYESIAN MODEL POOLING FOR THE ANALYSIS OF GENERALIZED LINEAR MODELS WITH MISSING-NOT-AT-RANDOM COVARIATES

submitted by **SEZGIN ÇIFTÇI** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Statistics Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Ayşen Dener Akkaya
Head of Department, **Statistics** _____

Assoc. Prof. Dr. Zeynep Işıl Kalaylıoğlu
Supervisor, **Statistics, METU** _____

Examining Committee Members:

Prof. Dr. Serpil Cula
Insurance, Baskent University _____

Assoc. Prof. Dr. Zeynep Işıl Kalaylıoğlu
Statistics, METU _____

Assoc. Prof. Dr. Haydar Demirhan
Mathematical Sciences, RMIT University _____

Assoc. Prof. Dr. Ceylan Talu Yozgatlıgil
Statistics, METU _____

Assoc. Prof. Dr. Özge Sezgin Alp
Accounting and Finance Management, Baskent University _____

Date: 22.09.2020

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Sezgin Çiftçi

Signature :

ABSTRACT

BAYESIAN MODEL POOLING FOR THE ANALYSIS OF GENERALIZED LINEAR MODELS WITH MISSING-NOT-AT-RANDOM COVARIATES

Çiftçi, Sezgin
Ph.D., Department of Statistics
Supervisor: Assoc. Prof. Dr. Zeynep Işıl Kalaylıoğlu

September 2020, 173 pages

Missing data in modeling is one of the main problems of researchers. If the missingness mechanism is related to the value of the variable itself, then the missing is defined as not-at-random (MNAR). Additionally, finding the best model in statistics has always been the focus of researchers. However, it's known from examples that there may be any other model with the same best fit but has different estimated coefficients. Hence, an uncertainty about model selection that may be problematic in estimation especially when there is also uncertainty about the accuracy of the estimations caused by MNAR mechanism can cause misleading inferences. In this thesis, a hybrid Bayesian method is developed for the analysis of a generalized linear model (GLM) with MNAR covariates. In the analysis of GLM with MNAR covariates, main response and missingness probabilities of the MNAR covariates are modeled jointly. In our approach, we create a model space as a set of the best fits among all possible joint models. This is accomplished by the Reversible Jump Monte Carlo Markov Chain (RJMCMC) approach that is adopted here. Coefficient estimates are obtained by pooling the posterior estimations of each model with the posterior model probabilities, which are also calculated within RJMCMC algorithm.

Keywords: Missing-Not-At-Random, Bayesian Approach, Generalized Linear Models, Reversible Jump Monte Carlo Markov Chain, Bayesian Model Averaging

ÖZ

RASTGELE OLMAYAN KAYIP VERİLİ DEĞİŞKENLER İÇEREN GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLERİN ANALİZİ İÇİN BAYESİYEN MODEL HAVUZU

Çiftçi, Sezgin
Doktora, İstatistik Bölümü
Tez Yöneticisi: Doç. Dr. Zeynep Işıl Kalaylıoğlu

Eylül 2020 , 173 sayfa

Kayıp veri, modelleme ve istatistiksel analizlerde araştırmacıların en temel sorunlarından biridir. Kayıp veri mekanizması değişkenin kendisine bağlı ise, rastgele olmayan kayıptır (MNAR). Ayrıca, istatistikte en iyi modeli bulmak her zaman araştırmacıların odak noktasıdır. Bununla birlikte, aynı en iyi uyuma sahip ancak katsayı tahminlerinde farklılık gösteren modellerin olabileceği literatürdeki bazı örneklerden bilinmektedir. Dolayısıyla model seçimi konusunda bir belirsizlik vardır ve bu belirsizlik özellikle MNAR mekanizmasının neden olduğu tahminlerin isabeti konusundaki belirsizlikle birlikte model katsayısı tahminlerinde sorun yaratabilir. Bu tezde, MNAR'a sahip ortak değişkenleri bulunan genelleştirilmiş doğrusal bir modelin (GLM) analizi için Bayes Yaklaşımlı Model Ortalaması (BMA) ve Geriye sıçramalı Monte Carlo Markov Zinciri (RJMCMC) metotlarını kullanarak oluşturulan hibrit bir Bayes yaklaşımlı model havuzlama sistemi (BMP) önerilmiştir. MNAR'a sahip verilerde değişken hakkında göz ardı edilemeyecek ek bilgiler olduğundan, ana model (GLM) ve kayıp veri mekanizması birlikte modellenmelidir. Tam koşullu mo-

del ve Bayes yaklaşımının özellikleri kullanılarak birleşik model adayları arasından en iyi uyumluları içeren uygun bir model uzayı tanımlanır. Ardından, model uzayındaki her bir model için model katsayılarının sonsal dağılımları RJMCMC yaklaşımı kullanılarak elde edilir. Son olarak, nihayi katsayı tahminleri, her model için bulunan sonsal dağılımlardan elde edilen tahminler, yine RJMCMC algoritmasında hesaplanan sonsal model olasılıklarıyla ağırlıklandırılarak elde edilir.

Anahtar Kelimeler: Rastgele Olmayan Kayıp Veri, Bayesiyen Yaklaşım, Genelleştirilmiş Doğrusal Modeller, Geriye Sıçramalı Monte Carlo Markov Zinciri, Bayesiyen Model Ortalama

To my beloved family

ACKNOWLEDGMENTS

First of all, I would like to acknowledge my supervisor Assoc. Prof. Dr. Zeynep Işıl Kalaylıoğlu, whose brilliant suggestions, help, ideas and endless patience made this study a reality. Her guidance and persistent dialogue provided me great motivation and courage throughout.

Furthermore, I am very grateful to Assoc. Prof. Dr. Haydar Demirhan and Assoc. Prof. Dr. Ceylan Talu Yozgatlıgil for their presence and inspiring suggestions and contributions during my PhD Examining Committee meetings.

I'd especially like to thank Prof. Dr. Recai Yücel whose lectures and supports gave me the opportunity to learn more about Bayesian Data Analysis when I attended the State University at Albany, NY, USA as a research scholar in the fall semester of 2017 - 2018. Also I would also like to thank TÜBİTAK-BİDEB for their support with scholarship providing this visit.

I would like thank Prof. Dr. Geert Molenberghs for his experienced guidance and advises about the study when I was in Hasselt University in the spring semester of 2016 - 2017 for Erasmus Exchange Program.

In addition, I thank Prof. Dr. Fatma Zehra Muluk, Prof. Dr. Serpil Cula and Assist. Prof. Dr. Şeref Hoşgör in Başkent University for their endless supports, patience and help during my research.

Moreover, my lovely girlfriend Hatice Hande Aykun deserves a special thanks for her patience, encouraging blessing and as well as for her dedicated contributions during the proof reading of the thesis.

My family deserves an additional expression of my gratefulness. Thank you for your existence, support and patience at home..

Thanks to all the brilliant people of the Department of Statistics and Institute of Nat-

ural Sciences for their help and understanding.

My thanks in the last place must go to the coronavirus for providing me the uninterrupted time period for writing this thesis during self isolation.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

BGR	Brooks, Gelman and Rubin
BIC	Bayesian Information Criteria
BMA	Bayesian Model Averaging
BMP	Bayesian Model Pooling
EM	Expectation-Maximization
ESS	Effective Sample Size
GLM	Generalized Linear Models
IQR	Interquartile Range
MAR	Missing At Random
MCAR	Missing Completely At Random
MCMC	Monte Carlo Markov Chain
MCSE	Monte Carlo Standard Error
M-H	Metropolis-Hastings
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MNAR	Missing Not At Random
RJMCMC	Reversible Jump Monte Carlo Markov Chain
SE	Standard Error
ZIP	Zero Inflated Poisson

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

Missing data in statistical analyses and modeling is one of the main problems of researchers. Although a variety of methods are used to overcome this issue, the uncertainty about the missingness mechanisms and thus the uncertainty about the accuracy of estimations made under uncertain assumptions still occur often. As Little and Rubin [46] define, missingness mechanisms are explained in three different types: Missing-Completely-At-Random (MCAR), Missing-At-Random (MAR) and Missing-Not-At-Random (MNAR). If the missingness does not depend on data values itself, then it is called MCAR. If the missingness depends on only the observed data, not the missing data itself, then it is called MAR [46]. When the missingness is independent from the missing data part, then it is called ignorable. When the missingness mechanism is MCAR or MAR, the missing parts can be ignored and only observed values are used or the missing parts are fulfilled by using imputation methods in order to overcome the uncertainty caused by missingness. On the other hand if the data are of an MNAR mechanism, which is non-ignorable, missingness depends on the missing values itself and it is not considered random anymore [46]. If the missingness mechanism is related to the value of the variable itself, then the missing is said to be not at random. It means that there is additional non-ignorable information in missingness aside from the observed part. Therefore the distribution of response and missingness mechanism should be modeled jointly. Some methods for factorization of this kind of joint distribution found in the literature are selection model [26], pattern-mixture model [46] and shared parameter model [46]. These methods are chosen according to the researchers' interest of the way of factoring the

joint distribution. The details will be given in preliminaries section.

There are also imputation methods for handling MNAR such as Complete-case analysis, Mean imputation, likelihood based multiple imputation, Bayesian iterative simulation approaches such as Data Augmentation, Gibbs sampling, Expectation - Maximization (EM), etc. However, for missingness models for MNAR mechanism that are weakly identified or for which there is a large fraction of missing information, convergence to a maximum may be very slow. An advantage of the Bayesian approach is that it mimics Maximum Likelihood (ML) inference in large samples, but also provides inference based directly on the posterior distribution without invoking large-sample normal approximations, which is likely to be superior to ML in small samples. However, each MNAR method consists of unstable assumptions since the uncertainty about the data mechanism. For example, in the study of Galimard [18], the multiple imputation method works only for MNAR in response variable, not in covariates. Also ignorable likelihood causes biased estimations [80]. Moreover, it is quite impossible to be sure about whether the missingness mechanism is MAR or MNAR without any additional information [51]. So, in addition to untestable assumptions there is also an uncertainty about ignorability. This problem may mislead the inference about estimations and modeling.

Studies for modeling MNAR with some approaches in the literature are likelihood-based with some other frequentest based approaches, based on sensitivity analysis [79] [75] [34] [74] and Bayesian approaches. Bayesian approach estimations for modeling MNAR can be seen in the literature in different types such as linear regression [54], generalized linear regression [31] [33], [44], [32], [70], [37], [81], [17], logistic regression [45] [40], survival models [27], spatial models [63]. For example, Mason et. al (2010) proposed a strategy that a base model is selected and then a sub-model to impute the missing covariates is added to it. They also indicate that the occurrence of missingness mechanism with regard to the response variable or covariates specifies the appropriateness of the analysis. They study missingness mechanism on response variables. Lastly, they argued that a strategy of modeling with MNAR covariates should be developed [49]. Nandram and Woo (2015) has developed a methodology in order to analyze the uncertainty of non-ignorable non-response with a Bayesian perspective and they have proposed a model for it [53].

In addition to all these, finding the best model in statistics is always the focus of the researchers in any scientific area. Fitting a best model for MNAR mechanism are studied by the researchers above and more. However, Hoeting et. al. ask the following question: what if there is any other model with the same best fit but has different estimated coefficients and standard errors [29]. Also note that literature is abundant with examples of models with the same good fit but different coefficients and standard errors [62] [48] [39] [59]. There is always a model uncertainty that should be in consideration especially when there is missingness in the data. Because in addition to the uncertainty, which is caused by missingness, there is another uncertainty occurs about selecting the missingness model. For minimizing the uncertainty of model selection,we believe it is better not to choose a final best model, but instead using multiple model approaches seems more reliable. Thus, the estimation of parameters can be done by using the tools of averaging or pooling for some specified models in order to reduce the uncertainty.

There are both frequentest and Bayesian approaches for model selection, but using only data and likelihood may mislead the selection in this study, since the data contains non-random missingness and trusting the data in hand only does not help to solve or minimize the uncertainty of missingness. On the other hand, as Ntzoufras (1999) mentioned, Bayesian model selection may offer better solutions by using a well known Bayesian approach called Markov chain Monte Carlo (MCMC) technique which can construct the model space, identify good models and estimate their posterior probabilities based on both prior knowledge and observed data [55]. By using careful diagnostic tools and sensitivity analysis which MCMC method offers, it might be possible to reduce the uncertainty of MNAR. Posterior probabilities are also very useful for model selection because it can be seen as the weights of the models in that model space as well.

Therefore, using a Bayesian model pooling system in MNAR modeling may reduce the uncertainty and increase the accuracy of the estimation of the parameters of interest in the base model. First of all, there should be a “well-constructed" model space in order to build a pooling system. “Well-constructed" can be taken as the ability to select or filter the most efficient models among all the potential models. There are several Bayesian methods used for model selection as well such as Gibbs, Metropolis-

Hastings, Bayesian Model Averaging, Reversible Jump Markov Chain Monte Carlo, etc. Naturally, in a finite countable set of candidate models, the dimensions of models may vary according to the parameters (coefficients) of the models. So, a standard Gibbs or Metropolis-Hastings algorithm for model selection may not work well. Instead, an algorithm which has flexible trans-dimensional ability among the models should be taken into account such as Reversible Jump Markov Chain Monte Carlo (RJMCMC). Green (1995) introduced RJMCMC as a method that provides a general outline for Markov Chain Monte Carlo (MCMC) simulation where the dimension of the parameter space may vary among the Markov Chain iterates [24]. This flexibility and the ability to move in different dimensions makes RJMCMC method a suitable candidate for model selection, or better, for creating a "well-constructed" model space for model pooling. For instance, Dey and Jhamb (2018) has used this algorithm for model selection of varying dimensions for user-movie recommendation [11]. Bouranis et. al. (2018) also used RJMCMC algorithm for model comparison [7]. Moreover, Kwak and Kim (2019) studied recently on finding the best model in multiple reggression by using RJMCMC [42]. For the first time, Hoegh et. al. (2017) has tried RJMCMC method for model selection containing MNAR components while examining possible changes to foxhound training enclosure policy and fox survival [28]. However, they have just focused on finding the best fit.

In this study, for the sake of simplicity, only some of the covariates are assumed to have MNAR mechanism while the others are fully observed in a base model, a set of missingness models, which are the elements of the model pooling system are taken and added jointly to the base model by using the properties of full conditional model and Bayesian approach in order to build the methodology. If the number of covariates contains missingness is just one, it might be easy to apply RJMCMC directly, since the model space will be in one dimensional space. It means that, let $\mathcal{M}$ be the model space where $\mathcal{M} \in \mathbb{R}$. But when the number of covariates containing missingness parts increases, the model space become enormous and the dimension of the model space increases at the same time. For a better expression, let $\mathcal{M}$ be the model space with 3 MNAR missingness mechanisms, then $\mathcal{M} \in \mathbb{R}^3$. When there are only a few potential covariates for the missingness mechanism, a 3 dimensional model space may not be a problem. For example, if there are 4 potential covariates

in which 3 of them have missingness mechanisms, then the model space will contain 2^4x2^4x2^4 = 16x16x16 = 4096 different missingness model trios. If there are 5 potential covariates there would be 32768 different missingness model trios. Therefore, before applying RJMCMC directly, some other model space reduction methods like Occam's razor and Occam's window as Hoeting et. al. (1999) suggest for model averaging [29], might be applied in order to reduce the model space. By applying RJMCMC, model posteriors can be achieved, and they can be used as weights of the parameter estimations for every model in the model space. Finally, by using a hybrid Bayesian modeling system, instead of a single model estimation, the weighted mean of all the parameter estimations can be assessed in order to reduce the uncertainty about the accuracy of estimations.

1.2 Contributions and Novelties

Contributions made in modeling of GLMs with covariates containing MNAR throughout the thesis are as follows:

- Reduction of uncertainty about missingness mechanism and the accuracy of parameter estimations by using RJMCMC method (see Section 3.2): Putting RJMCMC method in the center of the methodology provides a freedom to work with varying dimensional models. Hence, this benefits the reduction of uncertainty caused by both parameter estimates and MNAR mechanism.

- Definition of a "well-constructed" model space by using the T-step Occam's window method (see Section 3.3.1): Classical Occam's window method can not be used properly for filtering the best candidate models in the model space when MNAR mechanism occurs in some of the covariates. Therefore, it is possible to apply Occam's window method to these kind of data by using the T-step proposal given in methodology section.

- Usage of dimension difference impact on the construction of model transition probabilities (see Section 3.4): Model transitions within the same dimension are expected to be easier than the transitions among other dimensions. Also, as the dimension difference increases between two models, the transition prob-

ability is assumed to decrease evenly. So the impact of this effect should be taken into consideration when calculating the transition probabilities.

- Hybrid Bayesian Modeling system definition called Bayesian Model Pooling (BMP) (see Section 3): The combination of both RJMCMC and Bayesian Model Averaging (BMA) is considered to be a helpful system for reducing the uncertainty about model selection and MNAR mechanism at the same time. This system is studied for GLMs containing covaraites with MNAR.

- R code: Since the methodology is based on computational algorithms, the necessary genuine codes are written in R software consist of interpenetrated RJMCMC, Metropolis-Hastings sampling, Gibbs sampling, T-step Occam's window and BMA algorithms.

- Sensitivity analysis (see Section 4.1.7): The sensitivity of the methodology are assessed for parameter estimation in terms of assumptions.

1.3 The Outline of the Thesis

Handling both uncertainty about the missingness mechanism, especially with non-ignorable missing data in the covariates, and uncertainty about accuracy of parameter estimations are the main focuses of this study. The organization of this thesis is as follows: In Chapter 2, the concepts of missingness mechanisms (especially MNAR), Generalized linear models, modeling MNAR, methods of model selection and Bayesian inference methods are introduced briefly. In Chapter 3, the methodology of the thesis is described in detail with model definition, RJMCMC algorithm, model space construction, transition probability assessments for RJMCMC and the process of BMP for parameter estimations. In Chapter 4, the methodology is applied with a validation study on a simulated dataset with known true parameters by generating artificial data under the assumption of a GLM model contains covariates of which only one has an MNAR mechanism. The validation study on a simulated dataset is tested by using sensitivity analysis under different assumptions and scenarios such as missingness percentages and parameter assumptions. After the validation study, a real data application is conducted and the results of the analysis is compared with an-

other method, of which has the same data, in order to see the efficiency and accuracy of the proposed Bayesian pooling system. In Chapter 5, the study is concluded with some discussion.

CHAPTER 2

PRELIMINARIES

The thesis focus on estimation of coefficients in Generalized Linear Models of which some of the covariates contain MNAR mechanism. The literature and methods on missingness in GLM is vast and should be carefully reviewed first. In this section we provide literature review and preliminaries for the following:

1. Generalized linear models (GLM)
2. Missing data in GLM
3. Bayesian missing data analysis in GLM

2.1 Generalized Linear Models

Generalized Linear Models (GLM) are generalizations of ordinary linear regression models. While ordinary linear regression models are used under the assumption of normally distributed response variables, in GLM the response variable may be non-normal and therefore the model itself may be non-linear [1] [2]. GLM has three components: response variable (random component), explanatory variables (predictors or covariates) with parameter vector and a monotone link function [1] [2] [12]. It is defined in equation 2.1.

$$g(E(Y)) = \mathbf{X}\boldsymbol{\beta} \tag{2.1}$$

where Y is the response variable with a probability distribution belongs to the exponential family, $E(Y)$ is the expected value of Y, $\mathbf{X} = [X_1, X_2, ..., X_p]$ is the vector of predictor variables (covariates) with parameter vector $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_p]^T$, $g(\cdot)$ is

the monotone link function. The link function $g(\cdot)$ connects the predictors to the response variable. If $g(\cdot)$ is an identity function, then GLM turns into an ordinary linear regression model. GLM can be specialized according to specific response variables and link functions. For instance, when the response variable is binary (occurrence=1, non-occurrence=0) and the link function is the logarithm of odds, called logit link function, GLM is called as logistic regression [1]. If the link function is chosen as cumulative probability function for standard normal distribution, then it is termed as probit regression. Since our interest in this thesis lies in binary response variables, we will next review logistic and probit regressions.

2.1.1 Logistic Regression

If the response variable is binary then,

$$Y = \begin{cases} 1 & \text{if occurrence} \\ 0 & \text{if no occurrence} \end{cases} \tag{2.2}$$

with probabilities $P(Y = 1) = \pi$ and $P(Y = 0) = 1 - \pi$. When $\pi = \frac{exp(\mathbf{X}\boldsymbol{\beta})}{1+exp(\mathbf{X}\boldsymbol{\beta})}$ then the link function is defined as $g(\pi) = log\left(\frac{\pi}{1-\pi}\right)$ and the GLM is called logistic regression or logit regression [1].

$$logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = \mathbf{X}\boldsymbol{\beta} \tag{2.3}$$

Logistic regression is used when there is a non-linear relationship between the data X and $\pi(X)$ [1]. An example of logistic regression that shows this nonlinear relation is given in Figure 2.1.

2.1.2 Probit Regression

If the response variable is binary, as in logistic regression, the probabilities are defined as $P(Y = 1) = \pi$ and $P(Y = 0) = 1 - \pi$. As in Figure 2.1, the shape of regression when $\beta > 0$ looks like a cumulative distribution function. So, this time let us denote $\pi = \Phi(\mathbf{X}\boldsymbol{\beta})$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function. If the link function is defined as $g(\pi) = \Phi^{-1}(\pi)$ the GLM is called probit regression [1] [2].

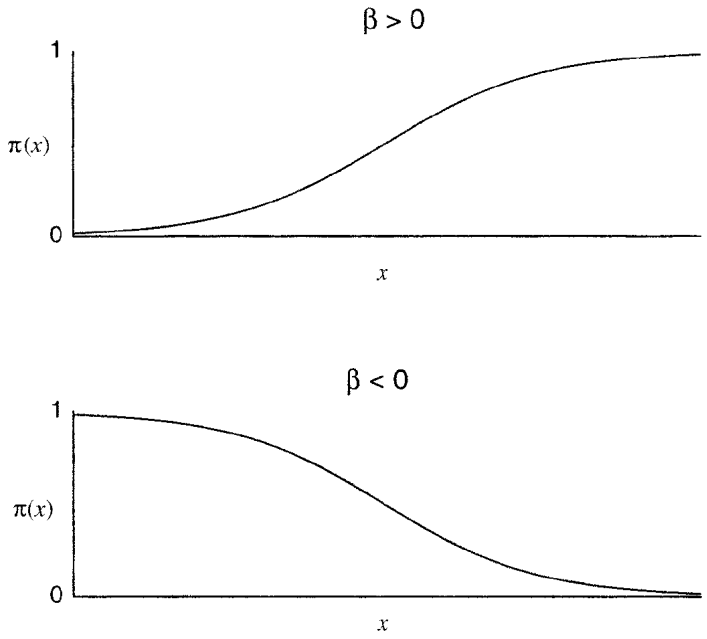

Figure 2.1: Logistic Regression Plot Example, *taken from the book "Categorical Data Analysis" by Agresti (2003)* [1]

$$\Phi^{-1}(\pi) = \mathbf{X}\boldsymbol{\beta} \tag{2.4}$$

In this thesis, MNAR mechanism is assumed to occur in some of the covariates of GLM with binary response. Both logistic and probit regression are used for modeling binary response data and both show very similar performance. However, in some areas one can be more favorable than the other. When odds ratio is meaningful for inference like in epidemiology, logistic regression is preferable, on the other hand, probit regression is more popular in econometric analysis. Concerning handling MNAR mechanism, Enders (2010) indicates that the classic selection model uses probit regression to predict missingness indicator R, although logistic regression is a common approach for analyzing binary outcomes [14]. In this thesis, logistic regression is applied for response variable and probit regression is applied for predicting missingness indicator as the selection model is chosen for the thesis' purposes.

2.2 Missing Data and Missingness Mechanisms

Incomplete or missing data occur if the data planned to be ascertained are not observed. This is a very common problem in practice for statistical inference. If the data to be analyzed are thought to be a rectangular block, missing data can be imagined as holes or/and non-filled columns in this block. In statistical inference, missingness may cause a bias on the inference of the parameters. Various methods are available to resolve this problem. The decision of which method to use is determined by the reason behind the missing data occurrence. The information of the mechanism of missing data is substantial in choosing the suitable missing data handling method. The related information can be gathered by asking the right questions, such as why the data is missing, whether it is randomly missing or not, and if there is a specific reason which depends on the data itself etc. As Rubin (1976) and later Little and Rubin (2002) defined, missingness mechanisms are basically categorized into three different types: Missing-Completely-At-Random (MCAR), Missing-At-Random (MAR) and Missing-Not-At-Random (MNAR) [68] [46]. All three mechanisms are inspected carefully in the following sections. The knowledge gathered about the mechanism of the missing data leads to suitable selection of the methods for handling the missingness, methods being as substantial as the missing mechanisms. These methods are applied accordingly and can be generally categorized into deletion methods, single imputation methods, multiple imputation methods and likelihood based methods including Bayesian methods in the literature. The following sections deal with the details of these methods. When the missingness is independent from the missing variable, then the impact of the missingness can be relatively small and it can be called as ignorable. So if the missingness mechanism is identified as MCAR or MAR, the missing parts in this case can be ignored. In order to overcome the uncertainty caused by the missingness, only observed values can be used or imputation methods can be carried out for the missing parts. On the other hand, if the data are of an MNAR mechanism, which is non-ignorable, missingness depends on the missing values itself and it is not random anymore [46]. In this particular case, traditional deletion or imputation methods don't function standalone to handle the missingness mechanism. The missing data and the mechanism should be taken into consideration jointly.

2.2.1 Missing-Completely-At-Random (MCAR)

If the missingness depends on the observed variables but does not depend on data values itself, then the related mechanism is called Missing-Completely-At-Random (MCAR) [46]. An electronic device that measures and records the seismic movements of the ground in an area would create a good example for MCAR. If that device is broken at some point, the device can not measure and record the rest of the seismic movements until it is fixed again. Therefore, the data will be MCAR. For a better expression, let us denote $D = (D^{obs}, D^{miss})$ is the data matrix where D^{obs} is the observed part and D^{miss} is the missing part. Also let us define the missingness indicator as $R = \{r_i,\ for\ i = 1, \ldots, n\}$ and

$$r_i = \begin{cases} 0\ if\ x_i \in D^{obs} \\ 1\ if\ x_i \in D^{miss} \end{cases} \tag{2.5}$$

where x_i is the i^{th} observation of variable X containing missing values. The missingness mechanism can be specified as the conditional distribution function $f(R|D, \theta)$ where θ is the unknown parameter. So, if

$$P(R|D, \theta) = P(R|\theta) \tag{2.6}$$

for all D and θ then the data are MCAR. If the mechanism is MCAR, it is assumed that there is no relation between complete cases and missing cases, which means the missingness is ignorable. So, handling the missingness can be done by omitting the missing parts with deletion methods such as listwise deletion (complete case analysis) or pairwise deletion. Let the columns of the data block are variables and rows of the data block are individuals. In listwise deletion, all of the information (whole row) of that individual in the data is deleted if any value is missing on that row. So only complete data is left at hand. In piecewise deletion method, the deletion is done for the particular analysis carried out. For example, some observed information in the row, containing missing parts, may contribute to some analysis. So the size of the data may differ in different analyses. Imputation methods can also solve the problem of MCAR mechanism. The details about missing data handling methods is given in the following sections.

2.2.2 Missing-At-Random (MAR)

If the missingness depends on the observed data and variables, but not the missing data itself, then it is called Missing-At-Random (MAR) [46]. In other words, the mechanism can be explained by the variables with full information. Let us think about the same example in MCAR case. The same electronic device can measure and record the seismic movements if the temperature is between $-10°$C and $30°$C. Thus, the data would be missing when the temperature is less than $-10°$C or more than $30°$C, meaning the missing data is MAR. The same notations in MCAR case can be used in here as well. The missingness mechanism is MAR if

$$P(R|D,\theta) = P(R|D^{obs},\theta) \tag{2.7}$$

for all D^{miss} and θ. MAR mechanism can also be ignorable and handling methods as deletion, imputation and likelihood based can be used in order to make reliable inferences about the parameters of interest.

2.2.3 Missing-Not-At-Random (MNAR)

If the missing mechanism (non-response) is related to the value of the response itself, then the missing is said to be not at random [46]. In other words, if the reason of missingness is related with the variable itself then the missing mechanism is Missing-Not-At-Random (MNAR). Additionally, the mechanism can be explained or modeled by the observed part of that variable and other variables as well. Let us give an example about a survey on teenage addiction of illegal drugs. If questions regarding drug usage are asked directly to teenagers, they will either not answer those questions or lie. Thus, the missing data depends on the question itself and we can say that it is MNAR. Let us use the same notations like in MCAR and MAR cases. The mechanism is MNAR if,

$$P(R|D,\theta) = P(R|D^{obs},D^{miss},\theta) \tag{2.8}$$

for all D^{miss}, D^{obs}, and θ. MNAR is non-ignorable, so using traditional methods (deletion, single imputation, etc.) may not be a good idea since the characteristics (mean, variance, etc.) of observed data and the missing data may not be the same. It

can be explained as follows:

$$\begin{aligned} Under\ MAR: &\quad P(D^{miss}, \theta | D^{obs}, R_i = 0) = P(D^{miss}, \theta | D^{obs}, R_i = 1) \\ Under\ MNAR: &\quad P(D^{miss}, \theta | D^{obs}, R_i = 0) \neq P(D^{miss}, \theta | D^{obs}, R_i = 1) \end{aligned} \tag{2.9}$$

Therefore, the traditional approaches may be invalid for handling MNAR. A joint model of the data D and the missingness mechanism indicator R as in the equation 2.10 can be proven helpful.

$$P(R, D|\theta) = P(R, D^{miss}|D^{obs}, \theta) \tag{2.10}$$

Deciding on the factorization of the joint models may differ. Three different factorization methods can be found in the literature: selection models, pattern-mixture models and shared parameter models. The details are given in the following sections.

2.3 Handling MNAR Mechanism

The best method for handling the missingness is the ability to prevent the possibility of missingness before the research. But in practice, fully preventing it would prove nearly impossible. Therefore, the main purpose lies in reducing the uncertainty and the bias for the inference of parameters of interest. Several different methods with different classifications are given in the literature over years [46] [14] [52]. Most studies can be categorized under three main methods; deletion methods, imputation methods and likelihood-based methods. As an example, Skarga-Bandurova et al. (2018) has classified methods as in Figure 2.2 [72]. According to the Figure 2.2, the main deletion methods are listwise deletion and pairwise deletion. Imputation methods vary by the research areas and the nature of the data. Mainly, it can be categorized as single and multiple imputation methods. Model-based methods are based on likelihood principle which also includes Bayesian methods. There are also some machine learning methods for handling the missingness mechanism, but it is out of the content of this thesis.

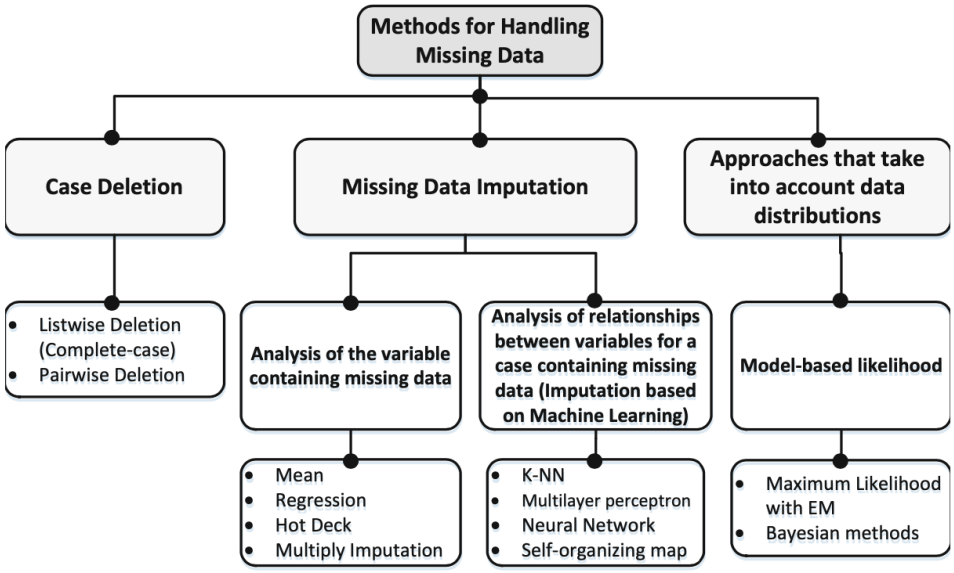

Figure 2.2: Classified Missing Data Methods of Skarga-Bandurova et al.(2018)

2.3.1 Deletion Methods

Deletion methods are based on the principle of omitting missingness. If the missing percentage is small and assumed to be ignorable, deletion of missing cases can help to handle the problem. Though in practice, this method should be chosen carefully since it may lead to biased inference for parameters of interest [14] [67]. Deletion can be performed in two different ways; listwise or pairwise.

Listwise Deletion (Complete-Case Analysis): It can be defined as basically the deletion of all cases that include missingness. [14]. That is to say, if the data is represented as a rectangular shaped block with multiple covariates (columns) and sample size of n (rows), any row that includes any missing cells will be deleted. Since only the complete cases (rows) are left in the data set, it is also called complete-case analysis. Therefore, loss of information can cause biased inferences about the data. That is why this method should not be preferred if the missingness mechanism is ignorable and the missing percentage is too high [67].

Pairwise Deletion (Available-Case Analysis): It is simply omitting the cases including missing data depending on the particular analysis carried out [14]. For ex-

ample, an individual answers all the questions in a survey except one. So omitting just that missing case and using the rest of the data provided by the individual can be useful for some analysis. By doing this, the sample size will differ for each analysis. Both listwise and pairwise deletion methods always cause a loss of information and hence misleading inference about the parameters of interest [67].

2.3.2 Imputation Methods

Imputation refers as placing new values instead of missing values [46] [14] [52]. In here, it is assumed that the imputed values should have been the real values if the missing data weren't missing. Of course, it is not possible to know the real value of a missing value in practice. Therefore, imputation methods still contain an uncertainty about missingness even though it may help with reducing it [13]. There are a lot of imputation methods according to the missingness mechanism and the nature of the data, however it can basically be categorized as single and multiple imputation methods.

Single Imputation Methods: They are done simply by imputation for once. The most well-known ones are mean imputation, regression imputation and hot-deck imputation. In longitudinal studies, another method called last observations carried forward can be found.

Mean Imputation can be defined as a method that all the missing values are replaced by the mean of the observed values of that variable [14] [46]. This method ensures to maintain the sample size but, the variability of the data reduces. Hence, the variance and standard error will be underestimated. Further, the inferences for parameters of interest may be biased, especially if the missingness mechanism is identified as MAR or MNAR [14] [13].

Regression Imputation is a method through which the missing values are replaced by the predicted values which are obtained from a regression model or a predictive distribution, that is obtained from the available data [14] [46]. In here, the predictions

are based on the regression model, which makes the imputations highly correlated with the regression model. As Enders (2010) indicated, this method may also cause underestimated variance and standard error although it reduces the bias of the estimates [14]. Stochastic regression models are suggested in order to reduce the correlations of imputations by adding normally distributed residual terms with mean zero and variance equal to the residual variance of the predicted value in regression model. This addition adds randomness to the imputations and reduces the correlations, however the standard error is underestimated [14].

Hot-Deck Imputation is the method applied when the missing value is replaced by the value of the other observed values in other variables of which are very similar or matching. Therefore, it is also called matching imputation [14] [46]. For a better explanation, if the columns of the data are variables and the rows are individuals, the missing value of an individual is replaced by the same value of another individual of which their other available values are matching or very close to each other. This imputation may also cause an underestimated standard error and variance [67].

Last Observation Carried Forward method is used in longitudinal studies. If the individual drops out from the research after a specific time, the last value observed from that individual is used as it is not changed until the end of the study [14].

Multiple Imputation Methods: They are the methods that missing values are replaced by new values for multiple times. This means that multiple data sets are obtained and used in order to reduce the uncertainty about the imputed values [68] [46]. These data sets are generated from predictive distributions or constructed regression models with available data. Rubin (1987) indicates the process of multiple imputation as imputation, analysis and pooling [68]. After the imputation, the necessary analysis can be performed with imputed data sets and inference for parameters of interest are obtained for each imputed data sets. Then, all of the inferences or estimations are combined to a single result with a pooling technique [68]. Therefore, the uncertainty about the missingness is naturally reduced. In single imputation methods, the imputed values are applied only once, so the missing value is treated as a known single

value which doesn't reflect the variability of the missing value and causes inaccurate estimates. Multiple imputation, on the other hand, handles this problem, although with a burden of more work such as multiple data sets, multiple analysis, etc. [46].

2.3.3 Likelihood-Based Methods

Handling missing data can also be performed by using likelihood principles. Likelihood functions for observed or complete data with some specific model assumptions are used in order to handle missingness [68] [46] [52]. These methods can be based on both Maximum Likelihood (ML) and Bayesian Approach.

Maximum Likelihood Methods: Generally, maximum likelihood methods are used when the missingness mechanism is ignorable. The most well-known method is an iterative approach called the Expectation Maximization (EM) Algorithm. It is a two-step algorithm of which the first step is Expectation (E-step) and the second step is Maximization (M-step). In E-step, the complete-data are obtained as the the missing values are replaced by the conditional means given the observed data and parameter estimates. In other words, E-step can be given in Equation 2.11 with the same notations in section 2.2.1.

$$D_i^{imputed} = E(D_i^{miss}|D_i^{obs}, \theta), \;\; i = 1, ..., n_{miss} \tag{2.11}$$

where $D_i^{imputed}$ is the imputed value of i^{th} missing value and n_{miss} is the size of the missing data. So the complete-data is $D^{complete} = (D_i^{imputed}, D_i^{obs})$. Then in M-step, the complete-data is maximized with ML principle [52]. The likelihood to be maximized is $L(\theta|D^{complete})$.

This method is said to be suitable in MCAR and MAR mechanisms, since the parameters estimation are unbiased, but it is not recommended for MNAR mechanisms, since the estimations tend to be biased [46] [14]. Little and Rubin (2002) also mentions that when the missing data proportion is large, the convergence may be very slow and additionally at times, a closed form of the likelihood can not be obtained in M step [46].

Bayesian Methods: Posterior distributions for parameters of interest can be obtained by combining priors for unknown parameters with the likelihood [46]. Bayesian methods use multiple imputation methods. Data Augmentation is one of the most popular Bayesian method for handling missing data.

Data Augmentation is an iterative method with 2 steps. The first step is called the Imputation step (I-step), in which the data imputation is done by using a predictive stochastic regression model by using the observed data and estimated parameters for the missing data [14]. This predictive stochastic regression model can be expressed as conditional distribution given the observed values and estimated parameters or posterior predictive distribution. The detailed information about posterior predictive distribution and other Bayesian Approach terms is given in Section 2.4.1. Imputed data sets are obtained from I-step and after that every data set is used to get likelihood functions in the second step called Posterior step (P-step), where the Bayesian analysis is applied. Also, prior distributions for the parameters are specified and posterior distributions are defined in order to get alternative parameter estimations. These two steps are applied for number of times by using another Bayesian approach method called Monte Carlo Simulation. Let us explain this part by using notation used in section 2.2.1;

$$\begin{aligned} I-Step: &\quad \boldsymbol{D}_1^{imputed} \sim P(D^{miss}|D^{obs}, \boldsymbol{\theta}_0^*) \\ P-Step: &\quad \boldsymbol{\theta}_1^* \sim P(\boldsymbol{\theta}|D^{obs}, \boldsymbol{D}_1^{imputed}) \end{aligned} \tag{2.12}$$

where $\boldsymbol{D}_1^{imputed}$ is the imputed data set of the trial 1, $P(D^{miss}|D^{obs}, \boldsymbol{\theta}_0^*)$ is the posterior predictive model for missing data and $\boldsymbol{\theta}_0^*$ is the initial parameters. This iteration is applied for t times as Monte Carlo Simulation. So the steps becomes as in Equation 2.13.

$$\begin{aligned} I-Step: &\quad \boldsymbol{D}_t^{imputed} \sim P(D^{miss}|D^{obs}, \boldsymbol{\theta}_{t-1}^*) \\ P-Step: &\quad \boldsymbol{\theta}_t^* \sim P(\boldsymbol{\theta}|D^{obs}, \boldsymbol{D}_t^{imputed}) \end{aligned} \tag{2.13}$$

where $\boldsymbol{D}_t^{imputed}$ is the imputed data set of the trial t, $P(D_i^{miss}|D_i^{obs}, \boldsymbol{\theta}_{t-1})$ is the posterior predictive model for missing data and $\boldsymbol{\theta}_{t-1}$ is the simulated values of parameters from trial $t-1$. So, every iteration uses the information of the previous iteration, which means every iteration depends on the previous one. That creates a chain, which is called Monte Carlo Markov Chain (MCMC). The detailed information is given in

Section 2.4.1. If the chain is long enough (t is large enough), the draws for the t^{th} imputation can be obtained from a posterior distribution that averages of the entire range of all possible values for the missing data [14]. This means the imputations may converge to proper values that represents the missing values at best. Convergence diagnostics for MCMC can also be used in order to see if the imputations are accurate. The meaning of convergence and convergence diagnostics are discussed on Section 2.4.1 as well. Data Augmentation method can further be taken as the first step of multiple imputation method. The imputed data sets in the chain can be taken and the multiple imputation steps (imputation, analysis and pooling) can be applied in order to handle missing data.

2.3.4 Joint Models for Dealing MNAR

When the data has MNAR mechanism, it is not possible to think and analyze the data and the missing mechanism separately. The joint model of the data and the missing mechanism should be considered together in order to handle the missingness. As it is stated in the Equation 2.10 in Section 2.2.3, the joint model should be factorized for analysis. There are three main factorizations for MNAR joint models; selection model, pattern-mixture model and shared parameter model. These factorization methods imply the same joint model with different assumptions, and that is why they can produce different estimates [14]. Therefore, sensitivity analysis with different assumptions to the same data is very crucial in handling MNAR mechanism.

Selection Model: It is a factorization method for joint models in MNAR mechanism [14] [52] [46]. It can be denoted as follows;

$$P(R, D|\beta, \phi) = P(R|D, \phi)P(D|\beta) \tag{2.14}$$

where R is the missingness indicator, D is the data, ϕ is the parameter vector for the missingness model and β is the parameter vector related with the data, $P(R|D, \phi)$ is the conditional distribution for missingness model indicator given the data and $P(D|\beta)$ is the marginal distribution of the data. When the data is separated as $D =$

(D^{miss}, D^{obs}) the factorization becomes as in the Equation 2.15.

$$\begin{aligned} P(R, D|\beta, \phi) &= P(R, D^{miss}|D^{obs}, \beta, \phi) \\ &= P(R|D^{miss}, D^{obs}, \phi)P(D^{miss}|D^{obs}, \beta) \end{aligned} \tag{2.15}$$

where $P(R, D^{miss}|D^{obs}, \beta, \phi)$ joint distribution of the missingness indicator and the missing data given the observed data, $P(R|D^{miss}, D^{obs}, \phi)$ is the conditional distribution for missingness model indicator given the data and $P(D^{miss}|D^{obs}, \beta)$ is the conditional distribution of the missing data given the observed data.

Pattern Mixture Model: It is another factorization method for joint models used in MNAR mechanism when the missingness mechanism has been divided into different stratifications or patterns [14] [52] [46]. It can be denoted as follows;

$$P(R, D|\beta, \phi) = P(D|R, \beta)P(R|\phi) \tag{2.16}$$

where R is the missingness indicator, D is the data, ϕ is the missingness indicator parameters and β is the parameters related with the data, $P(D|R, \beta)$ is the conditional distribution for the data under different patterns of missingness mechanism and $P(R|\phi)$ is the marginal distribution of the missingness mechanism. When the data is separated as $D = (D^{miss}, D^{obs})$ the factorization becomes as in the Equation 2.17.

$$\begin{aligned} P(R, D|\beta, \phi) &= P(R, D^{miss}|D^{obs}, \beta, \phi) \\ &= P(D^{miss}|R, D^{obs}, \beta)P(R|D^{obs}, \phi) \end{aligned} \tag{2.17}$$

where $P(R, D^{miss}|D^{obs}, \beta, \phi)$ joint distribution of the missingness indicator and the missing data given the observed data, $P(D^{miss}|R, D^{obs}, \beta)$ is the conditional distribution for the missing data under different missingness mechanism patterns and $P(R|D^{obs}, \phi)$ is the conditional distribution of the missingness mechanism given the observed data.

Shared Parameter Model: Sometimes MNAR mechanism may depend on an unknown latent variable. Therefore, latent variables or random effects are needed to be added the joint model as a new unknown parameter which is shared by both the missing data and missing mechanism. Shared Parameter Models are used in these cases when the factorization is not convenient by only selection or pattern-mixture

models [52]. It can be denoted with both selection model factorization and pattern-mixture model factorization as follows;

$$\begin{aligned} Selection: & P(R, D, Z|\beta, \phi, \xi) = P(R|D, Z, \phi)P(D|Z, \beta)P(Z|\xi) \\ Pattern - Mixture: & P(R, D, Z|\beta, \phi, \xi) = P(D|R, Z, \beta)P(R|Z, \phi)P(Z|\xi) \end{aligned} \tag{2.18}$$

where R is the missingness indicator, D is the data, Z is the latent variable or random effect variable, ϕ is the missingness indicator parameters and β is the parameters related with the data, ξ is the parameters related with the random effect, $P(R|D, Z\phi)$ is the conditional distribution for missingness model indicator given the data and the random effect, $P(D|Z, \beta)$ is the conditional distribution of the data given the random effect and $P(Z|\xi)$ is the marginal distribution of the random effect. When the data is separated as $X = (D^{miss}, D^{obs})$ the factorization becomes as in the Equation 2.19.

$$\begin{aligned} P(R, D, Z|\beta, \phi, \xi) &= P(R, D^{miss}, Z|D^{obs}, \beta, \phi, \xi) \\ Selection: \quad &= P(R|D^{miss}, D^{obs}, Z, \phi)P(D^{miss}|D^{obs}, Z, \beta)P(Z|D^{obs}, \xi) \\ Pattern - Mixture: \quad &= P(D^{miss}|R, D^{obs}, Z, \beta)P(R|D^{obs}, Z, \phi)P(Z|D^{obs}, \xi) \end{aligned} \tag{2.19}$$

where $P(R, D^{miss}, Z|D^{obs}, \beta, \phi, \xi)$ joint distribution of the missingness indicator, the missing data and the random effect given the observed data, $P(R|D^{miss}, D^{obs}, Z, \phi)$ is the conditional distribution for missingness model indicator given the missing data, observed data and the random effect, $P(D^{miss}|D^{obs}, Z, \beta)$ is the conditional distribution of the data given the observed data and the random effect, $P(D^{miss}|R, D^{obs}, Z, \beta)$ is the conditional distribution of the data given the missingness indicator, observed data and the random effect, $P(R|D^{obs}, Z, \phi)$ is the conditional distribution for missingness model indicator given the observed data and the random effect, $P(Z|\xi)$ is the conditional distribution of the random effect given the observed data. In this thesis, selection model is chosen for factorizing the joint distribution of the main GLM model and missingness models (also GLMs).

2.3.5 Selection Model For MNAR in GLM

Both main and the proposed missingness models are considered as GLM models and the joint distribution of them should be taken into account. The main model is assumed to be a logistic model with binary response variable Y, covariates $\boldsymbol{X} = (X_1, X_2, ..., X_p)$. If the missingness mechanism belongs to the response variable, then the

selection model will be as in Equation 2.20.

$$\begin{aligned} P(R, Y|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi}) &= P(R, Y^{miss}|Y^{obs}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi}) \\ &= P(R|Y^{miss}, Y^{obs}, \boldsymbol{X}, \boldsymbol{\phi}) P(Y^{miss}|Y^{obs}, \boldsymbol{X}, \boldsymbol{\beta}) \end{aligned} \tag{2.20}$$

where Y^{miss} is the missing part, Y^{obs} is the observed part of the response, $\boldsymbol{\beta}$ is the main model parameter vector and $\boldsymbol{\phi}$ is the missingness model parameter vector, $P(R, Y^{miss}|Y^{obs}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi})$ joint distribution of the missingness indicator and the missing data given the observed data, $P(R|Y^{miss}, Y^{obs}, \boldsymbol{X}, \boldsymbol{\phi})$ is the conditional distribution for missingness indicator model given the data and $P(Y^{miss}|Y^{obs}, \boldsymbol{X}, \boldsymbol{\beta})$ is the conditional distribution of the missing data given the observed data.

If the missingness mechanism belongs not to the response variable but to the covariates then the covariates are separated as $\boldsymbol{X} = (\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}})$ the factorization becomes as in the Equation 2.21.

$$\begin{aligned} P(R, Y, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, \boldsymbol{\beta}, \boldsymbol{\phi}) = & P(R|Y, \boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, \boldsymbol{\phi}) \\ & P(Y|\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, \boldsymbol{\beta}) \\ & P(\boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}) \end{aligned} \tag{2.21}$$

where $P(R, Y, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, \boldsymbol{\beta}, \boldsymbol{\phi})$ joint distribution of the missingness indicator, main model and the missing data given the observed data, $P(R|Y, \boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, \boldsymbol{\phi})$ is the conditional distribution for missingness indicator model given the data and $P(Y|\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, \boldsymbol{\beta})$ is the conditional distribution of the main model given the data and $P(\boldsymbol{X^{miss}}|\boldsymbol{X^{obs}})$ is the conditional distribution of the missing covariates given the observed data. The study proceeds with brief explanation of fitting methods of the main and missingness models.

2.4 Model Selection for Missingness Mechanism

As George E. P. Box once said "Essentially, all models are wrong but some are useful", model selection in statistical modeling is always challenging. Since modeling is a way of approximation of how the data behave, the true and correct model does not exist. The aim should be towards choosing the best model or models that fit the data. However, Hoeting et. al. ask a question like what if there is any other model with the

same best fit but has different estimated coefficients and standard errors [29]. And also, there are some examples of models with the same good fit but different coefficients and standard errors [62] [48] [39] [59]. All this implies to model selection's uncertainty and it may be problematic in estimation. We cannot know that in predictive modeling, the chosen model is definitely the best model. There is always a model uncertainty that should be considered. In addition to the uncertainty of choosing the best fit for the main model, there is also uncertainty about the MNAR mechanism and for choosing the best missingness indicator model. In this thesis, Bayesian inference and Bayesian model selection methods are used in order to take these uncertainties in to consideration together.

2.4.1 Bayesian Inference

Bayesian inference can be made with conditional probability for parameters of interest or for unobserved data given the observed data. This conditional probability is based on Bayes' Rule.

$$P(D|\theta) = \frac{P(D,\theta)}{P(\theta)} \tag{2.22}$$

Equation 2.22 is the simplest form of Bayes Rule where $P(D|\theta)$ is the probability distribution of the data D given the parameter of interest θ, $P(\theta, D)$ is the joint probability distribution and $P(\theta)$ is the marginal distribution for θ. So this equation can be rewritten as $P(D,\theta) = P(D|\theta)P(\theta)$. This expression can be used in the following equation:

$$P(\theta|D) = \frac{P(D,\theta)}{P(D)} \Rightarrow P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{2.23}$$

The Bayes' rule is applied as in the Equation 2.23 and the joint probability $P(D,\theta)$ is rewritten as $P(D|\theta)P(\theta)$. In here, $P(D) = \int_\theta P(D|\theta)P(\theta)d\theta$ is the integration. Since, $P(D)$ does not depend on θ and the data are observed, it can be considered as a normalizing constant. Therefore, Equation 2.23 can be written as follows:

$$P(\theta|D) \propto P(D|\theta)P(\theta) \tag{2.24}$$

Equation 2.24 is the basis of Bayesian inference and Bayesian data analysis. When dealing with modeling, the conditional probability of parameter of interest θ given the observed data D, $P(\theta|D)$, is called the posterior distribution (or density), the

conditional probability of observed data D given the parameter θ, $P(D|\theta)$, is called the likelihood function, and the marginal probability of θ, $P(\theta)$, is called as prior distribution (or density). Posterior distribution can be considered as updating the prior information about the parameter of interest, by using the likelihood function, which is the information from the data at hand. Regarding this, if this process is repeatedly performed as iterations, the posterior density of the previous iteration is actually the new prior of the next one.

In a single parameter model, likelihood function is basically the likelihood of the chosen model and the prior is the density function or density probability for that single parameter. Choice of prior is an important matter for Bayesian Analysis. Especially, when there is no information about the prior, the discussions continue about the subjectivity of the analyses.

Choosing Prior Distribution: Prior distribution can be chosen as informative or non-informative. If there is apriori information about the parameter, it should be written in a suitable probability function [41]. Conjugate prior is a suitable solution for this problem. If the prior has the same distributional form with the model, then it is called conjugate prior [41]. Non-informative priors, also called as weak, vague, nonsubjective, etc., can be chosen when there is no apriori knowledge about the parameter of interest. In order to put this lack of knowledge to the model properly, the prior should be chosen carefully. Uniform priors or flat priors are used generally as non-informative priors, although there are some problems of these priors like invariant reparameterization and improper posterior [19] [41]. Jeffreys (1946) solve the invariance problem by proposing a new prior, called Jeffreys prior, that is the square root of the Fisher Information [35]. Mathematically, it is given as follows:

$$P(\theta) = \left(-E\left(\frac{d^2 log P(D|\theta)}{d\theta^2} \right) \right)^{1/2} \tag{2.25}$$

where $E\left(\frac{d^2 log P(D|\theta)}{d\theta^2} \right)$ is the expected value of the second derivative of the log-likelihood function. Although, it solves the invariant reparameterization problem, it may also cause improper posteriors if it is not used properly [43]. Informative priors are also called as subjective priors. These priors are chosen when there is historical data at hand other than the present data, also when there is expert knowledge about

the prior.

After calculating the posterior distribution by using the prior and the likelihood (of the data), the inference for the parameter of interest can be made, however the estimation may not be easy to determine. The computation of the integration in the normalizing constant is usually hard to handle when the number of parameter of interest is large. Therefore, there are some numerical integration methods for Bayesian data analysis. The most well-known class of methods is called as Markov Chain Monte Carlo (MCMC) simulation. The details are given in the following section.

Markov Chain Monte Carlo (MCMC) Methods: Sampling from well-known densities are computationally available in statistical softwares. Markov Chain Monte Carlo (MCMC) methods can be defined as Bayesian sampling methods used when the form of the posterior density is not familiar [10] [43]. The chain starts with sampling from posterior density with pregiven initial values at time 0. Then, at time 1, new samples are drawn depending on the sample obtained from the posterior density at time 0. If the chain is long enough, the sample obtained from the posterior density at time t is taken as the estimates of real values. This sample can be the estimates of the parameter of interest or unobserved data. In Equation 2.26, the MCMC sampling process is given in a very simple form.

$$\begin{aligned} \theta^1 &\sim P(\theta^0|D) \\ \theta^2 &\sim P(\theta^1|D) \\ &\vdots \\ \theta^t &\sim P(\theta^{t-1}|D) \end{aligned} \tag{2.26}$$

where θ^0 is the initial value, θ^1 is the first sample from posterior density $P(\theta^0|D)$ and θ^t is the sample at time t. Obviously, all the sampling steps depends on the previous one, which may cause autocorrelation problems and it requires attention. At this point, it is good to mention about "burn-in" term. It is basically discarding the early iterations of the MCMC process [20]. Because of the dependency, and the effect of starting values, the early iterations may cause unstationarity in the chain. Discarding the part of the chain up to the first indications of convergence of the posterior is needed [10]. There are some MCMC diagnostics that show whether the posterior

distribution converges or not and also show the autocorrelation among the iterations. The details about the diagnostics are given at Section 2.4.3.

The most popular MCMC methods are Metropolis-Hastings Sampling and Gibbs Sampling. Details about these methods are given in the following sections.

Metropolis-Hastings Sampling is the extension of Metropolis Algorithm, proposed by Metropolis (1953) [50]. Hastings (1970) had extended this algorithm and afterwards the algorithm has been called Metropolis-Hastings Algorithm [25]. In this algorithm, candidate samples are obtained from a proposal density, and then the candidate samples are tested with an acceptance-rejection rule. Let us denote θ^t as the sample of parameter (or parameter vector) at time (iteration) t and θ^c as the candidate sample of parameter (or parameter vector) obtained from the proposal density $q(\theta^c|\theta^t)$. The proposal density is also called as jumping probability or transition kernel. The sample of the next iteration θ^{t+1} is obtained with a ratio tested by an acceptance-rejection rule. If the ratio is acceptable then $\theta^{t+1} = \theta^c$, otherwise $\theta^{t+1} = \theta^t$. The steps of Metropolis-Hastings algorithm can be illustrated as follows;

Step 1: $\theta^c \sim q(\theta^c|\theta^t)$

Step 2: $\theta^{t+1} = \theta^c$ with probability $\alpha(\theta^t, \theta^c)$

Step 3: Otherwise $\theta^{t+1} = \theta^t$.

In here, the acceptance probability $\alpha(\theta^t, \theta^c)$ is,

$$\alpha(\theta^t, \theta^c) = min\left\{\frac{P(\theta^c|D)q(\theta^t|\theta^c)}{P(\theta^t|D)q(\theta^c|\theta^t)}, 1\right\} \tag{2.27}$$

where $P(\theta^c|D)$ is the posterior probability of θ^c, $q(\theta^t|\theta^c)$ is the proposal density (transition kernel) for θ^t, $P(\theta^t|D)$ is the posterior probability of θ^t and $q(\theta^c|\theta^t)$ is the proposal density (transition kernel) for θ^c. The results of the Metropolis-Hastings algorithm can differ according to the use of proposal density [43].

Gibbs Sampling was first introduced by Geman and Geman (1984) [22]. It is an iterative algorithm used when the parameter of interest vector has more than one

dimension. The joint posterior density of the parameter vector $\boldsymbol{\theta} = (\theta_0, ..., \theta_n)$ is $P(\theta_0, ..., \theta_n|D)$ and it can be factorized by using Bayesian context as in Equation 2.28.

$$\begin{aligned} P(\theta_0, ..., \theta_k, ..., \theta_n|X) =& P(\theta_0|\theta_1 ..., \theta_n, D) \\ & \vdots \\ & P(\theta_k|\theta_0 ..., \theta_{k-1}, \theta_{k+1}, ..., \theta_n, D) \\ & \vdots \\ & P(\theta_n|\theta_0 ..., \theta_{n-1}, D) \end{aligned} \tag{2.28}$$

In Gibbs sampling, this factorization property is used. Every element in the right hand side of this equation is a marginal posterior density. Marginal posterior densities of parameters are also called as full conditional distributions or just full conditionals [43].

Gibbs sampling algorithm starts off with an initial value vector of $\boldsymbol{\theta^{(0)}} = (\theta_0^{(0)}, ..., \theta_n^{(0)})$ at time (iteration) 0. In iteration 1, the algorithm generate samples for each parameter in the vector, by using the initial values. Equation 2.29 shows the sampling algorithm at time 1.

$$\begin{aligned} \theta_0^{(1)} &\sim P(\theta_0|\theta_1^{(0)}, ..., \theta_n^{(0)}, D) \\ \theta_1^{(1)} &\sim P(\theta_1|\theta_0^{(1)}, \theta_2^{(0)}, ..., \theta_n^{(0)}, D) \\ &\vdots \\ \theta_k^{(1)} &\sim P(\theta_k|\theta_0^{(1)}, ..., \theta_{k-1}^{(1)}, \theta_{k+1}^{(0)}, ..., \theta_n^{(0)}, D) \\ &\vdots \\ \theta_n^{(1)} &\sim P(\theta_n|\theta_0^{(1)}, ..., \theta_{n-1}^{(1)}, D) \end{aligned} \tag{2.29}$$

where $\theta_k^{(1)}$ is the k^{th} parameter generated from full conditional density of θ_k at iteration 1. Equation 2.30 shows the general process of Gibbs sampling.

$$\begin{aligned} \theta_0^{(t)} &\sim P(\theta_0|\theta_1^{(t-1)}, ..., \theta_n^{(t-1)}, D) \\ \theta_1^{(t)} &\sim P(\theta_1|\theta_0^{(t)}, \theta_2^{(t-1)}, ..., \theta_n^{(t-1)}, D) \\ &\vdots \\ \theta_k^{(t)} &\sim P(\theta_k|\theta_0^{(t)}, ..., \theta_{k-1}^{(t)}, \theta_{k+1}^{(t-1)}, ..., \theta_n^{(t-1)}, D) \\ &\vdots \\ \theta_n^{(t)} &\sim P(\theta_n|\theta_0^{(t)}, ..., \theta_{n-1}^{(t)}, D) \end{aligned} \tag{2.30}$$

where $\theta_k^{(t)}$ is the k^{th} parameter generated from marginal posterior density of θ_k at iteration t.

Both Gibbs and Metropolis-Hastings sampling algorithms are iterative approaches and they can both run repeatedly. These repetitions are called chains of MCMC. The chains can run one after the other or run at the same time according to the software and the written codes. Deciding on which sampling algorithm to use, Gibbs or Metropolis-Hastings, depends on the structure and the aim of the analysis. If full conditionals for parameters are in forms, from which sampling is slightly easy, Gibbs sampling may be more preferable. When the proposal density for the parameter of interest in Metropolis-Hastings algorithm enables better and fast convergence, then this algorithm is more favorable.

These sampling methods are very practical, on the other hand, they are challenging in performance of convergence of the parameter of interest. Further, the accuracy of posterior results should be diagnosed with some visual and computational techniques. There are a lot of methods called MCMC Diagnostics for monitoring and improving the convergence and the performance of the sampling method. They are very crucial for a proper Bayesian analysis.

2.4.2 Bayesian Model Selection Methods

Model averaging approaches are useful to solve the uncertainty about model selection. There are both frequentest and Bayesian approaches for model selection and model averaging. Ntzoufras (1999) indicates that Bayesian model selection may offer better solutions by using a well-known Bayesian approach called Markov Chain Monte Carlo (MCMC) technique which can construct the model space, identify good models and estimate their posterior probabilities based on both prior knowledge and observed data [55]. The focus of this thesis is estimating the parameters of interest by using models in a model space constructed with Bayesian model selection methods. The model selection in Bayesian point of view is based on choosing the models of which the posterior distributions are most likely. In model averaging or model pooling, estimations are calculated from the models with high posterior probabilities by averaging or pooling with different methods. Giving more details about Bayesian

model selection methods helps for better comprehension.

Bayes Factor: Raftery and Kass (1995) proposed an odds ratio by using the Bayes' Formula properties as comparison of two hypotheses for a given data [39]. It is basically a quantity for comparing the models pair-wisely. Let D is the data, model M_i and model M_j are two models. The posterior probabilities for models are given in Equation 2.31.

$$\begin{aligned} P(M_i|D) &= \frac{P(D|M_i)P(M_i)}{P(D)} \\ P(M_j|D) &= \frac{P(D|M_j)P(M_j)}{P(D)} \end{aligned} \tag{2.31}$$

In here, $P(D|M_i) = \int P(D|M_i, \theta_i)P(\theta_i|M_i)d\theta_i$ is the marginal likelihood of the model M_i where θ_i is the parameter (or the vector of parameters) under model M_i and $P(\theta_i|M_i)$ is the prior density for the parameter (or the vector of parameters) [39]. Integrating $P(D|M_i)$ needs some numerical methods most of the time [4]. The odds for these two posterior probabilities is as follows in Equation 2.32.

$$\begin{aligned} \frac{P(M_i|D)}{P(M_j|D)} &= \frac{P(D|M_i)}{P(D|M_j)} \times \frac{P(M_i)}{P(M_j)} \\ Posterior\ Odds &= \mathbf{Bayes\ Factor} \times Prior\ Odds \end{aligned} \tag{2.32}$$

According to Equation 2.32, Bayes Factor B_{ij} can be calculated as in Equation 2.33.

$$B_{ij} = \frac{Posterior\ Odds(M_i, M_j)}{Prior\ Odds(M_i, M_j)} = \frac{P(M_i|D)/P(M_j|D)}{P(M_i)/P(M_j)} \tag{2.33}$$

The main purpose is comparing two models and the ratio should be interpreted according to some criterion. Jeffreys (1961) suggested how Bayes Factor should be interpreted as a scale of evidence [36]. The interpretation is given in Table 2.1.

When one of the priors or both models are improper, a problem occurs in calculating Bayes factors. There are alternative solutions for this problem such as Psuedo Bayes Factor.

Bayesian Information Criterion (BIC): Information criteria are used for model selection. Bayesian Information Criterion (BIC) was firstly introduced by Schwarz (1978) for model selection by using a penalty term for the parameters in the models [69]. Let $L(D|M_i)$ be the likelihood of the data D under the model M_i and θ_i is the

Table 2.1: Jeffreys Interpretation for Bayes Factor

Bayes Factor (B_{ij})	**Interpretation**
Less than 1	M_i is not favorable
Between 1 and 3	Barely worth to mention M_i is favorable
Between 3 and 10	Substantial evidence for favoring M_i
Between 10 and 100	Strong evidence for favoring M_i
More than 100	Decisive evidence for favoring M_i

parameter vector of M_i to be estimated, then the BIC of the M_i is given in Equation 2.34.

$$BIC_{M_i} = -2 \ln L(D|M_i) + \ length(\theta_i) \log(n) \tag{2.34}$$

where $length(\theta_i)$ is the dimension of the model, n is the sample size for data D. For this equation, the smaller the BIC, the better the model is. In Bayesian approach, BIC can be used in calculating posterior probability of a model. The details is given in following sections.

Bayesian Model Averaging (BMA): It is a method of model averaging with Bayesian perspective. Hoeting et al. (1999) proposed this method with a simple weighted averaging formula. The model is given in Equation 2.35 [29].

$$P(\theta|D) = \sum_{i=1}^{k} P(\theta|M_i, D)P(M_i|D) \tag{2.35}$$

where k is the length of the model space, $P(\theta|M_i, D)$ is the posterior distribution of the parameter of interest given the data D and the model M_i and $P(M_i|D)$ is the posterior probability of model M_i given the data. In here, $P(M_i|D)$ can be regarded as the weights for parameter estimations. However, calculating these weights can be challenging. $P(M_i|D)$ is expressed in Equation 2.36.

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{\sum_{j=1}^{k} P(D|M_j)P(M_j)} \tag{2.36}$$

where

$$P(D|M_i) = \int P(D|\theta_i, M_i)P(\theta_i|M_i)d\theta_i \tag{2.37}$$

is the integrated likelihood for model M_i and θ_i is the parameter vector for the model M_i. Calculating $P(D|M_i)$ is not always easy. As it is mentioned in Section 2.4.2, numerical methods are needed. The main problems of BMA can be stated as below;

1. If number of models are very large, the estimation process for the parameter of interest for every model, i.e. calculation of each $P(\theta|M_i, D)$, takes too long.
2. The integral for posterior distribution of model M_i, $P(M_i|D)$, is hard to compute.
3. Specification of the prior probability of the model M_i, $P(M_i)$, needs attention for the analysis.

There are some methods for reducing the number of models, stated as the first problem of BMA, such as Occam's Window, Occam's Razor, Leaps and Bounds and Monte Carlo Model Composition, etc. Occam's Window method is one of the most popular methods used for solving this problem. Madigan and Raftery (1994) proposed this method by defining a set for models, which are chosen according to a criteria. This set is given in Equation 2.38 [48].

$$\mathcal{A}' = \left\{ M_k : \frac{max_l\{P(M_l|D)\}}{P(M_k|D)} \leq C \right\} \tag{2.38}$$

where $max_l\{P(M_l|D)\}$ is the posterior probability of the most likely model M_l among the other models in the model space, C is the threshold value, chosen arbitrarily. Generally C is chosen 10, 20 or 100 but it may change up to the aims of the research.

About the second potential problem, as it is stated earlier, the computation of the posterior probabilities is challenging. As Raftery et al. (2003) indicates, the posterior probabilities can be computed with an easier way by using BIC or AIC [61]. In this thesis, BIC adaptation is used. So posterior probabilities of the models can be computed by using the formula in Equation 2.39.

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{\sum_{j=1}^{k} P(D|M_j)P(M_j)} \approx \frac{P(M_i)\exp(-\frac{1}{2}BIC_{M_i})}{\sum_{j=1}^{k} P(M_j)exp(-\frac{1}{2}BIC_{M_j})} \tag{2.39}$$

where

$$BIC_{M_i} = -2\ln L(D|M_i) + \ length(\vartheta_i)\log(n) \tag{2.40}$$

Lastly, the choice of the prior probabilities, $P(M_i)$, depends on the specifications of the study. If all the models in the model space are equally likely, then it can be omitted from the Equation 2.39.

After Occam's window, the model space is reduced to a decent size and then the parameter of interest can be estimated by using the BMA Equation 2.35, where the posterior distribution of that parameter given each model is $P(\theta|M_i, D)$ and the posterior probability of each model is $P(M_i|D)$.

Reversible Jump Monte Carlo Markov Chain (RJMCMC): Green (1995) introduced a new method called Reversible Jump Markov Chain Monte Carlo which provides a general outline for Markov Chain Monte Carlo (MCMC) simulation where the dimension of the parameter space may vary among the Markov Chain iterates. The methodology is suitable for very different areas including factorial experiments, variable selection in regression, non-nested regression models, mixture deconvolution with an unknown number of components, Bayesian choice between models with different numbers of parameters, multiple change-point problems, image segmentation, object recognition, layout design etc. [24]. The studies of Green [24], [64], Waagepetersen [76], Sisson [71], [16], Ai [3], Özmen & Demirhan [56], Lunn [47], Yeh et. al. [78], Baghfalaki and Jalali [5], Karakuş [38], Theorell and Nöh [73] about reversible jump method are examined carefully. There is also a table, in which the software programs, the researchers used, are stated in Sission's (2004) article on RJMCMC [71]. The reversible jump algorithm can be defined as an extension of the Metropolis-Hastings algorithm onto more general state spaces with different dimensions. Model selection is one of the functions of RJMCMC. In this thesis, the focus for RJMCMC will be the estimation of joint posteriors for parameters as well as model posteriors by using model selection techniques because of some practical advantages. The first advantage is using an MCMC method, in which model space can be simulated easily. Details about the model space is given in the following sections. The second advantage is that when dealing with the computations for joint models with varying dimensions, an MCMC approach provide convinience in computations and model selection. Suppose that a model space $\mathcal{M} = \{M_1, M_2, ..., M_k\}$ in which the models have different dimension of parameter spaces with vector of $\boldsymbol{\theta}_\mathbf{k}$. Dimen-

sions of the spaces where $\boldsymbol{\theta}_\mathbf{k}$'s lie are $\mathcal{R}^{n_k}$. So the joint distribution of $(M_k, \boldsymbol{\theta}_\mathbf{k}, D)$, where D is the data, is given as;

$$P(M_k, \boldsymbol{\theta}_\mathbf{k}, D) = P(D|\boldsymbol{\theta}_\mathbf{k}, M_k)P(\boldsymbol{\theta}_\mathbf{k}|M_k)P(M_k) \tag{2.41}$$

where $P(D|\boldsymbol{\theta}_\mathbf{k}, M_k)$ is the likelihood under model M_k, $P(\boldsymbol{\theta}_\mathbf{k}|M_k)$ is the conditional probability of the parameter vector of interest under model M_k and $P(M_k)$ is the prior for model M_k. Markov Chain Monte Carlo (MCMC) approach allows to reach a target density by using transition kernels of a current state to the next state of the simulation. In reversible jump case, the transition can be among any state with different parameter spaces. Let $q(x, dx')$ is the transition kernel where, x is the current state while x' is the next state. The reversibility basically means that the transition can be done from both states, which means $q(x', dx) \neq 0$, where x' is the new current state and x is the new next state. According to this information with the MCMC Approach, Reversible Jump algorithm is said to be the extension of Metropolis-Hastings Algorithm. An acceptance probability, which is used for every proposed value, is given as,

$$\alpha(x, x') = min\left\{1, \frac{p(dx')q(x', dx)}{p(dx)q(x, dx')}\right\} \tag{2.42}$$

where $p(dx')$ is the target density while $p(dx)$ is the density of the current state.

In RJMCMC, there are 3 types of moves: moving to a state with higher dimension, moving to a state with lower dimension and moving to a state with the same dimension (Metropolis-Hastings). So, let $(M_k, \boldsymbol{\theta}_\mathbf{k})$ be the current state and $(M_{k'}, \boldsymbol{\theta}_\mathbf{k})$ be the next state.

Higher Move: If the dimension of $M_{k'}$, is higher than the one with M_k ($|M_{k'}| > |M_k|$), a new vector **u** with dimension $|M_{k'}| - |M_k|$ should be generated from a proposed prior $h(u)$ and therefore the new parameter space, $\boldsymbol{\theta}_{\mathbf{k'}}$, will be equal to the bijection of $\boldsymbol{\theta}_\mathbf{k}$ and **u**, which can be defined as $\boldsymbol{\theta}_{\mathbf{k'}} = g(\boldsymbol{\theta}_\mathbf{k}, \mathbf{u})$. The acceptance probability for the new state is given as,

$$\alpha(M_k, \boldsymbol{\theta}_\mathbf{k}, M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}}) = min\left\{1, \frac{P(D|M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}})P(\boldsymbol{\theta}_{\mathbf{k'}}|M_{k'})P(M_{k'})q(h, M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}})}{P(D|M_k, \boldsymbol{\theta}_\mathbf{k})P(\boldsymbol{\theta}_\mathbf{k}|M_k)P(M_k)q(h, M_k, \boldsymbol{\theta}_\mathbf{k})h(\mathbf{u})} \mid J_u \mid\right\} \tag{2.43}$$

Here, $P(D|M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}})P(\boldsymbol{\theta}_{\mathbf{k'}}|M_{k'})P(M_{k'})$ is given as in Equation 2.41 and $\mid J_u \mid$ is the Jacobian for u. Dimension of $\theta_{k'}$ is $dim(\theta_{k'}) = dim(\theta_k) + dim(u)$, so for dimension matching, determinant of the Jacobian, $\mid J_u \mid$, is needed where $\mid J_u \mid = \left|\frac{\partial g(\boldsymbol{\theta}_\mathbf{k}, \mathbf{u})}{\partial(\boldsymbol{\theta}_\mathbf{k}, \mathbf{u})}\right|$.

Lower Move: If the dimension of $M_{k'}$, is lower than the one with M_k ($|M_{k'}| < |M_k|$), the vector **u'** with dimension $|M_k| - |M_{k'}|$ should be discarded from the new

parameter vector, $\boldsymbol{\theta}_{\mathbf{k}}$. It is the inverse transformation of the bijection $g^{-1}(\boldsymbol{\theta}_{\mathbf{k}}) = (\boldsymbol{\theta}_{\mathbf{k'}}, \mathbf{u'})$.

$$\alpha(M_k, \boldsymbol{\theta}_{\mathbf{k}}, M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}}) = min\left\{1, \frac{P(D|M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}})P(\boldsymbol{\theta}_{\mathbf{k'}}|M_{k'})P(M_{k'})q(d, M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}})h(\mathbf{u'})}{P(D|M_k, \boldsymbol{\theta}_{\mathbf{k}})P(\boldsymbol{\theta}_{\mathbf{k}}|M_k)P(M_k)q(d, M_k, \boldsymbol{\theta}_{\mathbf{k}})} \mid J_{u'} \mid\right\} \quad (2.44)$$

Here, dimension of θ_k is $dim(\theta_k) = dim(\theta_{k'}) + dim(u')$, so for dimension matching, determinant of the Jacobian, $\mid J_{u'} \mid$, is needed where $\mid J_{u'} \mid = \left|\frac{\partial g^{-1}(\boldsymbol{\theta}_{\mathbf{k}})}{\partial(\boldsymbol{\theta}_{\mathbf{k'}}, \mathbf{u'})}\right|$.

Current Move: If the dimension of $M_{k'}$, is equal to the one with M_k ($|M_{k'}| = |M_k|$), the classical Metropolis-Hastings algorithm is applied.

$$\alpha(M_k, \boldsymbol{\theta}_{\mathbf{k}}, M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}}) = min\left\{1, \frac{P(D|M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}})P(\boldsymbol{\theta}_{\mathbf{k'}}|M_{k'})P(M_{k'})q(s, M_{k'}, \boldsymbol{\theta}_{\mathbf{k'}})}{P(D|M_k, \boldsymbol{\theta}_{\mathbf{k}})P(\boldsymbol{\theta}_{\mathbf{k}}|M_k)P(M_k)q(s, M_k, \boldsymbol{\theta}_{\mathbf{k}})}\right\} \quad (2.45)$$

The algorithm follows with obtaining the visiting probabilities, which are calculated by counting the total visits to that model during the process. Visiting probabilities is a useful tool to pick the best models since, they can be used as model posterior probabilities $P(M_k|D)$. These posterior probabilities can be used in Equation 2.35 for averaging the parameter of interests, $P(\boldsymbol{\theta}|D)$.

RJMCMC algorithm should continue until the joint posterior converges. There are specific indicators, measures and charts in order to detect the convergence, stationarity and autocorrelation of the Markov Chain called MCMC Diagnostics. Burn-in and iteration sizes can be decided and the convergence can be seen according to these Diagnostics.

2.4.3 MCMC Diagnostics

Improving the convergence and the performance of MCMC methods means that when the sampling is slow or having problems regarding converging, MCMC diagnostics can detect these and moreover, they can help to speed up the convergence. Clarification of the term convergence in MCMC context would be helpful at this point. It should not be understood as the convergence of the parameter of interest to its MLE or any unbiased estimator. Convergence describes how the sampling algorithm gets close to the true posterior distribution [43]. MCMC diagnostics can be examined in two parts: convergence diagnostics and accuracy diagnostics.

Convergence Diagnostics

Trace Plots: They are basically scattered line plots showing the value of the parameter in each iteration. Iterations are shown on the x-axis, while the values of the parameter of interest are on the y-axis [10]. In Figure 2.3, a simple example of a trace plot for one parameter and one chain is provided. The example is plotted with R software for 20000 iterations. Trace plots are very basic forms for understanding the

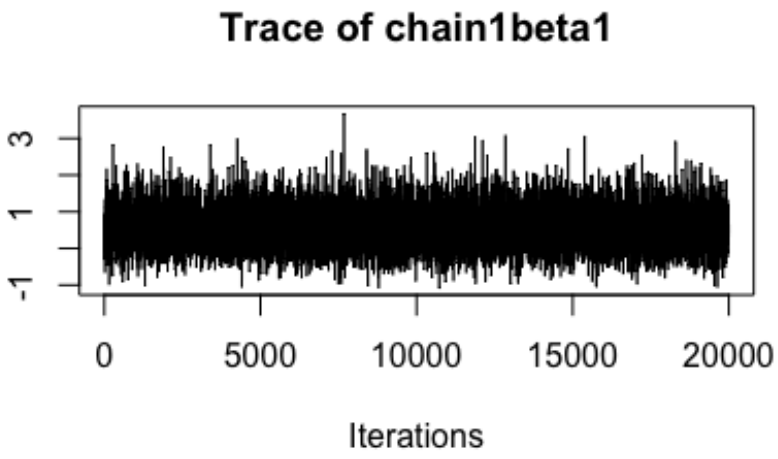

Figure 2.3: Trace plot example

stationary and convergence. If the moves of the parameter are not distinguishable, the trace plot looks like a thick line, which indicates stationarity [43]. Any change that breaks the stationarity can be monitored with this graph. If the samples of the parameter of interest do not change by iteration, some horizontal lines appear, which shows slow convergence and unstationarity. In addition, trace plots are useful to identify the mixing behavior of the chains. When there is more than one chain in the sampling algorithm, the same stationarity should be monitored at all chains. Trace plots can be plotted with chains all together, in order to see the chains are intermixing after some point.

Autocorrelation Plots: This graph is for measuring and identifying how the samples are correlated with each other by iteration. At it is stated in Section 2.4.1, every

sample depends on the previous one. This dependence causes slow convergence of the posterior. The distance among iterations is called lag. The autocorrelation is the measure of the correlation between parameter θ^t and θ^{t+k} for lag k [43] [10]. The formula for autocorrelation is given in Equation 2.46.

$$\rho_t = \frac{\sum_{i=1}^{N-t}(\theta_i - \bar{\theta})(\theta_{i+t} - \bar{\theta})}{\sum_{i=1}^{N}(\theta_i - \bar{\theta})^2} \tag{2.46}$$

where ρ_t is the autocorrelation for parameter of interest θ at lag-t [8]. So, autocorrelation plots show which lag or distance the autocorrelation of the iteration continues up to. Autocorrelation level is expected to drop around zero immediately when the lags start to increase. If it does not drop immediately and decrease gradually, it is a sign of slow convergence. Though paying attention is recommended at this point, since it does not mean that the posterior does not converge.

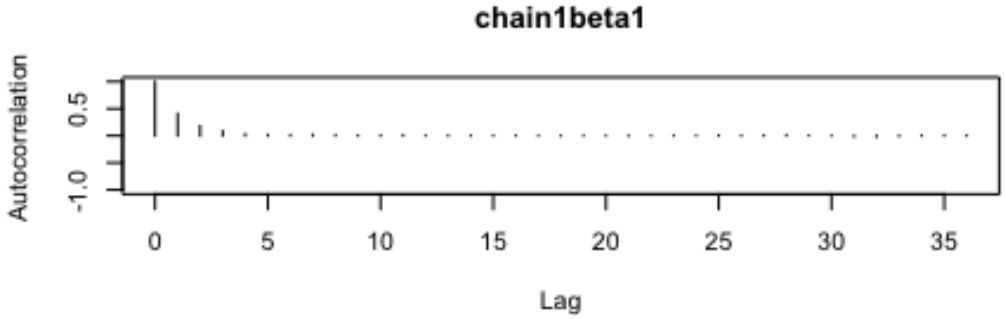

Figure 2.4: Autocorrelation plot example

In Figure 2.4, there is a simple autocorrelation plot, constructed by R software. After lag 5, the autocorrelation drops to zero.

Acceptance Rate in Metropolis-Hastings: The percentage of the accepted samples in Metropolis-Hastings algorithm is called Acceptance Rate. It measures the speed of convergence. According to Gelman and Rubin (2013) good acceptance rate is between 20% and 45% [20]. Low acceptance rate indicates most of the sampling is rejected and convergence is slow. High acceptance rate indicates, most of the sampling are accepted and thus Markov chain explores the posterior slowly, which is also an indication of slow convergence [43].

Monte Carlo Standard Error (MCSE): It is a measure of performance of the MCMC process. MCSE measures the error of sample values of the parameter of interest (mean, median, etc.) estimated in the posterior of the MCMC process and shows the convergence performance of the posterior. There are some methods to calculate MCSE such as batch means, spectral variance etc. [43]. The most popular calculation technique is the batch means method which divides the chain into b batches, then calculate the means for every batch $\bar{\theta}_b$ and also the overall mean of the chain $\bar{\theta}$. Then MCSE can be calculated by using the Equation 2.47 [43].

$$MCSE = S_\theta^b = \sqrt{\frac{\sum_{b=1}^{b}(\bar{\theta}_b - \bar{\theta})^2}{(b-1)b}} \tag{2.47}$$

If MCSE is small enough, this would mean that the posterior converges. On the other hand, higher MCSE indicates the need of longer chain size. As a rule of thumb, the MCMC process should continue up until MCSE is less than $\frac{1}{20}$ of the estimated posterior standard deviation for the parameter of interest [10].

Effective Sample Size (ESS): It is a measure to see the number of independent samples in an MCMC process according to the autocorrelation. ESS can be calculated by the formula given in Equation 2.48 [8] [66].

$$ESS = N_{eff} = \frac{N}{1 + 2\sum_{i=1}^{\infty}\rho_i(\theta)} \tag{2.48}$$

where N is the total iteration (sample) size, $\rho_i(\theta)$ is the autocorrelation for the parameter of interest at lag i. Lower ESS indicates more dependent samples and high autocorrelation, thus a slow convergence and mixing. In order to increase ESS, the methods for reducing autocorrelation can be applied and the size of the chain should be increased.

Accuracy Diagnostics Examining the accuracy of desired posterior parameters are as important as monitoring the stationarity and convergence of the MCMC process. There are a variety of diagnostics for this purpose, although these diagnostics not only examine the accuracy, but also they can monitor the convergence and stationarity as well.

Geweke Diagnostics: It is an MCMC diagnostic that helps to detect the stationarity and provides valuable insight to detect the size of the burn-in part of the chain. Geweke (1991) proposed this diagnostic as identifying the stationary of the Markov chain [23]. After discarding the burn-in part, the iterations are divided in to two parts, early half A and late half B. After taking the means of the parameter values ($\bar{\theta}_A$, $\bar{\theta}_B$) in these parts, Geweke diagnostic actually uses a well-known unpaired Z-test for two pairs [43]. However, Casella and Robert (2004) states that the Geweke diagnostic in R software uses a t-test rather than a Z-test [66]. The equations and test statistics are stated in both Lesaffre et. al (2012) and Casella and Robert (2004). Having said this, eventually the process is the same. The significance of Geweke statistic indicates stationarity, otherwise it implies that the burn-in part or the whole size of the chain is short [43]. The same diagnostic can be applied by slicing the iterations into k parts and testing it with pairs. This method is called "dynamic Geweke" diagnostic [43].

Raftery-Lewis Diagnostic: When the posterior distribution is skewed, posterior median is preferable than the posterior mean. This diagnostic is for ensuring the accuracy of the quantiles with a probability [43]. Raftery and Lewis (1992) proposed this diagnostic in two steps that in the first step it calculates the burn-in size and necessary size of the iterations in order to get the accurate convergence of the quantile with a probability, in the second step it calculates the minimum number of iterations to calculate the accurate quantile [60]. If the quantile is denoted by U and at the first step the diagnostic calculates the minimum burn-in size M, the necessary chain size N for $P(U < u|data) = q$ where q is the desired probability for accurate quantile. At the second step the diagnostic calculates the minimum chain size N_{min} and a ratio $I = \dfrac{M+N}{N_{min}}$ called "dependence factor". I is expected to be close to 1. According to Raftery and Lewis, if I is more than 5, than the chain implementation is problematic because of high autocorrelation or bad starting values [60].

The Brooks, Gelman,and Rubin (BGR) Diagnostic: This diagnostic was proposed for detecting a problem of being stuck around a local mode in a chain, a convergence problem, occurring especially when the chain is single and the posterior is multi-modal [43]. In order to avoid this convergence problem, at least two chains are

applied with different starting values. This diagnostic compares the chains with different starting values within themselves. Gelman and Rubin (1992) proposed a convergence diagnostic at first and then Brooks and Gelman (1998) have extended this diagnostic in multivariate cases [21] [9]. When different starting values are assigned to at least two chains, the posterior means by chains and an overall mean are calculated in order to calculate the with-in and between chain variability like in ANOVA analysis. Let $\bar{\theta}_c = \frac{1}{N}\sum_{i=1}^{N}\theta_{ci}$ is the posterior mean of the c^{th} chain, $\bar{\theta} = \frac{1}{C}\sum_{c=1}^{C}\bar{\theta}_c$ is the overall posterior mean and $S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\theta_{ci} - \bar{\theta}_c)^2$ is the posterior variance of the c^{th} chain. In Equation 2.49 calculations of within and between-chain variances are given [20] [43].

$$\begin{aligned} Between-chain\ Variability \quad B &= \frac{N}{C-1}\sum_{c=1}^{C}(\bar{\theta}_c - \bar{\theta})^2 \\ Within-chain\ Variability \quad W &= \frac{1}{C}\sum_{c=1}^{C}S_c^2 \end{aligned} \tag{2.49}$$

Based on these equations, Gelman and Rubin (1995) and later Lesaffre and Lawson (2012) gave the unbiased estimator (under stationarity condition) of posterior variance as a weighted average of W and B in Equation 2.50.

$$\widehat{V} = \widehat{var}(\theta|X) = \frac{N-1}{N}W + \frac{1}{N}B \tag{2.50}$$

$\hat{V}$ is said to be overestimated when the bad starting values occur and mixing of the chains are slow [20] [43]. Also as it can be seen in Equation 2.50, when $n \to \infty$, $var(\theta|X)$ approaches to W. It means that if the chain length is long enough, the posterior converges, and all the chains are mixed and their posterior parameters will be the same. Therefore, Gelman and Rubin (1992) proposed a ratio called $\widehat{R}$, and it is given in Equation 2.51 [20].

$$\widehat{R} = \frac{\widehat{V}}{W} \tag{2.51}$$

$\widehat{R}$ is called as the estimated potential scale reduction factor (PSRF) [43]. It should approach to 1 when $N \to \infty$, which means the chain size is long enough, the posterior converges and all the chains are mixed well. The MCMC process is acceptable when $\widehat{R}$ is less than 1.1 [20]. Lesaffre and Lawson (2012) also indicates a corrected version of $\widehat{R}$ for taking into the sampling variability of the variance estimates as $\widehat{R}_c = \left(\frac{\widehat{d}+3}{\widehat{d}+3}\right)R$ where $\widehat{d} = \frac{2\widehat{V}}{\widehat{var}(\widehat{V})}$. This time the process is acceptable when

$\widehat{R}_c$ is less than 1.1 or 1.2 [43]. The dynamic and graphical version of BGR Diagnostic is proposed by Brooks and Gelman (1998). It is basically dividing the chains in batches indexed by b for all chains and calculating $\widehat{R}_b$ values for every batch. This provides graphing the $\widehat{R}_b$ values and it is possible to monitor how and when $\widehat{R}$ values converges around 1 [9]. In Figure 2.5 an example of BGR dignostic graph is given.

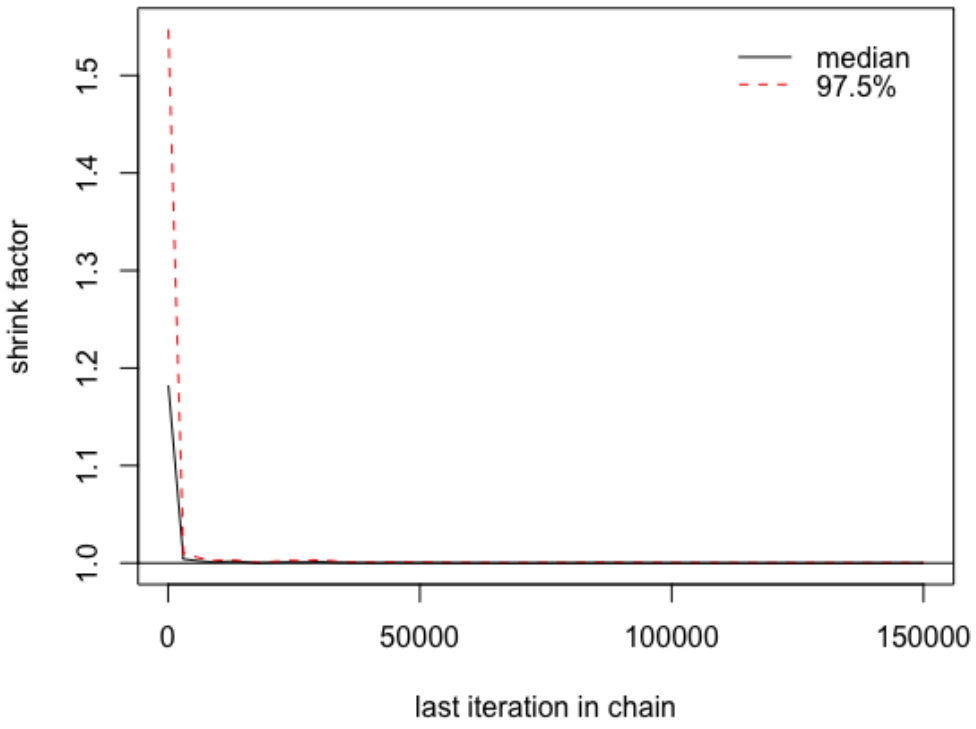

Figure 2.5: BGR Diagnostic plot example

It can be seen that the shrink factor $\widehat{R}$ declines around 1 immediately after the iteration starts and it declines to 1 totally before 50000^{th} iteration. This indicates well mixed chains and convergence after the 50000^{th} iteration.

MCMC diagnostics are the tools for monitoring how good is the process about the convergence of the posterior, accuracy of the posterior parameter estimates, etc. There are also some methods for accelerating the convergence of the posterior.

Acceleration of The Convergence Lesaffre and Lawson (2012) indicate that although there are a lot of methods to accelerate, but none of them guarantees the acceleration in all circumstances [43].

The first method is choosing better starting values. If the starting values can be chosen close to area where the posterior probability is higher, the convergence will be faster.

Another method is the transforming the data. When there is high multicollinearity among the variables in regression, the convergence is slow [43]. In order to solve the multicollinearity problem, transformation methods can be used to accelerate the convergence.

The other method for acceleration is called thinning. Autocorrelation causes slow convergence as it is stated in section 2.4.3, thus lowering the autocorrelation provides more independent sampling and accelerated convergence. Thinning is the solution for this problem. It is basically picking every m^{th} iteration in the chain. For example, with 1000 iterations in a chain, picking every 5^{th} iteration and getting a sample size of 200, is a thinning process with size 5. It lowers the autocorrelation but also reduces MCSE since the size of the chain decreases. If the chain size is long enough to get acceptable MCSE even by thinning, it is a useful acceleration method [43]. It may cause to monitor high acceptance rate in Metropolis-Hastings Algorithm since thinned samples include more accepted samples than its original chain.

Blocking is yet another method for accelerating the convergence. This method is applied by taking the parameters of interest in blocks and block by block sampling, through which convergence can be accelerated. When sampling of parameters is done from similar densities, they can be gathered as a block. For example, in normal linear regression the conditional posterior of the coefficients is a multivariate normal. So, instead of sampling from normal distribution one by one, all the coefficients can be sampled from a multivariate normal distribution. This may accelerate the convergence [43].

Reparameterization of the parameters can also accelerate the convergence. For example a logarithmic transformation of the parameters can simplify the sampling and it may provide an avoidance on generating unrealistic sample values and thus it can

accelerate the convergence [43].

In this chapter, literature review and preliminaries for Generalized linear models (GLM), Missing data in GLM and Bayesian missing data analysis in GLM are provided. In the next chapter, the methodology of the thesis based on this knowledge is given in detail.

CHAPTER 3

METHODOLOGY

In this chapter, the methodology of hybrid Bayesian modeling system called Bayesian Model Pooling (BMP) for GLM with MNAR covariates is explained in detail. The reason of calling this method as "hybrid" can be basically explained as this methodology is interpreted as a combination of RJMCMC and BMA methods. A brief summary of this hybrid Bayesian modeling system can be helpful for better comprehension. Firstly, a model space containing joint models of the main model, missing covariates and the candidate missingness models is constructed in order to build an environment for RJMCMC process. Then, model transition properties are defined confirming to the dimensions of the models (i.e. number of parameters), and transition probabilities among models are assessed accordingly. These transition probabilities are used in RJMCMC algorithm. During the RJMCMC process, parameters are estimated in each iteration for a different candidate model. RJMCMC algorithm should continue until the joint posterior converges which can be confirmed according to MCMC Diagnostics. After RJMCMC process is finished, model posteriors are taken as model weights, and posterior estimates of the parameters for every model in the model space are pooled by using the model weights with BMA methods. Details are given in the following sections.

3.1 The Structure of The Joint Model in a Model Space

Before construction of the model space, the joint models in the model space should be defined properly. Let Y_i for $i = 1, \ldots, n$ be the binary response variable, let $\boldsymbol{X}_i = (X_{1,i}, X_{2,i}, \ldots, X_{p,i})$ for $i = 1, \ldots, n$ be the covariate matrix of the main

model. The covariate matrix can be divided into two parts; the ones being fully observed and the ones containing MNAR mechanism. Let $\boldsymbol{X}_i = (\boldsymbol{X^{miss}}_i, \boldsymbol{X^{obs}}_i)$ where $\boldsymbol{X}_i^{miss} = (X_{1,i}, X_{2,i}, \ldots, X_{r,i})$. Accordingly, let $\boldsymbol{R}_i = (R_{1,i}, R_{1,i}, \ldots, R_{r,i})$ for $i = 1, \ldots, n$ be the binary missingness indicator matrix generated according to the covariates containing MNAR mechanism. Generation of $\boldsymbol{R}_i$ is defined in Equation 2.5. The main model is assumed to be a logistic model, while the missingness models are assumed as probit models. Regarding this, $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_p\}$ is the parameter vector for the main model, $\boldsymbol{\phi} = \{\phi_1, \ldots, \phi_v\}$ is the parameter vector for the missingness model. Lastly, let $\boldsymbol{M} = \{M_1, \ldots, M_{n_m}\}$ be the vector of model indices for the model space.

The joint posterior distribution of $(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}})$ is then given as following;

$$\begin{aligned} P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}} | \boldsymbol{X^{obs}}, Y, \boldsymbol{R}) &= \frac{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, \boldsymbol{R})}{P(\boldsymbol{X^{obs}}, Y, \boldsymbol{R})} \\ &\propto P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, \boldsymbol{R}) \\ &= P(\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, \boldsymbol{R} | M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}) P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}) \end{aligned} \tag{3.1}$$

where M_k is the k^{th} model index, $\boldsymbol{\phi_k}$ is the parameter vector of k^{th} missingness model, $\boldsymbol{\beta_k}$ is the the parameter vector of the main model under k^{th} model index, $P(\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, \boldsymbol{R} | M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k})$ is the full data likelihood, and $P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k})$ is the joint prior distribution.

Further, the full data likelihood, $P(\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, \boldsymbol{R} | M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k})$, can be factorized by using selection model and it is given by,

$$\begin{aligned} P(\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, \boldsymbol{R} | M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}) =& P(\boldsymbol{R} | \boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, M_k, \boldsymbol{\phi_k}) \\ & * P(Y | \boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, M_k, \boldsymbol{\beta_k}) \\ & * P(\boldsymbol{X^{miss}} | \boldsymbol{X^{obs}}, M_k) \end{aligned} \tag{3.2}$$

where $P(\boldsymbol{R} | \boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, M_k, \boldsymbol{\phi_k})$ is the likelihood function of the missingness indicator vector $\boldsymbol{R}$ that belongs to k^{th} model, $P(Y | \boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, M_k, \boldsymbol{\beta_k})$ is the likelihood function of main model given the model M_k, and $P(\boldsymbol{X^{miss}} | \boldsymbol{X^{obs}}, M_k)$ is the conditional distribution for missing covariates $\boldsymbol{X^{miss}}$.

Further, the joint prior in Equation 3.1 can now be factorized as in Equation 3.3.

$$P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}) = P(\boldsymbol{\beta_k} | M_k) P(\boldsymbol{\phi_k} | M_k) P(M_k) \tag{3.3}$$

where $P(\boldsymbol{\beta_k}|M_k)$ is the prior for the parameter vector of main model under k^{th} model, $P(\boldsymbol{\phi_k}|M_k)$ is the prior for the parameter vector of k^{th} missingness model and $P(M_k)$ is the priors for the k^{th} model.

The likelihood of the main model is $P(Y|\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, M_k, \boldsymbol{\beta_k})$ and the one for missingness mechanism models is $P(\boldsymbol{R}|\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, M_k, \boldsymbol{\phi_k})$. Both models are considered as GLM models with binary responses. Thus, their likelihoods can be written explicitly as in Equations 3.4 and 3.6.

$$P(Y = y_i|\boldsymbol{X}^{obs}, \boldsymbol{X}^{miss}, M_k, \boldsymbol{\beta_k}) = \prod_{i=1}^{n} {p_{y_i}}^{y_i}(1 - p_{y_i})^{1-y_i} \tag{3.4}$$

where $y_i = 0, 1$ and $p_{y_i} = g(\boldsymbol{X}_i^T \boldsymbol{\beta_k})$ is the logistic link function with $\boldsymbol{X}_i$ being the covariate vector for i^{th} observation. Also,

$$P(R_r = r_{r_i}|\boldsymbol{X}^{obs}, \boldsymbol{X}^{miss}, Y, M_k, \boldsymbol{\phi_k}) = \prod_{i=1}^{n} {p_{r_{r_i,k}}}^{r_{ri}}(1 - p_{r_{r_i,k}})^{1-r_{ri}} \tag{3.5}$$

where $r_{r_i} = 0, 1$ and $p_{r_{r_i,k}} = g(\boldsymbol{W}_{i,k}^T \boldsymbol{\phi_k})$ is the probit link function where $\boldsymbol{W}_{i,k}$ is the covariate vector for k^{th} missingness model. At this point, it should be emphasized that $\boldsymbol{W}_k$ must contain the covariate vector possessing the MNAR mechanism, because MNAR mechanism depends on the covariate itself. It may also contain other covariates as well. If there are r covariate with MNAR, then the full conditional likelihood for the missingness indicator vector $\boldsymbol{R}$ can be factorized as in Equation 3.6.

$$\begin{aligned} P(\boldsymbol{R}|\boldsymbol{X}^{obs}, \boldsymbol{X}^{miss}, Y, M_k, \boldsymbol{\phi_k}) =& P(R_1|\boldsymbol{X}^{obs}, \boldsymbol{X}^{miss}, Y, M_k, \boldsymbol{\phi_k}) \\ & * P(R_2|\boldsymbol{X}^{obs}, \boldsymbol{X}^{miss}, Y, M_k, \boldsymbol{\phi_k}) \\ & \vdots \\ & * P(R_r|\boldsymbol{X}^{obs}, \boldsymbol{X}^{miss}, Y, M_k, \boldsymbol{\phi_k}) \\ =& \prod_{j=1}^{r} \prod_{i=1}^{n} {p_{r_{j_i,k}}}^{r_{j_i}}(1 - p_{r_{j_i,k}})^{1-r_{j_i}} \end{aligned} \tag{3.6}$$

Moreover, the last part of the full data likelihood in the Equation 3.2 is the conditional distribution of the missing covariates $\boldsymbol{X}^{miss} = (X_1^{miss}, \ldots, X_r^{miss})$ given the

observed covariates $\boldsymbol{X}^{obs} = (X_1^{obs}, \ldots, X_r^{obs})$ under the k^{th} missingness model M_k. The explicit form of $P(\boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, M_k)$ is given in Equation 3.7.

$$\begin{aligned} P(\boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, M_k) =& P(X_1^{miss}|X_1^{obs}, M_k) \\ & * P(X_2^{miss}|X_2^{obs}, M_k) \\ & \vdots \\ & * P(X_r^{miss}|X_r^{obs}, M_k) \end{aligned} \tag{3.7}$$

Every missing value belonging to any missing covariate can be considered as an unknown parameter, thus, every one of them also has a prior distribution. Flat distributions, conjugate distributions, or distributions which fit to the observed part of the covariate with MNAR can be chosen as priors. If the prior distribution of j^{th} missing covariate is $P(X_j^{miss}|X_j^{obs}, M_k)$ then,

$$P(X_j^{miss}|X_j^{obs}, M_k) = \prod_{i=1}^{n_j^{miss}} f(x_{ji}, \boldsymbol{\theta_{x_j}}) \tag{3.8}$$

where $f(x_{ji}, \boldsymbol{\theta_{x_j}})$ is the prior distribution for the missing value x_{ji}, n_j^{miss} is the number of the missing values in the covariate X_j, and $\boldsymbol{\theta_{x_j}}$ is the parameter vector for missing covariate X_j. If the Equation 3.7 is rewritten byh using the expression in the Equation 3.8, then the prior distribution for all the missing cases can be obtained as,

$$P(\boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, M_k) = \prod_{i=1}^{n_1^{miss}} \cdots \prod_{i=1}^{n_r^{miss}} f(x_{1i}, \boldsymbol{\theta_{x_1}}) \cdots f(x_{ri}, \boldsymbol{\theta_{x_r}}) \tag{3.9}$$

where the total number of the missing values is $n^{miss} = n_1^{miss} + \cdots + n_r^{miss}$.

After explaining the full data likelihood, every element of the joint prior in the Equation 3.3 should be interpreted in detail. First of all, prior distributions for the parameter vector of the main model $\boldsymbol{\beta}$ given the missingness model M_k is stated in Equation 3.10.

$$\begin{aligned} P(\boldsymbol{\beta_k}|M_k) =& P(\beta_{k0}|M_k) P(\beta_{k,2}|M_k) \cdots P(\beta_{k,p}|M_k) \\ =& \prod_{i=1}^{p} f(\beta_{k,i}, \boldsymbol{\theta_{\beta_{k,i}}}) \end{aligned} \tag{3.10}$$

where $f(\beta_{k,i}, \boldsymbol{\theta_{\beta_{k,i}}})$ is the prior for the i^{th} β coefficient under k^{th} missingness model.

Secondly, prior distributions for the parameter vector of the missingness models $\boldsymbol{\phi_k}$

given the missingness model M_k is stated in Equation 3.11.

$$\begin{aligned} P(\boldsymbol{\phi_k}|M_k) =& P(\phi_{k,0}|M_k)P(\phi_{k,2}|M_k)\cdots P(\phi_{k,v}|M_k) \\ =& \prod_{i=1}^{v} f(\phi_{k,i}, \boldsymbol{\theta_{\phi_{k,i}}}) \end{aligned} \tag{3.11}$$

where $f(\phi_{k,i}, \boldsymbol{\theta_{\phi_{k,i}}})$ is the prior for the i^{th} ϕ coefficient of k^{th} missingness model.

As the last part of the joint prior in Equation 3.3, prior distribution for M_k, which is the probability of choosing the k^{th} model, can be non-informative equally likely, or some models can be highly likely than the others if there is an expert or pre-knowledge present about these models. If the model probabilities are assumed to be equally likely, then the probability of the k^{th} model is given as,

$$P(M_k) = \frac{1}{n_m} \tag{3.12}$$

where n_m is the size of the model space.

After the structure of the joint model is explained in detail, the essentials of RJMCMC approach for estimation of posteriors for parameters of interest and how the joint posterior is used in RJMCMC are provided in the following section.

3.2 RJMCMC For Modeling

The details of RJMCMC method are already given in Section 2.4.2. In this study, it is proposed that the inclusion of RJMCMC method and it's model selection features can be useful for obtaining the parameter posteriors for each model in a model space. As a reminder, RJMCMC is the generalized version of the Metropolis-Hastings method used when the dimension of the models is varied. Since the computation of the joint posterior, $P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})$, is challenging, and thus some numerical methods like MCMC are needed, RJMCMC is very suitable even when the model dimensions are varied.

With the intent of using the model selection features of RJMCMC, the algorithm runs as follows: While the process is at an initial model M_k for iteration t, a randomly chosen candidate model $M_{k'}$ in the model space is whether accepted with an acceptance

probability, given in Equation 3.13, or not, thus making the process stay at the initial model for the next iteration.

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{P(M_{k'}, \boldsymbol{\phi_{k'}}, \boldsymbol{\beta_{k'}}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})q(M_{k'}, M_k)}{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})q(M_k, M_{k'})}\right\} \quad (3.13)$$

where $P(M_{k'}, \boldsymbol{\phi_{k'}}, \boldsymbol{\beta_{k'}}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})$ is the joint posterior of the candidate model, $P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})$ is the joint posterior of the initial model, $q(M_{k'}, M_k)$ is the transition probability from the candidate model $M_{k'}$ to the initial model M_k and $q(M_k, M_{k'})$ is the transition probability from the initial model M_k to the candidate model $M_{k'}$.

The acceptance probability is determined according to the move types, specified in Section 2.4.2, since the dimension difference of the models in comparison should be taken into account and some additional terms should be used in order to equalize the dimensions.

Moving to a higher dimension: If the dimension of the candidate model $M_{k'}$ is higher than the initial model M_k ($|k'| > |k|$), the acceptance probability for the new state is given as follows;

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{P(M_{k'}, \boldsymbol{\phi_{k'}}, \boldsymbol{\beta_{k'}}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})q(M_{k'}, M_k)}{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})q(M_k, M_{k'})h(u)} \times \mid J_u \mid\right\} \quad (3.14)$$

Moving to a lower dimension: If the dimension of the candidate model $M_{k'}$ is lower than the initial model M_k ($|k'| < |k|$), the acceptance probability for the new state is given as follows;

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{P(M_{k'}, \boldsymbol{\phi_{k'}}, \boldsymbol{\beta_{k'}}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})q(M_{k'}, M_k)h(u')}{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})q(M_k, M_{k'})} \times \mid J_{u'} \mid\right\} \quad (3.15)$$

Moving to an equal dimension: If the dimension of the candidate model $M_{k'}$ is equal to the initial model M_k ($|k'| = |k|$), the acceptance probability for the new state is given as follows;

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{P(M_{k'}, \boldsymbol{\phi_{k'}}, \boldsymbol{\beta_{k'}}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})q(M_{k'}, M_k)}{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})q(M_k, M_{k'})}\right\} \quad (3.16)$$

The determinants of Jacobians in Equations 3.14 and 3.15 are all equal to 1, since the RJMCMC process is applied on a model space and the dimensions of models

are defined as the number of the covariates that are independent. For example, when the move type is a higher move, $\boldsymbol{\beta_{k'}} = g(\boldsymbol{\beta_k}, \mathbf{u})$, then $dim(\boldsymbol{\beta_{k'}}) = dim(\boldsymbol{\beta_k}) + dim(u)$. Since dimension of models indicates the number of coefficients in this case, it can be considered as adding extra coefficients of the independent covariates as many as $|dim(\boldsymbol{\beta_{k'}}) - dim(\boldsymbol{\beta_k})|$. The determinant of the Jacobian $\mid J_u \mid$ is presented by $\left|\dfrac{\partial g(\boldsymbol{\beta_k}, \mathbf{u})}{(\boldsymbol{\beta_k}, \mathbf{u})}\right|$ and written as in Equation 3.17.

$$\left|\frac{\partial g(\boldsymbol{\beta_k}, \mathbf{u})}{(\boldsymbol{\beta_k}, \mathbf{u})}\right| = \begin{vmatrix} \frac{\partial \beta_0}{\partial \beta_0} & \cdots & \frac{\partial \beta_{n_\beta}}{\partial \beta_0} & \frac{\partial u_1}{\partial \beta_0} & \cdots & \frac{\partial u_n}{\partial \beta_0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \beta_0}{\partial \beta_{n_\beta}} & \cdots & \frac{\partial \beta_{n_\beta}}{\partial \beta_{n_\beta}} & \frac{\partial u_1}{\partial \beta_{n_\beta}} & \cdots & \frac{\partial u_n}{\partial \beta_{n_\beta}} \\ \frac{\partial \beta_0}{\partial u_1} & \cdots & \frac{\partial \beta_{n_\beta}}{\partial u_1} & \frac{\partial u_1}{\partial u_1} & \cdots & \frac{\partial u_n}{\partial u_1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \beta_0}{\partial u_n} & \cdots & \frac{\partial \beta_{n_\beta}}{\partial u_n} & \frac{\partial u_1}{\partial u_n} & \cdots & \frac{\partial u_n}{\partial u_n} \end{vmatrix} = \begin{vmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 1 \end{vmatrix} = 1 \quad (3.17)$$

Lastly, u and u', stated in Equations 3.14 and 3.15, stand for balancing the dimension of the models. Since the variation in dimension is only in missingness models in this case, $h(u)$ and $h(u')$ are chosen exactly the same as the prior chosen for ϕ. Let $u = (u_1, \dots, u_d)$ and $u' = (u'_1, \dots, u'_d)$ be the vectors for additional terms, $h(u)$ and $h(u')$ are given in Equation 3.18.

$$\begin{aligned} h(u) = h(u_1, \dots, u_d) = h(u_1) \dots h(u_d) = [P(\boldsymbol{\phi_k}|M_k)]^d \\ h(u') = h(u'_1, \dots, u'_d) = h(u'_1) \dots h(u'_d) = [P(\boldsymbol{\phi_k}|M_k)]^d \end{aligned} \quad (3.18)$$

After defining the acceptance probabilities of RJMCMC for different move types, the construction of a model space and assessing method for transition probabilities of models in the model space should be explained in detail.

3.3 Construction of A Model Space

The first step for applying RJMCMC is to identify the model space along with model transitions with different dimensions. Here, the model space refers to the missingness model space. Models in the model space differ according to variations of missingness models. Therefore, let us denote the model space of the r^{th} missingness indicator for

the r^{th} covariate with MNAR as $\mathbb{M}_r = (M_{r_1}, M_{r_2}, ..., M_{r_{n_{m_r}}})$ with a size of n_{m_r}. $\mathbb{M}_r$ can be constructed manually with expert knowledge on the variable with MNAR. However, most of the time it may not be possible to find the most suitable candidate models manually through expert knowledge. If there is a lack of expert knowledge, one should always search for an alternative method for choosing. In this case, the model space is constructed by considering particularly the most suitable candidates for "missingness models". One of the most logical ways to construct a model space is by using the subsets of the variable set of the main model including covariates ($\boldsymbol{X}$), response (Y) and/or interactions ($\boldsymbol{T}$). Let $\mathbb{V} = \{\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, \boldsymbol{T^{miss}}, \boldsymbol{T^{obs}}\}$ be the variable set in which $\boldsymbol{X^{miss}}$ is the covariate vector of variables with MNAR, $\boldsymbol{X^{obs}}$ is the variable vector of fully observed variables, $\boldsymbol{T^{miss}}$ is the vector of variable interactions, including variables with MNAR, $\boldsymbol{T^{obs}}$ is the vector of variable interactions, not including variables with MNAR. Let the size of $\mathbb{V}$ be n_v, $2^{n_v} - 1$ subsets are to be found excluding the null set. There are other subsets that should also be excluded from the model space. Since MNAR mechanism depends on the missing variable itself, all the subsets must include the missing variable of interest or it's interaction. That is to say, if a model space is constructed for the r^{th} missingness indicator, R_r is the missingness indicator and, let $\mathbb{S}_r$ be the set of subsets of $\mathbb{V}$ where $\mathbb{S}_r = \bigcup_{k=1}^{n_m} \mathbb{S}_{r,k}$ and $\mathbb{S}_{r,k} \subseteq \mathbb{V}$ for all k.

$$\begin{aligned}\mathbb{S}_{r,k} =& \{(\mathbb{S}_{r1,k}, \mathbb{S}_{r2,k}) \in \mathbb{S}_{r,k} \mid \mathbb{S}_{r1,k} \subseteq (X_r \cup \boldsymbol{T_{X_r}}) \\ & ; \mathbb{S}_{r2,k} = \emptyset \vee \mathbb{S}_{r2,k} \subseteq (\boldsymbol{X_{-r}} \cup Y \cup \boldsymbol{T_{-X_r}}) \\ & ; \mathbb{S}_{r1,k} \cup \mathbb{S}_{r2,k} = \mathbb{S}_{r,k}\}\end{aligned} \tag{3.19}$$

Here, $\mathbb{S}_{r1,k}$ is the part of the k^{th} subset that includes the r^{th} covariate with MNAR (X_r) or the interaction of it such as X_rX_i, X_rY or X_r^2. $\mathbb{S}_{r2,k}$ is the remaining part of the subset, $\boldsymbol{T_{X_r}}$ is the vector of interactions of X_r, $\boldsymbol{X_{-r}}$ is the vector of covariates excluding X_r, $\boldsymbol{T_{-X_r}}$ is the vector of interaction terms excluding the ones containing X_r. Therefore, $\mathbb{S}_{r,k}$ is actually the covariate vector of the k^{th} model for the missingness mechanism R_r. Hence, the missingness model, given in Equation 3.6 would be equal to the expression given in Equation 3.20.

$$P(R_r = r_{r_i} | \boldsymbol{X}^{obs}, \boldsymbol{X}^{miss}, Y, M_{r_k}, \boldsymbol{\phi_k}) = \prod_{i=1}^{n} g(\mathbf{S}_{r,k}^T \boldsymbol{\phi_k})^{r_{r_i}} (1 - g(\mathbf{S}_{r,k}^T \boldsymbol{\phi_k}))^{1-r_{r_i}} \tag{3.20}$$

where $\mathbf{S}_{r,k}$ is the vector form of subset $\mathbb{S}_{r,k}$.

All in all, there should be n_r model spaces with different sizes for modeling n_r missingness indicators. All of these sets are given in Equation 3.21.

$$\begin{aligned}
\mathbb{S}_1 &= \{\mathbb{S}_{1,1} \cup \mathbb{S}_{1,2} \cup ... \cup \mathbb{S}_{1,n_{m_1}}\} = \bigcup_{k=1}^{n_{m_1}} \mathbb{S}_{1,k} \\
\mathbb{S}_2 &= \{\mathbb{S}_{2,1} \cup \mathbb{S}_{2,2} \cup ... \cup \mathbb{S}_{2,n_{m_2}}\} = \bigcup_{k=1}^{n_{m_2}} \mathbb{S}_{2,k} \\
&\vdots \\
\mathbb{S}_r &= \{\mathbb{S}_{r,1} \cup \mathbb{S}_{r,2} \cup ... \cup \mathbb{S}_{r,n_{m_r}}\} = \bigcup_{k=1}^{n_{m_r}} \mathbb{S}_{r,k} \\
&\vdots \\
\mathbb{S}_{n_r} &= \{\mathbb{S}_{n_r,1} \cup \mathbb{S}_{n_r,2} \cup ... \cup \mathbb{S}_{n_r,n_{m_{n_r}}}\} = \bigcup_{k=1}^{n_{m_{n_r}}} \mathbb{S}_{n_r,k}
\end{aligned} \tag{3.21}$$

After constructing the subsets $\mathbb{S}$, the model spaces $\mathbb{M}$, can now be constructed by using these subsets. Model space for the r^{th} missingness indicator is denoted as $\mathbb{M}_r$ and it can be constructed by choosing the best candidate models, of which the variable sets belong to the subsets of $\mathbb{S}_r$ in Equation 3.21.

Some methods for filtering the model space (i.e. choosing the best candidate models for the model space) are already mentioned in the previous sections such as Occam's window, Occam's razor, etc. The adaptation of the Occam's window for the model space of r^{th} missingness indicator is given in Equation 3.22 by using the Equation 2.38.

$$\mathbb{M}_r = \left\{ M_{r_k} : \frac{max_l\{P(M_{r_l}, |\boldsymbol{D_r}, \boldsymbol{\phi_{r_l}}, \boldsymbol{\beta_{r_l}})\}}{P(M_{r_k}|\boldsymbol{D_r}, \boldsymbol{\phi_{r_k}}, \boldsymbol{\beta_{r_k}}, \boldsymbol{X^{miss}})} \leq C \right\} \tag{3.22}$$

where $max_l\{P(M_{r_l}, |\boldsymbol{D_r}, \boldsymbol{\phi_{r_l}}, \boldsymbol{\beta_{r_l}})\}$ is the model posterior with highest probability, $P(M_{r_k}|\boldsymbol{D_r}, \boldsymbol{\phi_{r_k}}, \boldsymbol{\beta_{r_k}}, \boldsymbol{X^{miss}})$ is the model posterior of the k^{th} model in the model space and C is the arbitrarily chosen threshold value for the ratio.

Estimation of model posteriors are obtained by using the Equation 2.39 in Section 2.4.2. It is to be seen that, the model posterior contains unknown model parameters $\boldsymbol{\phi_{r_l}}$, $\boldsymbol{\beta_{r_l}}$ and missing values $\boldsymbol{X^{miss}}$ (can also be considered as unknown parameters) which should be imputed by using full conditional distributions of these unknown parameters with proper sampling methods. When the size of unknown parameters

are large, the posterior becomes very sensitive to these imputations. Therefore, the model posterior with the highest probability can easily be affected by imputation. In order to reduce this sensitivity, the classic Occam's Window procedure is adjusted and this alternative method called "T-step Occam's Window" procedure is proposed in the following section.

3.3.1 T-Step Occam's Window Method

The alternative method proposed in this thesis for reducing the size of the model space by filtering is based on the Occam's Window method. Since high sensitivity of unknown parameters in imputation, and the model posterior with the highest probability can differ easily, first step to take in this alternative method is to keep the threshold value C small, so that the model space would be large. Afterwards, the model space of the r^{th} missingness indicator can be created with the following steps:

Step 1: Impute unknown parameters with a suitable imputation method in order to compute the model posterior in the next step.

Step 2: Calculate all the model posteriors in the model space.

Step 3: Apply the Occam's window process.

Step 4: Repeat the Occam's Window process (from Step 1 to Step 3) by t times and obtain t model spaces with different sizes.

Step 5: Construct a frequency table for model spaces altogether in order to identify the common models occurring in model spaces.

Step 6: Choose the s most frequent models as candidate models among these t model spaces.

Accordingly, these t model spaces with different sizes for the r^{th} missingness indica-

tor are given by,

$$\begin{aligned}
\mathbb{M}_r^{(1)} &= \left\{ M_{r_k}^{(1)} : \frac{max_l\{P(M_{r_l}, |\boldsymbol{D_r}, \boldsymbol{\phi_{r_l}}, \boldsymbol{\beta_{r_l}})\}}{P(M_{r_k}|\boldsymbol{D_r}, \boldsymbol{\phi_{r_k}}, \boldsymbol{\beta_{r_k}}, \boldsymbol{X^{miss}})} \leq C \right\} = \{M_{r_1}^{(1)}, M_{r_2}^{(1)}, ..., M_{r_{n_1}}^{(1)}\} \\
\mathbb{M}_r^{(2)} &= \left\{ M_{r_k}^{(2)} : \frac{max_l\{P(M_{r_l}, |\boldsymbol{D_r}, \boldsymbol{\phi_{r_l}}, \boldsymbol{\beta_{r_l}})\}}{P(M_{r_k}|\boldsymbol{D_r}, \boldsymbol{\phi_{r_k}}, \boldsymbol{\beta_{r_k}}, \boldsymbol{X^{miss}})} \leq C \right\} = \{M_{r_1}^{(2)}, M_{r_2}^{(2)}, ..., M_{r_{n_2}}^{(2)}\} \\
&\vdots \\
\mathbb{M}_r^{(t)} &= \left\{ M_{r_k}^{(t)} : \frac{max_l\{P(M_{r_l}, |\boldsymbol{D_r}, \boldsymbol{\phi_{r_l}}, \boldsymbol{\beta_{r_l}})\}}{P(M_{r_k}|\boldsymbol{D_r}, \boldsymbol{\phi_{r_k}}, \boldsymbol{\beta_{r_k}}, \boldsymbol{X^{miss}})} \leq C \right\} = \{M_{r_1}^{(t)}, M_{r_2}^{(t)}, ..., M_{r_{n_t}}^{(t)}\}
\end{aligned} \tag{3.23}$$

Naturally, a candidate model may be chosen in more than one model space, therefore, common models may occur among model spaces. The most frequent s common models among model spaces are chosen as the best candidate models in order to reduce the sensitivity of imputation for unknown parameters. The size of t can be decided arbitrarily, to the size of the unknown parameters and of course in relation to the strength of computational resources. For example, when $t = 100$, the common models with a frequency s greater than 10 can be chosen as the best candidate models. However, the size of this chosen model space should be optimal for RJMCMC and model pooling. If it is too small, the uncertainty about missingness models cannot be reduced or if the size of the model space is very large, RJMCMC process takes too much time computationally.

3.4 Model Transitions and Transition Probability Matrix

Model transition can be defined as jumping from a model to another one with a probability. In Metropolis-Hastings algorithm, transition is done in the same dimension, and in this case the dimension corresponds to the model dimension. Number of parameters of the model determines it's dimension. Since, model dimensions in a model space are mostly different, RJMCMC method is widely considered appropriate use as a model selection method among varying dimensions. In addition, as it is stated in Section 2.4.2, transitions from a model to another one should also be reversible. Furthermore, assumptions about transitions can highly affect the process of obtaining posterior distributions for each model. For example, if all the model transitions in a model space are assumed to be equally likely, then the effect of model dimensions is eliminated significantly. For a better comprehension, let the transition probability from model M_k to $M_{k'}$ be $q(M_k, M_{k'})$, given in Equation 3.13, and

the transition probability from model $M_{k'}$ to M_k be $q(M_{k'}, M_k)$. Then, if the transitions are assumed equally likely, the probabilities would be equal to each other (i.e. $q(M_{k'}, M_k) = q(M_k, M_{k'})$). So, the acceptance probability in Equation 3.13 is simplified as in Equation 3.24.

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{P(M_{k'}, \boldsymbol{\phi_{k'}}, \boldsymbol{\beta_{k'}}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})}{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})}\right\} \tag{3.24}$$

As it can be seen in Equation 3.24, RJMCMC acceptance probability depends only on the joint posterior values under initial and candidate models. This applies for the move types as well. The acceptance probabilities in Equations 3.14 and 3.15 depend on the joint posteriors, on the function which equalizes the dimensions, and on the Jacobian. Since assumption of equal transition reduces the effect of model dimensions, further assumptions taking the dimension difference into consideration are evaluated for this thesis. Transition for only one dimension up or down among models can be considered as well, however this restricts the mobility, and the transitions cannot be done from some models to other ones of which the dimension difference is greater than one. So taking into account all of these, transition assumptions, that takes both full mobility on the model space and also the effect of dimension difference into consideration, are shared below.

1. Every model is to be considered as the center of the model space and the probabilities are calculated for every model. Meaning that, n_{m_r} (size of the model space of r^{th} covariate with MNAR) transition probabilities are assigned to every single model, of which the total probability equals to 1. This allows assigning different probabilities to model transitions and their reversible transitions.

2. Transition ability of two models with the same dimension in a model space is assumed to be the highest. So, transition probabilities decrease as the absolute dimension difference increases between models.

Transition probabilities of a model space are shown in a matrix form with a dimensions $n_{m_r} \times n_{m_r}$ at Table 3.1.

Every row of the table is a vector of transition probabilities from an initial model to other models and the total probability of every row equals to 1. With this matrix

Table 3.1: Transition Probabilities of a Current State to Next State for Covariate X_r with MNAR

For Covariate X_r	M_1	M_2	...	M_{n_m}	Row Total
M_1	$q(M_1, M_1)$	$q(M_1, M_2)$	...	$q(M_1, M_{n_m})$	1
M_2	$q(M_2, M_1)$	$q(M_2, M_2)$	...	$q(M_2, M_{n_m})$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
M_{n_m}	$q(M_{n_m}, M_1)$	$q(M_{n_m}, M_2)$	...	$q(M_{n_m}, M_{n_m})$	1

setup, it is crucial to state that, the transition probability is not guaranteed to be equal to its reversible transition probability (i.e. $q(M_k, M_{k'}) \neq q(M_{k'}, M_k)$).

The second assumption is defining the impact of a dimension difference. Let $d(M_k)$ be the dimension of model M_k, $d(M_{k'})$ be the dimension of model $M_{k'}$, the absolute difference of two dimensions be defined as $\mid d(M_k) - d(M_{k'}) \mid = d_{kk'}$, and the largest absolute difference between dimensions be $max\{d_{kk'}\} = n_d$. The impacts of dimensions are given in Table 3.2.

Table 3.2: Impacts of Dimension Difference

$d_{kk'}$	**Impact Value ($i_{kk'}$)**
0	$n_d + 1$
1	n_d
$\vdots$	$\vdots$
n_d	1

These impact values can be utilized in order to calculate transition probabilities. In Table 3.3, the related dimension impacts are shown in a matrix form.

The transition probabilities are merely the weighted average of the row vector of that model. The formula is given in Equation 3.25.

$$q(M_k, M_{k'}) = \frac{i_{kk'}}{I_k} \tag{3.25}$$

Table 3.3: Dimension Differences of The Models for Covariate X_r with MNAR

For Covariate X_r	M_1	M_2	...	M_{n_m}	Row Total
M_1	i_{11}	i_{12}	...	i_{1n_m}	I_1
M_2	i_{21}	i_{22}	...	i_{2n_m}	I_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
M_{m_n}	i_{n_m1}	i_{n_m2}	...	$i_{n_mn_m}$	I_{n_m}

where $i_{kk'}$ is the impact value of model k to k', and I_k is the total impact of model k to other models.

3.5 Estimation of Parameters

The last part of the methodology is pooling the estimations of parameters by using the properties of BMA. When RJMCMC process is finished, parameter estimations are obtained under each model in the model space. Also counts of visits to every model are recorded each time the process visits a model during the iterations. Thus, the frequencies of visits for models at the end of the process can be considered as posterior model probabilities $P(M_k|D)$. Parameters of interest, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)$, belong to the main model. For recalling, the joint posterior used in RJMCMC is $P(\boldsymbol{M}, \boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})$. Here, it should be noted that parameters of interest $\boldsymbol{\beta}$ and parameters for the missingness indicators $\boldsymbol{\phi}$ are independent. Therefore, using only the joint posterior of the parameters of interest and models **M**, given as $P(\boldsymbol{\beta}, \boldsymbol{M}|\boldsymbol{D})$, is enough for estimation of β posteriors. As a remainder, $\boldsymbol{D} = (\boldsymbol{X^{miss}}, \boldsymbol{X^{obs}}, Y, \boldsymbol{R})$ is the data and $\boldsymbol{M} = (M_1, \ldots, M_{n_m})$ is the model space. Since all the β coefficients are independent, marginal posterior distributions of parameters of interest can be stated as in Equation 3.26.

$$P(\beta_0, \ldots, \beta_p|\boldsymbol{M}, \boldsymbol{D}) = P(\beta_0|\boldsymbol{M}, \boldsymbol{D}) \ldots P(\beta_p|\boldsymbol{M}, \boldsymbol{D}) \tag{3.26}$$

Let $\hat{\beta}_{ik}$ be the estimator of i^{th} β coefficient under k^{th} model, M_k. In this case, $\hat{\beta}_{ik}$ is calculated from the posterior distribution $P(\beta_i|M_k, \boldsymbol{D})$ which is obtained through

RJMCMC process. Maximum Likelihood Estimators (MLE) can be taken as estimators of β coefficients under every model. Besides, medians are more suitable choices as estimators if the posterior distribution is skewed. The BMA formula stated in 2.35 is adapted to this problem, and it is provided in Equation 3.27.

$$\begin{aligned}
\hat{\beta}_0 &= \sum_{k=1}^{n_m} \hat{\beta}_{0_k} P(M_k|D) \\
\hat{\beta}_1 &= \sum_{k=1}^{n_m} \hat{\beta}_{1_k} P(M_k|D) \\
&\vdots \\
\hat{\beta}_p &= \sum_{k=1}^{n_m} \hat{\beta}_{p_k} P(M_k|D)
\end{aligned} \tag{3.27}$$

where $P(M_k|D)$ is the posterior probability of the k^{th} model, calculated by using the visiting frequency of RJMCMC. At this point, let N be the iteration size and N_v be the size of the jump counts (total counts of visits). It is assumed that in N, there are iterations staying in a model for some amount of time, but in N_v only visits are counted. Additionally, f_{M_k} is the frequency of the visits of M_k in N_v. A detail should be mentioned here: sometimes the jump process can be immovable on a model for some time, while iterations continue to be counted. In such a case, iterations, counted "during" this immovable period, are not counted as visits, only each new visit to that model are to be counted as visits. With this information provided here, the posterior model probability of M_k is calculated as in Equation 3.28.

$$P(M_k|D) = \frac{f_{M_k}}{N_v} \tag{3.28}$$

Hereby, Bayesian Model Pooling system is finished with this last step and parameter of interest estimations are obtained.

3.6 Algorithm of The Hybrid Bayesian Modeling System

Before moving forward with the validation study on a simulated dataset, the entire algorithm of the methodology is summarized step by step for clarity. It consists of the following steps:

Step 1: Definition of covariates and response of the main model.

Step 2: Generation of missingness indicators, $\boldsymbol{R}$, according to the covariates with MNAR.

Step 3: Setting all possible candidate variables for modeling missingness indicators of the covariates with MNAR.

Step 4: Construction of all possible missingness indicator models for all missing indicators of the covariates with MNAR.

Step 5: Application of T-Step Occam's Window process for construction of the model space by using the joint posteriors, $P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, \boldsymbol{X^{miss}}|\boldsymbol{X^{obs}}, Y, \boldsymbol{R})$ for every model M_k.

Step 6: Construction of transition probability matrices for model transitions $q(M_k, M_{k'})$ for different missingness indicators of covariates with MNAR.

Step 7: Application of RJMCMC process.

Step 8: Application of MCMC diagnostics and acceleration of convergence through thinning and discarding burn-in part.

Step 9: Calculation the estimators of parameters of interest, $\hat{\beta}_{i_k}$, for every model M_k.

Step 10: Calculation of posterior model probabilities $P(M_k|D)$.

Step 11: Application of Bayesian Model Pooling (BMP) by using $\hat{\beta}_i = \sum_{k=1}^{n_m} \hat{\beta}_{i_k} P(M_k|D)$ for $i = 1, \ldots, n_\beta$.

The following chapter deals with this proposed algorithm through a validation study on a simulated dataset with known true parameters, and a real data application.

CHAPTER 4

APPLICATION OF THE METHODOLOGY

In this section, the proposed methodology of hybrid Bayesian Modeling system called Bayesian Model Pooling (BMP) is applied in two steps; a validation on a simulated dataset with known true parameters and a real data analysis. Firstly, a validation study on a simulated data set is designed and carried out in order to investigate the performance of the methodology under different scenarios. Then, the efficiency and the accuracy of the system are tested with several MCMC diagnostic tools. Afterwards, a sensitivity analysis is conducted in consideration of the results of the MCMC diagnostics. The algorithm is enhanced and calibrated by the results of the sensitivity analysis. After the calibration of the algorithm, a real data example with MNAR mechanism is taken and the same algorithm used in the validation study is adapted. The results from real data analysis are compared with other results of the studies, the purpose of which is the same for the same data set in the literature. Finally, a sensitivity analysis is also applied to the real data analysis.

4.1 Validation Study On A Simulated Data With Known True Parameters

In the validation study of this thesis, we consider fully observed binary response variable Y and two covariates X_1 and X_2. The covariate X_1 contains MNAR mechanism while X_2 is fully observed. The relevant data are generated by using the distributions

given in the Equation 4.1.

$$\begin{aligned} x_{1i} &\sim Normal(\mu_{x_1}, \sigma^2_{x_1}) \\ x_{2i} &\sim Normal(\mu_{x_2}, \sigma^2_{x_2}) \\ y_i|p_{y_i} &\sim Bernoulli(p_{y_i}) \\ logit(p_{y_i}) &= \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} \end{aligned} \tag{4.1}$$

where $logit(p_{y_i})$ is the logit link function.

Two covariates, X_1 with MNAR, fully observed X_2, and a binary response variable Y are generated with a sample size of $n = 250$. The fully observed X_2 variable is generated from $X_2 \sim \mathcal{N}(\alpha_{21}; \alpha^2_{22})$ with the vector $(\alpha_{21}; \alpha_{22}) = (1; 1)$ and the fully observed X_1 variable is generated from $X_1 \sim \mathcal{N}(\alpha_{11} + \alpha_{12}X_2; \alpha^2_{13})$ with the vector $(\alpha_{11}; \alpha_{12}; \alpha_{13}) = (-1, 5; 0, 5; 0, 75)$. In addition, the binary response variable Y is generated from $Y \sim \mathcal{B}inomial(n = 250; p_y)$ of which $p_y = \dfrac{e^{\beta_0+\beta_1X_1+\beta_2X_2}}{1 + e^{\beta_0+\beta_1X_1+\beta_2X_2}}$ is the logit link function according to the chosen true values for β coefficients with $(\beta_0; \beta_1; \beta_2) = (2; 1; -1)$.

Further, a binary missingness indicator is also needed and defined here as R, generated from $R \sim \mathcal{B}inomial(n = 250; 1 - p_r)$ where $p_r = \dfrac{e^{\phi_0+\phi_1Y+\phi_2X_2+\phi_3X_1}}{1 + e^{\phi_0+\phi_1Y+\phi_2X_2+\phi_3X_1}}$ and the true parameter settings are given in Equation 4.2 for various different missingness percentages. As seen below, four different missingness percentages are considered for the validation study.

$$(\phi_0; \phi_1; \phi_2; \phi_3) = \begin{cases} (-5, 85; 1; 1; 1) & \text{if Missing Rate} = 1\% \\ (-4, 40; 1; 1; 1) & \text{if Missing Rate} = 3\% \\ (-4, 20; 1; 1; 1) & \text{if Missing Rate} = 5\% \\ (-3, 40; 1; 1; 1) & \text{if Missing Rate} = 10\% \end{cases} \tag{4.2}$$

It is to be noted that ϕ_0 values are determined with a function given in Appendix A.

Now that the data set is generated, as the next step, the analysis of this simulated data is conducted by using the joint posterior distribution given in Equation 3.1. The contribution is rewritten according to the needs of the analysis and for the k^{th} model as $(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, {X_1}^{miss}|{X_1}^{obs}, X_2, Y, R)$. So if $D^{obs} = ({X_1}^{obs}, X_2, Y, R)$, the joint

posterior distribution is given in Equation 4.3.

$$\begin{aligned} P(M_k, \phi_k, \beta_k, {X_1}^{miss}|D^{obs}) &= \frac{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, {X_1}^{miss}, D^{obs})}{P(D^{obs})} \\ &\propto P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, {X_1}^{miss}, D^{obs}) \\ &= P(D^{obs}, {X_1}^{miss}|M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}) P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}) \end{aligned} \tag{4.3}$$

where $P(D_{obs}, {X_1}^{miss}|M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k})$ stands for the full data likelihood, $P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k})$ is the joint prior distribution, D^{obs} is the observed part of the data, ${X_1}^{miss}$ is the missing part of the data, M_k is the k^{th} model index, $\boldsymbol{\phi_k}$ is the parameter vector of k^{th} missingness model, $\boldsymbol{\beta_k}$ is the the parameter vector of the main model in k^{th} model index. In the following subsections 4.1.1 and 4.1.2 the prior and the likelihood are investigated and described thoroughly.

4.1.1 Prior

The joint prior $P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k})$ is shown in Equation 4.4.

$$P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}) = P(\boldsymbol{\beta_k}|M_k) P(\boldsymbol{\phi_k}|M_k) P(M_k) \tag{4.4}$$

where $P(\boldsymbol{\beta_k}|M_k) = P(\beta_{k_0}, \beta_{k_1}, \beta_{k_2}|M_k)$. In this equation, β_{k_0} is the intercept, β_{k_1} is the coefficient of X_1, and β_{k_2} is the coefficient of X_2 under M_k. Similarly, the joint prior for coefficients of the missingness model under M_k is $P(\boldsymbol{\phi_k}|M_k) = P(\phi_{k_0}, \ldots, \phi_{k_{n_k}}|M_k)$ in which the size of $\boldsymbol{\phi_k}$ varies by model, and $P(M_k)$ is the prior probability of model M_k. All of these models and prior distributions are described below.

The prior for missing observations of the variable X_1: It is assumed to be i.i.d. normally distributed. Consequently, the explicit form of the prior distribution for

X_1^{miss} is given in Equation 4.5 where $x_{1,i}^{miss} \sim N(\mu_{x_1}, \sigma_{x_1}^2)$.

$$\begin{aligned} P({X_1}^{miss}|{X_1}^{obs}) &= P(x_{1,1}^{miss}, \ldots, x_{1,n_{miss}}^{miss}|{X_1}^{obs}, M_k) \\ &= P(x_{1,1}^{miss}|{X_1}^{obs}, M_k) \ldots P(x_{1,n_{miss}}^{miss}|{X_1}^{obs}, M_k) \\ &= \prod_{i=1}^{n_{miss}} \frac{1}{\sqrt{2\pi\sigma_{x_1}^2}} e^{-\frac{(x_{1i}^* - \mu_{x_1})^2}{2\sigma_{x_1}^2}} \end{aligned} \tag{4.5}$$

where μ_{x_1} is the assumed expected value of X_1^{miss} and σ_{x_1} represents the assumed standard deviation of X_1^{miss}. x_{1i}^* is the i^{th} imputed ${X_1}^{miss}$ value, which is generated through Gibbs sampling by using the full conditional distribution of ${X_1}^{miss}$.

Prior distribution for main model coefficients $\boldsymbol{\beta_k}$: It is also considered to be normally distributed, hence, $\beta_{k_j} \sim N(\mu_{\beta_{kj}}, \sigma_{\beta_{kj}}^2)$ with $j = 0, 1, 2$. The explicit form of the prior distribution is given as follows;

$$P(\boldsymbol{\beta_k}|M_k) = P(\beta_{k_0}|M_k)P(\beta_{k_1}|M_k)P(\beta_{k_2}|M_k) = \prod_{j=0}^{2} \frac{1}{\sqrt{2\pi\sigma_{\beta_{kj}}^2}} e^{-\frac{(\beta_{kj}^* - \mu_{\beta_{kj}})^2}{2\sigma_{\beta_{kj}}^2}} \tag{4.6}$$

Here, $\mu_{\beta_{kj}}$ stands for the proposed expected value of j^{th} β_k coefficient while $\sigma_{\beta_{kj}}$ being the proposed standard deviation of j^{th} β_k coefficient. $\beta_{k_j}^*$ values are generated with Gibbs sampling by using the full conditional distribution of β_k.

Prior distribution for missingness model coefficients $\boldsymbol{\phi_k}$: Considered to be normally distributed, $\phi_{k_j} \sim N(\mu_{\phi_{kj}}, \sigma_{\phi_{kj}}^2)$ with $j = 1, \ldots, n_{\phi_k}$, its explicit form is given as follows;

$$P(\boldsymbol{\phi_k}|M_k) = P(\phi_{k_0}|M_k) \ldots P(\phi_{k_{n_{\phi_k}}}|M_k) = \prod_{j=1}^{n_{\phi_k}} \frac{1}{\sqrt{2\pi\sigma_{\phi_{kj}}^2}} e^{-\frac{(\phi_{kj}^* - \mu_{\phi_{kj}})^2}{2\sigma_{\phi_{kj}}^2}} \tag{4.7}$$

Seen in Equation 4.7, $\mu_{\phi_{kj}}$ represents the proposed expected value of j^{th} ϕ_k coefficient, and $\sigma_{\phi_{kj}}$ stands for the proposed standard deviation of j^{th} ϕ_k coefficient. ${\phi_k}^*$ values are also generated with Gibbs sampling by using the full conditional distribution of ϕ_k.

Prior distribution for M_k: All the models in the model space are considered apriori equally likely, hence the discrete uniform prior, shared in in Equation 4.8, is used.

$$P(M_k) = \frac{1}{n_m} \tag{4.8}$$

where n_m is the size of the constructed model space.

Prior assumptions for coefficients β, ϕ and missing variable X_1 are provided below.

$$\begin{aligned} \beta &\sim \mathcal{N}(\mu_\beta = 0, \sigma^2_\beta = 4^2) \quad or \quad \beta \sim \mathcal{N}(\mu_\beta = 0, \sigma^2_\beta = 6^2) \\ \phi &\sim \mathcal{N}(\mu_\phi = 0, \sigma^2_\phi = 4^2) \quad or \quad \phi \sim \mathcal{N}(\mu_\phi = 0, \sigma^2_\phi = 6^2) \\ X_1 &\sim \mathcal{N}(\mu_{x_1} = -1, \sigma^2_{x_1} = 1.5^2) \quad or \quad X_1 \sim \mathcal{N}(\mu_{x_1} = -1, \sigma^2_{x_1} = 1^2) \end{aligned} \tag{4.9}$$

These alternative assumptions are utilized for getting more accurate estimations. At this point, it would be useful to mention the reason behind choosing these alternative assumptions for the missing covariate X_1. The generated data for the validation study lead to the summary statistics, given in Table 4.1, for the observed part of the covariate with MNAR. Looking at the Table 4.1, it can be seen that the mean of the observed cases falls around -1 with standard deviations falling between 0,80 - 0,85. As previously stated in Section 2.2.3, the observed parts and the missing parts may show different characteristics. Therefore, the assumption for the mean is fixed at -1 while the variance assumptions are taken larger.

Table 4.1: Summary Statistics For Covariate X_1

	Missing Rate			
Summary	**1%**	**3%**	**5%**	**10%**
$Mean_{obs}$	-0.89	-0.98	-1.06	-1.15
$Std.Dev_{obs}$	0.83	0.86	0.85	0.79
n_{obs}	247	242	237	225
n_{miss}	3	8	13	25
n_{total}	250	250	250	250

4.1.2 Likelihood

After giving details about priors used in the validation study, the likelihood is explained in detail in this sub-section. Full data likelihood, given in Equation 4.3, can be factorized by using selection model given in Equation 4.10.

$$\begin{aligned} P(D^{obs}, {X_1}^{miss}|M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}) =& P(R|{X_1}^{miss}, D^{obs}, M_k, \boldsymbol{\phi_k}) \\ & * P(Y|{X_1}^{miss}, D^{obs}, M_k, \boldsymbol{\beta_k}) \\ & * P({X_1}^{miss}|{X_1}^{obs}, M_k) \end{aligned} \tag{4.10}$$

where $P(R|{X_1}^{miss}, D^{obs}, M_k, \boldsymbol{\phi_k})$ is the missingness model for the k^{th} model index, $P(Y|{X_1}^{miss}, D^{obs}, M_k, \boldsymbol{\beta_k})$ is the main model for the k^{th} model index, and prior for missing observations is $P({X_1}^{miss}|{X_1}^{obs}, M_k) = P(x_{1,1}^{miss}, \ldots, x_{1,n_{miss}}^{miss}|{X_1}^{obs}, M_k)$.

The main model: It is considered to be a logistic model thus the reformulation of the likelihood function of the main model in Equation 4.10. The reformulated function is shared in Equation 4.11.

$$P(Y = y_i|{X_1}^{obs}, {X_1}^{miss}, X_2, M_k, \boldsymbol{\beta_k}) = \prod_{i=1}^{n} {p_{y_i}}^{y_i}(1 - p_{y_i})^{1-y_i} \tag{4.11}$$

where $logit(p_{y_i}) = \mathbf{X}^t\boldsymbol{\beta_k}$ and $p_{y_i} = \dfrac{e^{\mathbf{X}^t\boldsymbol{\beta_k}}}{1 + e^{\mathbf{X}^t\boldsymbol{\beta_k}}}$ is the logit link function with $\mathbf{X} = (x_{1i}, x_{2i})$ being the covariate vector for the main model.

Missingness models: Since R is dichotomous, the missingness models are chosen to be probit models. The likelihood for missingness indicator R in Equation 4.10 is formulated in Equation 4.12.

$$P(R = r_i|{X_1}^{obs}, {X_1}^{miss}, X_2, Y, M_k, \boldsymbol{\phi_k}) = \prod_{i=1}^{n} p_{r_{ki}}^{r_i}(1 - p_{r_{ki}})^{1-r_i} \tag{4.12}$$

where $p_{r_{ki}} = \Phi(\mathbf{W}_{ki}^t\boldsymbol{\phi_k})$ is the probit function in which $\mathbf{W}_{ki}$ is the covariate vector for the k^{th} missingness model. Every missingness model in the model space should contain X_1 or some sort of it's interaction, since MNAR mechanism depends on the variable itself. Based on this knowledge, the variables X_1, X_2, Y, X_1^2, X_1X_2, X_2^2,

X_1Y and X_2Y are taken as candidate covariates for modeling missingness indicator R. Accordingly, $\mathbf{W}_{ki}$ must contain at least one of the covariates X_1, X_1^2, X_1X_2 or X_1Y, as a result of the missingness mechanism depending on X_1 itself.

4.1.3 Gibbs Sampling

Regression coefficients $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ and missing observations $x_{1_i}^{miss}$ should be imputed for a proper Bayesian framework. Therefore, an iterative Gibbs sampling algorithm is applied for imputation. After the imputation, the imputed values for parameters are used in priors. The algorithm of Gibbs sampling updates is given step by step in the following:

Step 1: Draw initials

$$\beta_{k_j}^{(0)} \sim N(\mu_{\beta_{k_j}}, \sigma^2_{\beta_{k_j}}) \quad j = 0, 1, 2$$
$$\phi_{k_j}^{(0)} \sim N(\mu_{\phi_{k_j}}, \sigma^2_{\phi_{k_j}}) \quad j = 0, 1, \ldots, n_{\phi_k}$$

Step 2: Continue the process with a *for* loop from $t = 1$ to n_t

$$X_1^{miss,(t)} \sim \text{Metropolis-Hastings step}$$
$$\beta_{k_j}^{(t)} \sim f(\beta_{k_j}|X_1^{miss,(t)}, D_{obs}, M_k, \mu_{\beta_{k_j}}, \sigma^2_{\beta_{k_j}}) \quad j = 0, 1, 2$$
$$\phi_{k_j}^{(t)} \sim f(\phi_{k_j}|X_1^{miss,(t)}, D_{obs}, M_k, \mu_{\phi_{k_j}}, \sigma^2_{\phi_{k_j}}) \quad j = 0, 1, \ldots, n_{\phi_k}$$

Step 3: Set $(X_1^{*miss}, \beta_{k_j}^*, \phi_{k_j}^*) = (X_1^{miss,(n_t)}, \beta_{k_j}^{(n_t)}, \phi_{k_j}^{(n_t)})$

(4.13)

At this point, since the full conditional distribution of X_1^{miss} does not have a familiar form, a Metropolis-Hastings Step is added to the Gibbs sampling algorithm for generating $x_{1_i}^{miss}$ values with $i = 1, \ldots, n_{miss}$. The M-H algorithm is given as follows:

Step 1: Draw initial $x_{1_0}^{miss} \sim N(\mu_{x_1}, \sigma^2_{x_1})$

Step 2: Continue the process with a *for* loop from $i = 1$ to n_{miss}

Draw $x_{1_{proposed}}^{miss} \sim N(\mu_{x_1}, \sigma^2_{x_1})$ and $U \sim Unif(0, 1)$

Compute $\alpha(x_{1_{proposed}}^{miss}, x_{1_{i-1}}^{miss}) = min\left\{1, \dfrac{f(x_{1_{proposed}}^{miss}|D_{obs_i}, \boldsymbol{\beta_k}^{(t-1)}, \boldsymbol{\phi_k}^{(t-1)}, \mu_{x_1}, \sigma^2_{x_1})}{f(x_{1_{i-1}}^{miss}|D_{obs_i}, \boldsymbol{\beta_k}^{(t-1)}, \boldsymbol{\phi_k}^{(t-1)}, \mu_{x_1}, \sigma^2_{x_1})}\right\}$

Set $x_{1_i}^{miss,(t)} = \begin{cases} x_{1_{proposed}}^{miss} & \text{if } U < \alpha(x_{1_{proposed}}^{miss}, x_{1_{i-1}}^{miss}) \\ x_{1_{i-1}}^{miss} & \text{otherwise} \end{cases}$

Step 3: Set $X_1^{miss,(t)} = (x_{1_1}^{miss,(t)}, \ldots, x_{1_{n_{miss}}}^{miss,(t)})$

(4.14)

where $f(x_{1_i}^{miss}|D_{obs_i}, \boldsymbol{\beta_k}, \boldsymbol{\phi_k}, \mu_{x_1}, \sigma^2_{x_1})$ is the full conditional distribution for the i^{th}

missing observation. The explicit form of it is given in Equation 4.15.

$$\begin{aligned} f(x_{1_i}^{miss}|D_{obs_i}, \boldsymbol{\beta_k}, \boldsymbol{\phi_k}, \mu_{x_1}, \sigma_{x_1}^2) =& (p_{y_i})^{y_i}(1-p_{y_i})^{1-y_i} \\ & * (p_{r_{k_i}})^{r_i}(1-p_{r_{k_i}})^{1-r_i} \\ & * \frac{1}{\sqrt{2\pi\sigma_{x_1}^2}} e^{-\frac{(x_{1i}-\mu_{x_1})^2}{2\sigma_{x_1}^2}} \end{aligned} \tag{4.15}$$

where p_{y_i} and $p_{r_{k_i}}$ are the same as in Equation 4.12 and 4.11.

Moreover, in the Gibbs sampling, $f(\beta_{k_j}|X_1^{miss,(t)}, D_{obs}, M_k, \mu_{\beta_{k_j}}, \sigma_{\beta_{kj}}^2)$ stands for the full conditional distribution of j^{th} β coefficient under k^{th} model for iteration (t). This full conditional distribution can be factorized as in Equation 4.16.

$$f(\boldsymbol{\beta_k}|X_1^{miss}, D_{obs}, M_k) = L(\boldsymbol{\beta_k}|{X_1}^{miss}, D_{obs}, M_k)P(\boldsymbol{\beta_k}|M_k) \tag{4.16}$$

where $L(\boldsymbol{\beta_k}|{X_1}^{miss}, D_{obs}, M_k)$ is the marginal likelihood, $P(\boldsymbol{\beta_k}|M_k)$ being the prior. $\boldsymbol{\beta}_k \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta^2)$ of which $\boldsymbol{\mu}_\beta$ is the mean vector, and $\boldsymbol{\Sigma}_\beta^2$ is the variance covariance matrix for coefficients $\boldsymbol{\beta_k} = (\beta_{k_0}, \beta_{k_1}, \beta_{k_2})$. Since the main model for the response is logistic and the prior is normal as given above, conditional conjugacy is lost. Holmes and Held (2006) indicate a useful full conditional distribution for coefficients of a logistic regression model. According to their article, when an additional variable λ is added as the scale of error terms of auxiliary variable Z (the details are given below), only then it is easier to sample by using the full conditional distribution of β coefficients of logistic regression [30]. This helpful knowledge is adapted for this study as follows:

An auxiliary variable $\boldsymbol{Z}_{\boldsymbol{\beta_k}} = (z_{\beta_{k,1}}, \ldots, z_{\beta_{k,n}})$ is defined as

$$y_i = \begin{cases} 1 & \text{if } z_{\beta_{k,i}} > 0 \\ 0 & \text{if } z_{\beta_{k,i}} < 0 \end{cases} \tag{4.17}$$

In Equation 4.17, $z_{\beta_{k,i}} = \boldsymbol{x_i}\boldsymbol{\beta}_k + \epsilon_i$ with $\epsilon_i \sim N(0, \lambda_i)$ and $i = 1, \ldots, n$. After this definition, sampling can be done by using the full conditional distribution of $\boldsymbol{\beta}_k$ provided in Equation 4.18.

$$\begin{aligned} \boldsymbol{\beta}_k|X_1^{miss}, M_k, \boldsymbol{Z}_{\boldsymbol{\beta_k}}, \boldsymbol{\lambda} &\sim N(\tilde{\boldsymbol{\mu}}_\beta, \tilde{\boldsymbol{\Sigma}}_\beta) \\ \tilde{\boldsymbol{\mu}}_{\beta_k} &= \tilde{\boldsymbol{\Sigma}}_{\beta_k}(\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta + \boldsymbol{X^t B Z}_{\boldsymbol{\beta_k}}) \\ \tilde{\boldsymbol{\Sigma}}_{\beta_k} &= (\boldsymbol{\Sigma}_\beta^{-1} + \boldsymbol{X^t B X})^{-1} \end{aligned} \tag{4.18}$$

with $\boldsymbol{X} = (X_1^{*miss}, X_1^{obs}, X_2)$ and $\boldsymbol{B} = diag(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$.

Additionally, sampling of each auxiliary value $z_{\beta_{k,i}}$ from full conditional distribution can be done using a truncated normal distribution as described in Equation 4.19 [65].

$$z_{\beta_{k,i}}|\boldsymbol{\beta_k}, \boldsymbol{x_i}, y_i, \lambda_i \propto \begin{cases} N(\boldsymbol{x_i}\boldsymbol{\beta_k}, \lambda_i)I(z_i > 0) & \text{if } y_i = 1 \\ N(\boldsymbol{x_i}\boldsymbol{\beta_k}, \lambda_i)I(z_i \leq 0) & \text{if } y_i = 0 \end{cases} \tag{4.19}$$

Lastly, the full conditional distribution of the j^{th} ϕ coefficient under k^{th} model for iteration (t) is defined as $f(\phi_{k_j}|X_1^{miss,(t)}, D_{obs}, M_k, \mu_{\phi_{kj}}, \sigma^2_{\phi_{kj}})$. A more general form of this distribution can be factorized as given in Equation 4.20.

$$f(\boldsymbol{\phi_k}|X_1^{miss}, D_{obs}, M_k) = L(\boldsymbol{\phi_k}|{X_1}^{miss}, D_{obs}, M_k)P(\boldsymbol{\phi_k}|M_k) \tag{4.20}$$

where $L(\boldsymbol{\phi_k}|{X_1}^{miss}, D_{obs}, M_k)$ is the marginal likelihood, $P(\boldsymbol{\phi_k}|M_k)$ is the prior. $\boldsymbol{\phi}_k \sim N(\boldsymbol{\mu}_\phi, \boldsymbol{\Sigma}^2_\phi)$ where $\boldsymbol{\mu}_\phi$ is the mean vector and $\boldsymbol{\Sigma}^2_\phi$ is the variance covariance matrix for coefficients $\boldsymbol{\phi_k} = (\phi_{k_0}, \ldots, \phi_{k_{n_\phi}})$. Since missingness models are probit models, $P(\boldsymbol{\phi_k}|M_k)$ can be considered as the conjugate prior with a normal distribution. Therefore, the full conditional distribution given in the article of Holmes and Held (2006) is adapted using the following formula [30]:

An auxiliary variable $\boldsymbol{Z_{\phi_k}} = (z_{\phi_{k,1}}, \ldots, z_{\phi_{k,n}})$ is defined as

$$r_i = \begin{cases} 1 & \text{if } z_{\phi_{k,i}} > 0 \\ 0 & \text{if } z_{\phi_{k,i}} < 0 \end{cases} \tag{4.21}$$

In here, r_i is the i^{th} missingness indicator, $z_{\phi_{k,i}} = \boldsymbol{w_{k_i}}\boldsymbol{\phi}_k + \epsilon_i$ where w_{k_i} is the covariate vector of k^{th} missingness model, $\epsilon_i \sim N(0, 1)$, and $i = 1, \ldots, n$. After this definition, sampling can be done by using the full conditional distribution of $\boldsymbol{\phi}_k$ given in Equation 4.22.

$$\begin{aligned} \boldsymbol{\phi}_k|X_1^{miss}, M_k, \boldsymbol{Z_{\phi_k}} &\sim N(\tilde{\boldsymbol{\mu}}_{\phi_k}, \tilde{\boldsymbol{\Sigma}}_{\phi_k}) \\ \tilde{\boldsymbol{\mu}}_{\phi_k} &= \tilde{\boldsymbol{\Sigma}}_{\phi_k}(\boldsymbol{\Sigma}^{-1}_{\phi_k}\boldsymbol{\mu}_{\phi_k} + \boldsymbol{W_k^t Z_{\phi_k}}) \\ \tilde{\boldsymbol{\Sigma}}_{\phi_k} &= (\boldsymbol{\Sigma}^{-1}_{\phi_k} + \boldsymbol{W_k^t W_k})^{-1} \end{aligned} \tag{4.22}$$

where $\boldsymbol{W_k}$ is the covariate matrix of the k^{th} missingness model.

Again, sampling of each auxiliary value $z_{\phi_{k,i}}$ from full conditional distribution can be done using a truncated normal distribution as described in Equation 4.23 [65].

$$z_{\phi_{k,i}}|\boldsymbol{\phi_k}, \boldsymbol{w_{k,i}}, r_i \propto \begin{cases} N(\boldsymbol{w_{k,i}}\boldsymbol{\phi_k}, 1)I(z_i > 0) & \text{if } r_i = 1 \\ N(\boldsymbol{w_{k,i}}\boldsymbol{\phi_k}, 1)I(z_i \leq 0) & \text{if } r_i = 0 \end{cases} \tag{4.23}$$

4.1.4 RJMCMC Adaptation

The next step of the validation study, after the data generation and joint posterior definition, is the adaptation of RJMCMC process. All the acceptance ratios in Section 3.2 are adapted according to the validation assumptions. If the dimension of the candidate model $M_{k'}$ is higher than the initial model M_k ($|k'| > |k|$), then the acceptance probability for the new state is given as follows;

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{P(M_{k'}, \boldsymbol{\phi_{k'}}, \boldsymbol{\beta_{k'}}, X_1^{miss}|X_1^{obs}, X_2, Y, R)q(M_{k'}, M_k)}{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, X_1^{miss}|X_1^{obs}, X_2, Y, R)q(M_k, M_{k'})h(u)} \times \mid J_u \mid\right\} \tag{4.24}$$

If the dimension of the candidate model $M_{k'}$ is lower than the initial model M_k ($|k'| < |k|$), the acceptance probability for the new state is given as follows;

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{P(M_{k'}, \boldsymbol{\phi_{k'}}, \boldsymbol{\beta_{k'}}, X_1^{miss}|X_1^{obs}, X_2, Y, R)q(M_{k'}, M_k)h(u')}{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, X_1^{miss}|X_1^{obs}, X_2, Y, R)q(M_k, M_{k'})} \times \mid J_{u'} \mid\right\} \tag{4.25}$$

If the dimension of the candidate model $M_{k'}$ is equal to the initial model M_k ($|k'| = |k|$), the acceptance probability for the new state is given as follows;

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{P(M_{k'}, \boldsymbol{\phi_{k'}}, \boldsymbol{\beta_{k'}}, X_1^{miss}|X_1^{obs}, X_2, Y, R)q(M_{k'}, M_k)}{P(M_k, \boldsymbol{\phi_k}, \boldsymbol{\beta_k}, X_1^{miss}|X_1^{obs}, X_2, Y, R)q(M_k, M_{k'})}\right\} \tag{4.26}$$

At this point, determinants of the Jacobian terms $\mid J_u \mid$ and $\mid J_{u'} \mid$, already provided in Equation 3.17 in the previous chapter, are calculated as 1. As indicated in Section 3.2, with the dimension change happening only in missingness models, u and u' stand for coefficients of missingness models ϕ, since the dimension change happens only in missingness models. Therefore, $h(u)$ and $h(u')$ are chosen exactly the same as the prior of ϕ coefficients in Equation 3.11. With $u = (u_1, \ldots, u_{n_u})$ and

$u' = (u'_1, \ldots, u'_{n_{u'}})$, $h(u)$ and $h(u')$ are given in Equation 4.27.

$$
\begin{aligned}
& u_j^* \sim N(\mu_u, \sigma_u^2) \quad \text{and} \quad u_j'^* \sim N(\mu_{u'}, \sigma_{u'}^2) \\
& h(u) = h(u_1, \ldots, u_{n_u}) = h(u_1) \ldots h(u_n) = \prod_{j=1}^{n_u} \frac{1}{\sqrt{2\pi\sigma_u^2}} e^{-\frac{(u_j^* - \mu_u)^2}{2\sigma_u^2}} \\
& h(u') = h(u'_1, \ldots, u'_{n_{u'}}) = h(u'_1) \ldots h(u'_{n_{u'}}) = \prod_{j=1}^{n_{u'}} \frac{1}{\sqrt{2\pi\sigma_{u'}^2}} e^{-\frac{(u_j'^* - \mu_{u'})^2}{2\sigma_{u'}^2}}
\end{aligned}
\tag{4.27}
$$

where μ_u and $\mu_{u'}$ are the proposed expected value of u and u', while σ_u and $\sigma_{u'}$ being the proposed standard deviation of u and u'.

4.1.5 Construction of The Model Space

The methodology for constructing model spaces is explained in Section 3.3. For the validation study, it is assumed that there is only one missing variable and thus one missingness indicator R. The model space is defined as $\mathbb{M} = \{M_1, \ldots, M_{n_m}\}$. Models in this particular model space differ only through different variations of missingness models. The variable set for missingness models is described as

$$\mathbb{V} = (X_1, X_2, Y, X_1^2, X_1X_2, X_2^2, X_1Y, X_2Y)$$

with 8 variables and $2^8 - 1 = 255$ subsets (candidate models) excluding the null set. The ones not containing X_1, X_1^2, X_1X_2 or X_1Y should be excluded from the subsets due to the MNAR mechanism. Let $\mathbb{S}$ be a set consisting of subsets of the variable sets defined in $\mathbb{V}$ and its every subset can be described as $\mathbb{S} = \bigcup_{k=1}^{n_m} \mathbb{S}_k$ with $\mathbb{S}_k \subseteq \mathbb{V}$. $\mathbb{S}_k$ is given in Equation 4.28.

$$
\begin{aligned}
\mathbb{S}_k = & \{(\mathbb{S}_{k1}, \mathbb{S}_{k2}) \in \mathbb{S}_k \mid \mathbb{S}_{k1} \subseteq (X_1 \cup X_1^2 \cup X_1X_2 \cup X_2^2 \cup X_1Y) \\
& ; \mathbb{S}_{k2} = \emptyset \vee \mathbb{S}_{k2} \subseteq (X_2 \cup Y \cup X_2^2 \cup X_2Y) \\
& ; \mathbb{S}_{k2} \cup \mathbb{S}_{k2} = \mathbb{S}_k\}
\end{aligned}
\tag{4.28}
$$

where, $\mathbb{S}_{k1}$ is the part of the k^{th} subset that includes X_1, the covariate with MNAR, or the interaction of it. $\mathbb{S}_{k2}$ is the remaining part of the subset. After the process of excluding, there are 240 subsets left in $\mathbb{S}$ and its volume is still large for a model

space. Therefore, it should be filtered and reduced by using the adapted Occam's Window method provided in detail in Section 3.3.1.

The size of t is chosen as 200, which means the Occam's Window process is to be repeated 200 times. Now, 200 model spaces with different sizes are present. Among these 200 model spaces the first 15 common models with the highest frequency are taken as candidate models for model space. Occam's Window repetitions and model spaces are given in Equation 4.29.

$$\begin{aligned}
\mathbb{M}^{(1)} &= \left\{ M_k^{(1)} : \frac{max_l\{P(M_l|D)\}}{P(M_k|D)} \leq C \right\} = \{M_1^{(1)}, \ldots, M_{n_1}^{(1)}\} \\
&\vdots \\
\mathbb{M}^{(100)} &= \left\{ M_k^{(100)} : \frac{max_l\{P(M_l|D)\}}{P(M_k|D)} \leq C \right\} = \{M_1^{(100)}, \ldots, M_{n_{100}}^{(100)}\}
\end{aligned} \tag{4.29}$$

This process is applied under different missingness rates and parameter assumptions. The relevant numerical computations are given in the following sections. After construction of the model space for every covariate with MNAR, model transitions are defined and transition probabilities among models are calculated. Calculations of transition probabilities and parameter estimations by missing rates and by different underlying assumptions are given in the following sections as well. The same algorithm, given in Chapter 3, is adapted for the validation study by including data generation and by making some adjustments, which are already mentioned above.

RJMCMC process and estimation of the parameters of interest are realized via R software. Before coding, it would be useful to make it clear the steps of the algorithm. The algorithm is divided into its respective steps below:

Step 1: Data generation:

$x_{2_i} \sim \mathcal{N}(\alpha_{21}; \alpha_{22}^2)$

$x_{1_i} \sim \mathcal{N}(\alpha_{11} + \alpha_{12} x_{2_i}; \alpha_{13}^2)$

$y_i \sim \mathcal{B}inomial(n = 250; p_{y_i})$ where $logit(p_{y_i}) = \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i}$

Step 2: Set missingness rate 1%, 3%, 5% or 10%.

Step 3: Generate missingness mechanism R according to missing rate set.

Step 4: Reset X_1 according to the missing rate.

Step 5: Set other covariates $X_1^2, X_2^2, X_1X_2, X_1Y, X_2Y$.

Step 6: Construct the matrices of all potential covariates for all possible missingness models (i.e. model space). 240 candidate models are present.

Step 7: Apply adapted Occam's Window process by using BIC values joint posteriors. Set $t = 200$ and $c = 20$. Take the first 15 or 16 common models having the highest frequency for every missing rate.

Step 8: Construct transition probability matrix for model transitions $q(M_k, M_{k'})$. Transitions can be done according to the information provided in Section 3.4.

Step 9: Set initial model index $M_{initial}$ arbitrarily.
Initial parameter generation:
Apply Gibbs algorithm in 4.13.
Set $(\boldsymbol{\beta_k}, \boldsymbol{\phi_k}, M_k) = (\boldsymbol{\beta_k}^{(n_t)}, \boldsymbol{\phi_k}^{(n_t)}, M_{initial})$

Step 10: With $i = 1$;
Set $(\boldsymbol{\beta}^{(i)}, \boldsymbol{\phi}^{(i)}, M^{(i)}) = (\boldsymbol{\beta_k}, \boldsymbol{\phi_k}, M_k)$
Apply a ***for loop*** from $i = 2$ to $i = N + B$
where $B = n_{burn-in}$ and $N = n_{iterations}$
Set proposed model $M_{proposed}$ arbitrarily.
Proposed parameter generation:
Apply Gibbs algorithm in 4.13.
Set $(\boldsymbol{\beta_{k'}}, \boldsymbol{\phi_{k'}}, M_{k'}) = (\boldsymbol{\beta_{k'}}^{(n_t)}, \boldsymbol{\phi_{k'}}^{(n_t)}, M_{proposed})$
Draw $U \sim Unif(0, 1)$
If $U < \alpha(M_k, M_{k'})$ (acceptance probability in RJMCMC)
Set $(\boldsymbol{\beta}^{(i)}, \boldsymbol{\phi}^{(i)}, M^{(i)}) = (\boldsymbol{\beta_{k'}}, \boldsymbol{\phi_{k'}}, M_{k'})$
Else
Set $(\boldsymbol{\beta}^{(i)}, \boldsymbol{\phi}^{(i)}, M^{(i)}) = (\boldsymbol{\beta}^{(i-1)}, \boldsymbol{\phi}^{(i-1)}, M^{(i-1)})$

Step 11: Apply MCMC diagnostics and acceleration tools, such as thinning.

Step 12: Calculate posterior model probabilities $P(M_k|D)$.

Step 13: Calculate medians of parameters of interest, $\hat{\beta}_{i_k}$, for every model M_k.

Step 14: Apply Pooling for estimations by using posteriors as $\hat{\beta}_i = \sum_{k=1}^{n_m} \hat{\beta}_{i_k} P(M_k|D)$ for $i = 0, 1, 2$.

The validation study on a simulated dataset is conducted with `RStudio Version 1.2.5001`, based on `R` programming software. More than 60 trials were run with fluctuating durations from 15 hours to 144 hours (1 - 6 days). These fluctuations are a result of using 5 separate computers with different technical specs along with different simulation sizes to get convergence. Eventually, for all simulations, the replication size N is taken either as 1.500.000 or as 1.800.000 with 2 chains. Burn-in size B is taken 200.000 or 400.000 according to the results of MCMC diagnostics.

The results of estimations under different assumptions are given in Section 4.1.6 named Results. Models in the model space (as dots represent the covariates in each model) with their respective posterior model probabilities are tabulated. The ones with 1% missing rate under assumptions $\beta, \phi \sim \mathcal{N}(0, 4^2)$ and $X_1 \sim \mathcal{N}(-1, 1.5^2)$ are listed in Table 4.2, while the ones for other missing rates and under different prior assumptions are provided in Appendix B.1. For Table 4.2, the most saturated model is the 15^{th} model with a dimension of 5 (including intercept). Additionally, the model posteriors of the 1^{st} model containing only X_1 equals to 0 for all missing rates, meaning that only observed X_1 values would not be enough to explain the missingness indicator of MNAR. It is also important to note that model posterior probabilities increase as the models become more complex. Models with IDs between 10 and 15 can be considered as more preferable.

After the construction of the model spaces for different missing rates, transition probabilities as matrix forms are computed. All the transition probability matrices by different missing rates are constructed according to the model spaces given in Table 4.2 and also the ones given in Appendix B.1.

In Table 4.3, transition probabilities are given for a 1% missing rate under prior assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$. The rest of the transition probability matrices for other missing rates are available in Appendix B.2. Each cell in Table 4.3 shows the transition probability from a model to another model. For instance, the transition probability from model 3 to model 4 is found to be

Table 4.2: Model Space with Model Posterior Probabilities For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0, 4^2)$ and $X_1 \sim \mathcal{N}(-1, 1.5^2)$

	Covariate								
Model ID	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$\mathbf{P(M_k\|D)}$
1	•								0.0000
2				•					0.0007
3							•		0.0014
4	•		•						0.0216
5	•						•		0.0228
6		•					•		0.0215
7			•				•		0.0213
8				•			•		0.0219
9				•				•	0.0212
10	•		•	•					0.1478
11	•		•					•	0.1471
12	•						•	•	0.1467
13		•	•				•		0.1472
14			•				•	•	0.1488
15	•		•				•	•	0.1299

$q(M_3, M_4) = 0.073$, but transition probability from model 4 to model 3 equals to $q(M_4, M_3) = 0.060$. The reason behind this is the changing transition ability of models due to their dimensions, explained in Section 3.4. After calculating the transition probabilities by missing rate, RJMCMC algorithm can now be applied.

Another issue worth mentioning here is a problem occuring while calculating the joint posterior $P(M_{k'}, \phi_{k'}, \beta_{k'}, {X_1}^{miss}|D_{obs})$ in simulation runs. Since the joint posterior contains a lot of unknown parameters, the likelihood values become too small (for example e^{-250}). This indicates a numerical overflow problem in likelihood. Therefore, the acceptance probabilities are re-constructed into log-scale and very small epsilon

Table 4.3: Transition Probability Matrix for 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.098	0.098	0.098	0.073	0.073	0.073	0.073	0.073	0.073	0.049	0.049	0.049	0.049	0.049	0.024
2	0.098	0.098	0.098	0.073	0.073	0.073	0.073	0.073	0.073	0.049	0.049	0.049	0.049	0.049	0.024
3	0.098	0.098	0.098	0.073	0.073	0.073	0.073	0.073	0.073	0.049	0.049	0.049	0.049	0.049	0.024
4	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.080	0.060	0.060	0.060	0.060	0.060	0.040
5	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.080	0.060	0.060	0.060	0.060	0.060	0.040
6	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.080	0.060	0.060	0.060	0.060	0.060	0.040
7	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.080	0.060	0.060	0.060	0.060	0.060	0.040
8	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.080	0.060	0.060	0.060	0.060	0.060	0.040
9	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.080	0.060	0.060	0.060	0.060	0.060	0.040
10	0.043	0.043	0.043	0.064	0.064	0.064	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064
11	0.043	0.043	0.043	0.064	0.064	0.064	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064
12	0.043	0.043	0.043	0.064	0.064	0.064	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064
13	0.043	0.043	0.043	0.064	0.064	0.064	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064
14	0.043	0.043	0.043	0.064	0.064	0.064	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064
15	0.029	0.029	0.029	0.059	0.059	0.059	0.059	0.059	0.059	0.088	0.088	0.088	0.088	0.088	0.118

values (0.00001) are added in order to eliminate infinite results in log-scale. The new acceptance probabilities for RJMCMC are given below.

- If the dimension of $M_{k'}$, is higher than the one with M_k ($|k'| > |k|$), the acceptance probability for the new state is given as follows;

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{logP((M_{k'}, \phi_{k'}, \beta_{k'}, {X_1}^{miss}|D_{obs}) + \epsilon) + log(q(M_{k'}, M_k))}{log(P(M_k, \phi_k, \beta_k, {X_1}^{miss}|D_{obs}) + \epsilon) + log(q(M_k, M_{k'})) + log(h(u))}\right\} \quad (4.30)$$

- If the dimension of $M_{k'}$, is lower than the one with M_k ($|k'| < |k|$), the acceptance probability for the new state is given as follows;

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{logP((M_{k'}, \phi_{k'}, \beta_{k'}, {X_1}^{miss}|D_{obs}) + \epsilon) + log(q(M_{k'}, M_k)) + log(h(u'))}{log(P(M_k, \phi_k, \beta_k, {X_1}^{miss}|D_{obs}) + \epsilon) + log(q(M_k, M_{k'}))}\right\} \quad (4.31)$$

- If the dimension of $M_{k'}$, is equal to the one with M_k ($|k'| = |k|$), the acceptance probability for the new state is given as follows;

$$\alpha(M_k, M_{k'}) = min\left\{1, \frac{logP((M_{k'}, \phi_{k'}, \beta_{k'}, {X_1}^{miss}|D_{obs}) + \epsilon) + log(q(M_{k'}, M_k))}{log(P(M_k, \phi_k, \beta_k, {X_1}^{miss}|D_{obs}) + \epsilon) + log(q(M_k, M_{k'}))}\right\} \quad (4.32)$$

where $P(M_{k'}, \phi_{k'}, \beta_{k'}, {X_1}^{miss}|D_{obs})$ is the joint posterior distribution of proposed model M'_k while $P(M_k, \phi_k, \beta_k, {X_1}^{miss}|D_{obs})$ is the joint posterior of the initial model

M_k, $q(M_{k'}, M_k)$ is the transition probability from the proposed model $M_{k'}$ to the initial model M_k while $q(M_k, M_{k'})$ is the transition probability from the initial model M_k to the proposed model $M_{k'}$, and finally $h(u)$ and $h(u')$ are the densities in order to balance the dimensions of parameter spaces of missingness models.

4.1.6 Results

The last part of the validation study is completed with the results of the BMP process. Median is chosen as estimators of coefficients since the distribution of the coefficients are skewed (see Appendix B.3.1).

The results of the BMP are evaluated in reference to the best case scenario in which fitted models are the same as the underlying data generating models given in Section 4.1, except probit link is used for the main response model. These are given in Table 4.4 under the column "reference model". Below in Table 4.4, the relative bias values with standard errors (SE) and Monte Carlo Standard Errors (MCSE) of the estimations from RJMCMC process of the proposed hybrid method and also of a single reference model are given. The same iteration sizes, thinning sizes under different prior assumptions are applied for the reference model and the estimates are obtained by using OpenBUGS 3.2.3.

Relative bias, SE and MCSE of estimates are calculated with the formulas given in Equation 4.33.

$$\begin{aligned} \textit{Relative Bias}_{\hat{\beta}_j} &= \frac{\hat{\beta}_j - \beta_j}{\beta_j} \\ SE_{\hat{\beta}_j} &= \sqrt{\frac{\sum_{i=1}^{N} \hat{\beta}_j^{(i)2}}{N} - \left(\frac{\sum_{i=1}^{N} \hat{\beta}_j^{(i)}}{N}\right)^2} \\ MCSE_{\hat{\beta}_j} &= \sqrt{\frac{\sum_{l=1}^{b} (\hat{\beta}_{jl} - \hat{\beta}_j)^2}{(b-1)b}} \end{aligned} \tag{4.33}$$

where $\hat{\beta}_j$ is the estimate and β_j is the true value of j^{th} coefficient, $\hat{\beta}_j^{(i)}$ is the j^{th} coefficient estimate at the i^{th} iteration, N is the iteration size after thinning, $\hat{\beta}_{jl}$ is the j^{th} coefficient estimate of the l^{th} batch, and b is the size of the batch. As a remainder, the true values of the parameters are $(\beta_0, \beta_1, \beta_2) = (2, 1, -1)$.

Table 4.4: Relative Bias, Standard Errors (SE) and Monte Carlo Standard Errors (MCSE) of Estimations Under Different Prior Assumptions by Missing Rate

		Relative Bias (SE) MCSE					
		β_0		β_1		β_2	
Missing Rate	**Prior Assumptions**	**Proposed Method**	**Reference Model**	**Proposed Method**	**Reference Model**	**Proposed Method**	**Reference Model**
1%	A: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1^2)$	0.1767 (0.0291) 0.0104	-0.4145 (0.2695) 0.0007	0.1308 (0.0271) 0.0103	-0.3924 (0.1631) 0.0005	-0.0889 (0.0274) 0.0101	-0.3964 (0.1374) 0.0004
	B: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.1767 (0.0291) 0.0105	-0.4145 (0.2694) 0.0008	0.0782 (0.0028) 0.0052	-0.4000 (0.1633) 0.0005	-0.0411 (0.0030) 0.0058	-0.4022 (0.1368) 0.0004
	C: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1^2)$	0.1767 (0.0291) 0.0106	-0.5145 (0.2439) 0.0006	0.6380 (0.0315) 0.0116	-0.4903 (0.1518) 0.0004	0.3830 (0.0086) 0.0117	-0.4798 (0.1280) 0.0003
	D: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.1767 (0.0291) 0.0107	-0.5153 (0.2436) 0.0007	0.5842 (0.0069) 0.0080	-0.4894 (0.1516) 0.0004	0.3675 (0.0095) 0.0080	-0.4849 (0.1278) 0.0004
3%	A: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1^2)$	0.1767 (0.0291) 0.0108	-0.3585 (0.2528) 0.0005	0.0903 (0.0053) 0.0063	-0.2733 (0.1539) 0.0003	0.0872 (0.0041) 0.0064	-0.2952 (0.1477) 0.0003
	B: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.1767 (0.0291) 0.0109	-0.3670 (0.2553) 0.0005	0.0296 (0.0019) 0.0033	-0.2891 (0.1567) 0.0003	0.0483 (0.0013) 0.0033	-0.3149 (0.1470) 0.0003
	C: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1^2)$	0.1767 (0.0291) 0.0110	-0.4590 (0.2297) 0.0005	0.6000 (0.0286) 0.0139	-0.3727 (0.1437) 0.0003	0.6517 (0.0436) 0.0126	-0.3885 (0.1383) 0.0003
	D: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.1767 (0.0291) 0.0111	-0.4660 (0.2307) 0.0005	0.4889 (0.0508) 0.0141	-0.3856 (0.1454) 0.0003	0.5458 (0.0457) 0.0132	-0.4028 (0.1375) 0.0003
5%	A: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1^2)$	0.1767 (0.0291) 0.0112	-0.3895 (0.2562) 0.0005	0.0401 (0.0058) 0.0067	-0.3773 (0.1523) 0.0003	-0.0494 (0.0053) 0.0065	-0.2376 (0.1484) 0.0003
	B: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.1767 (0.0291) 0.0113	-0.4080 (0.2575) 0.0005	-0.1745 (0.0014) 0.0034	-0.4053 (0.1553) 0.0003	0.1254 (0.0010) 0.0034	-0.2683 (0.1468) 0.0003
	C: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1^2)$	0.1767 (0.0291) 0.0114	-0.4870 (0.2334) 0.0005	0.3469 (0.0032) 0.0065	-0.4677 (0.1430) 0.0003	0.7125 (0.0059) 0.0067	-0.3282 (0.1383) 0.0003
	D: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.1767 (0.0291) 0.0115	-0.5003 (0.2341) 0.0005	0.2192 (0.0115) 0.0089	-0.4911 (0.1450) 0.0003	0.6729 (0.0169) 0.0084	-0.3499 (0.1369) 0.0003
10%	A: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1^2)$	0.1767 (0.0291) 0.0116	-0.4410 (0.2673) 0.0008	-0.2537 (0.0027) 0.0059	-0.4371 (0.1626) 0.0005	-0.1552 (0.0018) 0.0059	-0.3799 (0.1414) 0.0004
	B: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.1767 (0.0291) 0.0117	-0.4940 (0.2749) 0.0008	-0.3969 (0.0106) 0.0078	-0.5245 (0.1698) 0.0005	-0.2460 (0.0103) 0.0093	-0.4365 (0.1377) 0.0004
	C: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1^2)$	0.1767 (0.0291) 0.0118	-0.5328 (0.2424) 0.0007	0.0245 (0.0505) 0.0231	-0.5282 (0.1508) 0.0004	0.2203 (0.0600) 0.0202	-0.4565 (0.1318) 0.0004
	D: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.1767 (0.0291) 0.0119	-0.5719 (0.2470) 0.0007	-0.0886 (0.0035) 0.0058	-0.5963 (0.1557) 0.0004	0.1047 (0.0044) 0.0061	-0.4968 (0.1290) 0.0004

According to Table 4.4, all of the estimates of the reference model are underestimated. Estimation performance and accuracy of β_1 is the most important indicator for this study since the MNAR mechanism belongs to the variable X_1, there is a positive relative bias to be found but remains comparably very small for missing rates of 1% and 3% under prior assumptions A and B. For 5% missing rate, there is negative relative bias with a high percentage (-17.45%) under parameter assumption B. The relative bias is only very small (4.01%) for prior assumption A. Under prior assumptions C and D, relative bias values of the proposed method for 1%, 3% and 5% missing rates are quite higher, which is an indication of overestimation. However, the trend is different than this for 10% missing rate. For this missing rate, highly negative relative bias ratios (-25.37% and -39.69%) are observed under the assumptions A and B, which are still lower than ones of the reference model. This time, under the assumptions C and D, the relative bias values decrease to 2.45% and -8.86% respectively. The accuracy of the intercept coefficient β_0 and the accuracy of β_2 are similar with β_1 in the most of the scenarios and also the behaviour of them is very similar to the behavioural pattern of β_1 coefficient. Only when the missing rate is at 10% and under the assumption C, the relative bias is very low in β_1 but higher at the other coefficients, e.g. 24.21% for β_0 and 22.03% for β_2. Moreover, the Standard Errors (SE) of estimations in process are quite small, indicating the convergence of the posterior since most of the MCSE values are very small and less than $\frac{1}{20}$ of standard errors of the estimations. The MCSE values for the reference models are much more smaller than the proposed method since there is only a single model is simulated for the estimations, however most of the relative bias ratios of the proposed method are smaller than the reference model. All in all, the higher the missing rate is, the more sensitive the estimations become to the prior assumptions.

4.1.6.1 Convergence Diagnostics Results

The accuracy and the convergence of the RJMCMC process can be argued upon by checking the convergence diagnostics given in Section 2.4.3. Firstly, density plots given in Appendix B.3.1 provide information about the distributions of coefficients. All the plots indicate that the densities of β_0 and β_1 coefficients are right skewed while β_2's density is left skewed. These results strengthen the assumption of choosing

medians as estimators since the mean values are highly affected from skewness.

The stationarity of the MCMC process by missing rate can be seen in the traceplots given in Appendix B.3.3. In Table 4.5, the summary of some MCMC diagnostics results are given. Acceptance rates, $\hat{R}$ values of BGR diagnostic, thinning values and dependence factors (I) of Raftery-Lewis diagnostic are given under different parameter assumptions by missing rate.

Table 4.5: MCMC Diagnostics For Validation Study Under Different Assumptions by Missing Rate

Missing Rate	**Prior Assumptions**	**Acceptance Rate**	$\hat{R}$ **Value**	**Thinning**	**Raftery Dependence Factor (chain 1 - chain2)**		
					β_0	β_1	β_2
1%	A: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1^2)$	0.169	1	50	1.10 - 1.11	1.17 - 1.15	1.12 - 1.14
	B: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.480	1	20	1.22 - 2.08	1.18 - 2.17	1.23 - 2.04
	C: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1^2)$	0.380	1	30	1.19 - 1.17	1.20 - 1.23	1.19 - 1.15
	D: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.685	1	20	1.01 - 0.99	0.99 - 1.01	1.00 - 1.02
3%	A: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1^2)$	0.543	1	20	1.14 - 1.17	1.15 - 1.16	1.15 - 1.16
	B: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.710	1	10	1.01 - 1.00	1.01 - 1.00	1.00 - 1.00
	C: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1^2)$	0.279	1	40	1.18 - 1.16	1.14 - 1.20	1.10 - 1.17
	D: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.319	1	40	1.15 - 1.18	1.17 - 1.12	1.13 - 1.12
5%	A: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1^2)$	0.419	1	20	1.23 - 1.24	1.25 - 1.24	1.26 - 1.28
	B: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.712	1	10	1.01 - 1.00	1.00 - 1.01	1.01 - 1.00
	C: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1^2)$	0.558	1	10	1.05 - 1.04	1.04 - 1.05	1.04 - 1.04
	D: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.413	1	20	1.01 - 1.01	1.03 - 1.01	1.03 - 1.02
10%	A: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1^2)$	0.649	1	20	0.99 - 1.00	1.01 - 1.01	1.00 - 1.01
	B: $\beta, \phi \sim N(0, 4^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.430	1	40	1.07 - 1.05	1.04 - 1.03	1.05 - 1.04
	C: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1^2)$	0.610	1	100	0.99 - 1.01	1.02 - 0.99	1.01 - 0.99
	D: $\beta, \phi \sim N(0, 6^2)$ & $X_1 \sim N(-1, 1.5^2)$	0.595	1	10	1.02 - 1.02	1.04 - 1.05	1.05 - 1.04

Examining the Table 4.5, acceptance rates are varying between 16.9% and $71,2\%$, but most of the acceptance rates are within the range of the suggestions of Gelman and Rubin [20]. Higher or lower acceptance rates indicates slow convergence. Thus, it can be said that the RJMCMC algorithm slowly converges for missing rate 1% under assumptions A and D, for 3% under assumptions A and B, for 5% under assumptions B and C, and for 10% A, B and D. Most of these scenarios are also the ones indicating smaller relative bias ratios. Although the convergence is slow, it is established. All the $\hat{R}$ values are 1 and this is considered to be a sign of convergence and proper mix of chains. In addition to these diagnostics, BGR plots and Geweke plots in accordance with different missing rates are given in Appendix B.3.4 and B.3.5, and they

verify the results shared here. For convergence and autocorrelation, Raftery-Lewis diagnostic is used. Thinning values are determined in order to reduce the autocorrelation and dependence factor values. The dependence factors (I) of Raftery-Lewis diagnostics in Table 4.5 are given for two chains and by different missing rates. According to the table, the dependence factors are around 1, which is a sign of almost no autocorrelation and posterior convergence except the values for 1% missing rate under assumptions B. Those dependence factors are also lower than 5, therefore it can still be considered as an indication of low autocorrelation and posterior convergence. Furthermore, autocorrelation plots given in Appendix B.3.2 verify these indications.

4.1.7 Sensitivity Analysis

Since the RJMCMC algorithm in the validation study is based on a large amount of assumptions and since there are a lot of unknown parameters to be estimated, a sensitivity analysis needs to be performed in order to see the sensitive assumptions of the algorithm. During the validation study, more than 60 runs of algorithm are conducted with different scenarios and assumptions. The first substantial observation about sensitivity is to be the assumption of missing covariate characteristics. A small change in assumption affects the results remarkably. As it can be seen in Table 4.4, the relative bias of β_1 estimation for 5% missing rate decreases effectively when the variance is taken as 1^2 instead of 1.5^2. In addition to that, when the mean was assumed to be 0 for X_1 during the simulation trials, the estimations were significantly underestimated, therefore, these results of them were not considered or put in this study. Moreover, it is discovered that the assumptions about coefficients are critical as a means of getting accurate estimations. As the missing rate increases, the larger variances of β and ϕ coefficients are needed for lower relative bias. For instance, in Table 4.4, the relative bias of β_1 coefficient for 10% missing rate decreases as the variance assumption of β and ϕ gradually increase from 4^2 to 6^2. At this point, additional attention is needed regarding the convergence and stationarity of the RJMCMC algorithm since the ability of algorithm's movement among models can be affected by this variation. During the simulation trials, large variances are observed to cause slow convergence because of the large variety of proposed coefficient values, as small variances can also induce slow convergence since it causes a stuck around a set of similar proposed

values. Another observation regarding sensitivity is about the choice of models. The main model is taken as both logit and probit. Logit models are observed to perform better than the probit ones. Moreover, the construction of the model space is also very effective on the estimations. The higher the repetitions in the T-Step Occam's Window method used for filtering the model space, the more convenient models are filtered. The results of the validation study indicates that the algorithm becomes more sensitive to prior assumptions as the missing rate increases. Table 4.4 points out the need of calibration in the variance of coefficients, and the variance of covariate X_1 is order to get more accurate estimations as the missing rate increase.

4.2 Real Data Application

In this section, the methodology is adapted and applied utilizing a real data containing MNAR mechanism. The results are then compared with the results of other studies for the same data in order to see the differences and similarities of the estimations.

4.2.1 Data Description

The data set is taken from Pan-London Assertive Outreach Study, and consists of 3 parts, all of which are conducted in 2003 [77] [6] [57] with two additional studies in 2004 [58] and 2007 [15]. The first study in 2003 (part 1) deals with the characterisation of assertive outreach teams on mental health field in the United Kingdom (UK) and determination of possible distinction within groups [77]. The second study in 2004 (part 2) is conducted in order to observe the staff experience of assertive outreach teams on this field and compare it with the community mental health teams and other different types of assertive outreach teams [6]. The third study in 2007 (part 3) dwells on the characteristics of mental health patients and the effects of the assertive outreach team application in the UK on patients [57]. The additional study conducted in 2004 focuses on the characteristics of staff and patients in order to determine the predictors of voluntary and compulsory admissions of these services in the UK [58]. The last study concentrates on the role of the therapeutic relationship to the outcomes of assertive outreach service in order to provide and keep a sustainable service [15].

The details of the study and the data set are specified in those articles.

The data set is not used entirely in this thesis. The main focus is on modeling the hospitalization of patients with mental illness at the end of a 9-month follow-up. The model and the candidate variables are taken as the same with the study of Kalaylioglu and Ozturk (2013) [37], so a comparison can be made among the estimations in order to evaluate the methodology in this thesis. Therefore, the response variable is binary where $y_i = 1$ stands for the i^{th} patient who is hospitalized at the end of 9-month follow-up period with $y_i = 0$ denoting the i^{th} patient who is not hospitalized. Six candidate variables $(x_1, \ldots, x_6)$ are chosen for modeling hospitalization. x_1 is a discrete variable denoting the number of in-patient days of hospitalization in the last two years prior to baseline assessment, x_2 is the dichotomous variable that standing for the patient status ($1 = new$ or $2 = established$), x_3 is continuous and represents the age, x_4 is another dichotomous variable denoting the gender ($0 = male$ or $1 = female$), x_5 is 4 level categorical variable standing for the ethnicity ($1 = white$, $2 = black\ Caribbean$, $3 = black\ African$ and $4 = other$) of patients, while x_6 being 3 level categorical variable implying the cluster of the assertive outreach team serving patients ($1 = Team\ A$, $2 = Team\ B$ and $3 = Team\ C$). The variable with MNAR mechanism is considered as x_1, since the hospitalized patients with longer in-patient days for two years are expected to hide such information. For the purpose of leaving only one variable with missing values, some simple refinements are applied to other variables. There are 10 missing information pieces in the real data set about ethnicity of the patients, so they are categorized under other status to that end. Also there are 5 missing values concerning the age of patients, which are excluded. There are 50 missing observations for x_1 variable with the other variables being fully observed. The missing rate is $8.7\% \approx 9\%$. It is determined in the 3rd part of the study (2003) that, there is a statistically significant difference between patient status (new or established) and also the established patients in the team of cluster C have not been hospitalized as the established patients in the team of cluster A and B. Therefore, the interaction between patient status and cluster of the assertive outreach team is also taken as a candidate variable.

4.2.2 Construction of The Model

Since there is only one variable with MNAR mechanism, the model consists of two parts, the main model and the missingness model. While the missingness models differ, the main model remains the same for every model in model space.The main model is taken as a logistic model where response variable is Y and covariates $\mathbf{X} = (X_1, X_2, X_3, X_4, X_{51}, X_{52}, X_{53}, X_{61}, X_{62}, X_2X_{61}, X_2X_{62})$ given in Equation 4.34.

$$\begin{aligned} y_i|p_{y_i} \sim & Bernoulli(p_{y_i}) \\ logit(p_{y_i}) = & \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + \beta_3 x_{3_i} + \beta_4 x_{4_i} \\ & + \beta_5 x_{51_i} + \beta_6 x_{52_i} + \beta_7 x_{53_i} \\ & + \beta_8 x_{61_i} + \beta_9 x_{62_i} \\ & + \beta_{10} x_{2_i} x_{61_i} + \beta_{11} x_{2_i} x_{62_i} \end{aligned} \tag{4.34}$$

where X_{51}, X_{52} and X_{53} are the dummy variables for ethnicity white, black Caribbean and black African respectively, X_{61} and X_{62} are the dummy variables for cluster of teams A and B respectively and X_2X_{61} and X_2X_{61} are the interaction terms of patient status and cluster of teams.

Missingness models are also taken as logistic models. Covariates of missingness models differ by models in the model space. The general form of the missingness model is shared in Equation 4.35.

$$\begin{aligned} r_i|p_{r_{k_i}} \sim & Bernoulli(p_{r_{k_i}}) \\ logit(p_{r_{k_i}}) = & \mathbf{w}_{k_i}^t \boldsymbol{\phi_k} \end{aligned} \tag{4.35}$$

where $R = r_i$ is the missingness indicator ($r_i = 0$ if missing and $r_i = 1$ if observed), $\mathbf{W}_k$ is the covariate vector, and $\boldsymbol{\phi_k}$ is the coefficient vector of k^{th} missingness model.

The joint posterior given in Equation 4.3 can be applied with the same factorization and setup for the real data study. Only D^{obs} vector is formed additionally for the real data and it consists of all the covariates below:
$D^{obs} = ({X_1}^{obs}, X_2, X_3, X_4, X_{51}, X_{52}, X_{53}, X_{61}, X_{62}, X_2X_{61}, X_2X_{62}, Y, R)$.

The likelihoods for missingness indicator R and for the response Y in Equation 4.3

can be composed as in Equation 4.36 for two logistic regression models.

$$
\begin{aligned}
P(R = r_i | {X_1}^{miss}, D^{obs}, M_k, \boldsymbol{\phi_k}) &= \prod_{i=1}^{573} p_{r_{k_i}}^{r_i} (1 - p_{r_{k_i}})^{1-r_i} \\
P(Y = y_i | {X_1}^{miss}, D^{obs}, M_k, \boldsymbol{\beta_k}) &= \prod_{i=1}^{573} p_{y_{k_i}}^{y_i} (1 - p_{y_{k_i}})^{1-y_i}
\end{aligned}
\tag{4.36}
$$

Prior distribution of X_1 is described as
$P({X_1}^{miss}|D^{obs}, M_k) = P({x_{1,1}}^{miss}, \ldots, x_{1,n_{miss}}^{miss}|D^{obs}, M_k)$. With X_1 being discrete, number of in-patient days is considered to fit a Poisson distribution. Since there are too many zero values available in the data, prior distribution for X_1 is considered to be a Zero Inflated Poisson (ZIP) model, of which distribution is given in Equation 4.37.

$$
\begin{aligned}
x_{1_i} &\sim ZIP(\hat{\pi}, \hat{\lambda}) \\
P(X_1 = x_{1_i}) &= \begin{cases} \hat{\pi} + (1 - \hat{\pi}) e^{-\hat{\lambda}} & x_{1_i} = 0 \\ (1 - \hat{\pi}) \dfrac{\hat{\lambda}^{x_{1_i}} e^{-\hat{\lambda}}}{x_{1_i}!} & x_{1_i} > 0 \end{cases}
\end{aligned}
\tag{4.37}
$$

where $\hat{\pi}$ is the estimated mean of zero inflation probability, and $\hat{\lambda}$ is the estimated mean of average number of in-patient days for missing part of X_1. $\hat{\pi}$ and $\hat{\lambda}$ are calculated for $n_{miss} = 50$ with the following formulas:

$$
\hat{\pi} = \frac{\sum_{i=1}^{n_{miss}} logit(\pi_i)}{n_{miss}} \text{ and } logit(\pi_i) = \mathbf{x}_i^t \boldsymbol{\gamma} \tag{4.38}
$$

$$
\hat{\lambda} = \frac{\sum_{i=1}^{n_{miss}} log(\lambda_i)}{n_{miss}} \text{ and } log(\lambda_i) = \mathbf{x}_i^t \boldsymbol{\alpha} \tag{4.39}
$$

Age (X_3) and gender (X_4) of patients are taken as covariates for the ZIP model (i.e $\mathbf{x}_i = (x_{3,i}, x_{4,i})$) in order to estimate the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. They are estimated by using the complete part of the data set and by using the same ZIP model given in Equation 4.37. The model summary is provided in Table 4.6.

Table 4.6: ZIP Model Coefficients by Using Observed Data

	Coefficient	
	$\boldsymbol{\gamma}$	$\boldsymbol{\alpha}$
(Intercept)	-2.0086	4.8050
x_3	0.0335	0.0066
x_4	-0.1878	0.0348

According to Table 4.6, $logit(\pi_i)$ and $log(\lambda_i)$ model coefficients are structured below:

$$\begin{aligned} log(\lambda_i) &= \gamma_0 + \gamma_1 x_{3_i} + \gamma_2 x_{4_i} \\ &= -2,0086 + 0,0335 x_{3_i} - 0,1878 x_{4_i} \\ logit(\pi_i) &= \alpha_0 + \alpha_1 x_{3_i} + \alpha_2 x_{4_i} \\ &= 4,8050 + 0,0066 x_{3_i} + 0,0348 x_{4_i} \end{aligned}$$

Prior distributions for $\boldsymbol{\beta_k}$, $\boldsymbol{\phi_k}$ and M_k are taken exactly the same as in the validation study on a simulated dataset. See the Equations 3.10, 3.11 and 4.8. Only the sizes of the coefficients are adapted accordingly for this analysis.

Full conditional distributions of parameters X_1, β_k and ϕ_k used for Gibbs sampling algorithm in validation study (Equation 4.13) are updated according to the real data analysis. Updated full conditional distribution of X_1 is given as follows:

$$\begin{aligned} f(x_{1_i}^{miss} | D_{obs_i}, \boldsymbol{\beta_k}, \boldsymbol{\phi_k}, \mu_{x_1}, \sigma_{x_1}^2) =& (p_{y_i})^{y_i} (1-p_{y_i})^{1-y_i} \\ & * (p_{r_{k_i}})^{r_i} (1-p_{r_{k_i}})^{1-r_i} \\ & * \begin{cases} \hat{\pi} + (1-\hat{\pi}) e^{-\hat{\lambda}} & \text{if } x_{1_i}^* = 0 \\ (1-\hat{\pi}) \dfrac{\hat{\lambda}^{x_{1_i}^*} e^{-\hat{\lambda}}}{x_{1_i}^*!} & \text{if } x_{1_i}^* > 0 \end{cases} \end{aligned} \tag{4.40}$$

Since the main model for the response is logistic, full conditional distribution for β coefficients remains the same as in the validation study, however the missingness models are taken as logistic in the real data analysis, even though they were considered probit in the validation study. Therefore, full conditional distribution for ϕ_k is updated in the way of the full conditional distribution of β.

The same auxiliary variable, used in the validation study, $\boldsymbol{Z_{\phi_k}} = (z_{\phi_{k,1}}, \ldots, z_{\phi_{k,n}})$ is defined in Equation 4.41 in which $z_{\phi_{k,i}} = \boldsymbol{w_{k_i}}\boldsymbol{\phi}_k + \epsilon_i$ with $\boldsymbol{w_{k_i}}$ being the covariate vector of k^{th} missingness model along with $\epsilon_i \sim N(0, \lambda_i)$ and $i = 1, \ldots, n$.

$$\begin{aligned} \boldsymbol{\phi}_k | X_1^{miss}, M_k, \boldsymbol{Z_{\phi_k}} &\sim N(\tilde{\boldsymbol{\mu}}_{\phi_k}, \tilde{\boldsymbol{\Sigma}}_{\phi_k}) \\ \tilde{\boldsymbol{\mu}}_{\phi_k} &= \tilde{\boldsymbol{\Sigma}}_{\phi_k}(\boldsymbol{\Sigma}_\phi^{-1}\boldsymbol{\mu}_\phi + \boldsymbol{W_k^t B Z_{\phi_k}}) \\ \tilde{\boldsymbol{\Sigma}}_{\phi_k} &= (\boldsymbol{\Sigma}_\phi^{-1} + \boldsymbol{W_k^t B W_k})^{-1} \end{aligned} \tag{4.41}$$

where $\boldsymbol{B} = diag(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$.

Acceptence probabilities for RJMCMC are constructed to be the same as the validation part in Section 4.1.4. The model space construction is also the same as its counterpart in the validation study in Section 4.1.5. Model space table, containing 13 models, is shared in Table 4.7.

Table 4.7: Model Space with Model Posterior Probabilities of the Real Data Analysis

	Covariate												
Model No	X_1	X_2	X_3	X_4	X_{51}	X_{52}	X_{53}	X_{61}	X_{62}	X_2X_{61}	X_2X_{62}	Y	$\mathbf{P(M_k\|D)}$
1	•												0.0000
2	•				•								0.0241
3	•						•						0.0470
4	•							•					0.0471
5	•								•				0.0472
6	•									•			0.0478
7	•		•		•								0.1199
8	•			•	•								0.1198
9	•			•						•			0.1201
10	•				•		•						0.1198
11	•				•			•					0.1201
12	•				•					•			0.1206
13	•				•		•			•			0.0666

Covariates of the models are represented through dots along with the computations of

model posterior probabilities ($\mathbf{P(M_k|D)}$) for each model. In Table 4.7, it can be seen that, the model posterior probability of the 1^{st} model containing only the covariate X_1 is zero, meaning that only observed X_1 values would not be enough to explain the missingness indicator of MNAR. Although the most saturated model is the 13^{th} model with 4 covariates (5 coefficients including the intercept), posterior probability of the models through 7 to 12 are higher. After the construction of the model space, transition probabilities are computed according to that model space and it is provided in a matrix form in Table 4.8.

Table 4.8: Transition Probability Matrix for The Model Space of the Real Data Analysis

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.1250	0.0938	0.0938	0.0938	0.0938	0.0938	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0313
2	0.0698	0.0930	0.0930	0.0930	0.0930	0.0930	0.0698	0.0698	0.0698	0.0698	0.0698	0.0698	0.0465
3	0.0698	0.0930	0.0930	0.0930	0.0930	0.0930	0.0698	0.0698	0.0698	0.0698	0.0698	0.0698	0.0465
4	0.0698	0.0930	0.0930	0.0930	0.0930	0.0930	0.0698	0.0698	0.0698	0.0698	0.0698	0.0698	0.0465
5	0.0698	0.0930	0.0930	0.0930	0.0930	0.0930	0.0698	0.0698	0.0698	0.0698	0.0698	0.0698	0.0465
6	0.0698	0.0930	0.0930	0.0930	0.0930	0.0930	0.0698	0.0698	0.0698	0.0698	0.0698	0.0698	0.0465
7	0.0455	0.0682	0.0682	0.0682	0.0682	0.0682	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0682
8	0.0455	0.0682	0.0682	0.0682	0.0682	0.0682	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0682
9	0.0455	0.0682	0.0682	0.0682	0.0682	0.0682	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0682
10	0.0455	0.0682	0.0682	0.0682	0.0682	0.0682	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0682
11	0.0455	0.0682	0.0682	0.0682	0.0682	0.0682	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0682
12	0.0455	0.0682	0.0682	0.0682	0.0682	0.0682	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0682
13	0.0303	0.0606	0.0606	0.0606	0.0606	0.0606	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.1212

After calculating the transition probabilities, RJMCMC algorithm can now be applied. The algorithm used in the validation study is adapted to the real data application with minor changes but the structure of the algorithm remains the same.

4.2.3 Results

BMP algorithm of real data analysis is conducted with `RStudio` software, based on `R` programming software as used in the validation study. Two chains are obtained in order to assure convergence under different prior assumptions (A: $\beta, \phi \sim \mathcal{N}(0, 1^2)$ or B: $\beta, \phi \sim \mathcal{N}(0, 1.5^2)$ or C: $\beta, \phi \sim \mathcal{N}(0, 2^2)$). Iteration size is taken between

300,000 and 500,000 with additional 30,000 or 50,000 burn-in size. Only the results of the estimations under prior assumptions A are consistent with the results of the reference study, the results of B and C are tabulated in the sensitivity analysis part. For assumption A, every 5^{th} draws are taken for posterior inference, therefore 60,000 posterior draws are used in total. Initial values of β_j coefficients are taken as the point estimates given in Kalaylioglu and Ozturk (2013) [37]. MCMC diagnostics are carried out to ensure the convergence and accurate estimates. For assumption A, the acceptance ratio of RJMCMC is observed 84.3% and $\hat{R}$ values for all coefficients equal to 1. All the Raftery Dependency Factor (I) values for coefficients are found within the range of 1.04 and 1.07, indicating that there is hardly any autocorrelation among posterior draws. Additional to these diagnostics, trace plots, autocorrelation plots, BGR plots and MCSE values of coefficients are provided in Appendix C.

Table 4.9: Real Data Analysis Estimates (95% Posterior Interval Estimates)

	Estimate (95% Posterior Interval Estimates)		
	Proposed Method (BMP)	**Only Observed Data**	**Kalaylioglu & Ozturk (2013)[1]**
β_0	-0.6979 (-1.462, 0.067)	-0.9738 (-1.784, -0.164)	-1.360 (-2.317, -0.428)
$\beta_1^\star$	0.0019 (-0.004, 0.008)	0.0023 (0.001, 0.004)	0.003 (0.002, 0.005)
β_2	1.1623 (-0.680, 3.005)	1.3918 (0.495, 2.289)	0.777 (-0.439, 1.941)
β_3	-0.0005 (-0.021, 0.020)	-0.0037 (-0.020, 0.013)	-0.012 (-0.029, 0.006)
β_4	0.2987 (-0.387, 0.985)	0.1229 (-0.259, 0.505)	0.213 (-0.197, 0.634)
β_5	-0.0550 (-0.864, 0.754)	0.1847 (-0.292, 0.662)	0.200 (-0.390, 0.799)
β_6	0.2794 (-0.182, 0.741)	0.2067 (-0.293, 0.707)	0.013 (-0.675, 0.723)
β_7	-0.1089 (-0.883, 0.665)	0.1439 (-0.482, 0.770)	0.069 (-0.715, 0.854)
β_8	-0.1196 (-0.629, 0.390)	0.1518 (-0.378, 0.682)	0.713 (0.076, 1.373)
β_9	0.5693 (-0.450, 1.589)	0.4446 (-0.246, 1.136)	0.941 (0.151, 1.747)
β_{10}	-0.5500 (-2.294, 1.194)	-0.738 (-1.758, 0.282)	-0.195 (-1.471, 1.105)
β_{11}	-1.3055 (-3.383, 0.772)	-1.3451 (-2.690, 0.001)	-0.918 (-2.486, 0.678)

* Coefficient belonging covariate with MNAR

[1] Results are taken from Table 4 for Model I [37].

Results of the real data analysis are given in Table 4.9 together with the estimations of only observed data and the estimations given in the article of Kalaylioglu and Ozturk (2013). The same data set is used for these estimations [37].

Only β_1 coefficient belongs to the covariate with MNAR, the rest belongs to the fully observed covariates. MNAR mechanism of in-patient days of hospitalization is based on the opinion that the hospitalized patients with longer in-patient days for two years are prone to hide such information. Therefore, the impact of this covariate to the response is likely to be less, which means the β_1 coefficient is expected to be lower than the estimates from the observed part. This opinion is supported with the results too: β_1 estimate of BMP (0.0019) is less than its counterpart one with an observed part (0.0023). In the study of Kalaylioglu and Ozturk (2013), the same coefficient is estimated higher (0.003) than BMP. The reason of this may be the difference between the sample sizes ($n_{Kalaylioglu} = 476 < n_{BMP} = 573$). Nevertheless, the point estimates of BMP are in the ranges of 95% posterior interval estimates given in Kalaylioglu and Ozturk. Moreover, the results can be compared in terms of ranges of posterior interval estimates. The ranges in BMP are broader than the ones of Kalaylioglu and Ozturk in most of the posterior interval estimates, especially the one belonging MNAR covariate is three times wider.

4.2.4 Sensitivity Analysis for Real Data Application

As shown in the validation study, the higher the missing rate, the more sensitive is the method to the prior assumptions. Therefore in real data applications of BMP, we recommend a sensitivity analysis and illustrate it here on the dataset at hand. In the real data application, the main model contains 12 coefficients, while the missingness models contain at most 5 coefficients. This results in a maximum of 67 unknown parameters to be estimated including 50 missing values. When the total number of parameters to be estimated and missing values is high with respect to the size of the sample, it is observed that the computation duration increases significantly. 3 different prior assumptions on model coefficients β and ϕ are considered: A: $\beta, \phi \sim \mathcal{N}(0, 1^2)$, B: $\beta, \phi \sim \mathcal{N}(0, 1.5^2)$ and C: $\beta, \phi \sim \mathcal{N}(0, 2^2)$. The convergence diagnostics of the chains under the three different priors show similar behaviours and they converge. However, the acceptance ratios for B and C are lower than that for A. The results are provided in Table 4.10.

Accordingly, the posterior results, particularly for β_1 (the MNAR coefficient), are

Table 4.10: Estimates with MCSE of The Real Data Analysis under 3 Different Prior Assumptions

	Estimates (MCSE)		
	Prior Assumption		
	A: $\beta, \phi \sim \mathcal{N}(0, 1^2)$	B: $\beta, \phi \sim \mathcal{N}(0, 1.5^2)$	C: $\beta, \phi \sim \mathcal{N}(0, 2^2)$
$\hat{\beta}_0$	-0.6979 (0.0011)	-1.2555 (0.0114)	-1.3264 (0.0111)
$\hat{\beta}_1$	0.0019 (0.0000)	0.9622 (0.0076)	1.3337 (0.0082)
$\hat{\beta}_2$	1.1623 (0.0027)	0.8629 (0.0119)	0.7847 (0.0123)
$\hat{\beta}_3$	-0.0005 (0.0000)	-0.0489 (0.0081)	0.0579 (0.0091)
$\hat{\beta}_4$	0.2987 (0.0010)	0.1500 (0.0110)	0.2442 (0.0117)
$\hat{\beta}_5$	-0.0550 (0.0012)	0.1757 (0.0126)	0.1832 (0.0113)
$\hat{\beta}_6$	0.2794 (0.0007)	0.0641 (0.0111)	0.0132 (0.0118)
$\hat{\beta}_7$	-0.1089 (0.0011)	0.1123 (0.0110)	0.1084 (0.0118)
$\hat{\beta}_8$	-0.1196 (0.0008)	0.6942 (0.0105)	0.7272 (0.0108)
$\hat{\beta}_9$	0.5693 (0.0016)	0.9656 (0.0111)	0.9319 (0.0114)
$\hat{\beta}_{10}$	-0.5500 (0.0025)	-0.1288 (0.0107)	-0.1921 (0.0112)
$\hat{\beta}_{11}$	-1.3055 (0.0031)	-0.9768 (0.0114)	-0.9413 (0.0121)

sensitive to the prior distributions on β and ϕ. Furthermore, the MCSEs being considerably lower for A than B and C indicates that prior variance should be kept at a minimum level in such datasets.

Lastly, the repetitions of the T-Step Occam's window method should be much higher when the unknown parameters are as many as in this case. However, this time, the computation duration increases highly, in addition to the increase with respect to the unknown parameters.

CHAPTER 5

CONCLUSION AND DISCUSSION

In this thesis, a methodology of a hybrid Bayesian modeling system, called Bayesian Model Pooling (BMP), is introduced for handling the uncertainty caused by Missing-Not-At-Random (MNAR) mechanism in covariates of a GLM model in addition to the uncertainty in model selection within the frame of Bayesian inference. When the missingness mechanism of a variable is MNAR, an attention should be paid to the inference about that variable in order to obtain accurate estimations, since MNAR is known to cause bias on estimations. Moreover, determining the best fit for a model containing covariates with MNAR also increases the uncertainty about the model selection which causes another problem to be carefully examined. In this study, the variables with MNAR are considered to be the covariates of a generalized linear model (GLM) with a dichotomous response, and the missingness mechanism is handled with a model space consisting GLM candidates too. Structure of the joint model containing the main model and the missingness model, construction of the model space and all the inference for variables in the joint model are completed in Bayesian framework. Firstly, a proper model space containing the best candidate models is filtered according to an adaptive Occam's window method called T-Step Occam's Window. Then, posterior distributions of model coefficients for every model in the model space are achieved by using Reversible Jump Markov Chain Monte Carlo (RJMCMC) approach, which is an extension of classical Metropolis-Hastings (M-H) sampling algorithm for varying model dimensions. After that, coefficient estimates are obtained by pooling the posterior estimations for each model, of which their posterior probabilities are also calculated within RJMCMC algorithm.

In Chapter 2, preliminaries for constructing the methodology of the BMP are elu-

cidated. First of all, the definitions of generalized linear models, logistic and probit regression are given. Later, the definition of missing data, types of missingness mechanisms and handling methods for those types are provided in detail. Next, Bayesian inference is elaborated including the basic Bayesian approach, Markov Chain Monte Carlo (MCMC) approach with its diagnostics and Bayesian model selection methods such as Bayes factor, Bayesian Information Criteria (BIC), Bayesian model averaging (BMA) and RJMCMC.

In Chapter 3, the structure of the proposed hybrid Bayesian modeling system (BMP) methodology is provided in detail. Firstly, the joint posterior model, containing the main model (GLM), the missingness models (GLM) for the missingness indicators, conditional priors of covariates with MNAR, priors for the unknown model parameters and model priors in the model space, is constructed in accordance with underlying concepts given in Chapter 2. Selection model method is chosen for factorization of the joint posterior. After that, the constructions of model spaces are described for all missingness indicators. Only a certain amount of best fits are filtered for model spaces according to the T-Step Occam's Window algorithm in which the BIC values are used in order to attain model probabilities. Later, the method of constructing transition probability matrices, used in RJCMCM algorithm, are shared in detail. Then, RJMCMC algorithm are adapted as a model selection method, and it is followed by a clear explanation of the usage of joint posteriors along with the transition probabilities in this algorithm in order to achieve posterior distributions of coefficients. Lastly, the methodology is completed by providing how the coefficient estimates are pooled by using the posterior model probabilities as model weights.

In Chapter 4, a validation study on a simulated dataset with known true parameters and a real data analysis are performed. The validation study on a simulated dataset is conducted under different scenarios in order to examine the performance of the proposed methodology. It is noticed that the performance highly depends on missing rate and assumptions for parameters of missing values and model coefficients. The results of the sensitivity analysis indicate that the higher the missing rate is, the more sensitive the algorithm becomes. The reason of this is the increase in the number of unknown parameters, since every missing value is treated as an unknown parameter. Thus, the algorithm tries to estimate additional unknown parameters as

the missing rate increases. This result leads us to pay attention to other assumptions in the validation study. As the missing rate increases, the variance assumptions of the missing values and model coefficients should decrease and they should be adjusted carefully to get accurate estimations. The results are considered satisfying when the assumptions are set carefully, however computation duration may be a disadvantage, since a simulation run took around 20 - 24 hours with a mid-range built PC (AMD Ryzen 3600 Processor running 3.6GHz, using 16GB RAM, running Windows 10) in order to assure posterior convergence. A further study can be designed focusing on the enhancement of the algorithm performance and acceleration of convergence. Furthermore, the data analysis of a real life problem is conducted to ensure that the methodology could contribute solving of a real life experiment. All the setup of the analysis are based on the study of Kalaylioglu and Ozturk (2013) for comparison. The results are also found satisfying and relatively close to the reference study, although the sample sizes of analyses were not the same.

For further study, the methodology can be extended to more complex model spaces including different kinds of models such as generalized mixed models, longitudinal models, etc. A model space constructed with expert knowledge can be more useful in real life problems in order to get more accurate and realistic results. Also, this study can be extended for handling more than one covariate with MNAR or it can be adapted for handling response variable with MNAR.

REFERENCES

[1] A. Agresti. *Categorical data analysis*, volume 482. John Wiley & Sons, 2003.

[2] A. Agresti. *Foundations of Linear and Generalized Linear Models*. Wiley Series in Probability and Statistics. Wiley, 2015.

[3] J. Ai. Reversible-jump mcmc methods in bayesian statistics (unpublished master's thesis). Master's thesis, 2012.

[4] T. Ando. *Bayesian model selection and statistical modeling*. CRC Press, 2010.

[5] T. Baghfalaki and E. F. Jalali. Reversible jump mcmc to identify dropout mechanism in longitudinal data. *Communications in Statistics-Theory and Methods*, 48(11):2717–2733, 2019.

[6] J. Billings, S. Johnson, P. Bebbington, A. Greaves, S. Priebe, M. Muijen, I. Ryrie, J. Watts, I. White, and C. Wright. Assertive outreach teams in london: Staff experiences and perceptions: Pan-london assertive outreach study, part 2. *The British Journal of Psychiatry*, 183(2):139–147, 2003.

[7] L. Bouranis, N. Friel, and F. Maire. Model comparison for gibbs random fields using noisy reversible jump markov chain monte carlo. *Computational Statistics & Data Analysis*, 128:221–241, 2018.

[8] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. Handbook of markov chain monte carlo: Hardcover: 619 pages publisher: Chapman and hall/crc press (may 2011) language: English isbn-10: 1420079417, 2012.

[9] S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

[10] M. Cowles. *Applied Bayesian Statistics: With R and OpenBUGS Examples*. Springer Texts in Statistics. Springer New York, 2013.

[11] A. K. Dey and H. Jhamb. A novel empirical bayes with reversible jump markov chain in user-movie recommendation system. *arXiv preprint arXiv:1808.05480*, 2018.

[12] A. J. Dobson and A. G. Barnett. *An introduction to generalized linear models*. Chapman and Hall/CRC, 2018.

[13] I. Eekhout, R. M. de Boer, J. W. Twisk, H. C. de Vet, and M. W. Heymans. Missing data: a systematic review of how they are reported and handled. *Epidemiology*, 23(5):729–732, 2012.

[14] C. K. Enders. *Applied missing data analysis*. Guilford press, 2010.

[15] W. K. Fakhoury, I. White, S. Priebe, P. S. Group, et al. Be good to your patient: how the therapeutic relationship in the treatment of patients admitted to assertive outreach affects rehospitalization. *The Journal of nervous and mental disease*, 195(9):789–791, 2007.

[16] Y. Fan and S. A. Sisson. Reversible jump mcmc. *Handbook of Markov Chain Monte Carlo*, pages 67–92, 2011.

[17] F. Fang, J. Zhao, and J. Shao. Imputation-based adjusted score equations in generalized linear models with nonignorable missing covariate values. *Statistica Sinica*, 28(4):1677–1701, 2018.

[18] J.-E. Galimard, S. Chevret, C. Protopopescu, and M. Resche-Rigon. A multiple imputation approach for mnar mechanisms compatible with heckman's model. *Statistics in medicine*, 35(17):2907–2920, 2016.

[19] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.

[20] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. CRC press, 2013.

[21] A. Gelman, D. B. Rubin, et al. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

[22] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984.

[23] J. Geweke et al. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, 1991.

[24] P. J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[25] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

[26] J. J. Heckman. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4*, pages 475–492. NBER, 1976.

[27] A. H. Herring, J. G. Ibrahim, and S. R. Lipsitz. Non-ignorable missing covariate data in survival analysis: a case-study of an international breast cancer study group trial. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(2):293–310, 2004.

[28] A. Hoegh, I. Crandell, S. Klopfer, and M. Fies. Model selection with missing covariates for policy considerations in fox enclosures. *Journal of Applied Statistics*, 44(15):2645–2658, 2017.

[29] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401, 1999.

[30] C. C. Holmes, L. Held, et al. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian analysis*, 1(1):145–168, 2006.

[31] L. Huang, M.-H. Chen, and J. G. Ibrahim. Bayesian analysis for generalized linear models with nonignorably missing covariates. *Biometrics*, 61(3):767–780, 2005.

[32] J. G. Ibrahim, M.-H. Chen, S. R. Lipsitz, and A. H. Herring. Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, 100(469):332–346, 2005.

[33] J. G. Ibrahim, S. R. Lipsitz, and M.-H. Chen. Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):173–190, 1999.

[34] I. Jansen, N. Hens, G. Molenberghs, M. Aerts, G. Verbeke, and M. G. Kenward. The nature of sensitivity in monotone missing not at random models. *Computational statistics & data analysis*, 50(3):830–858, 2006.

[35] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461, 1946.

[36] H. Jeffreys. *The theory of probability*. OUP Oxford, 1998.

[37] Z. Kalaylioglu and O. Ozturk. Bayesian semiparametric models for nonignorable missing mechanisms in generalized linear models. *Journal of Applied Statistics*, 40(8):1746–1763, 2013.

[38] O. Karakuş. *Generalised Bayesian model selection using reversible jump Markov chain Monte Carlo*. PhD thesis, 2017.

[39] R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[40] J. K. Kim and C. L. Yu. A semi-parametric approach to fractional imputation for nonignorable missing data. *Survey Research Methods Proceedings*, pages 2603–2610, 2009.

[41] R. King, B. Morgan, O. Gimenez, and S. Brooks. *Bayesian analysis for population ecology*. Chapman and Hall/CRC, 2009.

[42] M. Kwak and M. Kim. Finding best model in multiple regression applying reversible jump mcmc. *Journal of the Korean Data Analysis Society (August 2019)*, 21(4):1675–1692, 2019.

[43] E. Lesaffre and A. B. Lawson. *Bayesian biostatistics*. John Wiley & Sons, 2012.

[44] S. R. Lipsitz, J. G. Ibrahim, M.-H. Chen, and H. Peterson. Non-ignorable missing covariates in generalized linear models. *Statistics in medicine*, 18(17-18):2435–2448, 1999.

[45] S. R. Lipsitz, M. Parzen, and M. Ewell. Inference using conditional logistic regression with missing covariates. *Biometrics*, pages 295–303, 1998.

[46] R. J. Little and D. B. Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

[47] D. J. Lunn, J. C. Whittaker, and N. Best. A bayesian toolkit for genetic association studies. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, 30(3):231–247, 2006.

[48] D. Madigan, J. York, and D. Allard. Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*, pages 215–232, 1995.

[49] A. Mason, N. Best, S. Richardson, and I. PLEWIS. Strategy for modelling non-random missing data mechanisms in observational studies using bayesian methods. 2010.

[50] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[51] G. Molenberghs, C. Beunckens, C. Sotto, and M. G. Kenward. Every missing not at random model has got a missing at random bodyguard. *Journal of the Royal Statistical Society, Series B, Submitted*, 2007.

[52] G. Molenberghs and M. Kenward. *Missing data in clinical studies*, volume 61. John Wiley & Sons, 2007.

[53] B. Nandram and N. Woo. A bayesian uncertainty analysis for nonignorable nonresponse. *Current Trends in Bayesian Methodology with Applications*, pages 415–433, 2015.

[54] C. Niu, X. Guo, W. Xu, and L. Zhu. Empirical likelihood inference in linear regression with nonignorable missing response. *Computational Statistics & Data Analysis*, 79:91–112, 2014.

[55] I. Ntzoufras. Aspects of bayesian model and variable selection using mcmc. 1999.

[56] İ. Özmen and H. Demirhan. A bayesian approach for zero-inflated count regression models by using the reversible jump markov chain monte carlo method and an application. *Communications in Statistics—Theory and Methods*, 39(12):2109–2127, 2010.

[57] S. Priebe, W. Fakhoury, J. Watts, P. Bebbington, T. Burns, S. Johnson, M. Muijen, I. Ryrie, I. White, and C. Wright. Assertive outreach teams in london: patient characteristics and outcomes: Pan-london assertive outreach study, part 3. *The British Journal of Psychiatry*, 183(2):148–154, 2003.

[58] S. Priebe, W. Fakhoury, I. White, J. Watts, P. Bebbington, J. Billings, T. Burns, S. Johnson, M. Muijen, I. Ryrie, et al. Characteristics of teams, staff and patients: associations with outcomes of patients in assertive outreach. *The British Journal of Psychiatry*, 185(4):306–311, 2004.

[59] A. E. Raftery. Hypothesis testing and model. *Markov chain Monte Carlo in practice*, page 163, 1995.

[60] A. E. Raftery and S. M. Lewis. [practical markov chain monte carlo]: comment: one long run with diagnostics: implementation strategies for markov chain monte carlo. *Statistical science*, 7(4):493–497, 1992.

[61] A. E. Raftery and Y. Zheng. Discussion: Performance of bayesian model averaging. *Journal of the American Statistical Association*, 98(464):931–938, 2003.

[62] R. R. Regal and E. B. Hook. The effects of model selection on confidence intervals for the size of a closed population. *Statistics in Medicine*, 10(5):717–721, 1991.

[63] B. J. Reich and D. Bandyopadhyay. A latent factor model for spatial data with informative missingness. *The annals of applied statistics*, 4(1):439, 2010.

[64] S. Richardson and P. J. Green. On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)*, 59(4):731–792, 1997.

[65] C. P. Robert. Simulation of truncated normal variables. *Statistics and computing*, 5(2):121–125, 1995.

[66] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2005.

[67] P. L. Roth. Missing data: A conceptual review for applied psychologists. *Personnel psychology*, 47(3):537–560, 1994.

[68] D. B. Rubin. *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons, 2004.

[69] G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[70] S. K. Sinha. Robust methods for generalized linear models with nonignorable missing covariates. *Canadian Journal of Statistics*, 36(2):277–299, 2008.

[71] S. A. Sisson. Transdimensional markov chains: A decade of progress and future perspectives. *Journal of the American Statistical Association*, 100(471):1077–1089, 2005.

[72] I. Skarga-Bandurova, T. Biloborodova, and Y. Dyachenko. Strategy to managing mixed datasets with missing items. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 608–620. Springer, 2018.

[73] A. Theorell and K. Nöh. Reversible jump mcmc for multi-model inference in metabolic flux analysis. *Bioinformatics*, 36(1):232–240, 2020.

[74] K. Van Steen, G. Molenberghs, G. Verbeke, and H. Thijs. A local influence approach to sensitivity analysis of incomplete longitudinal ordinal data. *Statistical Modelling*, 1(2):125–142, 2001.

[75] G. Verbeke, G. Molenberghs, H. Thijs, E. Lesaffre, and M. G. Kenward. Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics*, 57(1):7–14, 2001.

[76] R. Waagepetersen and D. Sorensen. A tutorial on reversible jump mcmc with a view toward applications in qtl-mapping. *International Statistical Review*, 69(1):49–61, 2001.

[77] C. Wright, T. Burns, P. James, J. Billings, S. Johnson, M. Muijen, S. Priebe, I. Ryrie, J. Watts, and I. White. Assertive outreach teams in london: models of operation: Pan-london assertive outreach study, part 1. *The British Journal of Psychiatry*, 183(2):132–138, 2003.

[78] Y.-T. Yeh, L. Yang, M. Watson, N. D. Goodman, and P. Hanrahan. Synthesizing open worlds with constraints using locally annealed reversible jump mcmc. *ACM Transactions on Graphics (TOG)*, 31(4):1–11, 2012.

[79] Y. Yuan. Sensitivity analysis in multiple imputation for missing data. in proceedings of the sas global forum 2014 conference, 2014.

[80] N. Zhang. *Ignorable and Nonignorable Modeling in Regression with Incomplete Covariates.* PhD thesis, 2011.

[81] J. Zhao and J. Shao. Semiparametric pseudo-likelihoods in generalized linear models with nonignorable missing data. *Journal of the American Statistical Association*, 110(512):1577–1590, 2015.

Appendix A

APPENDIX: R CODES

A.1 ϕ_0 Determination

```
# necessery functions
# mode function
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
# find function
rfind <- function(x)seq(along=x)[as.logical(x)]

seeds  <- 689659;
seedsn <- 898734;

runif(seeds);
rnorm(seedsn);

n=250; # sample size
mcsize=1; # monte carlo simulation size

# initial parameters for data generation
beta = c(2, 1, -1);
alpha1 = c(-1.5, 0.5,0.75);
alpha2 = c(1, 1);
```

```
# initial parameters for 15% missingness
phi1 = c(1, 1, 1);
phi2 = c(1, 1, 1.0, 1, 1);

x2<-rnorm(n, alpha2[1],alpha2[2]);
mux1=alpha1[1]+alpha1[2]*x2;
x1 <- rnorm(n, mux1, alpha1[3]);

# generate y from binomial
numpy=exp( beta[1]+ beta[2]*x1 + beta[3]*x2 );
py=numpy/(1+numpy);
y=rbinom(n,1,py);

# missingness percentage
misperc=0.01 #or 0.03 or 0.05 or 0.10

phi0i = seq(-50, 50, by=0.05)
bestphi0=matrix(0, nrow = 1000, ncol = 1);

for(j in 1:1000){
missprob = matrix(0, nrow = length(phi0i), ncol = 1);
r = matrix(0, nrow = n, ncol = 1);
w = matrix(0, nrow = n, ncol = 1);

for(k in 1:length(phi0i)){
numpr = phi0i[k]+phi1[1]*y+phi1[2]*x2+phi1[3]*x1
w = rnorm(n,numpr,1);
r[rfind(w>=0)] = 1;
r[rfind(w<0)] = 0;
missprob[k] = mean(r);
}
```

```
misprobdif<-abs(missprob-misperc);
A=cbind(phi0i,misprobdif);
minA<-A[order(misprobdif)];
bestphi0[j]<-minA[1];
}
getmode(bestphi0)
```

Appendix B

APPENDIX: VALIDATION STUDY

B.1 Model Spaces with Model Posterior Probabilities By Missing Rate

Table B.1: Model Space with Model Posterior Probabilities For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1				•					0.0000
2							•		0.0004
3	•		•						0.0110
4	•			•					0.0094
5	•						•		0.0096
6	•							•	0.0137
7			•				•		0.0074
8				•			•		0.0094
9							•	•	0.0092
10	•		•	•					0.0675
11	•		•				•		0.0697
12	•			•			•		0.0640
13			•	•			•		0.0637
14	•		•	•			•		0.2412
15	•		•	•			•	•	0.4238

Table B.2: Model Space with Model Posterior Probabilities For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1	•								0.0000
2				•					0.0005
3							•		0.0010
4	•		•						0.0162
5	•			•					0.0163
6	•						•		0.0180
7	•							•	0.0156
8			•				•		0.0156
9				•			•		0.0167
10							•	•	0.0167
11	•		•	•					0.1440
12	•		•				•		0.1449
13	•			•			•		0.1453
14			•	•			•		0.1462
15			•				•	•	0.1432
16	•		•	•			•		0.1600

Table B.3: Model Space with Model Posterior Probabilities For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1	•								0.0000
2							•		0.0003
3	•	•							0.0186
4	•		•						0.0194
5	•			•					0.0184
6	•						•		0.0184
7	•							•	0.0191
8			•				•		0.0182
9				•			•		0.0187
10				•				•	0.0182
11	•		•	•					0.1560
12	•		•				•		0.1563
13	•		•					•	0.1532
14	•			•			•		0.1550
15			•	•			•		0.1528
16			•				•	•	0.0774

Table B.4: Model Space with Model Posterior Probabilities For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1	•								0.0000
2				•					0.0005
3							•		0.0013
4	•		•						0.0163
5	•			•					0.0162
6	•						•		0.0172
7			•				•		0.0166
8				•				•	0.0171
9	•		•	•					0.1350
10	•		•				•		0.1328
11	•		•					•	0.1324
12	•			•				•	0.1352
13			•	•			•		0.1357
14			•				•	•	0.1307
15		•	•				•	•	0.1129

Table B.5: Model Space with Model Posterior Probabilities For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$\mathbf{P}(\mathbf{M_k}\|\mathbf{D})$
1	•								0.0000
2				•					0.0007
3							•		0.0016
4	•		•						0.0228
5	•						•		0.0230
6	•							•	0.0230
7		•					•		0.0229
8			•				•		0.0228
9							•	•	0.0230
10	•	•	•						0.1555
11	•		•				•		0.1548
12	•		•					•	0.1571
13	•						•	•	0.1585
14		•					•	•	0.1559
15			•				•	•	0.0787

Table B.6: Model Space with Model Posterior Probabilities For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1	•								0.0000
2				•					0.0007
3							•		0.0009
4	•		•						0.0269
5	•			•					0.0289
6	•						•		0.0267
7	•							•	0.0270
8			•	•					0.0295
9			•				•		0.0275
10				•			•		0.0315
11							•	•	0.0258
12	•		•	•					0.1699
13	•			•			•		0.1799
14			•				•	•	0.1782
15	•		•	•			•		0.2467

Table B.7: Model Space with Model Posterior Probabilities For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1	•								0.0000
2				•					0.0006
3							•		0.0005
4	•		•						0.0231
5	•			•					0.0224
6	•						•		0.0211
7	•							•	0.0239
8			•	•					0.0206
9			•				•		0.0226
10				•		•			0.0231
11						•	•		0.0233
12							•	•	0.0227
13	•		•				•		0.1515
14	•		•					•	0.1492
15	•						•	•	0.1575
16			•	•			•		0.1557
17	•		•				•	•	0.1822

Table B.8: Model Space with Model Posterior Probabilities For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$\mathbf{P}(\mathbf{M_k}\|\mathbf{D})$
1	•								0.0000
2				•					0.0011
3							•		0.0015
4	•		•						0.0285
5	•			•					0.0289
6	•						•		0.0299
7			•	•					0.0289
8			•				•		0.0283
9				•			•		0.0307
10							•	•	0.0287
11	•		•				•		0.1599
12	•		•					•	0.1590
13	•			•			•		0.1624
14			•				•	•	0.1614
15	•		•	•			•		0.1507

Table B.9: Model Space with Model Posterior Probabilities For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$\mathbf{P}(\mathbf{M_k}\|\mathbf{D})$
1	•								0.0000
2							•		0.0008
3	•		•						0.0245
4	•			•					0.0250
5	•						•		0.0242
6	•							•	0.0247
7			•	•					0.0248
8			•				•		0.0251
9				•			•		0.0248
10							•	•	0.0240
11	•		•	•					0.1457
12	•		•				•		0.1475
13	•		•					•	0.1464
14	•			•			•		0.1447
15	•						•	•	0.1462
16			•				•	•	0.0716

Table B.10: Model Space with Model Posterior Probabilities For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1	•								0.0000
2				•					0.0007
3							•		0.0015
4	•		•						0.0352
5	•			•					0.0351
6	•						•		0.0345
7			•	•					0.0344
8			•				•		0.0344
9				•			•		0.0339
10				•				•	0.0360
11							•	•	0.0345
12	•		•	•					0.2065
13	•		•				•		0.2036
14			•	•			•		0.2060
15			•				•	•	0.1037

Table B.11: Model Space with Model Posterior Probabilities For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1	•								0.0000
2						•			0.0001
3							•		0.0003
4	•		•						0.0036
5	•					•			0.0039
6	•						•		0.0044
7		•					•		0.0038
8			•				•		0.0035
9	•		•				•		0.0433
10	•						•	•	0.0461
11		•	•				•		0.0463
12			•				•	•	0.0449
13	•		•			•	•		0.3158
14	•		•				•	•	0.3222
15		•	•				•	•	0.1619

Table B.12: Model Space with Model Posterior Probabilities For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$\mathbf{P}(\mathbf{M_k}\|\mathbf{D})$
1	•								0.0000
2				•					0.0010
3						•			0.0022
4							•		0.0024
5	•		•						0.0320
6	•					•			0.0316
7	•							•	0.0303
8			•				•		0.0318
9				•			•		0.0313
10							•	•	0.0308
11	•		•				•		0.1813
12	•		•					•	0.1761
13	•			•			•		0.1776
14			•	•			•		0.1788
15			•				•	•	0.0930

Table B.13: Model Space with Model Posterior Probabilities For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1	•								0.0000
2						•			0.0010
3							•		0.0023
4	•		•						0.0285
5	•						•		0.0275
6	•							•	0.0267
7			•				•		0.0298
8						•	•		0.0275
9						•		•	0.0277
10							•	•	0.0272
11	•		•			•			0.1629
12	•		•				•		0.1663
13	•						•	•	0.1609
14			•				•	•	0.1629
15	•		•				•	•	0.1488

Table B.14: Model Space with Model Posterior Probabilities For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$\mathbf{P}(\mathbf{M_k}\|\mathbf{D})$
1	•								0.0000
2							•		0.0004
3	•		•						0.0383
4	•			•					0.0378
5	•				•				0.0380
6	•					•			0.0349
7	•						•		0.0351
8			•	•					0.0360
9			•				•		0.0345
10				•		•			0.0349
11						•	•		0.0355
12							•	•	0.0341
13	•		•			•			0.1868
14	•		•				•		0.1823
15			•	•			•		0.1805
16			•				•	•	0.0909

Table B.15: Model Space with Model Posterior Probabilities For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

	Covariate								
Model No	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	$P(M\|D)$
1	•								0.0000
2				•					0.0006
3							•		0.0015
4	•		•						0.0340
5	•			•					0.0340
6	•					•			0.0332
7			•	•					0.0332
8			•			•			0.0335
9			•				•		0.0333
10				•			•		0.0338
11	•		•			•			0.2181
12	•		•				•		0.2177
13	•					•	•		0.2174
14			•		•	•			0.1095
15	•		•				•	•	0.1488

B.2 Validation Study: Transition Probabily Matrices By Missing Rate

Table B.16: Transition Probability Matrix For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

From \ To	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.094	0.094	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.057	0.057	0.057	0.057	0.038	0.019
2	0.094	0.094	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.057	0.057	0.057	0.057	0.038	0.019
3	0.063	0.063	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.063	0.063	0.063	0.063	0.047	0.031
4	0.063	0.063	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.063	0.063	0.063	0.063	0.047	0.031
5	0.063	0.063	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.063	0.063	0.063	0.063	0.047	0.031
6	0.063	0.063	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.063	0.063	0.063	0.063	0.047	0.031
7	0.063	0.063	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.063	0.063	0.063	0.063	0.047	0.031
8	0.063	0.063	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.063	0.063	0.063	0.063	0.047	0.031
9	0.063	0.063	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.063	0.063	0.063	0.063	0.047	0.031
10	0.049	0.049	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.082	0.082	0.082	0.082	0.066	0.049
11	0.049	0.049	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.082	0.082	0.082	0.082	0.066	0.049
12	0.049	0.049	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.082	0.082	0.082	0.082	0.066	0.049
13	0.049	0.049	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.082	0.082	0.082	0.082	0.066	0.049
14	0.040	0.040	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.100	0.080
15	0.027	0.027	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.081	0.081	0.081	0.081	0.108	0.135

Table B.17: Transition Probability Matrix For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

From \ To	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.091	0.091	0.091	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.045	0.045	0.045	0.045	0.045	0.023
2	0.091	0.091	0.091	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.045	0.045	0.045	0.045	0.045	0.023
3	0.091	0.091	0.091	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.045	0.045	0.045	0.045	0.045	0.023
4	0.056	0.056	0.056	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.056	0.056	0.056	0.056	0.056	0.037
5	0.056	0.056	0.056	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.056	0.056	0.056	0.056	0.056	0.037
6	0.056	0.056	0.056	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.056	0.056	0.056	0.056	0.056	0.037
7	0.056	0.056	0.056	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.056	0.056	0.056	0.056	0.056	0.037
8	0.056	0.056	0.056	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.056	0.056	0.056	0.056	0.056	0.037
9	0.056	0.056	0.056	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.056	0.056	0.056	0.056	0.056	0.037
10	0.056	0.056	0.056	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.056	0.056	0.056	0.056	0.056	0.037
11	0.040	0.040	0.040	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.060
12	0.040	0.040	0.040	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.060
13	0.040	0.040	0.040	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.060
14	0.040	0.040	0.040	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.060
15	0.040	0.040	0.040	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.080	0.080	0.080	0.080	0.080	0.060
16	0.028	0.028	0.028	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.111

Table B.18: Transition Probability Matrix For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

From \ To	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.107	0.107	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.036	0.036	0.036	0.036	0.036	0.036
2	0.107	0.107	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.036	0.036	0.036	0.036	0.036	0.036
3	0.050	0.050	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050
4	0.050	0.050	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050
5	0.050	0.050	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050
6	0.050	0.050	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050
7	0.050	0.050	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050
8	0.050	0.050	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050
9	0.050	0.050	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050
10	0.050	0.050	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050
11	0.028	0.028	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083
12	0.028	0.028	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083
13	0.028	0.028	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083
14	0.028	0.028	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083
15	0.028	0.028	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083
16	0.028	0.028	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083

Table B.19: Transition Probability Matrix For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

From \ To	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.100	0.100	0.100	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050	0.025
2	0.100	0.100	0.100	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050	0.025
3	0.100	0.100	0.100	0.075	0.075	0.075	0.075	0.075	0.050	0.050	0.050	0.050	0.050	0.050	0.025
4	0.061	0.061	0.061	0.082	0.082	0.082	0.082	0.082	0.061	0.061	0.061	0.061	0.061	0.061	0.041
5	0.061	0.061	0.061	0.082	0.082	0.082	0.082	0.082	0.061	0.061	0.061	0.061	0.061	0.061	0.041
6	0.061	0.061	0.061	0.082	0.082	0.082	0.082	0.082	0.061	0.061	0.061	0.061	0.061	0.061	0.041
7	0.061	0.061	0.061	0.082	0.082	0.082	0.082	0.082	0.061	0.061	0.061	0.061	0.061	0.061	0.041
8	0.061	0.061	0.061	0.082	0.082	0.082	0.082	0.082	0.061	0.061	0.061	0.061	0.061	0.061	0.041
9	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083	0.063
10	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083	0.063
11	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083	0.063
12	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083	0.063
13	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083	0.063
14	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083	0.063
15	0.029	0.029	0.029	0.057	0.057	0.057	0.057	0.057	0.086	0.086	0.086	0.086	0.086	0.086	0.114

Table B.20: Transition Probability Matrix For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.095	0.095	0.095	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048	0.048	0.048
2	0.095	0.095	0.095	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048	0.048	0.048
3	0.095	0.095	0.095	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048	0.048	0.048
4	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.059	0.059
5	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.059	0.059
6	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.059	0.059
7	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.059	0.059
8	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.059	0.059
9	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.059	0.059
10	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083
11	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083
12	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083
13	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083
14	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083
15	0.042	0.042	0.042	0.063	0.063	0.063	0.063	0.063	0.063	0.083	0.083	0.083	0.083	0.083	0.083

Table B.21: Transition Probability Matrix For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.093	0.093	0.093	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.047	0.047	0.047	0.023
2	0.093	0.093	0.093	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.047	0.047	0.047	0.023
3	0.093	0.093	0.093	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.047	0.047	0.047	0.023
4	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.058	0.058	0.058	0.038
5	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.058	0.058	0.058	0.038
6	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.058	0.058	0.058	0.038
7	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.058	0.058	0.058	0.038
8	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.058	0.058	0.058	0.038
9	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.058	0.058	0.058	0.038
10	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.058	0.058	0.058	0.038
11	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.058	0.058	0.058	0.038
12	0.044	0.044	0.044	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.089	0.089	0.089	0.067
13	0.044	0.044	0.044	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.089	0.089	0.089	0.067
14	0.044	0.044	0.044	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.089	0.089	0.089	0.067
15	0.031	0.031	0.031	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.094	0.094	0.094	0.125

Table B.22: Transition Probability Matrix For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0.083	0.083	0.083	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.042	0.042	0.042	0.042	0.021
2	0.083	0.083	0.083	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.042	0.042	0.042	0.042	0.021
3	0.083	0.083	0.083	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.042	0.042	0.042	0.042	0.021
4	0.051	0.051	0.051	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.051	0.051	0.051	0.051	0.034
5	0.051	0.051	0.051	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.051	0.051	0.051	0.051	0.034
6	0.051	0.051	0.051	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.051	0.051	0.051	0.051	0.034
7	0.051	0.051	0.051	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.051	0.051	0.051	0.051	0.034
8	0.051	0.051	0.051	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.051	0.051	0.051	0.051	0.034
9	0.051	0.051	0.051	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.051	0.051	0.051	0.051	0.034
10	0.051	0.051	0.051	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.051	0.051	0.051	0.051	0.034
11	0.051	0.051	0.051	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.051	0.051	0.051	0.051	0.034
12	0.051	0.051	0.051	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.051	0.051	0.051	0.051	0.034
13	0.038	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.058
14	0.038	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.058
15	0.038	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.058
16	0.038	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.058
17	0.027	0.027	0.027	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.081	0.081	0.081	0.081	0.108

Table B.23: Transition Probability Matrix For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.095	0.095	0.095	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048	0.024
2	0.095	0.095	0.095	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048	0.024
3	0.095	0.095	0.095	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048	0.024
4	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
5	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
6	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
7	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
8	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
9	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
10	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
11	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065
12	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065
13	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065
14	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065
15	0.030	0.030	0.030	0.061	0.061	0.061	0.061	0.061	0.061	0.061	0.091	0.091	0.091	0.091	0.121

Table B.24: Transition Probability Matrix For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.091	0.091	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.045	0.045	0.045	0.045	0.045	0.045
2	0.091	0.091	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.045	0.045	0.045	0.045	0.045	0.045
3	0.054	0.054	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.054	0.054	0.054	0.054	0.054	0.054
4	0.054	0.054	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.054	0.054	0.054	0.054	0.054	0.054
5	0.054	0.054	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.054	0.054	0.054	0.054	0.054	0.054
6	0.054	0.054	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.054	0.054	0.054	0.054	0.054	0.054
7	0.054	0.054	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.054	0.054	0.054	0.054	0.054	0.054
8	0.054	0.054	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.054	0.054	0.054	0.054	0.054	0.054
9	0.054	0.054	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.054	0.054	0.054	0.054	0.054	0.054
10	0.054	0.054	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.054	0.054	0.054	0.054	0.054	0.054
11	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077
12	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077
13	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077
14	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077
15	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077
16	0.038	0.038	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.077	0.077	0.077	0.077	0.077	0.077

Table B.25: Transition Probability Matrix For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.103	0.103	0.103	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.034	0.034	0.034	0.034
2	0.103	0.103	0.103	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.034	0.034	0.034	0.034
3	0.103	0.103	0.103	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.034	0.034	0.034	0.034
4	0.053	0.053	0.053	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053
5	0.053	0.053	0.053	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053
6	0.053	0.053	0.053	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053
7	0.053	0.053	0.053	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053
8	0.053	0.053	0.053	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053
9	0.053	0.053	0.053	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053
10	0.053	0.053	0.053	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053
11	0.053	0.053	0.053	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053
12	0.032	0.032	0.032	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.097	0.097	0.097	0.097
13	0.032	0.032	0.032	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.097	0.097	0.097	0.097
14	0.032	0.032	0.032	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.097	0.097	0.097	0.097
15	0.032	0.032	0.032	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.097	0.097	0.097	0.097

Table B.26: Transition Probability Matrix For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.105	0.105	0.105	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053	0.026	0.026	0.026
2	0.105	0.105	0.105	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053	0.026	0.026	0.026
3	0.105	0.105	0.105	0.079	0.079	0.079	0.079	0.079	0.053	0.053	0.053	0.053	0.026	0.026	0.026
4	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064	0.064	0.064	0.064	0.043	0.043	0.043
5	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064	0.064	0.064	0.064	0.043	0.043	0.043
6	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064	0.064	0.064	0.064	0.043	0.043	0.043
7	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064	0.064	0.064	0.064	0.043	0.043	0.043
8	0.064	0.064	0.064	0.085	0.085	0.085	0.085	0.085	0.064	0.064	0.064	0.064	0.043	0.043	0.043
9	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065	0.065	0.065
10	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065	0.065	0.065
11	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065	0.065	0.065
12	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065	0.065	0.065
13	0.027	0.027	0.027	0.054	0.054	0.054	0.054	0.054	0.081	0.081	0.081	0.081	0.108	0.108	0.108
14	0.027	0.027	0.027	0.054	0.054	0.054	0.054	0.054	0.081	0.081	0.081	0.081	0.108	0.108	0.108
15	0.027	0.027	0.027	0.054	0.054	0.054	0.054	0.054	0.081	0.081	0.081	0.081	0.108	0.108	0.108

Table B.27: Transition Probability Matrix For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.103	0.103	0.103	0.103	0.069	0.069	0.069	0.069	0.069	0.069	0.034	0.034	0.034	0.034	0.034
2	0.103	0.103	0.103	0.103	0.069	0.069	0.069	0.069	0.069	0.069	0.034	0.034	0.034	0.034	0.034
3	0.103	0.103	0.103	0.103	0.069	0.069	0.069	0.069	0.069	0.069	0.034	0.034	0.034	0.034	0.034
4	0.103	0.103	0.103	0.103	0.069	0.069	0.069	0.069	0.069	0.069	0.034	0.034	0.034	0.034	0.034
5	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083	0.056	0.056	0.056	0.056	0.056
6	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083	0.056	0.056	0.056	0.056	0.056
7	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083	0.056	0.056	0.056	0.056	0.056
8	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083	0.056	0.056	0.056	0.056	0.056
9	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083	0.056	0.056	0.056	0.056	0.056
10	0.056	0.056	0.056	0.056	0.083	0.083	0.083	0.083	0.083	0.083	0.056	0.056	0.056	0.056	0.056
11	0.032	0.032	0.032	0.032	0.065	0.065	0.065	0.065	0.065	0.065	0.097	0.097	0.097	0.097	0.097
12	0.032	0.032	0.032	0.032	0.065	0.065	0.065	0.065	0.065	0.065	0.097	0.097	0.097	0.097	0.097
13	0.032	0.032	0.032	0.032	0.065	0.065	0.065	0.065	0.065	0.065	0.097	0.097	0.097	0.097	0.097
14	0.032	0.032	0.032	0.032	0.065	0.065	0.065	0.065	0.065	0.065	0.097	0.097	0.097	0.097	0.097
15	0.032	0.032	0.032	0.032	0.065	0.065	0.065	0.065	0.065	0.065	0.097	0.097	0.097	0.097	0.097

Table B.28: Transition Probability Matrix For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.095	0.095	0.095	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048	0.024
2	0.095	0.095	0.095	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048	0.024
3	0.095	0.095	0.095	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048	0.024
4	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
5	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
6	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
7	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
8	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
9	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
10	0.059	0.059	0.059	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.059	0.059	0.059	0.059	0.039
11	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065
12	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065
13	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065
14	0.043	0.043	0.043	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.087	0.087	0.087	0.087	0.065
15	0.030	0.030	0.030	0.061	0.061	0.061	0.061	0.061	0.061	0.061	0.091	0.091	0.091	0.091	0.121

Table B.29: Transition Probability Matrix For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.100	0.100	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.033	0.033	0.033	0.033
2	0.100	0.100	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.033	0.033	0.033	0.033
3	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
4	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
5	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
6	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
7	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
8	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
9	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
10	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
11	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
12	0.048	0.048	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.048	0.048	0.048	0.048
13	0.029	0.029	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.088	0.088	0.088	0.088
14	0.029	0.029	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.088	0.088	0.088	0.088
15	0.029	0.029	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.088	0.088	0.088	0.088
16	0.029	0.029	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.088	0.088	0.088	0.088

Table B.30: Transition Probability Matrix For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

To / From	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0.111	0.111	0.111	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.037	0.037	0.037	0.037
2	0.111	0.111	0.111	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.037	0.037	0.037	0.037
3	0.111	0.111	0.111	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.037	0.037	0.037	0.037
4	0.057	0.057	0.057	0.086	0.086	0.086	0.086	0.086	0.086	0.086	0.057	0.057	0.057	0.057
5	0.057	0.057	0.057	0.086	0.086	0.086	0.086	0.086	0.086	0.086	0.057	0.057	0.057	0.057
6	0.057	0.057	0.057	0.086	0.086	0.086	0.086	0.086	0.086	0.086	0.057	0.057	0.057	0.057
7	0.057	0.057	0.057	0.086	0.086	0.086	0.086	0.086	0.086	0.086	0.057	0.057	0.057	0.057
8	0.057	0.057	0.057	0.086	0.086	0.086	0.086	0.086	0.086	0.086	0.057	0.057	0.057	0.057
9	0.057	0.057	0.057	0.086	0.086	0.086	0.086	0.086	0.086	0.086	0.057	0.057	0.057	0.057
10	0.057	0.057	0.057	0.086	0.086	0.086	0.086	0.086	0.086	0.086	0.057	0.057	0.057	0.057
11	0.034	0.034	0.034	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.103	0.103	0.103	0.103
12	0.034	0.034	0.034	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.103	0.103	0.103	0.103
13	0.034	0.034	0.034	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.103	0.103	0.103	0.103
14	0.034	0.034	0.034	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.103	0.103	0.103	0.103

B.3 MCMC Diagnostics For Validation Study

B.3.1 Density Plots of Validation Study

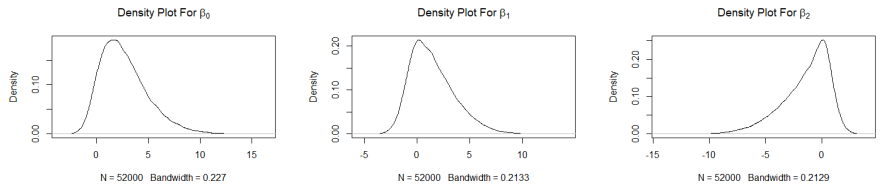

Figure B.1: Density Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

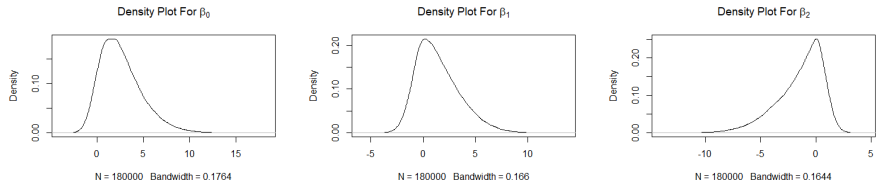

Figure B.2: Density Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

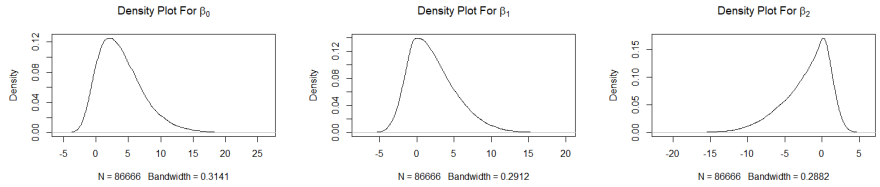

Figure B.3: Density Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

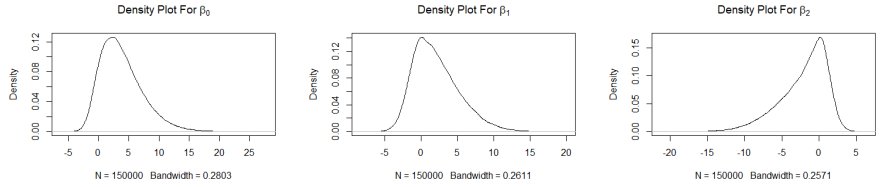

Figure B.4: Density Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

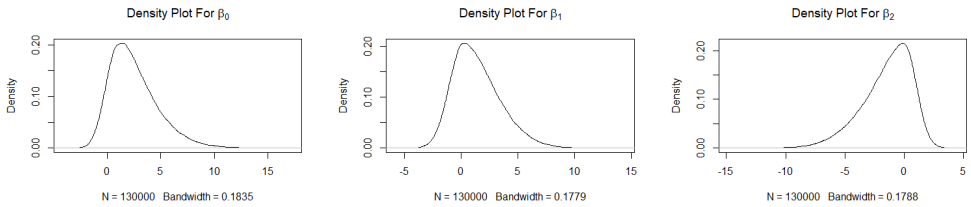

Figure B.5: Density Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

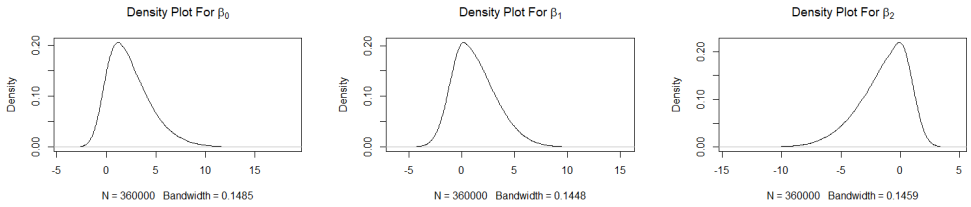

Figure B.6: Density Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

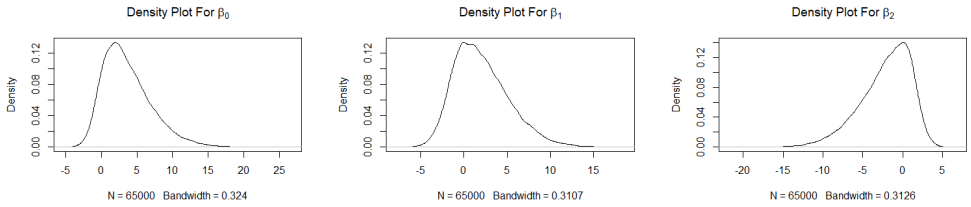

Figure B.7: Density Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

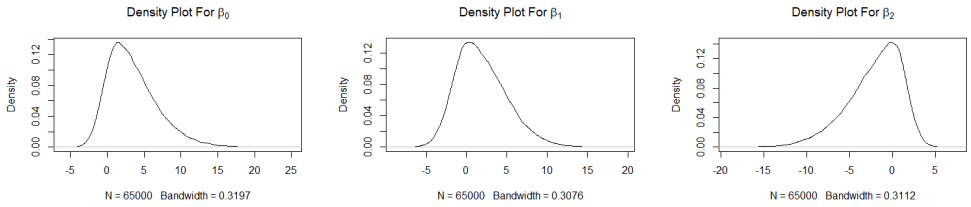

Figure B.8: Density Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

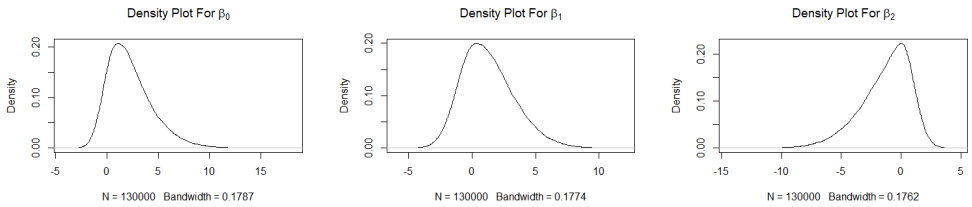

Figure B.9: Density Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

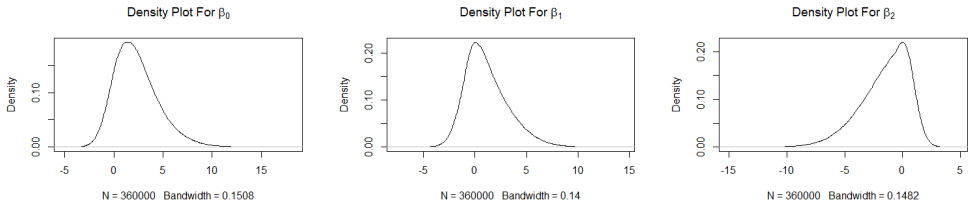

Figure B.10: Density Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

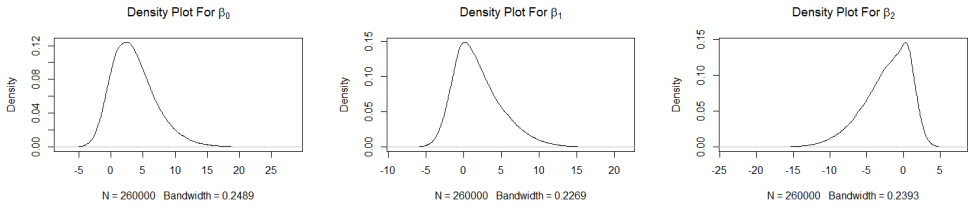

Figure B.11: Density Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

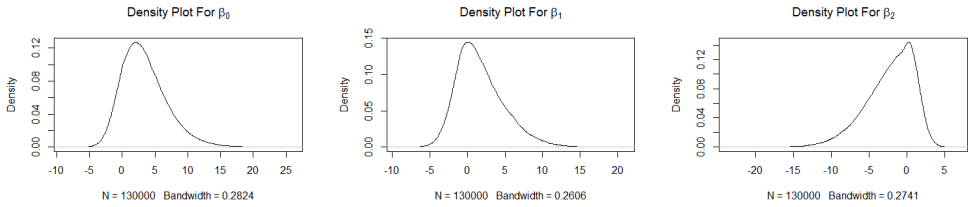

Figure B.12: Density Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

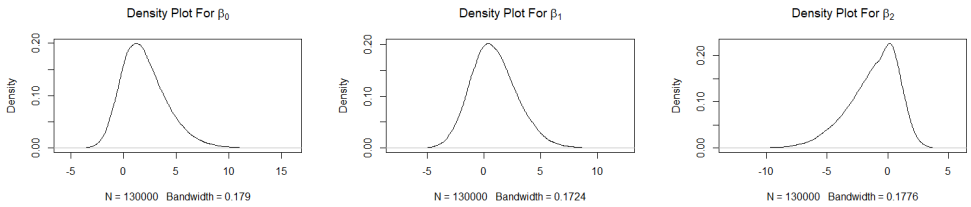

Figure B.13: Density Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

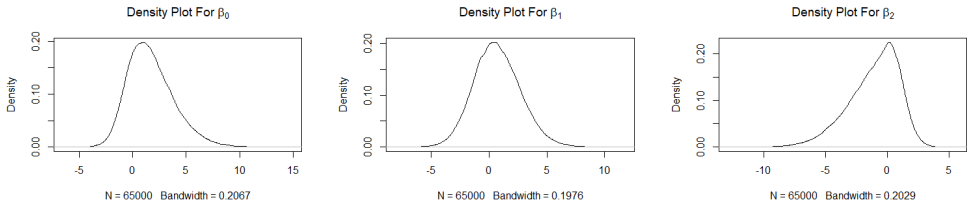

Figure B.14: Density Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

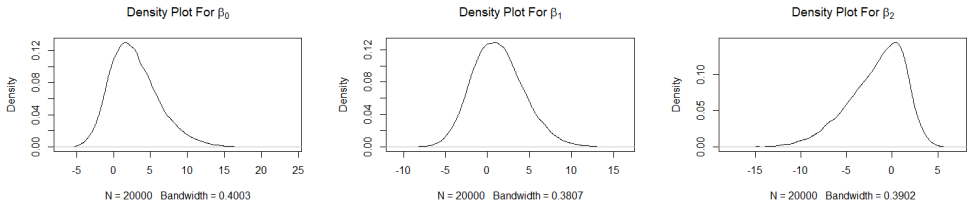

Figure B.15: Density Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

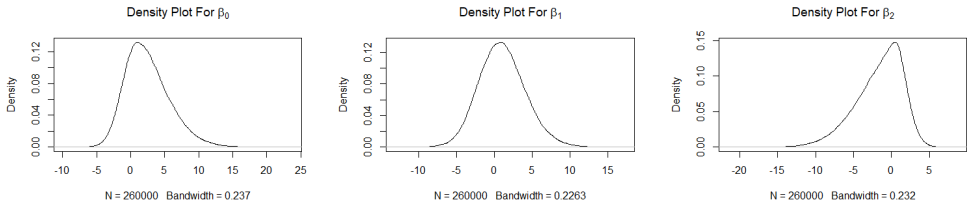

Figure B.16: Density Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

B.3.2 Autocorrelation Plots of The Simulation Study

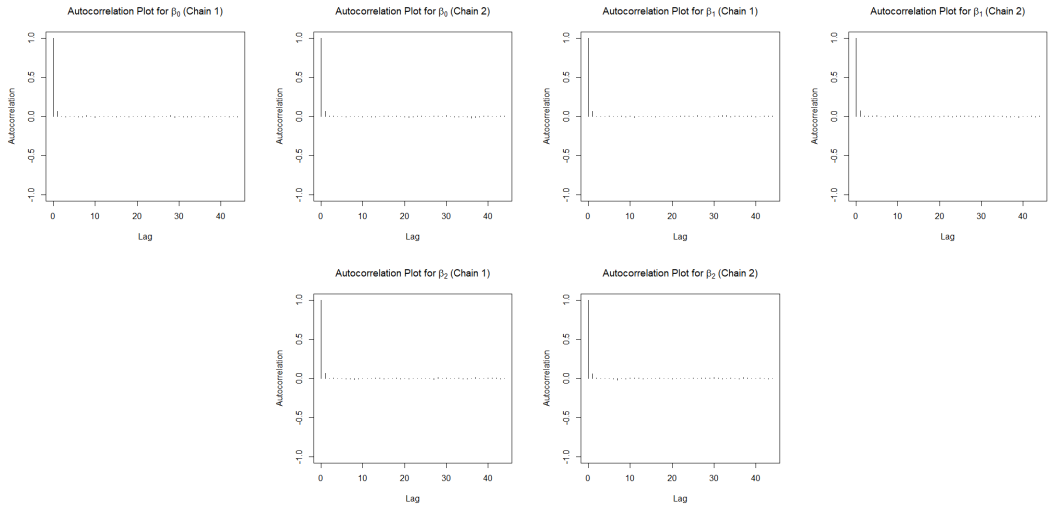

Figure B.17: Autocorrelation Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

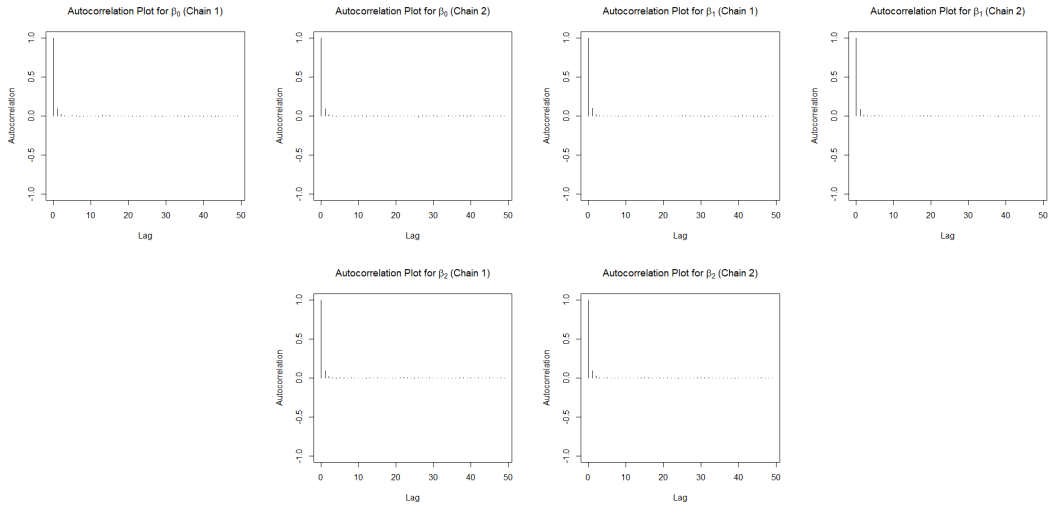

Figure B.18: Autocorrelation Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

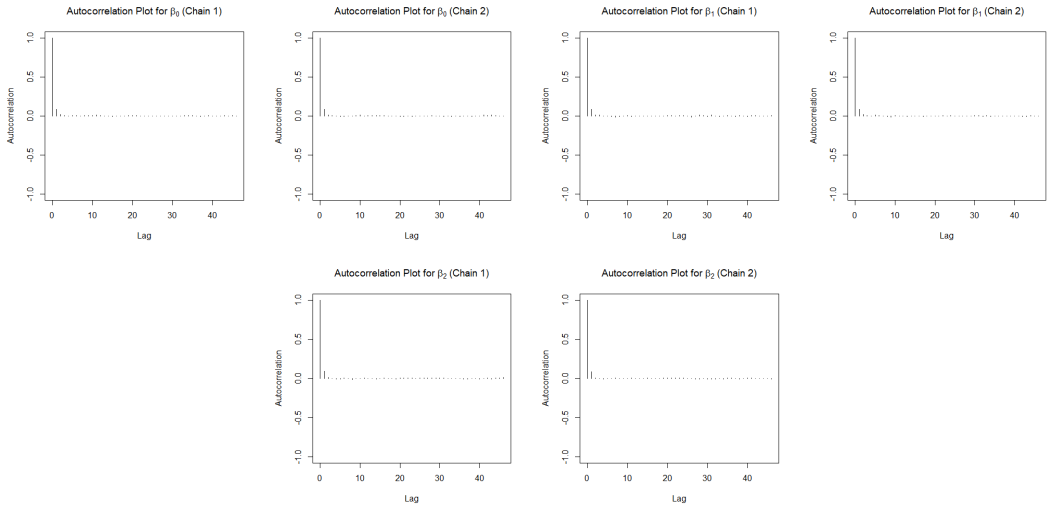

Figure B.19: Autocorrelation Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

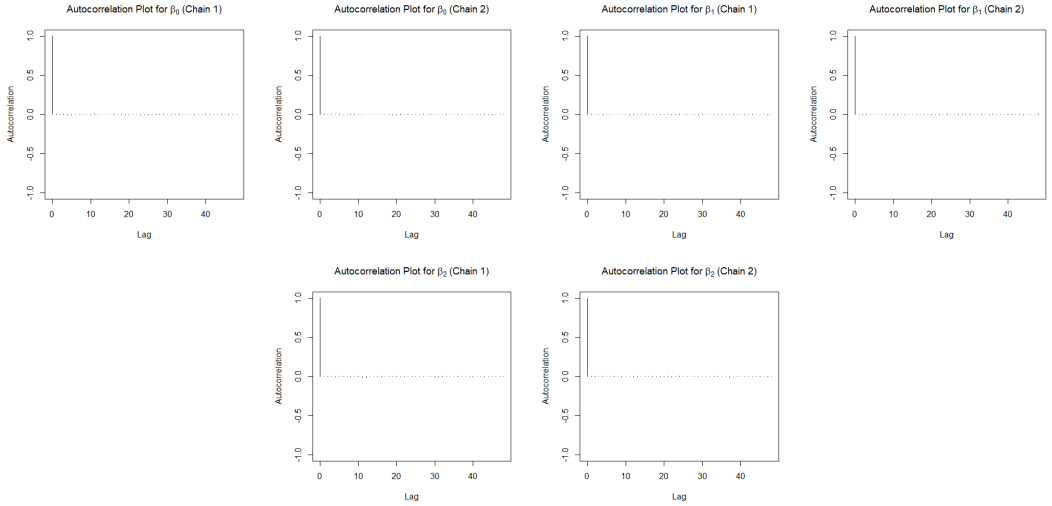

Figure B.20: Autocorrelation Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

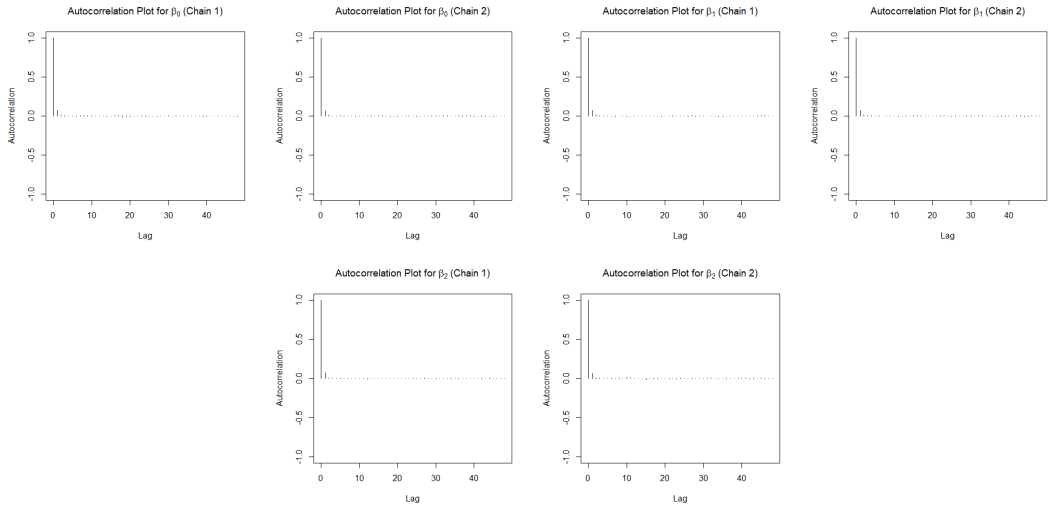

Figure B.21: Autocorrelation Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

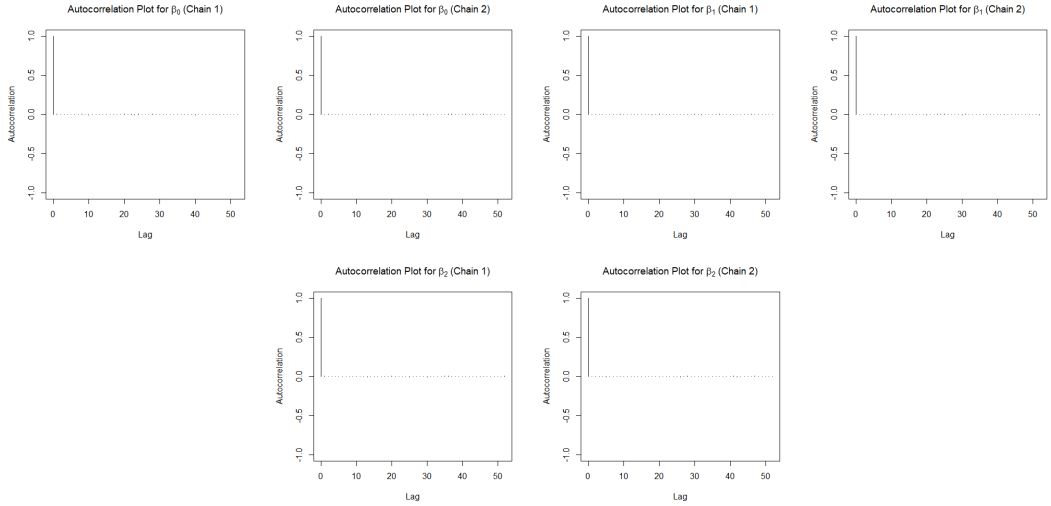

Figure B.22: Autocorrelation Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

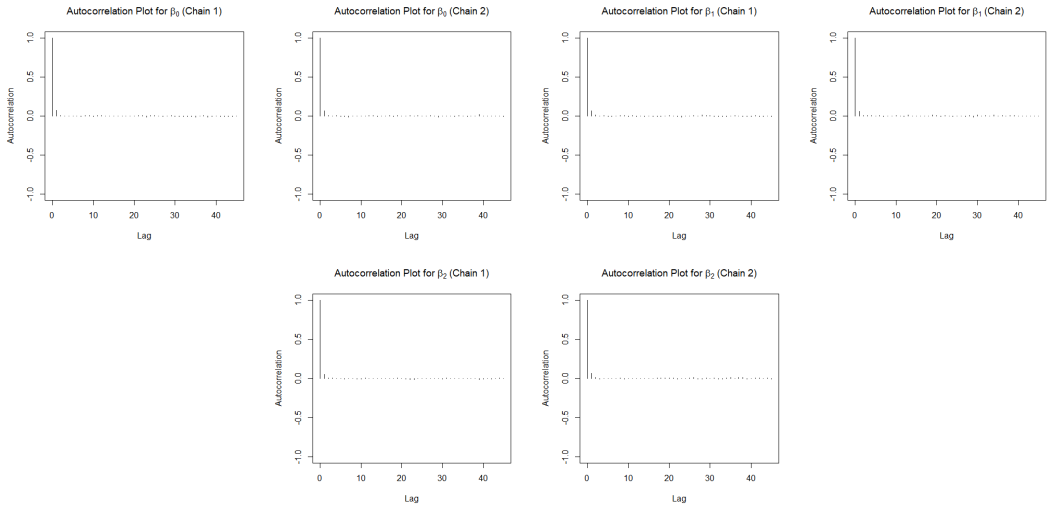

Figure B.23: Autocorrelation Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

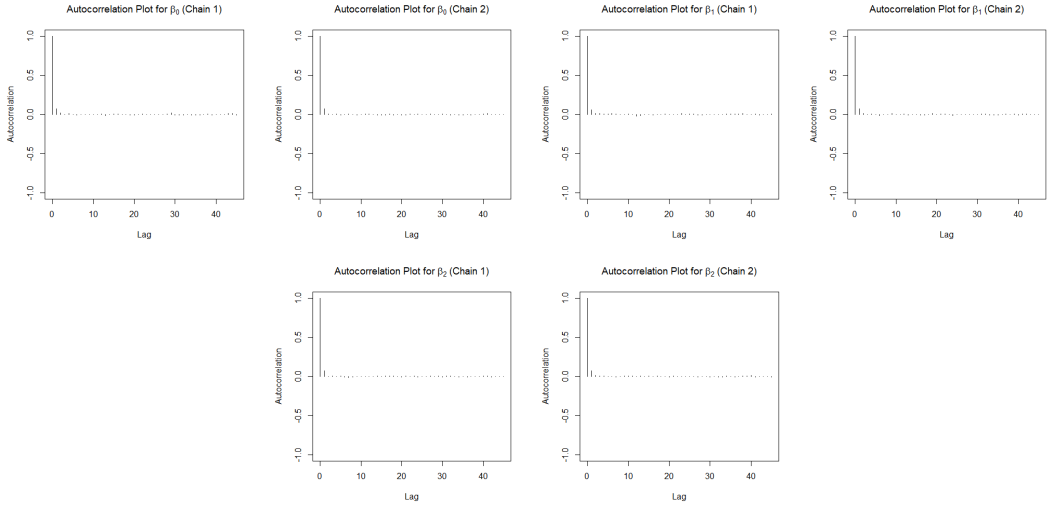

Figure B.24: Autocorrelation Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

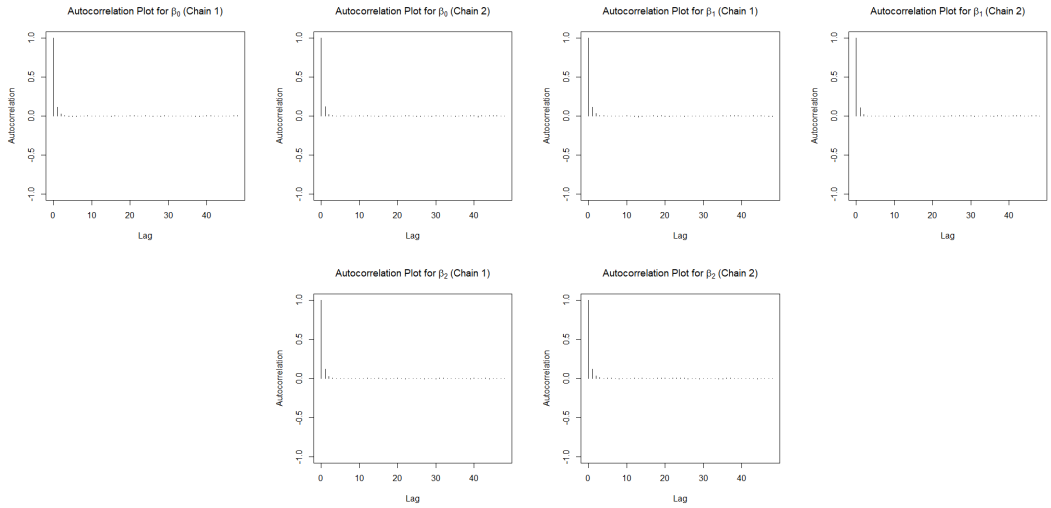

Figure B.25: Autocorrelation Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

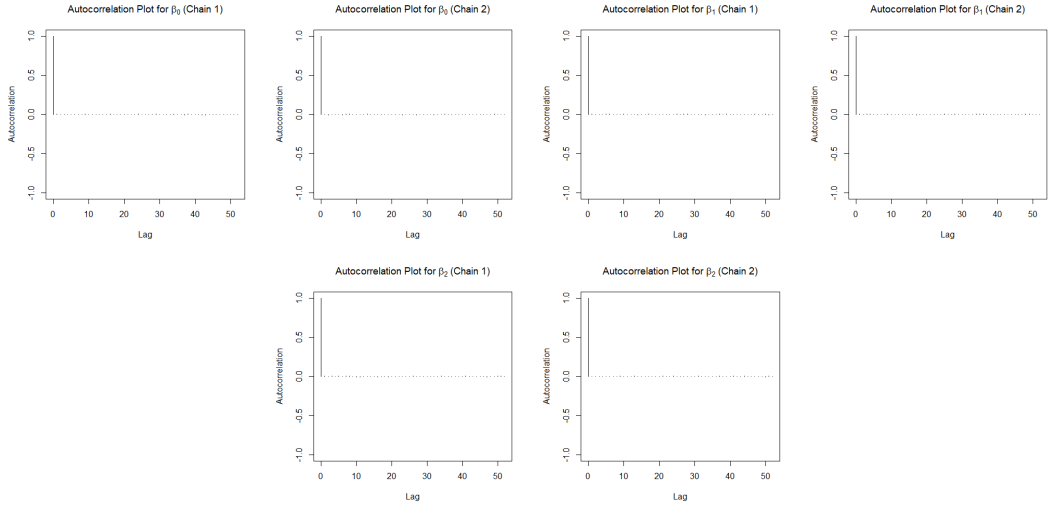

Figure B.26: Autocorrelation Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

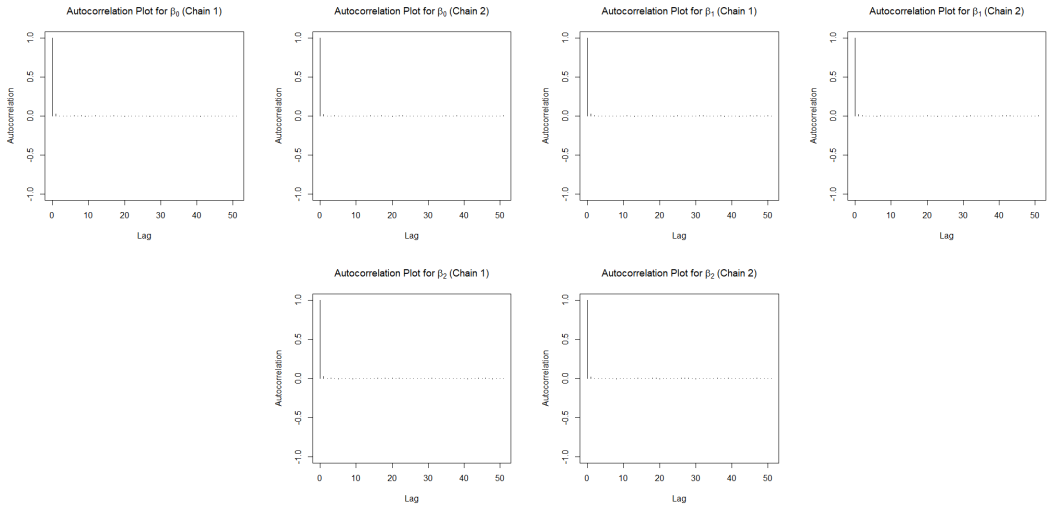

Figure B.27: Autocorrelation Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

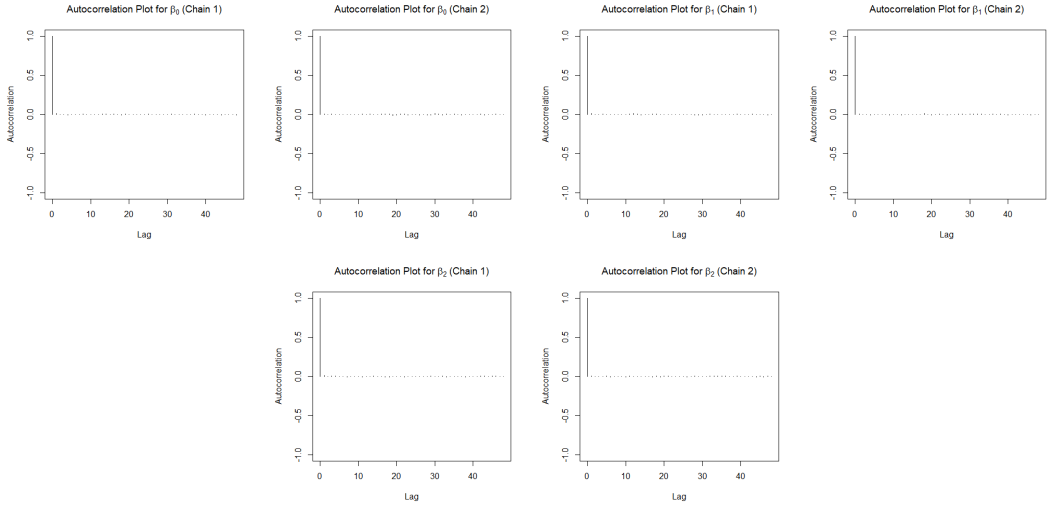

Figure B.28: Autocorrelation Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

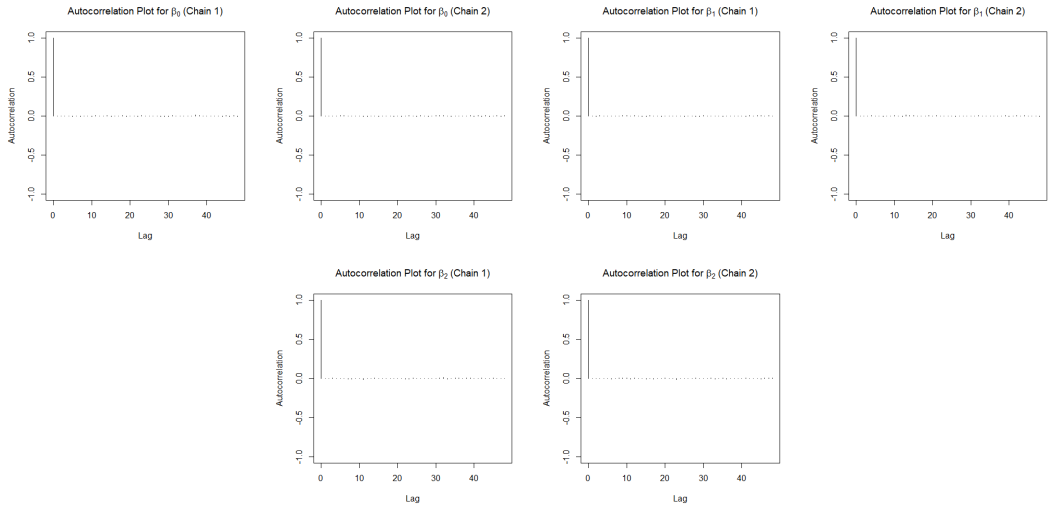

Figure B.29: Autocorrelation Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

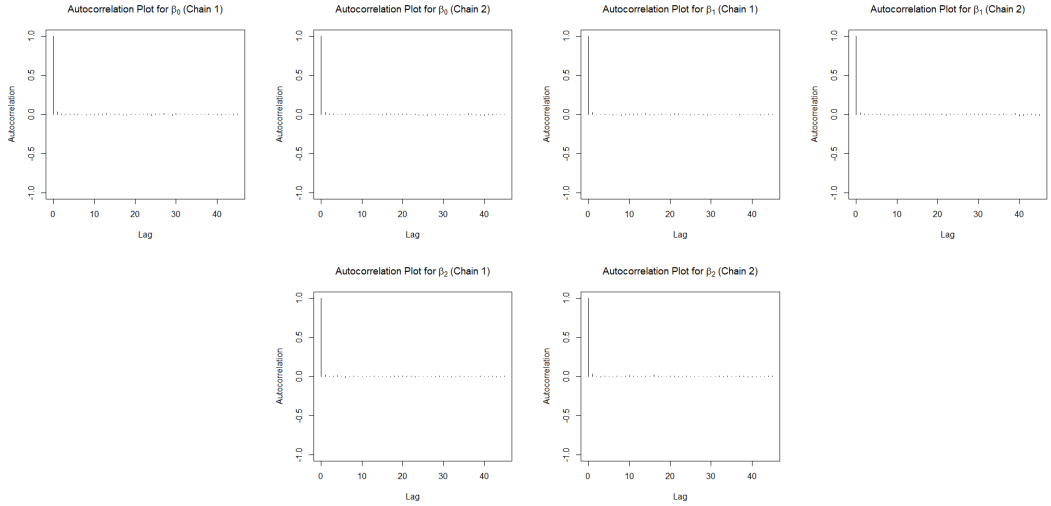

Figure B.30: Autocorrelation Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

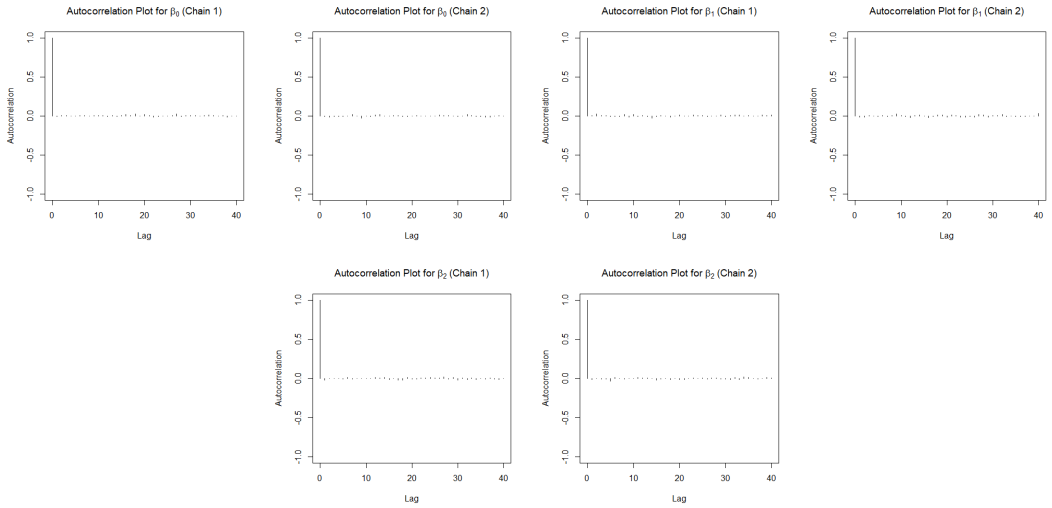

Figure B.31: Autocorrelation Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

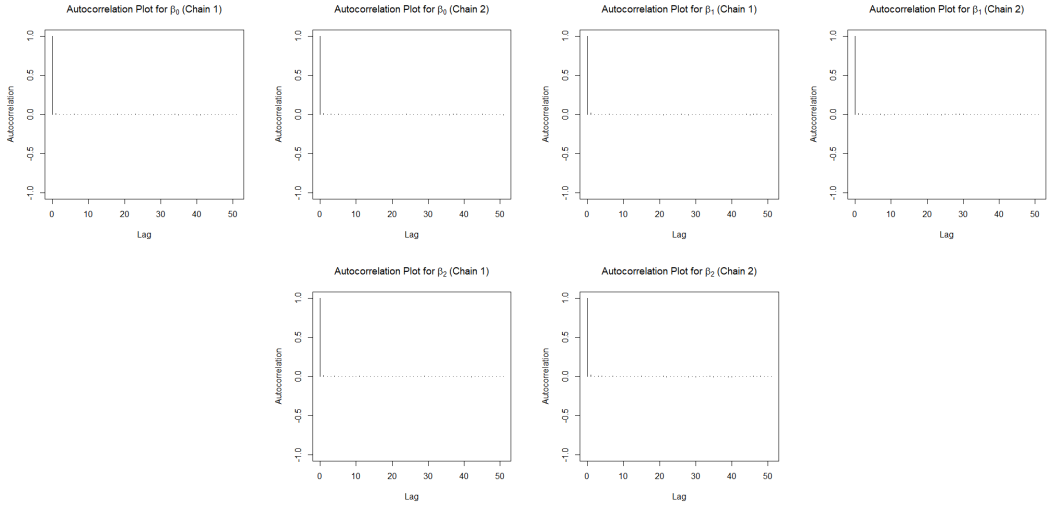

Figure B.32: Autocorrelation Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

B.3.3 Trace Plots of The Validation Study

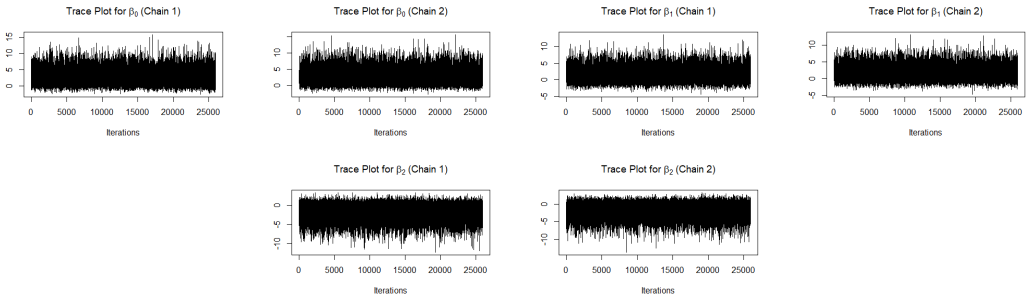

Figure B.33: Trace Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

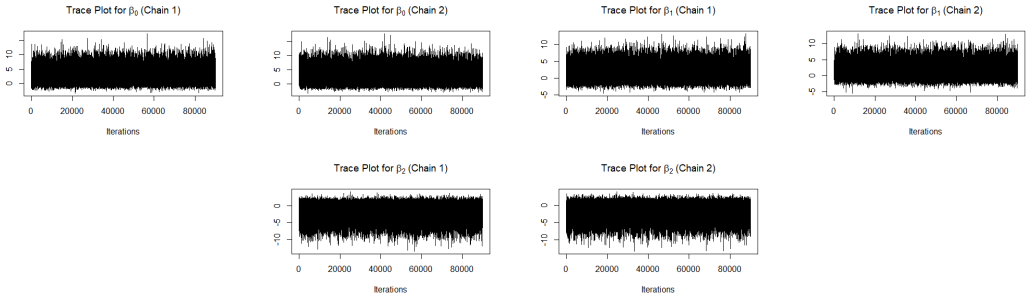

Figure B.34: Trace Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

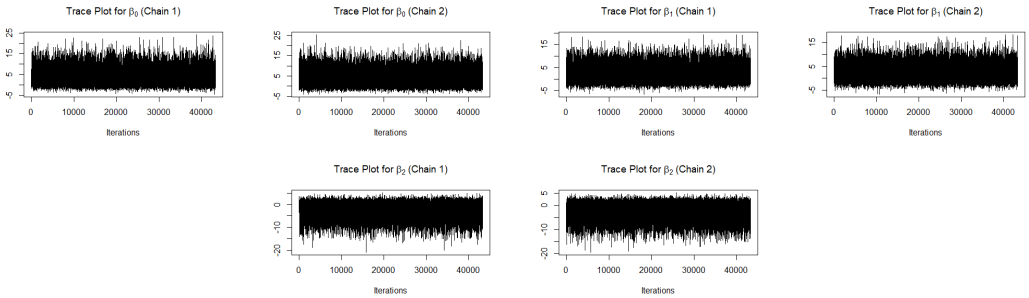

Figure B.35: Trace Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

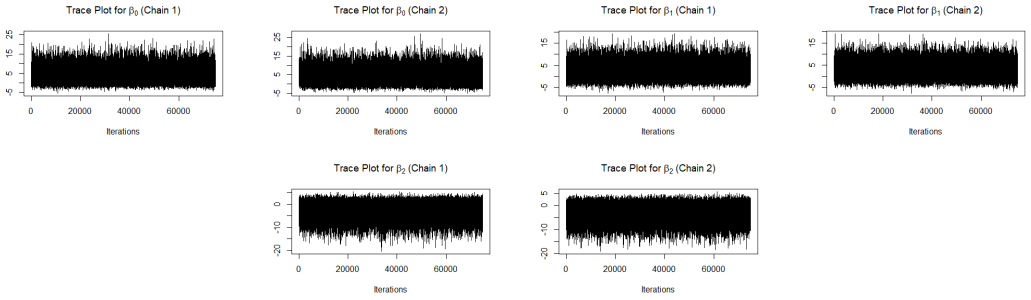

Figure B.36: Trace Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

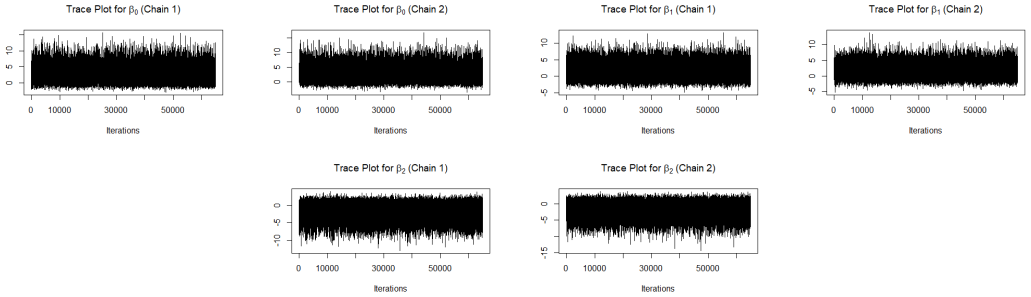

Figure B.37: Trace Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

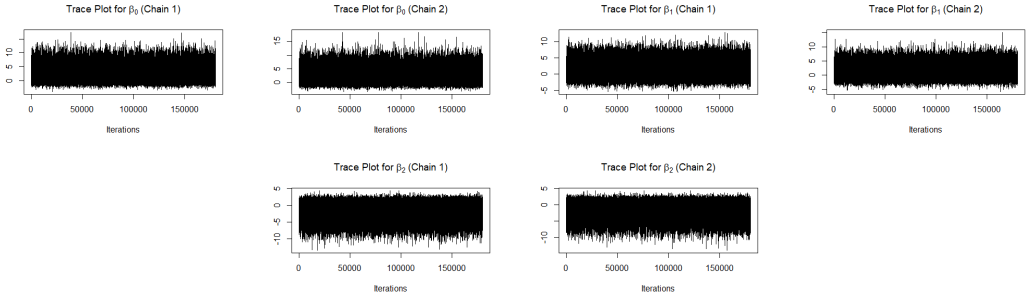

Figure B.38: Trace Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

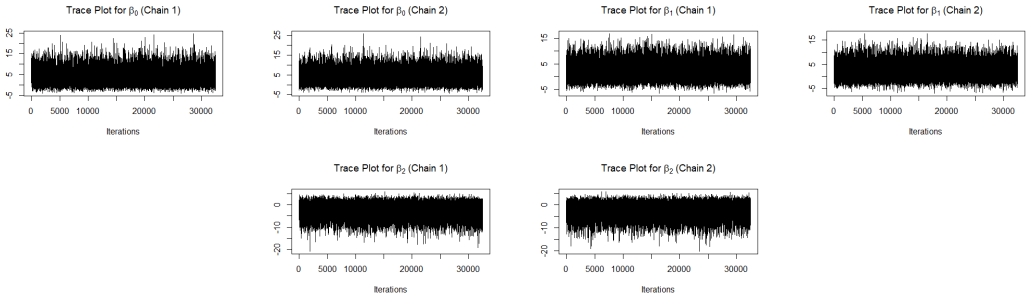

Figure B.39: Trace Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

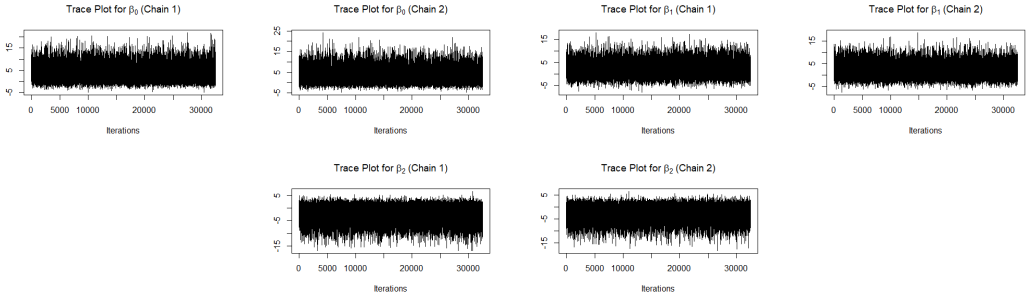

Figure B.40: Trace Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

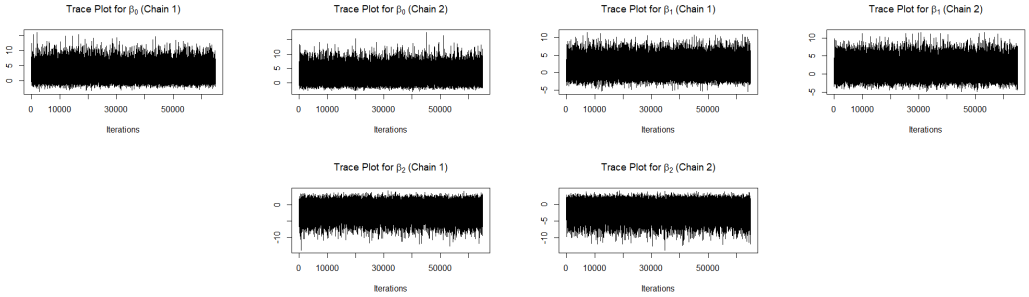

Figure B.41: Trace Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

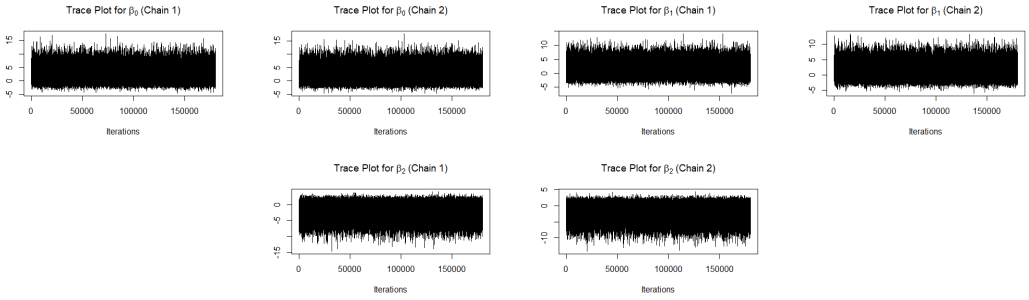

Figure B.42: Trace Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

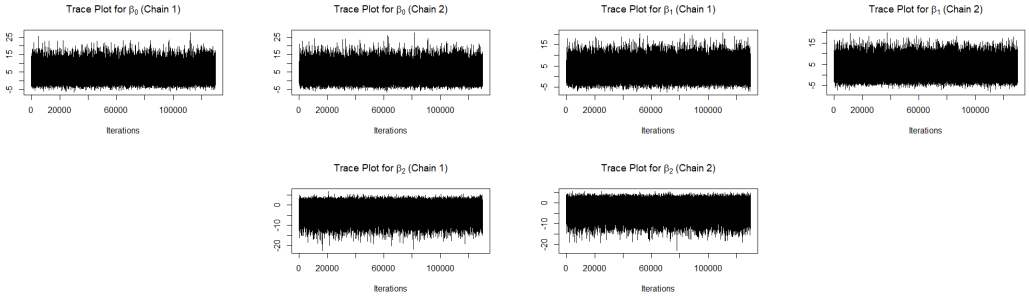

Figure B.43: Trace Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

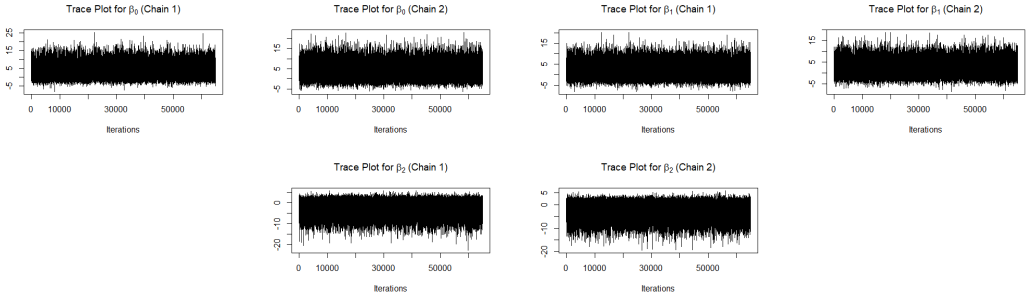

Figure B.44: Trace Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

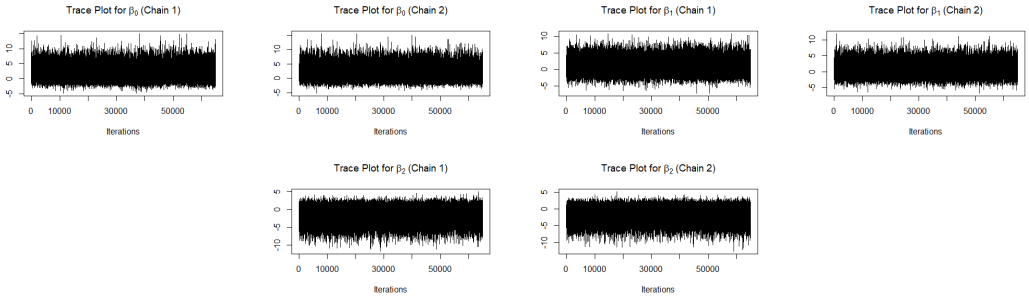

Figure B.45: Trace Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

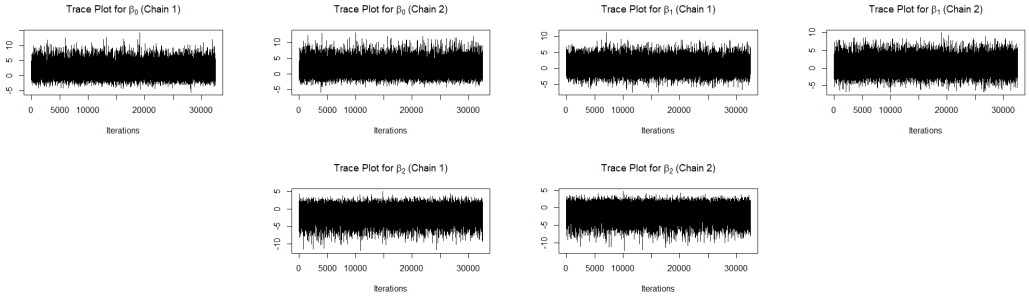

Figure B.46: Trace Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

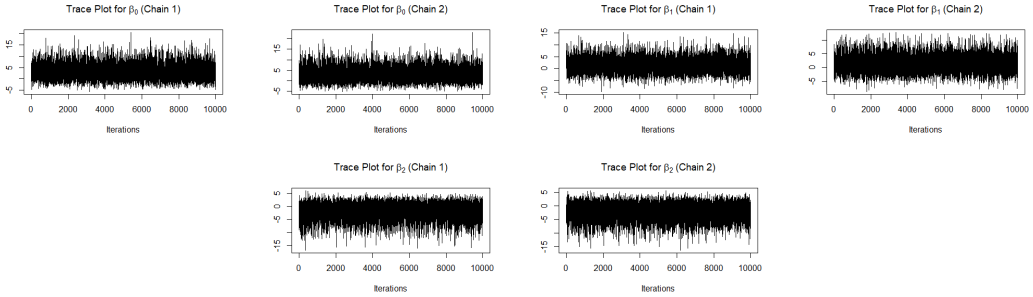

Figure B.47: Trace Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

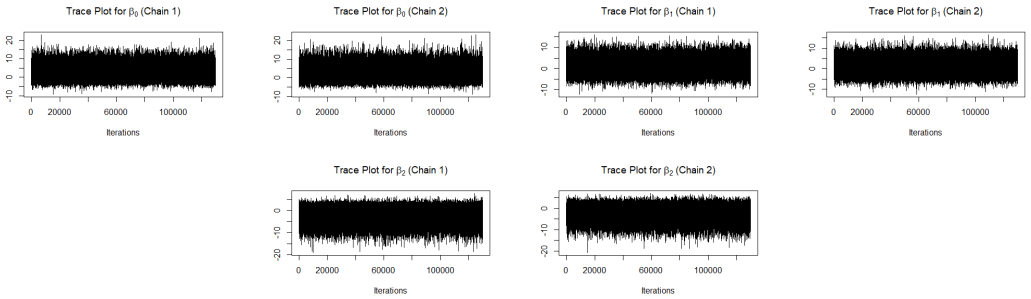

Figure B.48: Trace Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

B.3.4 Brooks-Gelman-Rubin Plots For Validation Study

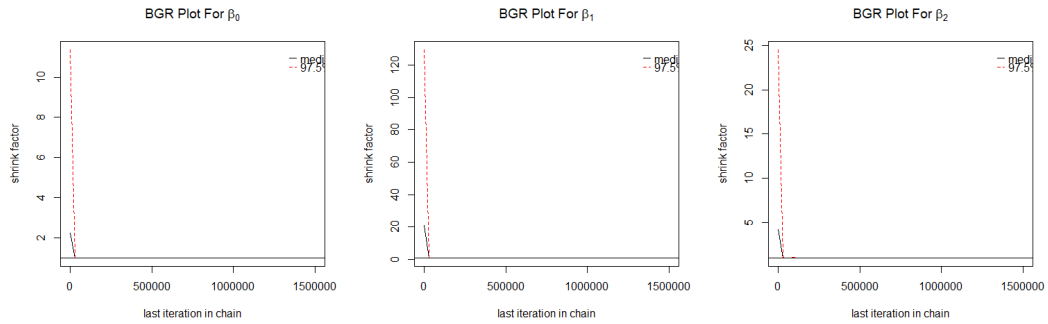

Figure B.49: BGR Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

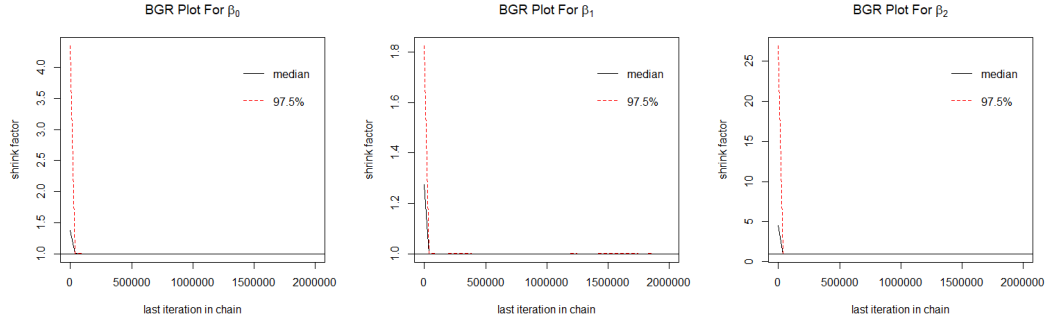

Figure B.50: BGR Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

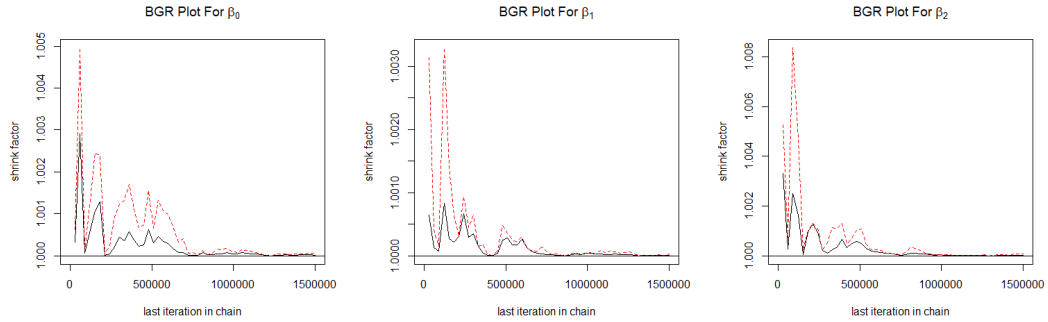

Figure B.51: BGR Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

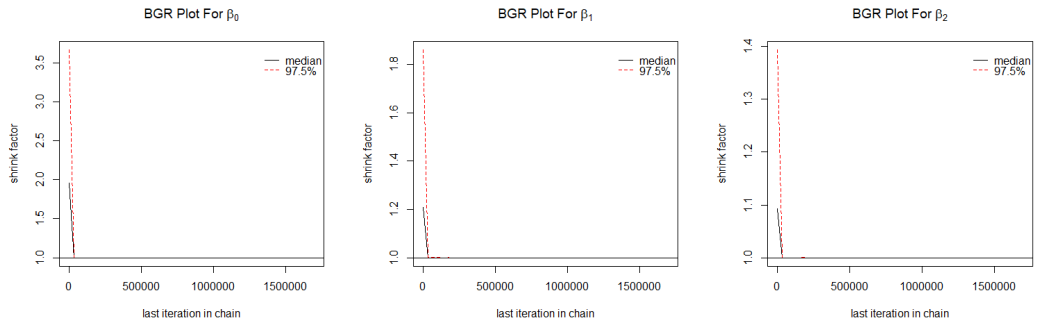

Figure B.52: BGR Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

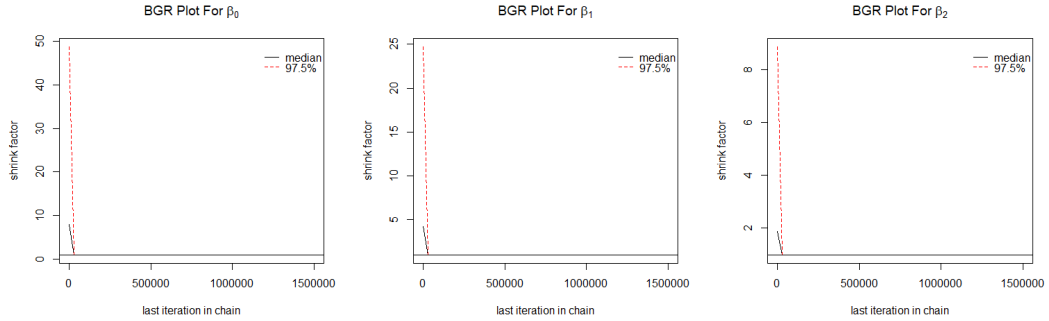

Figure B.53: BGR Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

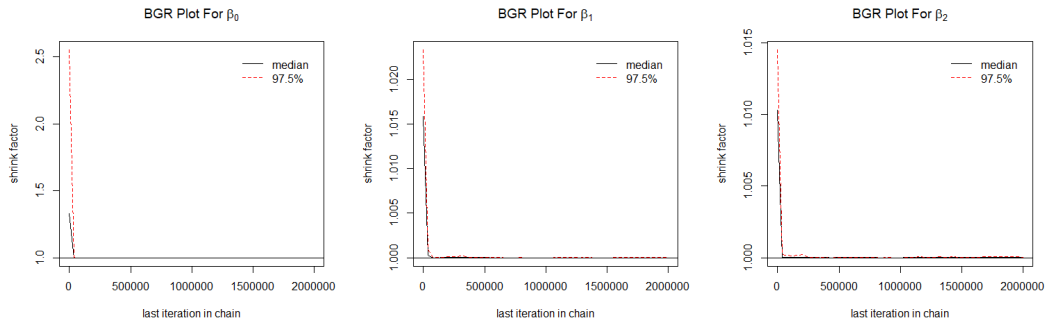

Figure B.54: BGR Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

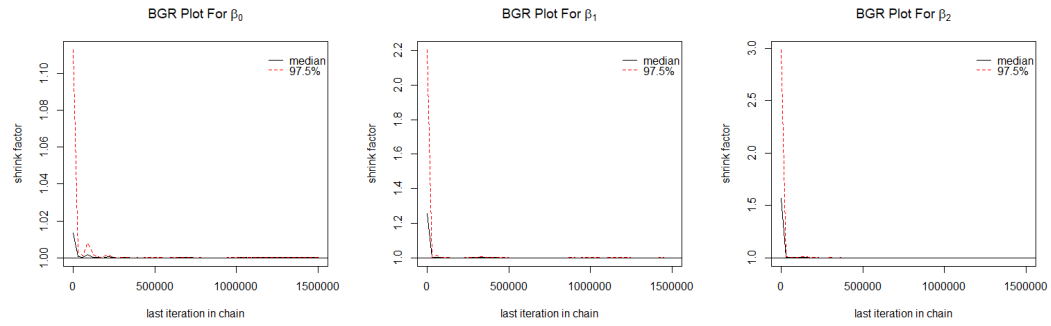

Figure B.55: BGR Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

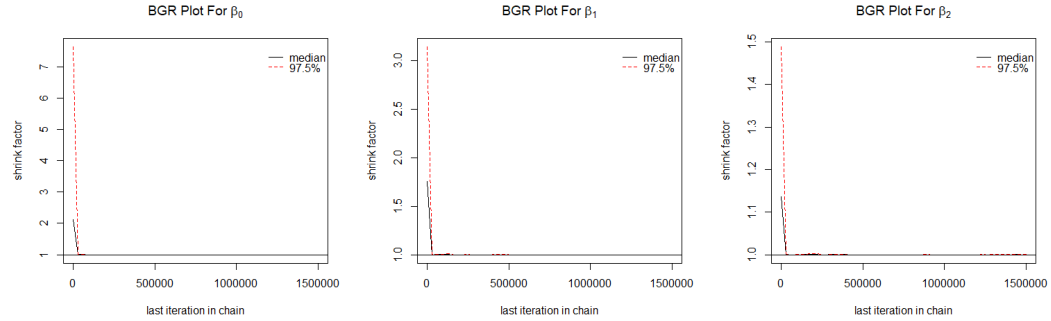

Figure B.56: BGR Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

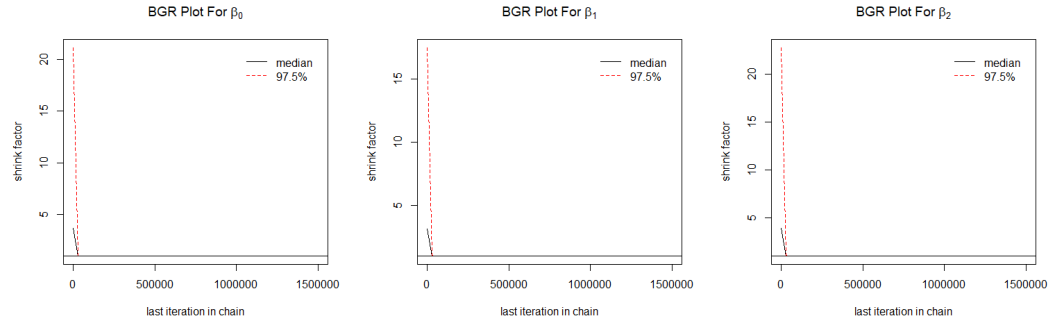

Figure B.57: BGR Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

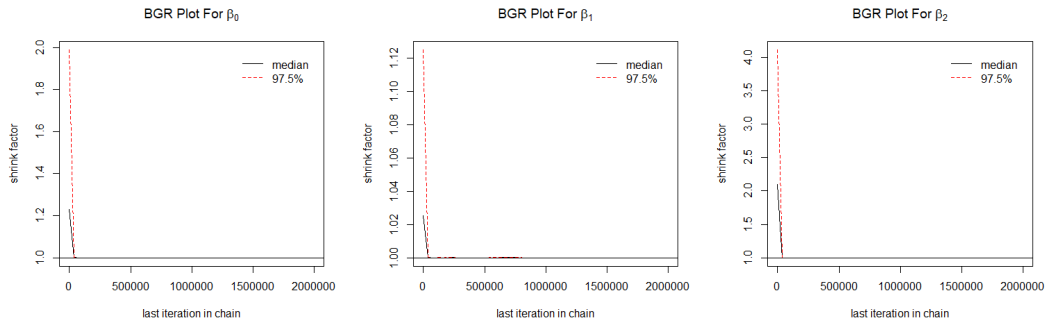

Figure B.58: BGR Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

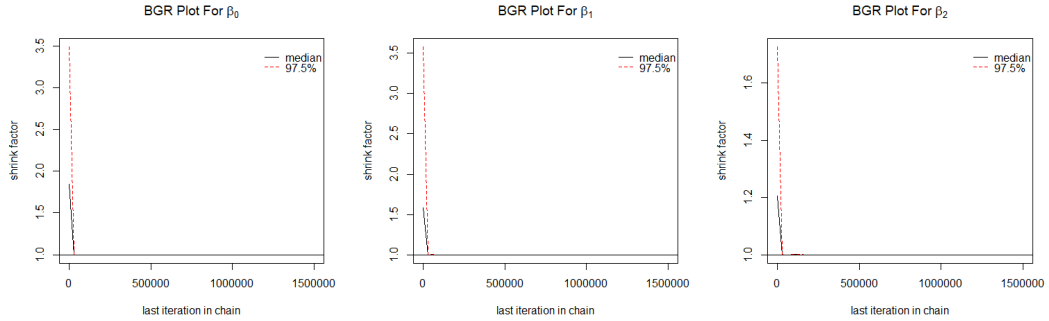

Figure B.59: BGR Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

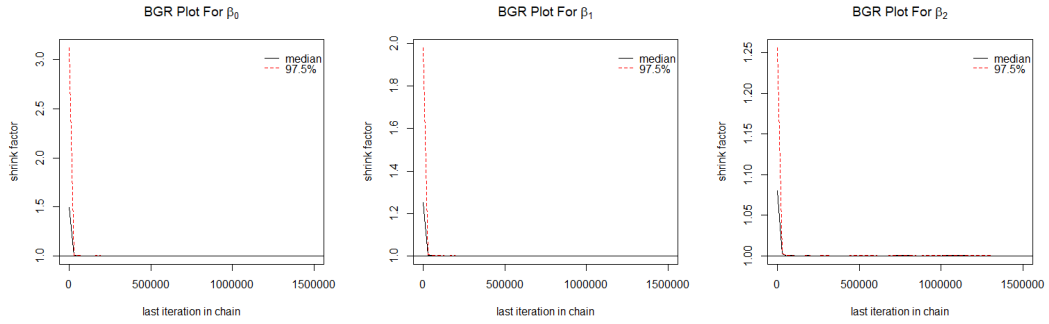

Figure B.60: BGR Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

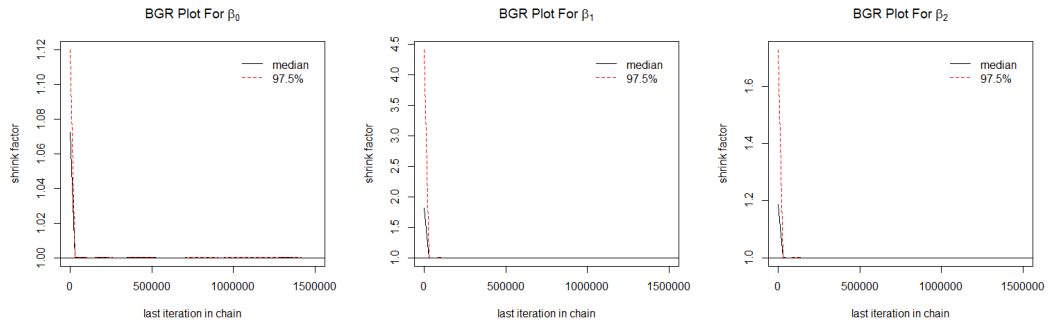

Figure B.61: BGR Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

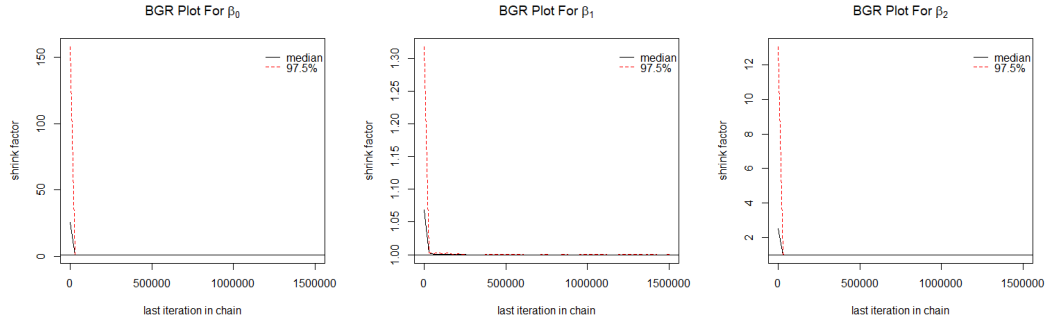

Figure B.62: BGR Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

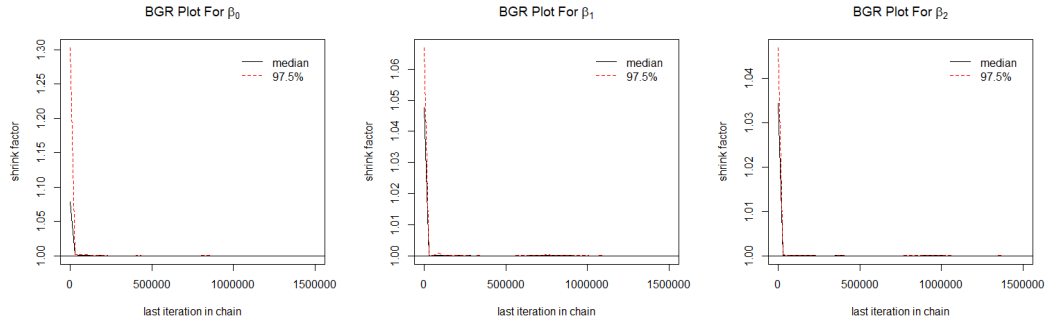

Figure B.63: BGR Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

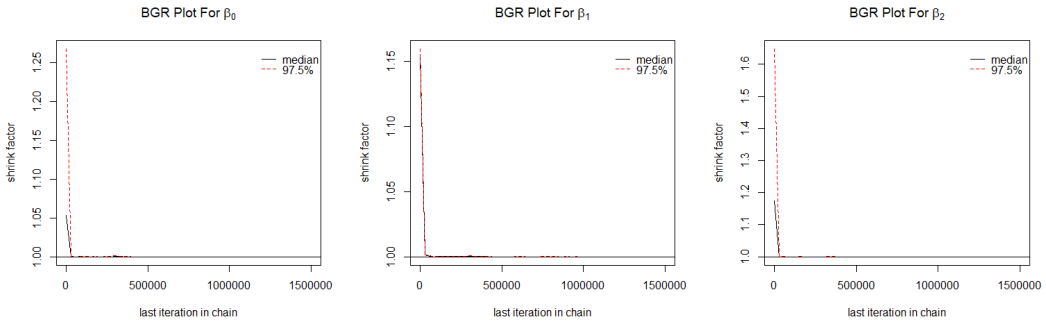

Figure B.64: BGR Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

B.3.5 Geweke Plots For Validation Study

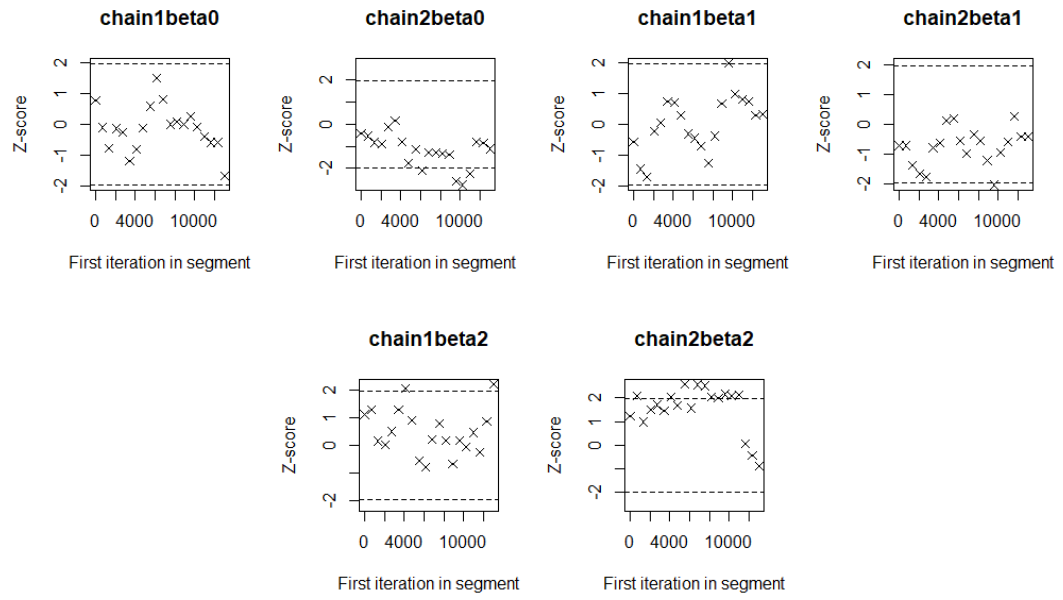

Figure B.65: Geweke Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

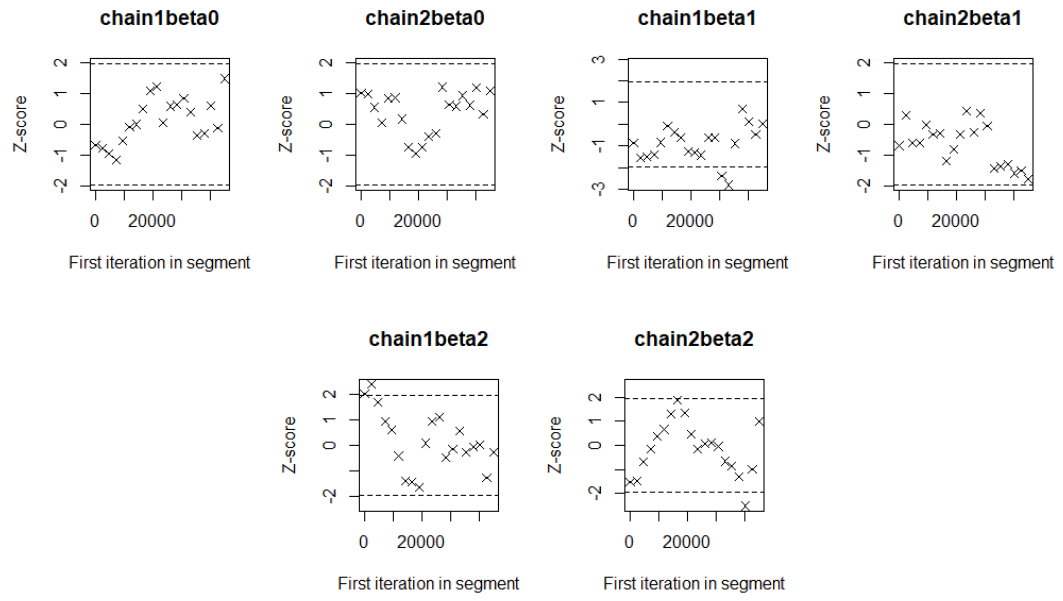

Figure B.66: Geweke Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

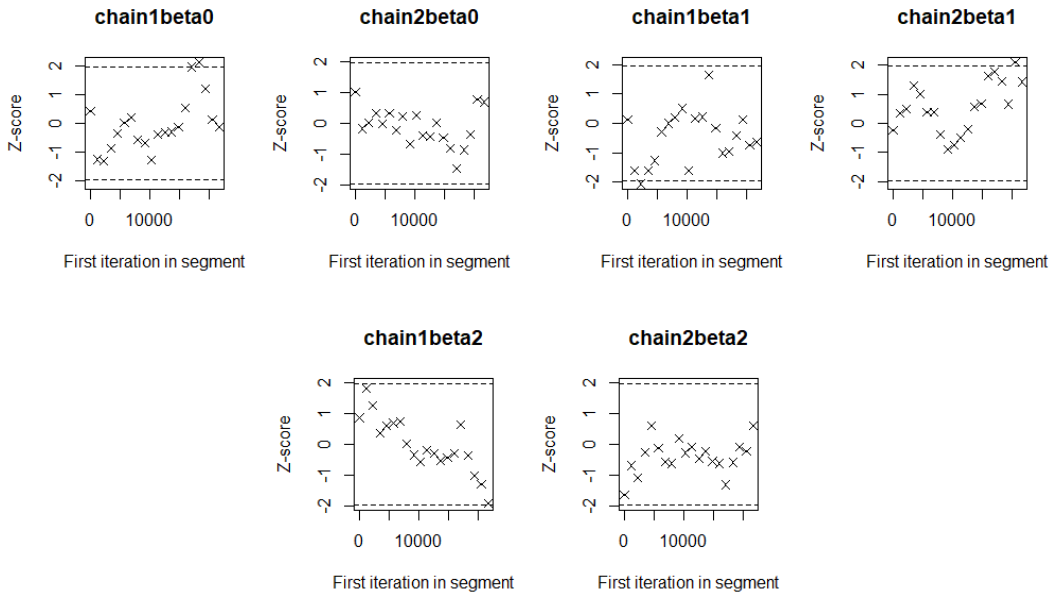

Figure B.67: Geweke Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

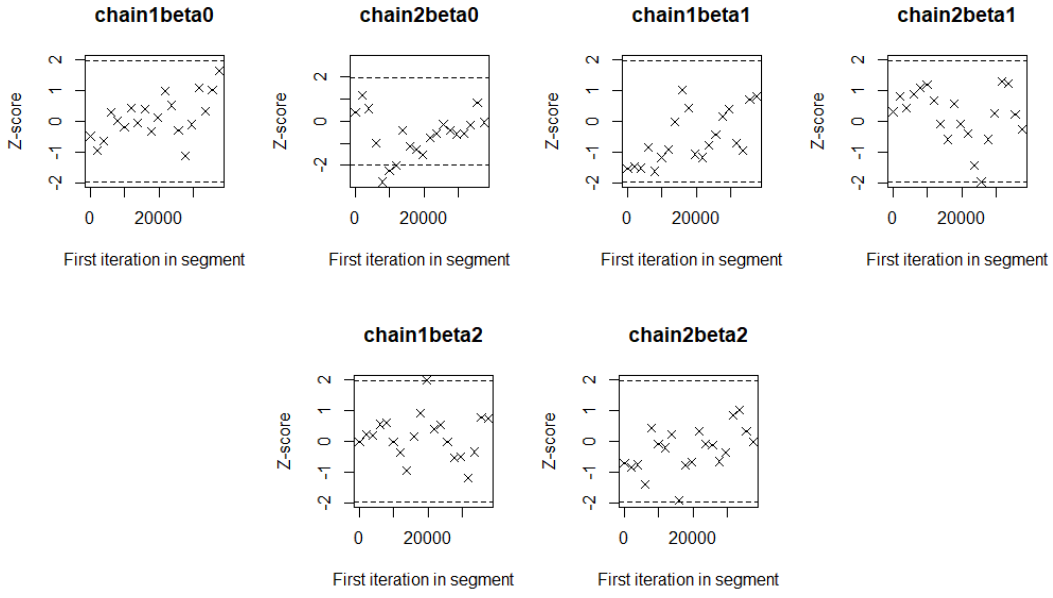

Figure B.68: Geweke Plots For 1% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

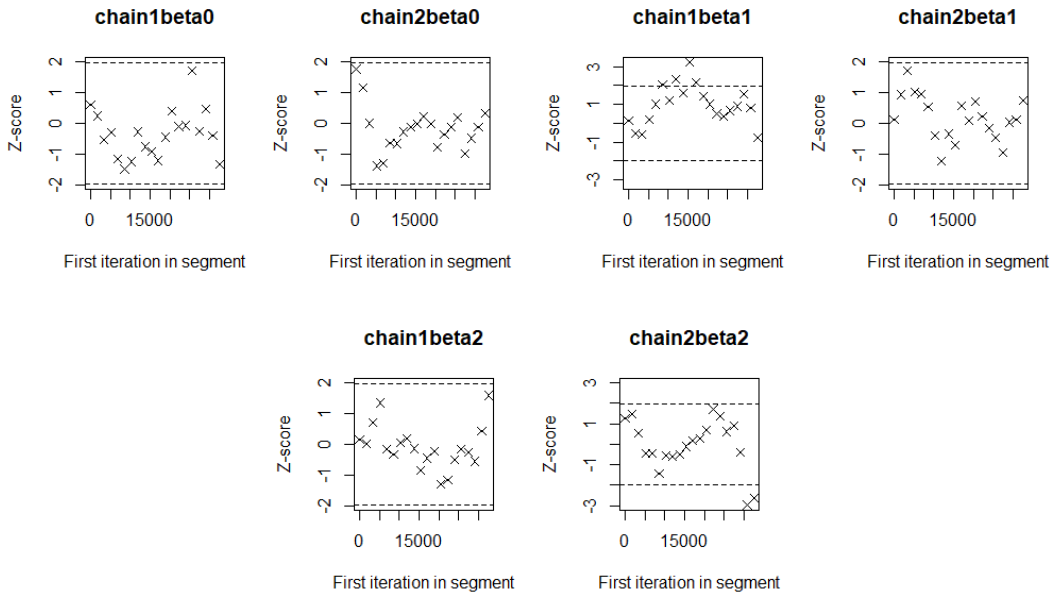

Figure B.69: Geweke Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

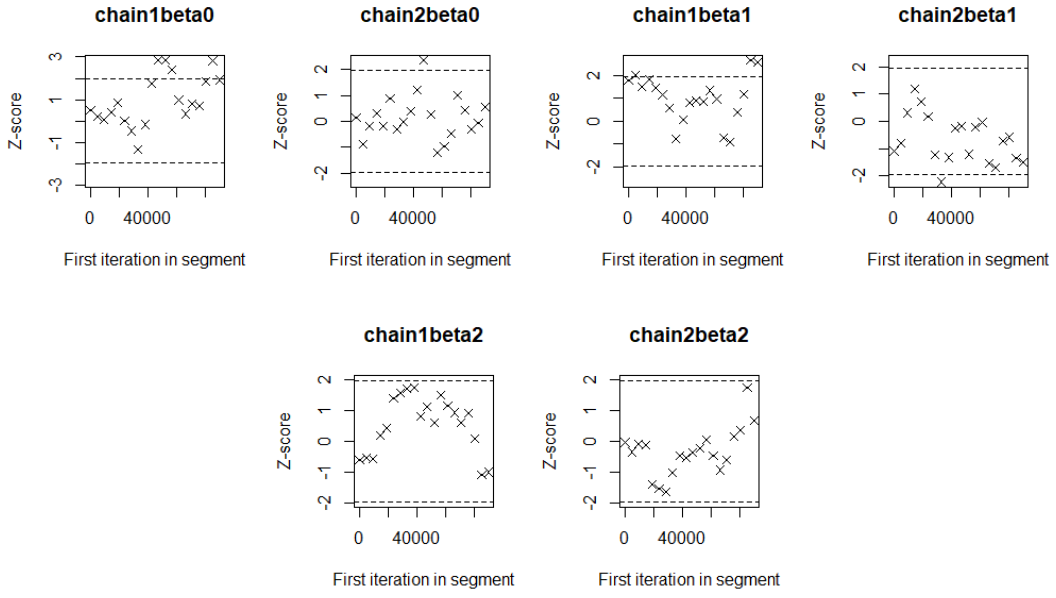

Figure B.70: Geweke Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

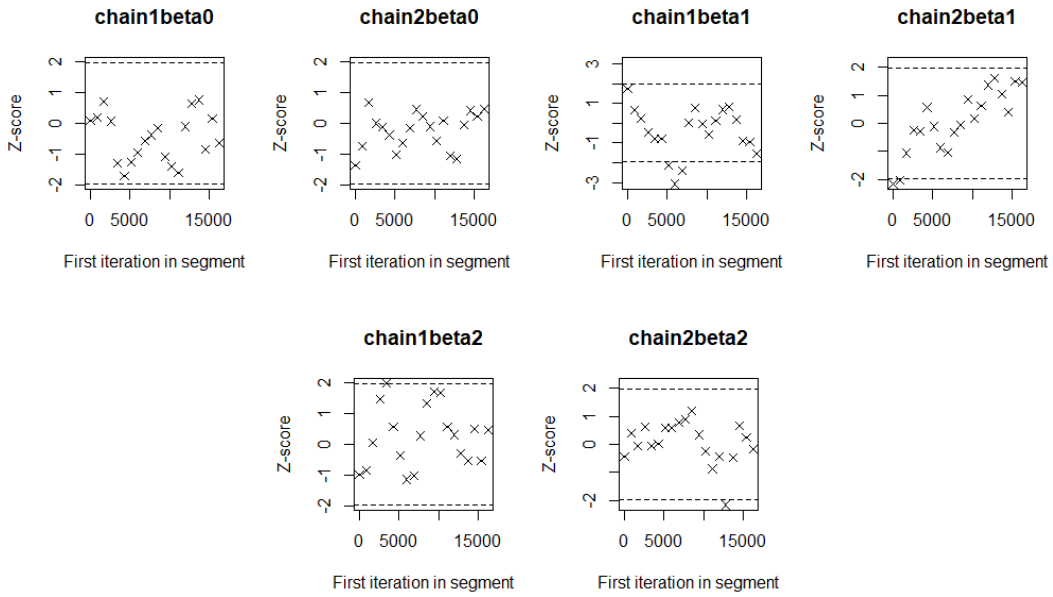

Figure B.71: Geweke Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

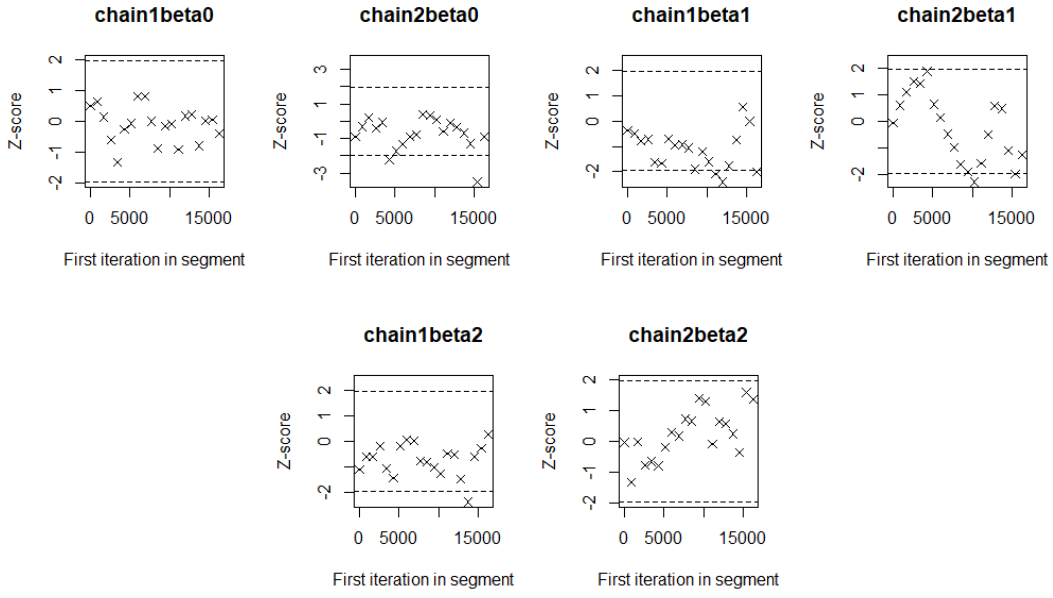

Figure B.72: Geweke Plots For 3% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

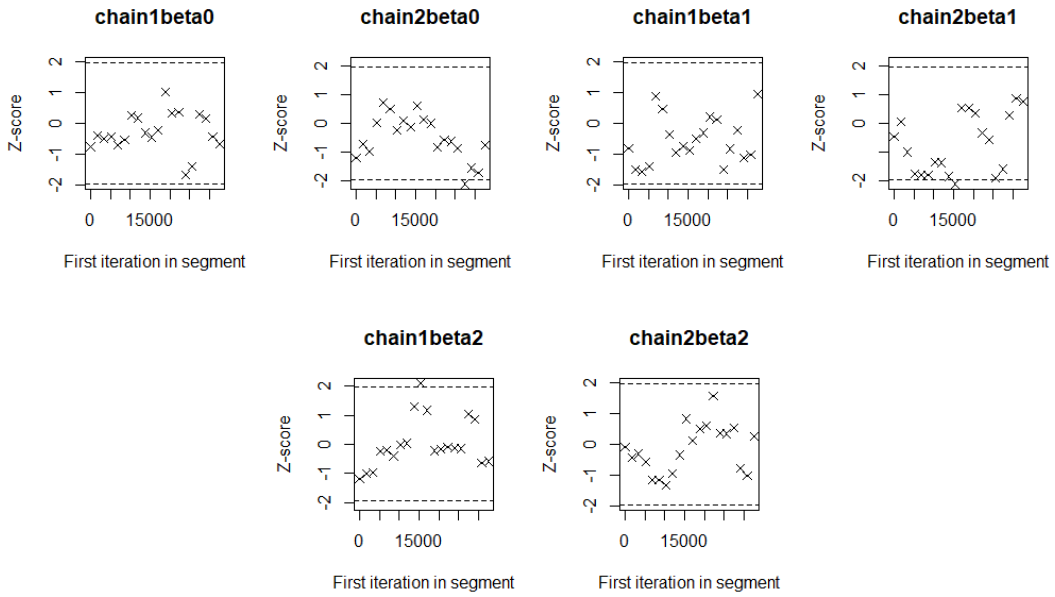

Figure B.73: Geweke Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

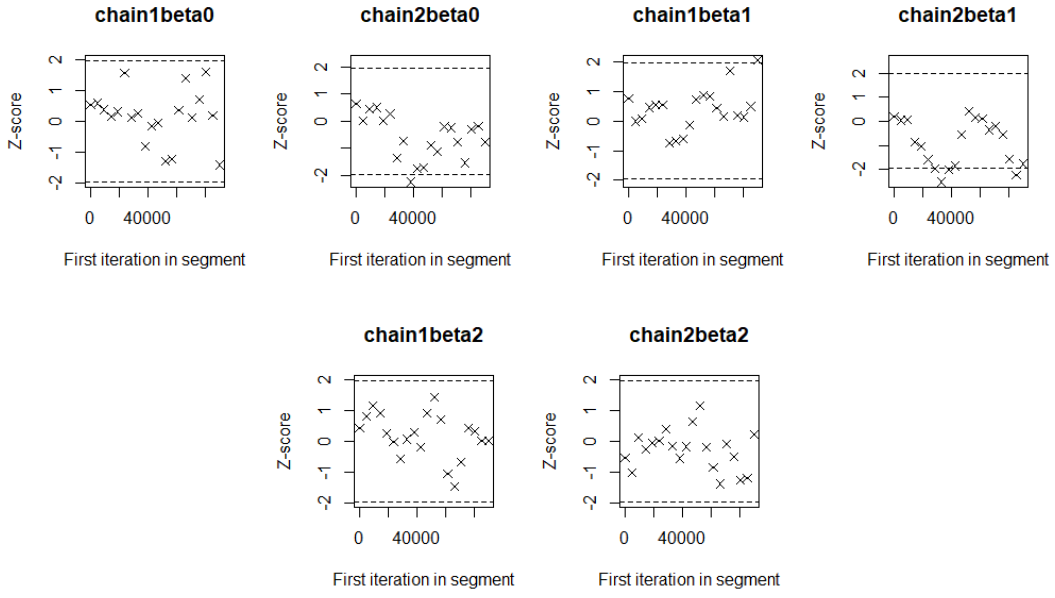

Figure B.74: Geweke Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

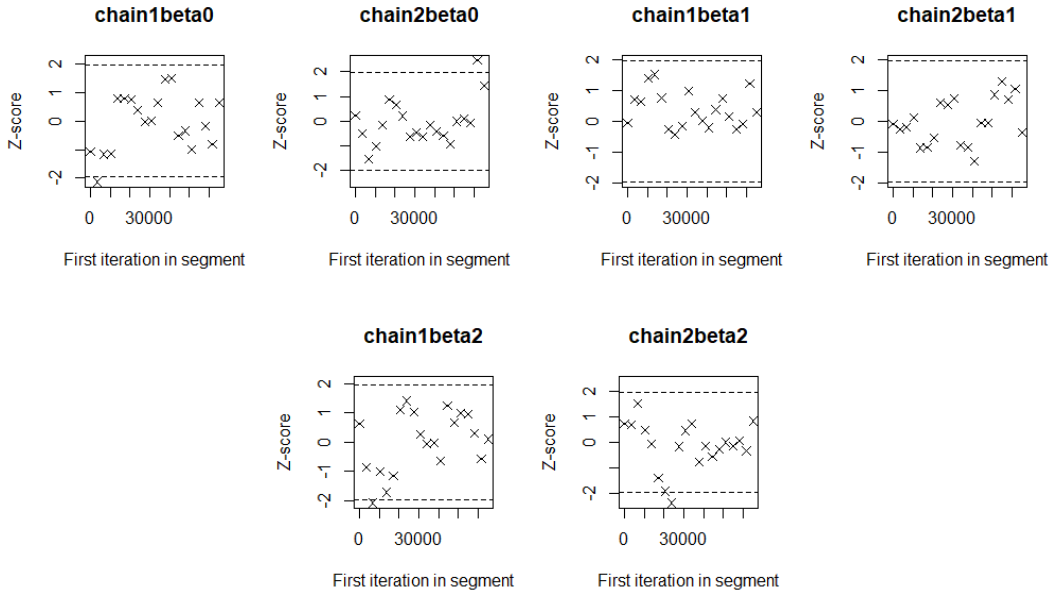

Figure B.75: Geweke Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

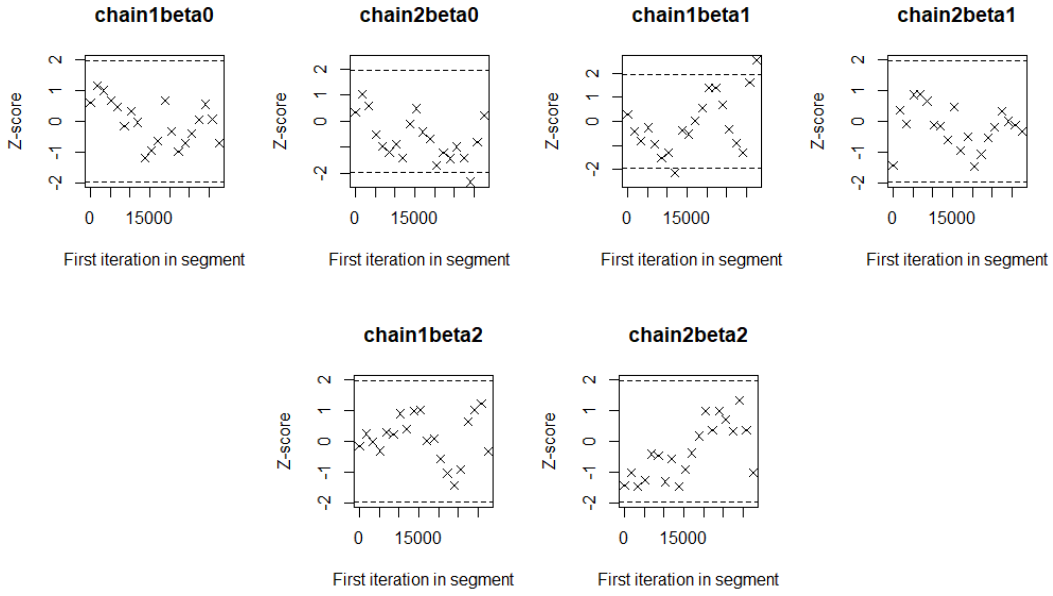

Figure B.76: Geweke Plots For 5% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

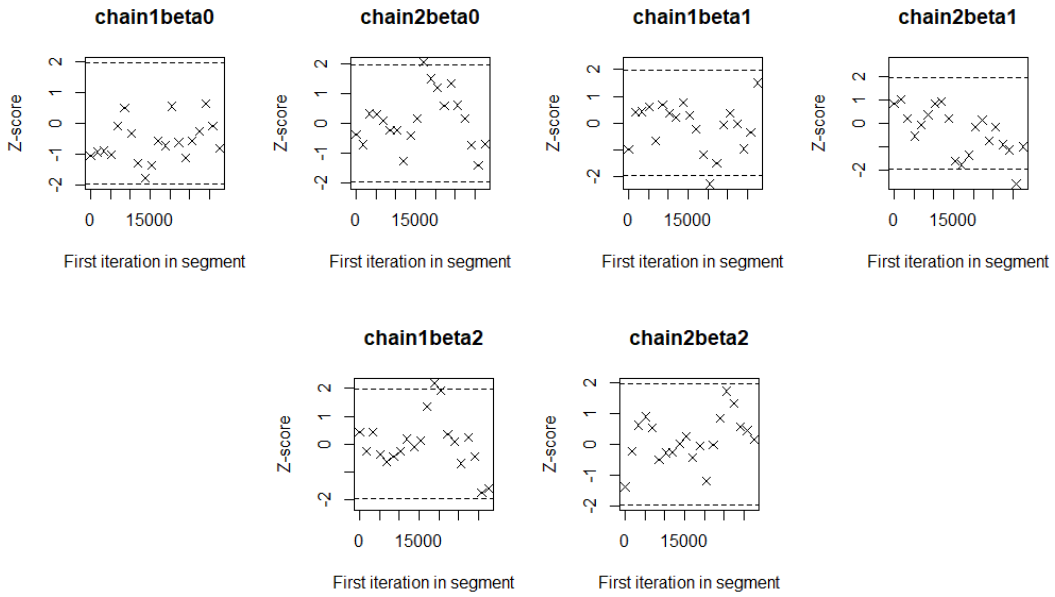

Figure B.77: Geweke Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

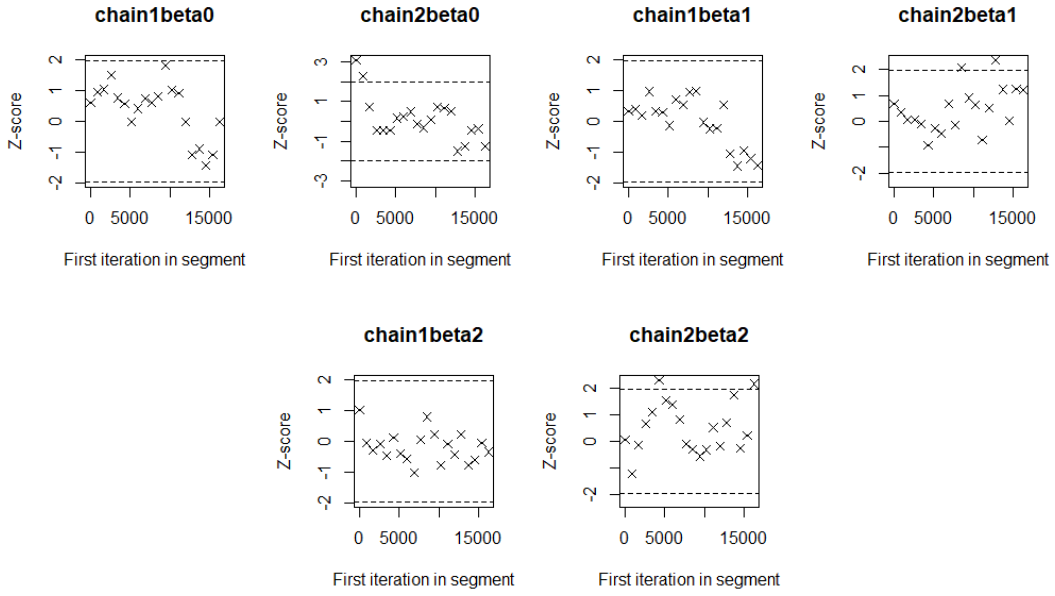

Figure B.78: Geweke Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 4^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

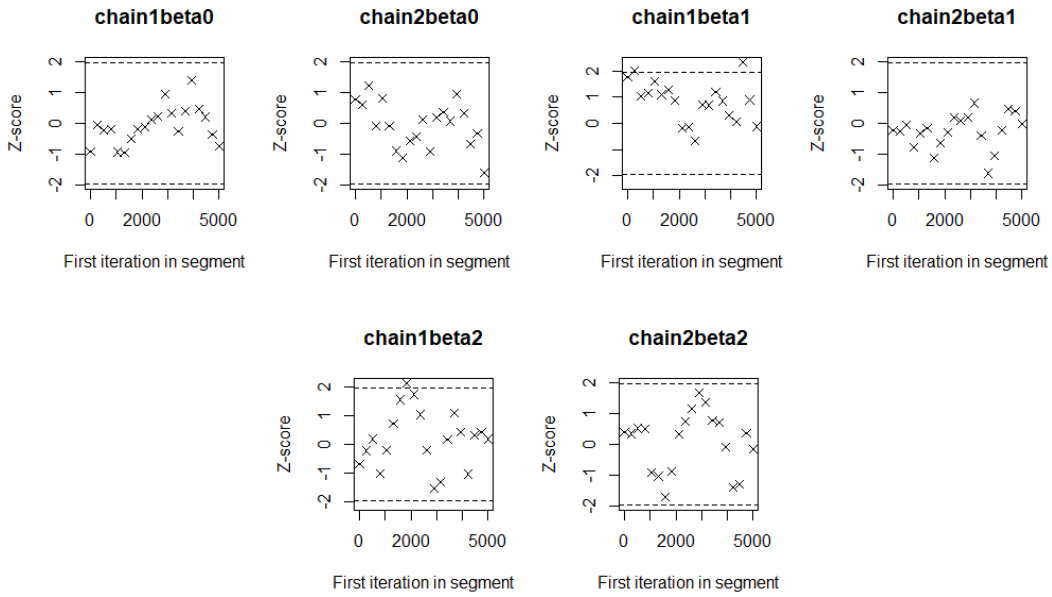

Figure B.79: Geweke Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1^2)$

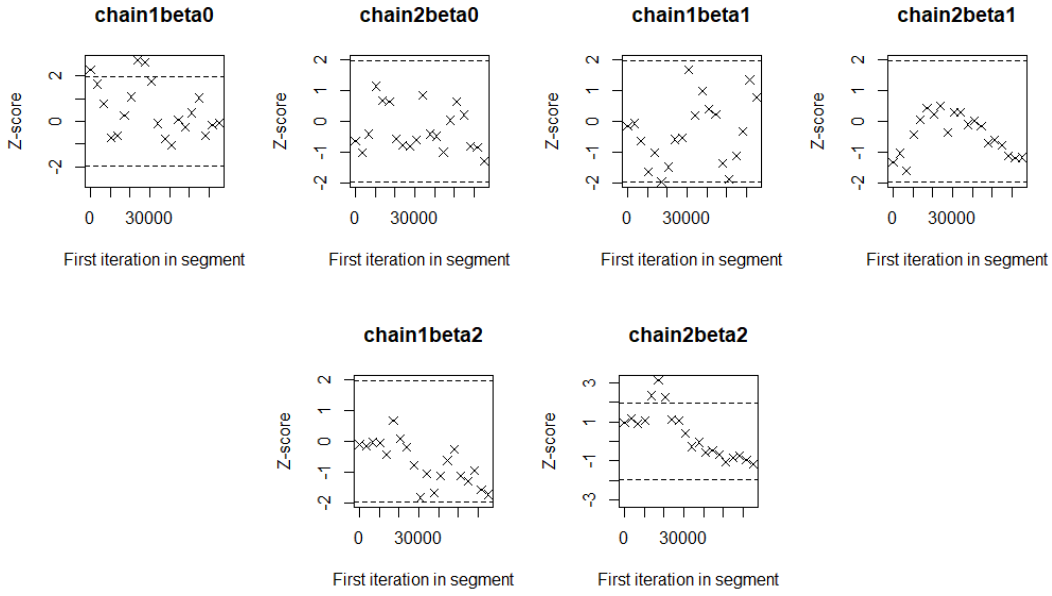

Figure B.80: Geweke Plots For 10% Missing Rate Under Prior Assumptions β and $\phi \sim \mathcal{N}(0; 6^2)$ and $X_1 \sim \mathcal{N}(-1; 1.5^2)$

Appendix C

APPENDIX: REAL DATA APPLICATION

C.1 MCMC Diagnostics For Real Data Application Under Prior Assumption A

C.1.1 Monte Carlo Standard Errors (MCSE) of Coefficients

Table C.1: Monte Carlo Standard Errors (MCSE) of Coefficients

Coefficient	**MCSE**	**Coefficient**	**MCSE**
$\hat{\beta}_0$	0.001130	$\hat{\beta}_6$	0.000725
$\hat{\beta}_1$	0.000009	$\hat{\beta}_7$	0.001144
$\hat{\beta}_2$	0.002685	$\hat{\beta}_8$	0.000781
$\hat{\beta}_3$	0.000031	$\hat{\beta}_9$	0.001552
$\hat{\beta}_4$	0.001043	$\hat{\beta}_{10}$	0.002546
$\hat{\beta}_5$	0.001208	$\hat{\beta}_{11}$	0.003082

C.1.2 BGR Plots of Coefficients

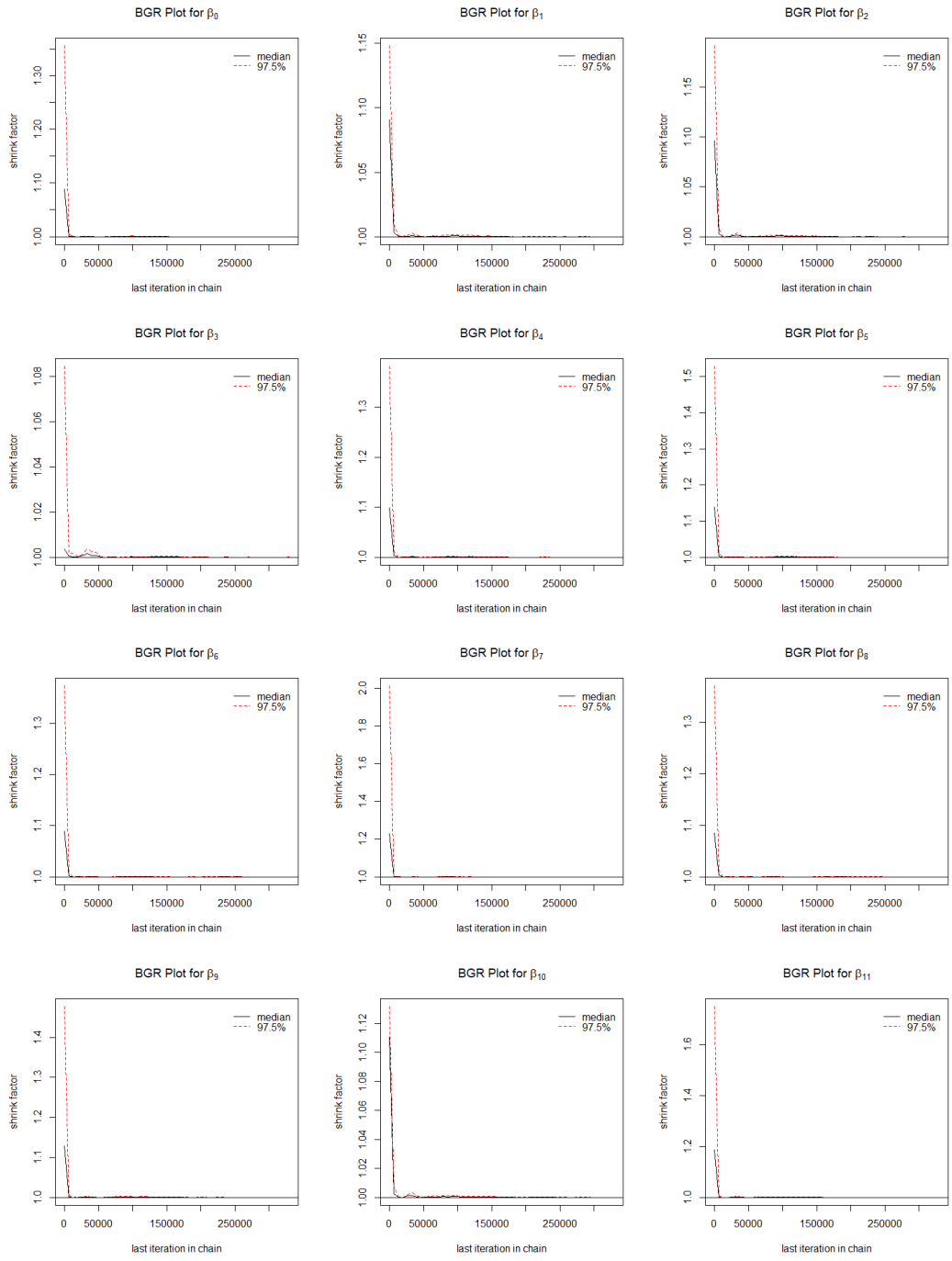

Figure C.1: Real Data Analysis: BGR Plots of Coefficients

C.1.3 Density Plots of Coefficients

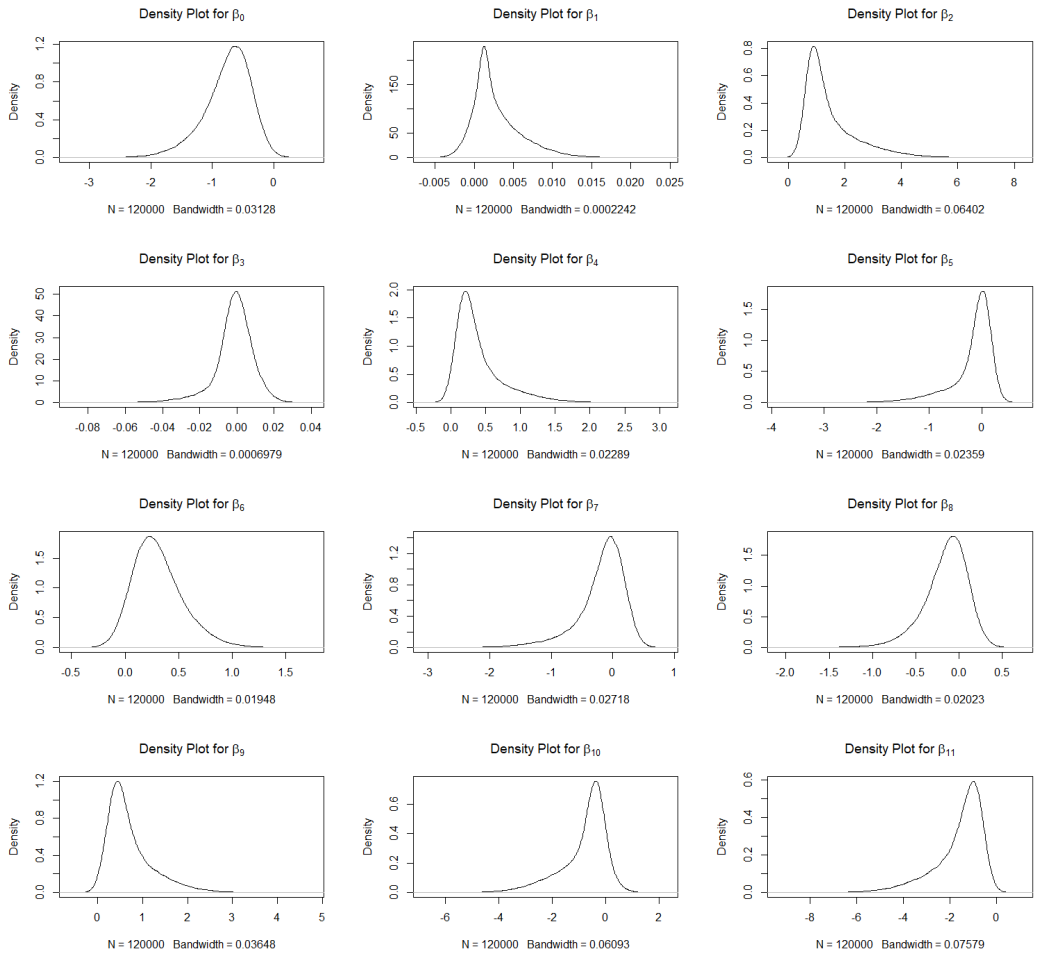

Figure C.2: Real Data Analysis: Density Plots of Coefficients

C.1.4 Trace Plots of Coefficients

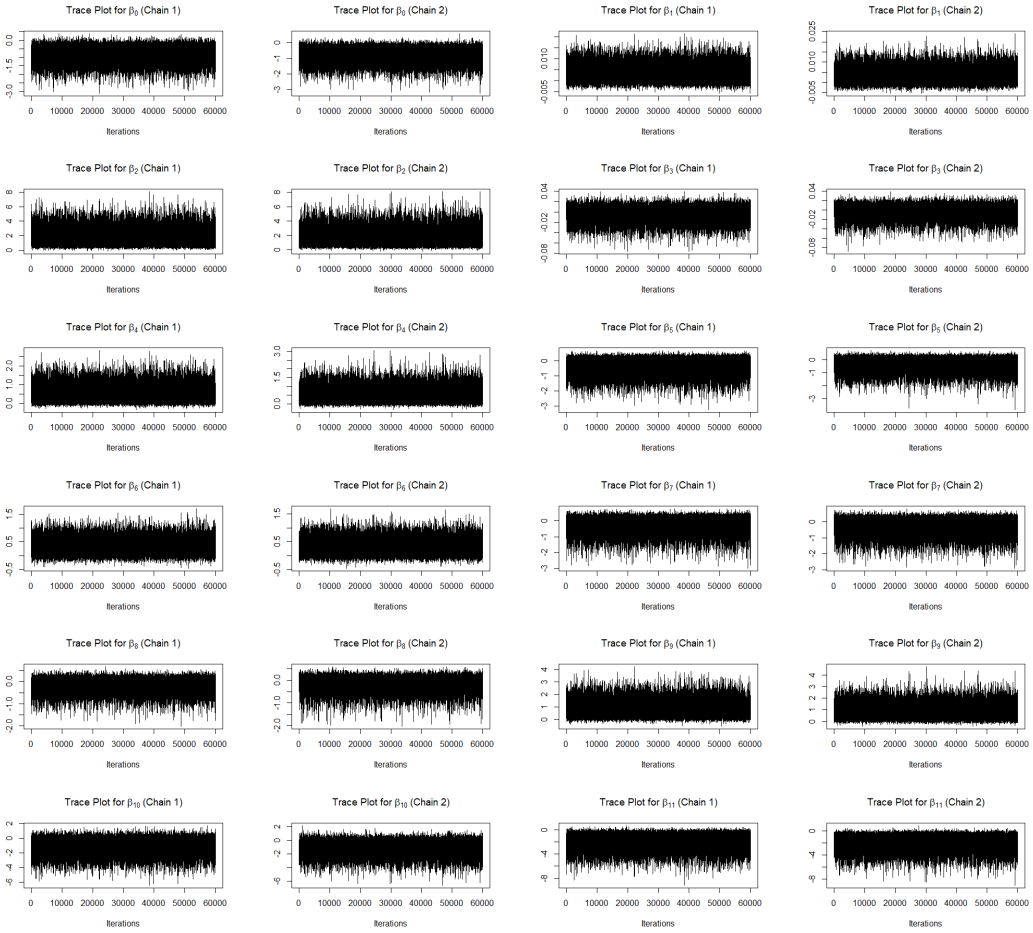

Figure C.3: Real Data Analysis: Trace Plots of Coefficients

C.1.5 Autocorrelation Plots of Coefficients

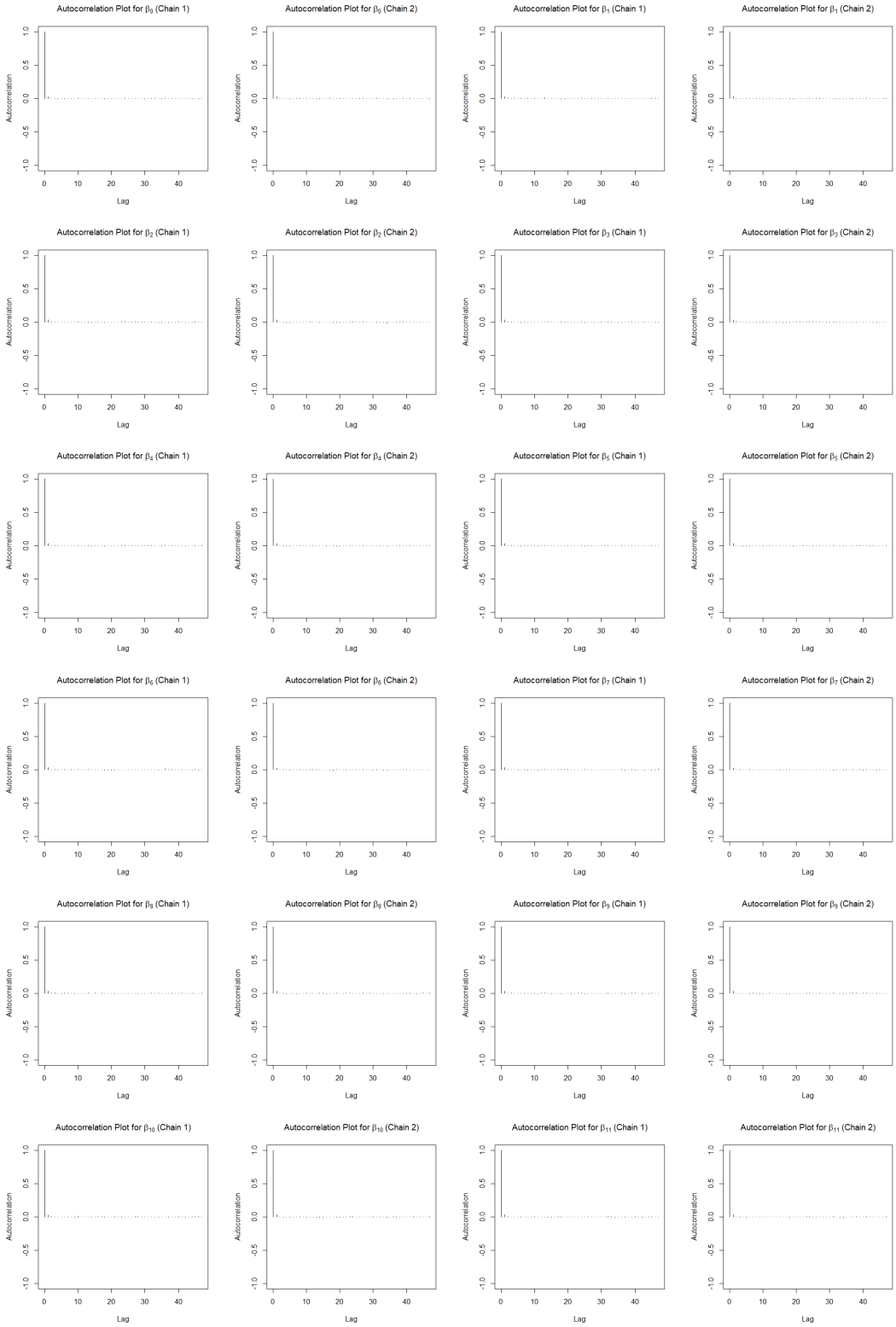

Figure C.4: Real Data Analysis: Autocorrelation Plots of Coefficients

C.1.6 Geweke Plots of Coefficients

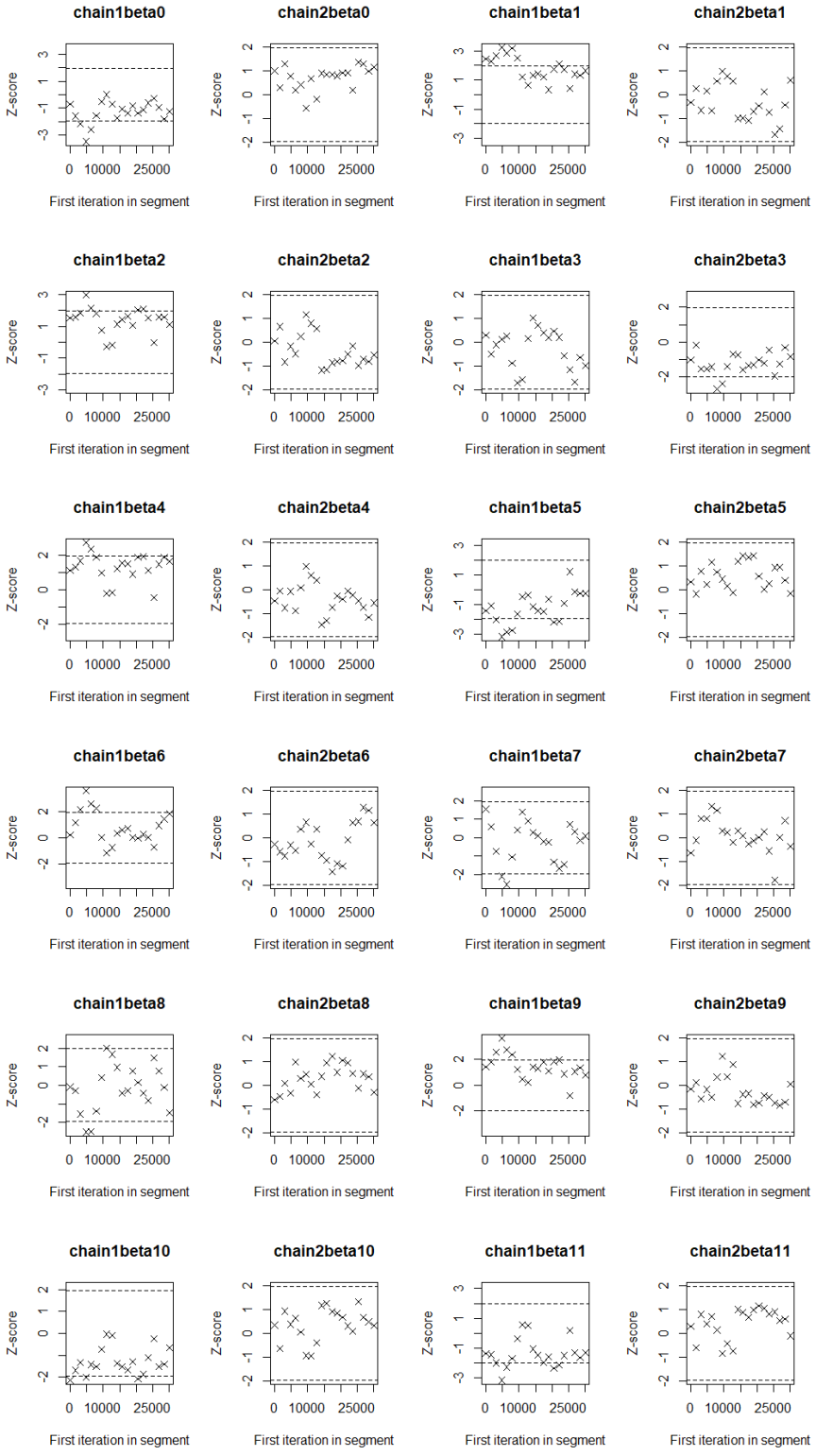

Figure C.5: Real Data Analysis: Geweke Plots of Coefficients

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Çiftçi, Sezgin
Nationality: Turkish
Date of Birth: 20.09.1986
Place of Birth: Ankara, Turkey
Marital Status: Single
Phone: +90 (505) 437-6570
E-mail: sezginciftci@yahoo.com

EDUCATION

Degree	Institution	Year of Graduation
M.Sc.	Actuarial Science, Inst. of App. Math., METU	2012
B.S.	Statistics, METU	2009
High School	Gazi Anatolian High School	2004

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2015 - Present	Dept. of Insurance and Risk Management, Başkent University	Lecturer
2011 - 2015	Dept. of Insurance and Risk Management, Başkent University	Research Assistant

PUBLICATIONS

1. Cula, S., Muluk, Z., **Çiftçi, S.** (2016). Sigorta Personelinin Sigorta Eğitimi Konusunda Görüşleri. *Journal of Insurance Research/Sigorta Arastirmalari Dergisi*, 12.

2. **Çiftçi, S.**, Düzgün, Ş., Erkan, B.B. (2014). Risk analysis based on spatial analysis of chronic obstructive pulmonary disease (COPD) with respect to provinces in Turkey. *Journal of Computational and Applied Mathematics.*

Posters

1. **Çiftçi, S.**, Evkaya, O. (2013). The Prediction of Turkish Local Derby Matches. 8th International Statistical Congress (ISKON), 27 – 30 October, 2013. Antalya, Turkey.

Presentations

1. Tuzun, I. K., Takay, B. A., Batum, U., **Ciftci, S.** (2015, April). A Research on Women Entrepreneurship in Turkey. In International OFEL Conference on Governance, Management and Entrepreneurship (p. 967). Centar za istrazivanje i razvoj upravljanja doo.

2. **Çiftçi, S.**, Düzgün, Ş., Erkan, B.B. (2014). Risk analysis based on spatial analysis of chronic obstructive pulmonary disease (COPD) with respect to provinces in Turkey. International Conference on Applied and Computational Mathematics (ICACM), October 3 – 6 2012, Middle East Technical University, Ankara, Turkey.

3. Açıkalın S, Kuyucuoğlu Ö, **Çiftçi S.** (2013). Siber Saldırılara Karşı Veri Koruma Sigortası ve Siber Saldırı Mali Sorumluluk Sigortasının Türkiye'deki Yeri ve Uygulaması. 1. Ulusal Sigorta ve Aktüerya Kongresi: Ankara, Türkiye; 06/06/2013 - 07/06/2013

Projects

1. Project Researcher, *Türkiye Hayat Tabloları Yenilenmesi Projesi*, Türkiye Sigorta Birliği, 2015-2017

2. Project Researcher, Batum U, Kalemci Tüzün İ, Araz Takay B., **Çiftçi S**, Kıral, S. ;*Ankara Kadın Girişimciliği; Öneriler ve Yol Haritası.* Ankara Kalkınma Ajansı, 31/12/2013.