5-2020

Bayesian Approach of Joint Models of Longitudinal Outcomes and Informative Time

Adbullah Ahmed Zaagan

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

BAYESIAN APPROACH OF JOINT MODELS OF LONGITUDINAL OUTCOMES AND INFORMATIVE TIME

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Abdullah Ahmed A Zaagan

College of Education and Behavioral Sciences
School of Educational Research Leadership and Technology
Department of Applied Statistics and Research Methods

May 2020

This Dissertation by: Abdullah Ahmed A Zaagan

Entitled: *Bayesian Approach of Joint Models of Longitudinal Outcomes and Informative Time*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in the College of Education and Behavioral Sciences in Department of Applied Statistics and Research Methods

Accepted by the Doctoral Committee

_____
Khalil Shafie Holighi, Ph.D., Research Advisor

_____
Trent L. Lalonde, Ph.D., Committee Member

_____
Han Yu, Ph.D., Committee Member

_____
Heng-Yu Ku, Ph.D., Faculty Representative

Date of Dissertation Defense _____

Accepted by the Graduate School

_____

Cindy K. Wesley, Ph.D.
Associate Provost and Dean
Graduate School and International Admission

ABSTRACT

Zaagan, Abdullah Ahmed. *Bayesian Approach of Joint Models of Longitudinal Outcomes and Informative Time.* Published Doctor of Philosophy dissertation, University of Northern Colorado, 2020.

Longitudinal studies are commonly encountered in a variety of research areas in which the scientific interest is in the pattern of change in a response variable over time. In longitudinal data analyses, a number of methods have been proposed. Most of the traditional longitudinal methods assume that the independent variables are the same across all subjects. It is commonly assumed that time intervals for collecting outcomes are predetermined and have no information regarding the measured variables. However, in practice, researchers might occasionally have irregular time intervals and informative time, which violate the above assumptions. Hence, if traditional statistical methods are used for this situation, the results would be biased.

The joint models of longitudinal outcomes and informative time are used as a solution to the above violations by using joint probability distributions, incorporating the relationships between outcomes and time. The joint models are designed to handle outcome distributions from a normal distribution with informative time following an exponential distribution.

Several studies used the maximum likelihood parameter estimates of the joint model. This study, however, presented an alternative method for parameters estimation, based on a Bayesian approach, with respect to joint models of longitudinal outcomes and

informative time. Using a Bayesian approach permitted the inclusion of knowledge of the observed data within the analysis through the prior distribution of unknown parameters.

In this dissertation, the prior distribution adopted three scenarios: (1) the prior distributions of all unknown parameters are noninformative prior, which will set to be vague but proper prior: Normal(0, 1e6). (2) The prior distributions of all unknown parameters are informative prior, which will be set to be normal for unrestricted parameters, and inverse gamma (IG) priors for positive parameters such as the variance σ^2. (3) A combination of two above scenarios, so the prior distributions of some unknown parameters are noninformative, and the others are informative.

The procedure for estimating the model parameters was developed via a Markov chain Monte Carlo method using the Metropolis-Hastings algorithm. The key idea was to construct the likelihood function, specify the prior information, and then calculate the posterior distribution. Simulated observations were generated by the MCMC technique from the posterior distribution.

Thus, the primary purpose of this study was to find Bayesian estimates for the unknown parameters in the joint model, with the assumptions of a normal distribution for the outcome process and an exponential distribution for informative time. The properties and merits of the proposed procedure were illustrated employing a simulation study through a written R program and OpenBUGS.

ACKNOWLEDGEMENTS

I'd like to sincerely thank almighty Allah for all his grants that he bestowed on me. Without His help, nothing can be accomplished. May Allah accept all my efforts and bless it, because it is not worth anything if it is not accepted and blessed by Him.

Next, I would like to express my sincere gratitude to my supervisor and my model professor Dr. Khalil Shafie for his excellent guidance, encouragement, caring, and patience. Since I have joined the department and up through writing this dissertation, he has never once been too busy to help and guide me through research and life challenges. He is mindful of training me to be an independent researcher. I have learned a lot from him during my academic career.

My sincere appreciation also goes to Dr. Trent Lalonde, who has provided me with invaluable chances to obtain hands-on research experiences in the computer lab. Very fortunately, it was his generous support and warmest encouragement that brought me onto my right academic track. I sincerely appreciate his quick and generous responses and his valuable comments and advice.

I am also grateful to the rest of my supervisory committee, Dr. Han Yu and Dr. Heng-Yu Ku, for their guidance, helpful suggestions, flexibility, and inspiration. I would also like to express my thanks for all those who have helped, guided, or encouraged me, in any way, during my years of graduate study in the department of Applied Statistics and Research Methods, especially Dr. Susan Hutchinson, Dr. Jay Shaffer, and Keyleigh Gurney.

I would like to present my deepest thankfulness to my dear mother, Fatimah Ahmed Almarai, and my deceased father, Ahmed Abdullah Zaagan, for their great role in my life and their numerous sacrifices for me and for my sisters and brothers. Many thanks for all my supportive family members for their continuous praying and sincere wishes for my success.

There are no proper words to convey my heartfelt gratitude and respect for my wonderful wife, Ashwag Helil Alziyad, for her love, continuous support, great patience, and understanding of my educational and career goals and the sacrifices she made so that I might obtain those goals. She is my rock, my inspiration, and my world, who was always there, cheering me up and stood by me through the good and bad times. Without her patience and support, I would have found it very difficult to overcome the ups and downs during the study. I owe her everything.

Furthermore, I would also like to thank my two wonderful sons, Rakan and Jaber, and my two lovely daughters, Rytal and Juanna, for filling my life with joy and happiness. Thank you for being good kids with your mother when I was busy and studying in the School.

Finally, I would like to thank Jazan University for giving me an opportunity to receive a scholarship and complete my Ph.D. study. Without this very important assistance, my dream can never become true. I am deeply grateful to some people who support, engorge, and put me on the right track, especially, Dr. Abdullah Alhussin. Also, I would like to thank the Saudi Arabia Culture Mission (SACM) for helping me during my study in the United States.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The term “longitudinal data” is a popular name that is used instead of repeated measurements when the outcomes on the same subjects or experimental units are collected or measured for a relatively long period of time to evaluate the changes over a period of time. (Hedeker & Gibbons, 2006). One key attribute of longitudinal study designs is the exclusion of between-subject variability, focusing on trends or patterns in changes that occur for the subjects. Thus, it increases its power over traditional cross-sectional designs in terms of the ability to capture the within-subject effect. In addition, because of the exclusion of between-subject variability, it is possible to calculate the estimate of within-subject effects with increased accuracy (Fitzmaurice, Laird, & Ware, 2012; Hedeker & Gibbons, 2006). Although longitudinal studies take a much longer time to complete, are more expensive, and can be harder to analyze, they have gained popularity because it is believed that the problem of causality can be solved (Twisk, 2013). Another attribute of longitudinal data is the correlation of all outcomes from one individual, requiring particular statistical methods to address this correlation. Therefore, it is particularly critical that the appropriate statistical methods are chosen to accommodate the specific type of outcome data and the covariates. For example, longitudinal data can take the form of continuous, binary, count, among others, and evaluation procedures need to fit the type of data.

A longitudinal study has the ability to observe an outcome and predictors simultaneously, which can help to define if there were changes in predictors before changes

show up in outcomes and whether those changes affected the outcome. Because of the different kinds of responses and independent variable types, a number of methods have been developed, which can range from simple to quite complex, depending on the study design and/or research objectives. Each method works under certain assumptions and for different types of situations; however, no one method can accommodate all of these situations. It is, therefore, incumbent upon the researcher to select the most appropriate method based on their research questions, purpose, and method of data collection.

An inherent problem with longitudinal methods is the large data to be analyzed, as well as the complexity of the data structure. First of all, a basic assumption is that the measurements are correlated in some way with each other since the outcomes are collected repeatedly at multiple time points from the same subjects. Traditional approaches to analysis such as the simple method, Analysis of Variance (ANOVA), and Multivariate Analysis of Variance (MANOVA), though commonly employed in longitudinal research, are inappropriate in many cases due to an unrealistic assumption that outcomes are independent of each other, or at least they cannot be too correlated to each other. Tabachnick and Fidell (2012) suggest that no correlation should be above ($r = 0.90$). However, ignoring the correlation among outcomes, which cannot be true in a longitudinal design.

Second of all, often, longitudinal studies take place over an extended period of time, during which many events may happen that affect the data collection but over which the researcher has no control. Participants may drop out for some reason or another or miss certain data collection points. Also, it is conceivable that individuals with poorer health outcomes will be asked to visit for more check-ups. Participants may not share common time occasions; instead, each participant may follow individual schedules for appointments

to gather data depending on their prior health outcomes (Lipsitz, Fitzmaurice, Ibrahim, Gelber, & Lipshultz, 2002). As a result, the final data bank may be incomplete or unbalanced or has missing data due to attrition, all of which add to the difficulties in analyzing the data, without any intention of the researcher initially.

Given the variety and complexity in the difficulties arising from the analysis of longitudinal data, there have been multiple proposed methods that purport to handle these problems (Hedeker & Gibbons, 2006). Unfortunately, most of these methods are limited to cases which do have a complete data set (Davis, 2002). To address this situation, the mixed-effects model or the generalized estimating equation (GEE) has been developed and is gaining popularity among some researchers. However, due to the complexity inherent in these methods of analysis, many researchers continue to use traditional methods that are within their scope of understanding due to their relative simplicity of computation. Both the mixed-effects model and the generalized estimating equations (GEE) are associated with highly math-oriented or computer-based techniques as a way to accommodate the special characteristics of longitudinal data: the correlation among outcomes, and unbalanced data structure, which is unequal time points in gathering data for each subject. As a result, even though these methods yield more accurate estimations of changes over time, their computational complexity and advanced computer skills often detract from their being considered in research designs. Nevertheless, fields such as biology, pharmaceuticals, and economics have influenced researchers to consider these models in light of their accuracy and efficiency of the estimation.

An added attraction of the GEE approach is its ability to handle binary and count outcomes as well as continuous outcomes. In addition, both the mixed-effects model and the GEE work under the assumption of fixed or predetermined times for data collection.

However, the estimators may be biased when the time points for data collection are irregular, or data unbalanced (Lin, Scharfstein, & Rosenheck, 2004). Thus, the studies continue for better and improved methods, as in finding an appropriate tool to analyze data when time is informative; that is, the upcoming time points for collecting measurements are adaptively determined based on the current outcomes for each subject. In this case, the methods mentioned above are not appropriate.

The Joint Model

In reality, there are occasions that alter unexpectedly the time points for collecting data, which can pose serious problems in longitudinal studies. Attrition is one of these occasions, along with situations that cause a delay or advance in time points due to sickness, family emergencies, vacation, and the like. Regardless of the reason, the effect is the creation of irregular time periods for collecting data, unbalanced data, and non-ignorable dropouts. These events mark the absence of important information and should not be ignored. Informative time needs to be integrated into models of analysis that can be calculated along with the longitudinal variable so as to yield the best possible inference. The models discussed above do not account for informative time and therefore are not appropriate for this study whereas the joint model, which can accommodate informative time, offers a better approach (Henderson, Diggle, & Dobson, 2000; Kim, Zeng, Chambless, & Li, 2012; Liang, Lu, & Ying, 2009; Lipsitz et al., 2002; Qiu, Stein, & Elston, 2013). Fundamentally, joint models are based on the joint distribution of outcomes and the time-related factor with maximum likelihood estimation and are versatile enough to apply to any kind of situation while providing more precise results (Qiu et al., 2013).

Making adaptations to the joint model to reflect iconic circumstances need to be done carefully, dutifully considering the assumptions embodied within the specific model.

An example of a researcher making such an adaptation is Bronsert (2009), who presented a joint mode named Gaussian-Exponential Model, in which normally distributed longitudinal responses and intermittent informative times following an exponential distribution are combined. In this way, he demonstrated that the joint model has a very good ability to accommodate longitudinal data analysis when compared to the mixed-effects model in his simulation study. A few years later, Lin (2011) extended Bronsert's study and showed that the parameters estimate to maintain the property of multivariate normality and that the joint model can be considered as an alternative method for the analysis of longitudinal data. Next, Seo (2015) developed an even further extension of Bronsert and Lin's models, showing that the joint model can be an alternate method for analyzing longitudinal data from a normal response, or in general, from an exponential family. Then, in 2017, Alomair adapted Bonsert's and Lin's joint model to be able to incorporate informative time and time-dependent covariates with a longitudinal response.

Purpose of the Study

This study presents an alternative method for parameter estimation, based on Bayesian estimation, with respect to joint models of longitudinal outcomes and informative time. Using a Bayesian approach permits the inclusion of knowledge of the observed data within the analysis through the prior distribution of unknown parameters. In this dissertation, the prior distribution consists of two types based on whether or not applied researchers possess the knowledge concerning the parameters of interest prior to conducting the research (informative and noninformative prior). The product of the prior distribution and the sampling distribution from the data will yield the posterior distribution, which is the distribution of interest. Then, a statistical inference will be made to include the estimation of parameters. In general, estimation of the parameters in the joint model of

longitudinal and informative time is more commonly conducted empirically, regardless of previous information, in contrast to Bayesian methods, which do include prior information. The primary difference between classical statistical theory and the Bayesian approach is that the latter considers the parameters as random variables that are characterized by a prior distribution (Ntzoufras, 2009).

Thus, the primary purpose of this study was to find Bayesian estimates for the unknown parameters in the joint model based on the three kinds of the prior distribution, (noninformative, informative, and semi-informative priors), with the assumptions of a normal distribution for the outcome process and an exponential distribution for informative time.

Definition of Terminology

The terminology used throughout this study is described below.

Longitudinal Data is a set of outcomes or observations measured repeatedly at multiple time points on the same subjects over a given period of time. In general, time points are determined by researchers before outcomes are collected.

Informative Time is the time period between each measurement for each individual. The next measurement is determined by the current outcome. Thus, all subjects may not share the common set of time intervals.

Posterior distribution refers to the conditional probability distribution of the unobserved quantities of ultimate interest, given the observed data (Gelman et al., 2014).

Prior distribution: refers to a probability distribution that treats parameter as a random variable, which may reflect previous information or belief as to what the true value of the parameter may be (Bain & Engelhardt, 1992).

Vague priors: Essentially, these are densities with high spread, such as a normal density with extremely large variance. These give similar prior values over a large range of parameter values.

BUGS: refers to the initials of the phrase “Bayesian inference Using Gibbs Sampling," which is a programming language-based software that is used to generate a random sample from the posterior distribution of the parameters of a Bayesian model. There are two main forms of BUGS, namely WinBUGS and OpenBUGS.

Research Questions

The question, “How will the Bayesian method be designed for estimating the parameters of the proposed joint models?”, was investigated in this dissertation through the following research questions.

Q1 How will the Bayesian method be designed for estimating the unknown parameters on the proposed joint model constructed by Lin (2011) for a longitudinal response variable with a set of informative time?

Q2 How are these Bayesian estimates of the proposed joint model influenced by a few select variations in subject sample size, types of design structures with a different number of observations for each subject, and the various parameter schemes, with three types of prior distribution on the parameters (noninformative, informative, and semi-informative priors)?

Q3 How will the developed R program work closely together with OpenBUGS for fitting Bayesian models? Could that support researchers obtain the Bayesian estimations for the unknown parameters on the proposed joint model?

Finally, the prior distribution in this dissertation will consist of three types based on whether or not applied researchers possess the information concerning the parameters of interest prior to conducting the research.

Limitations

The following limitations need to be taken into consideration by those researchers wishing to take advantage of this study:

1. This study will be limited to outcomes from a normal distribution with a single response variable. Thus, the model should not be applied to studies where outcomes are not normally distributed and/or contain multivariate responses.
2. In this study, time is assumed to be exponentially distributed and to be considered before applying the results to future studies, which may have different time factor distribution assumptions.
3. Moreover, currently, time and covariates will be assumed to be independent of each other.
4. In this dissertation, we will adopt three scenarios for the prior distribution:
   i. The prior distributions of all unknown parameters will be noninformative prior, which will be set to be a vague or flat prior.
   ii. The prior distributions of all unknown parameters will be informative prior, which will be set to be normal for unrestricted parameters, and inverse gamma (IG) priors for positive parameters such as the variance σ^2
   iii. A combination of two above scenarios, so the prior distributions of some unknown parameters are noninformative, and the others are informative.
5. The evaluation of parameter estimates will be limited to some convergence diagnostics (visual and quantitative).
6. Furthermore, prior distributions for all unknown parameters in the joint model will be assumed to be independent of each other.

Therefore, this model should not be applied to any study without considering this limitation.

Summary

The stellar characteristic of longitudinal data analysis is its ability to detect changes over time. Most frequently, the methods of analyzing longitudinal data assume or require that the time periods for data collection are fixed and predetermined by the researchers prior to the initiation of the study. However, there are certain circumstances where the time factor can be informative; i. e., the upcoming observation is determined by the previous outcome of the response variable. In these cases, approaches mentioned before are inappropriate and cannot be used as they will generate biased estimators. In order to combat this bias, Bronsert (2009) developed the joint model, combining the normally distributed longitudinal responses with the exponentially distributed informative time factor. This joint model was further refined by Lin (2011) and extended even more by Seo (2015) to present a joint model as an alternative method to analyze longitudinal data from a normal response or exponential family. The current study applied Bayesian approaches using simulation via Markov chain Monte Carlo (MCMC) methods to estimate the parameters of Lin's extended joint model.

CHAPTER II

LITERATURE REVIEW

A longitudinal study refers to a study where participant outcomes and possibly treatments or exposures are collected at multiple follow-up times. The outcome variables are measured repeatedly on the same cohort of individuals at multiple time-points, which provide the chance to observe individual patterns of change. The purpose of longitudinal research studies is to analyze data on growth, change, and development over time by measuring the change in outcomes at different time points on the same subjects. A variety of books recently have affirmed the nature of longitudinal data (Diggle, Heagerty, Liang, & Zeger, 2002; Fitzmaurice, Laird, & Ware, 2012; Hedeker & Gibbons, 2006; Verbeke & Molenberghs, 2000). In examining the historical background regarding the measurement of data on the same individual over time, termed continuous repeated measures data, the assumptions behind these measurements are that these data are gathered at regularly spaced observations times, and that of multivariate normality (Johnson & Wichern, 2007).

The beginning of the twenty-first century has seen more attention placed on Bayesian methods as helpful techniques in the estimation of many models. Especially, longitudinal models that are complex and which would be very difficult or even impossible to estimate using the currently-available MLE-based software (Seo, 2015). In spite of the power of Bayesian models and the diversity of analytic models, because of the computational complexities and demanding programming involved, these methods are often avoided by researchers. Taken along with the complexities of the models that apply

Bayesian methods, these approaches appear out of reach for many empirical researchers. This has caused in the progress of more practical and straightforward methods that are based on Bayesian analyses, particularly in cases where there are different types of longitudinal data (Azevedo, Fox, & Andrade, 2016; Quintana, Johnson, Waetjen, & Gold, 2016).

Traditional Methods for Longitudinal Data

The simplest method of longitudinal data is the paired *t*-test when there are only two measurement times. The paired *t*-test is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. In a paired sample *t*-test, each subject or entity is measured twice, resulting in pairs of observations. A common application of the paired *t*-test includes case-control studies. However, this method is not appropriate when there is more than two measurement time, which is generally the case in longitudinal studies.

Instead, the univariate (single variable) repeated measures analysis of variance (ANOVA) model can be used with more than two repeated measurements. “The univariate repeated measures ANOVA model provides a natural generalization of Student (1908) paired t-test to handle more than two repeated measurements, in addition to various between-subject factors” (Fitzmaurice, 2008). Traditionally, repeated measures ANOVA has been applied in the analysis of longitudinal data involving more than two-time points, and when covariates are considered (Stevens, 1998; Tabachnick, 2007). Repeated measures ANOVA assesses group differences over time. Group sizes may be different, but subjects must be measured at the same number of time points. Several assumptions need to be confirmed to use any statistical model. In ANOVA, the response or outcome variable has to be normally distributed with the assumptions of independence of observations between

subjects and homogeneity of variance being met. In addition, the important assumption of a repeated-measures ANOVA is sphericity, which refers to the condition where the variances of the differences between all possible pairs of within-subject observations are equal (Hox, 2002). Sphericity can be likened to the homogeneity of variances in a between-subjects ANOVA. Although all these assumptions are restrictive, the univariate repeated measures ANOVA model can be considered a pioneer of the more multilateral regression model for longitudinal data (Fitzmaurice, 2008),

While the assumption of normality of distribution may be overlooked from time to time due to the robustness of ANOVA, the assumption of sphericity in repeated measures ANOVA must be met. Violating this may severely compromise the interpretation of the test results. Mauchly's test can be used to checks this assumption (Davis, 2002). The univariate repeated measures ANOVA model can be written as:

$$Y_{ij} = X'_{ij}\beta + b_i + \varepsilon_{ij}\,, \tag{1}$$

where $i = 1, \ldots, m$ and $j = 1, \ldots, n$ and

Y_{ij} is the outcome of interest,

X'_{ij} is a design vector,

β is a vector of regression parameters,

b_i is the random effect,

ε_{ij} is the measurement error.

An unfortunate consequence of this ANOVA model is that, in spite of yielding an overall P-value, it is not possible to differentiate group means that are equal from those that are not equal (Fitzmaurice, 2008). The usual solution taken to answer this question is the performance of post hoc tests (comparisons made of either all possible combinations of means or of designated means that are of particular interest). However, this testing of

multiple null hypotheses creates an increased risk of a type I error if the significance criterion is not accordingly adjusted (Liu, Cripe, & Kim, 2010; Schober & Vetter, 2018) In addition, repeated measures ANOVA assumes the sphericity, which is the measured outcomes have equal variances, and covariances over time. This may be unrealistic since variances tend to increase with time, and covariances decrease with increasing intervals in time. Another disadvantage of repeated measures ANOVA is that time must be specified as a classification factor, and all-time points need to be fixed across all subjects, which is also considered unrealistic for the longitudinal data. Therefore, the repeated measures ANOVA model is inappropriate for our study on an informative time.

Instead, multivariate repeated measures analysis of variance (MANOVA) could be used with its more flexible variance-covariance assumption. MANOVA does not assume sphericity and/or specific correlation structure, which is considered an alternative to repeated-measures ANOVA (Everitt, 2006; Schober & Vetter, 2018). However, if sphericity holds, ANOVA is more powerful than MANOVA because the sphericity assumption increases degrees of freedom, which increases the power of ANOVA (Hedeker & Gibbons, 2006).

MANOVA still assumes the outcomes of the different subjects are independent, and the outcomes need to be multivariate normality distributed (Davis, 2002; Hedeker & Gibbons, 2006; Hox, 2002). MANOVA models treat the responses as a vector from the *i*th subject at time *j*,

$$\boldsymbol{Y}_i = \left(y_{i1}, y_{i2}, \dots, y_{in_i}\right)'. \tag{2}$$

where $i = 1, \dots, m$ and $j = 1, \dots, n_i$ and $\boldsymbol{Y}_i \sim N_t(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The one sample MANOVA model is given by

$$\boldsymbol{Y}_i = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_i \tag{3}$$

where, $\boldsymbol{\mu}$ is mean vector $(n \times 1)$ for timepoints, and $\boldsymbol{\varepsilon}_i$ is vector $(n \times 1)$ of errors, distributed as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in the population (Hedeker & Gibbons, 2006).

We note that MANOVA requires complete data as does ANOVA because these methods cannot handle unbalanced or incomplete data. The measurements must be available for each subject at each time point. A complete dataset is most probably unrealistic in longitudinal studies. Also, missing observations are quite common in repeated-measures designs (e.g., due to logistic reasons, withdrawal, or loss to follow-up). Therefore, MANOVA is not considered a good choice for longitudinal data (Hedeker & Gibbons, 2006; Ma, Mazumdar, & Memtsoudis, 2012).

In summary, the repeated measures ANOVA assumes that the variance and covariance of the dependent variable across time are equal (i.e., compound symmetry). In contrast, MANOVA for repeated measures only includes subjects with complete data across time, which focus on the estimation of group trends across time with a little help in understanding individuals change across time (Hedeker & Gibbons, 2006). In sum, traditional longitudinal methods have a requirement of no missing data, which is impractical in longitudinal observational studies (Cooley & Lohnes, 1971).

Modern Approaches for Longitudinal Data

The methods of handling longitudinal data commonly used today vary among one another in relation to the flexibility of these methods (Zeger, Liang, & Albert, 1988). Using standard regression methods requires the acceptance of an assumption of independence of all observations, and when this is extended to longitudinal outcomes, it may result in invalid standard errors. In order to make valid inferences about the average response over time, two approaches at present are most commonly used that take advantage of the flexibility of regression models. The first approach is called the Mixed Effects Regression;

the second approach is the Generalized Estimated Equations (GEE). The selection of which of these two approaches to employ is dependent on the desired interpretation of the estimated effects and the purpose of the research (Schober & Vetter, 2018).

Mixed-Effect Regression Models for Longitudinal Data

The mixed-effects model is a model with both fixed effects and random effects, which is a univariate regression analysis on correlated responses (Davis, 2002; Fitzmaurice, 2008; Pinheiro & Bates, 2000). A variety of names have been adopted in describing and developing mixed effect models: random-effects model (Diggle et al., 2002; Fitzmaurice et al., 2012; Laird & Ware, 1982), random regression models, random coefficient model (Leeuw & Kreft, 1986), mixed models (Longford, 1987; Wolfinger, 1993), multilevel model (Nash & Varadhan, 2011), hierarchical model (Lee & Nelder, 1996; Raudenbush & Bryk, 2002), and mixed-effect regression models (Hedeker & Gibbons, 2006).

The mixed-effects model has become popular for longitudinal data analyses for many reasons. One of them is the flexibility to handle unbalanced structures (incomplete data) and the ability to measure subjects at differently spaced time points (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). In addition, the mixed-effects model does not carry the assumption that all subjects are measured for outcomes at the same number of fixed time points. It is appropriate to model in cases where some subject is missing data or have incomplete data. By being able to include these subjects, the statistical power is thus increased (Hedeker & Gibbons, 2006). An added advantage of the mixed effect model is the ability to estimate change for each subject rather than obtaining an average change over time.

By using matrix notations, the mixed-effect model is given by:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{4}$$

where

$\boldsymbol{y}_i \ is \ n_i \times 1$ dependent variable vector for individual *i*,

$\boldsymbol{X}_i \ is \ n_i \times p$ covariate matrix for individual *i*,

$\boldsymbol{\beta} \ is \ p \times 1$ vector of fixed regression parameters,

$\boldsymbol{Z}_i \ is \ n_i \times r$ design matrix for random effect,

$\boldsymbol{\gamma}_i \ is \ r \times 1$ vector of random effects,

$\boldsymbol{\varepsilon}_i \ is \ n_i \times 1$ vector for error and with the assumptions of

$\boldsymbol{\varepsilon}_i \sim N\left(\mathbf{0}, \sigma^2 \boldsymbol{I}_{n_i}\right) \implies E(\boldsymbol{\varepsilon}_i) = 0$ and $Cov(\boldsymbol{\varepsilon}_i) = \sigma^2 \boldsymbol{I}_{n_i} = \boldsymbol{R}_i$,

$\gamma_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i) \implies E(\gamma_i) = 0$ and $cov(\gamma_i) = \boldsymbol{\Sigma}_i = \boldsymbol{G}_i$,

$\boldsymbol{\varepsilon}_i$ and γ_i are independent $\implies Cov(\boldsymbol{\varepsilon}_i, \gamma_i) = 0$.

Thus, the expectation and variance-covariance matrix of the model are given:

$$E(\boldsymbol{y}_i) = \boldsymbol{X}_i'\boldsymbol{\beta} \tag{5}$$

$$Cov(\boldsymbol{y}_i) = Cov(\boldsymbol{Z}_i\gamma_i + \boldsymbol{\varepsilon}_i) = \boldsymbol{Z}_i Cov(\gamma_i)\boldsymbol{Z}_i' + Cov(\boldsymbol{\varepsilon}_i) = \boldsymbol{Z}_i\boldsymbol{G}_i\,\boldsymbol{Z}_i' + \boldsymbol{R}_i. \tag{6}$$

Typically, the goal in mixed model analysis is to test and estimate the parameters in **β**, which require estimates of **G** and **R**. Davis (2002) suggested that the maximum likelihood (ML) estimation can be used to obtain parameter estimation for the random effects and fixed effects by using the numerical solution of a nonlinear optimization procedure. There is many statistical computing software that provides a variety of types of covariance structures for the G matrix, such as compound symmetry, unstructured, first-order autoregressive, or Toeplitz, etc., as an initial value of the iteration. Due to computational difficulties and bias for unbalanced designs in the ML estimation, the restricted maximum

likelihood (REML) approach was introduced as an alternative approach instead of the ML estimation (Davis, 2002; Patterson & Thompson, 1971).

The mixed-effects model for longitudinal data, in spite of its advantages, also has some drawbacks. One drawback is the nonstationary attribute of the covariance matrix structure of the y_i vector. Davis (2002) gave an example of the variance and covariance for collecting data from the same subjects at equally spaced time points are respectively:

$$Var(y_{ij}) = \sigma_\alpha^2 + 2j\sigma_{\alpha\beta} + j^2\sigma_\beta^2 + \sigma^2, \tag{7}$$

$$Cov(y_{ij}, y_{ij'}) = \sigma_\alpha^2 + (j + j')\sigma_{\alpha\beta} + jj'\sigma_\beta^2\,, \qquad j = i, \dots, n, \tag{8}$$

the general trends, thus, are

(1) the $Var(y_{ij})$ increases after time j when $j > -\frac{\sigma_{\alpha\beta}}{\sigma_\beta^2}$,

(2) the $Var(y_{ij})$ decreases up to time j when $j < -\frac{\sigma_{\alpha\beta}}{\sigma_\beta^2}$,

which is not actual longitudinal data. The other drawback, which was demonstrated in a simulation study that showed the quality of the mixed-effects model, is substantially affected by what the variance-covariance matrix structure is chosen (Davis, 2002). In addition, the mixed-effects model treats time as fixed; therefore, this model is inappropriate for our study, including longitudinal outcome and informative time.

Generalized Estimating Equations for Longitudinal Data

The marginal model utilizes an approach termed the generalized estimating equations (GEE) as a way to analyze repeated measurements and extends the generalized linear model to the processes of longitudinal data analyses by accounting for the within-subject correlation among the measurements. The marginal model using the GEE method to analyze repeated measurements was introduced and developed by Liang and Zeger

(1986). Circumvents many of the problems of previous models discussed inasmuch as the mean response is not influenced by previous responses or any random effects, and instead merely depends on the covariate ((Fitzmaurice et al., 2012; Seo, 2015).

The marginal models are a regression model for the response mean employing with a function that links the marginal mean response to the covariates at each event and aims to make inferences about population means (Fitzmaurice, 2008). An advantage of the marginal model is that it is not necessary to hold any assumptions regarding the distribution of outcomes; the model only requires assumptions regarding the mean of responses. Therefore, this GEE model is very useful and used for categorical and count outcomes as well as continuous outcomes (Hedeker & Gibbons, 2006). The marginal models for longitudinal data contain three parts:

1. The expected mean of each response given covariates,

$$E(y_{ij}|X_{ij}) = \mu_{ij}, \tag{9}$$

which can be rewritten with a link function as

$$g(\mu_{ij}) = \boldsymbol{x'_{ij}\beta}, \tag{10}$$

where:

y_{ij} is the response for subject i at time j,

x'_{ij} is $p \times 1$ vector of covariates,

$\boldsymbol{\beta}$ is $p \times 1$ vector of unknown parameters,

$g(.)$ is the link function,

μ_{ij} is the mean response.

2. The variance of the response given the covariates,

$$var(y_{ij}) = \phi v(\mu_{ij}), \tag{11}$$

where:

$v(.)$ is a known variance function, which is the relationship between the mean and the variance, expressing the variance as a function of the mean,

ϕ is a scale parameter that could be known or need to be estimated. The link and variance functions for normal outcomes are shown below:

$$g(\mu_{ij}) = \mu_{ij} = x'_{ij}\beta, \qquad var(\mu_{ij}) = 1, \qquad and \quad var(y_{ij}) = \phi v(\mu_{ij}) = \phi.$$

3. The 'within-subject association' among the responses is a function of the means and of additional parameters, say $\boldsymbol{\alpha}$, that may also need to be estimated.

Several advantages make the GEE method useful in the analysis of longitudinal data. The GEE method is flexible enough to allow the estimation of the mean and pairwise correlations among repeated measures. Also, it is able to accommodate situations where there is missing data or unbalanced data (Fitzmaurice, 2008). Additionally, subjects don't need to have the same number of outcomes that are gathered at the same time (Fitzmaurice, 2008).

There are, however, some disadvantages to the GEE method. One of which is that the GEE model is not an appropriate model when research questions to ask about estimating the variance-covariance structure. Another disadvantage lies in the assumption of this model that, although complete data across time for subjects is not necessary, there is an assumption that all the time points are fixed and that any missing responses must be missing completely at random (MCAR) (Hedeker & Gibbons, 2006). A third disadvantage is noted by Fitzmaurice (2008), who pointed out that the estimation of $\boldsymbol{\beta}$ that used in this model is not as efficient when compared to the maximum likelihood-based estimation because of the lack of assumption on the distribution. Finally, as discussed by Pepe and

Anderson (1994), because parameter estimates of $\widehat{\boldsymbol{\beta}}$ inconsistently estimate $\boldsymbol{\beta}$ in cases where time-varying covariances are present in the regression model, this GEE model is not considered as a right choice for this study including longitudinal and informative time.

Generalized Linear Model and Exponential Family

The class of generalized linear model (GLM) was introduced by Nelder and Wedderburn in 1972. GLM proved to be popular with both practitioners and statistical researchers because of its ability to deliver a unified class of models of regression analysis, regardless of discrete or continuous outcomes (Dobson, 2001; Fitzmaurice et al., 2012; Das & Dey, 2006, 2007). Three components have been delineated as forming GLM (Fitzmaurice et al., 2012):

(1) A random component (An observation model), which identifies the distribution of the outcome variable. GLM assumes that the outcome variable has a probability distribution from the exponential family of distributions and may include additional parameters γ.

$$p(y_i|\theta_i, \gamma), \tag{12}$$

for data y_i conditional on the expected response,

$$E[y_i] = \theta_i, \tag{13}$$

which is called the canonical parameter, at each design point, $i = 1, 2, \cdots, n$.
The variance of the outcome can be written as a product of a single scale or dispersion parameter, $\emptyset$, and it is called the variance function:

$$Var(\boldsymbol{Y}) = \emptyset\, v(\mu). \tag{14}$$

(2) a systematic component (The linear predictor), which identifies explanatory variables. These explanatory variables are combined into a linear format, and it is called the linear predictor:

$$\eta_i = \boldsymbol{X}_i'\boldsymbol{\beta} = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}, \quad (15)$$

where,

$\boldsymbol{X}_i'$ is $p \times 1$ vector of continuous or categorical covariates at the i^{th} design point,

$\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown parameters.

(3) A link function is a function that connects the linear predictor with the mean of the probability distribution. So, the function $g(.)$ must be inversely differentiable, and connects a random component to a systematic component, which can be written as

$$g(\eta_i) = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}\,, \quad (16)$$

the model is the relationship between the expected response and the linear predictor.

GLM has been used for modeling several types of data involving exponential family response with covariates. Typical examples include those for Poisson and binomial response data. A regression model determines the structure of the covariate information, where a link function specifies the relationship between the regression model and the expected values of the observation. However, a simple linear model with a normal error structure is a special case of GLM. All the distributions from the exponential family can be expressed as

$$f(y; \mu, \emptyset) = \exp\left(\frac{y\theta - b(\theta)}{\alpha(\emptyset)} + c(y\,, \emptyset)\right), \quad (17)$$

where,

θ is a canonical (natural) parameter,

ϕ is a scale dispersion parameter,

The commonly used distributions of the outcomes are the Gaussian for normally distributed outcomes, which can be rewritten in an exponential family form of

$$f(y;\mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

$$= \exp\left(-\tfrac{1}{2}\log(2\pi\sigma^2)\right)\exp\left(-\tfrac{(y-\mu)^2}{2\sigma^2}\right)$$

$$= \exp\left(-\frac{(y^2-2y\mu+\mu^2)}{2\sigma^2}-\frac{1}{2}\log(2\pi\sigma^2)\right)$$

$$= \exp\left(-\tfrac{(y\mu+\mu^2/2)}{\sigma^2}-\tfrac{1}{2}\left(\tfrac{y^2}{\sigma^2}+\log(2\pi\sigma^2)\right)\right), \quad (18)$$

with a canonical parameter $\theta = \mu$; and a dispersion parameter $\alpha(\emptyset) = \sigma^2$.

Joint Model for Longitudinal Data

Fundamentally, the joint model is based on the joint distribution of outcomes and the time-related factor with maximum likelihood estimation (combining longitudinal data and the time-related factor). More recently, a model referred to as the joint model has been gaining popularity as a way to deal with irregular occasions in the analysis process. The joint model is combining longitudinal data and time or other factors that the researcher is interested with fixed or random time (Diggle et al., 2002; Fitzmaurice, 2008; Henderson et al., 2000; Kim et al., 2012; Lin et al., 2004; Qiu et al., 2013; Wu, Liu, Yi, & Huang, 2012;). Recently, due to the importance of the joint model in studying and analyzing data for cases where the informative time and longitudinal outcomes were related, it has been used and developed among many current studies and research (Alomair, 2017; Bronsert, 2009; Huang, Wang, & Zhang, 2006; Liang et al., 2009; Lin, 2011; Lipsitz et al., 2002; Ryu, Sinha, Mallick, Lipsitz, & Lipshultz, 2007; Seo, 2015; Sun, Park, Sun, & Zhao, 2005).

The joint model for longitudinal data was developed by Lipsitz et al. (2002). Assumptions of the joint model include time points not being fixed or dependent on previous outcomes, and that repeated measurements conform to a multivariate normal

distribution. For example, in a longitudinal study, patients may be followed over long periods with differing lengths of time. Through the follow-up, many patients drop out for various reasons. Patients may leave a cohort and then return or be lost to follow-up, resulting in a different number of visits for each patient and different visit schedules among patients (Fitzmaurice, 2008).

Another example is when patients in the late stage of disease see doctors more times than those in early stages; ignoring the time informative leads to biased results (Song, Mu, & Sun, 2012). Previous examples explained cases for informative time with longitudinal outcomes where it is important to model both of them together in order to make a valid inference based on modeling the time distribution conditionally on the outcome measurements. The joint model grew out of the concept where time points depend only on previously observed data rather than the time points where the outcome measured.

In order to handle informative time and normally distributed longitudinal outcomes, Bronsert (2009) proposed and presented his variation of the joint model to handle informative time component and normally distributed longitudinal outcomes. Bronsert's Gaussian-Exponential model comes with assumptions of a normal distribution for the outcome process and an exponential distribution for informative time, whereas the repeated outcomes depend on the most recent outcome and current time point, which is given by:

$$f_{\theta}(\boldsymbol{y}_i, \boldsymbol{t}_i) = \frac{1}{\sqrt{2\pi(\sigma^2)}} exp\left(-\frac{1}{2}\frac{(y_{i1} - \boldsymbol{X}'_{i1}\boldsymbol{\beta})^2}{\sigma^2}\right) \times f(t_{i1})$$

$$\times \prod_{j=2}^{n_i} \left\{ \frac{1}{\sqrt{2\pi(\sigma^2)}\sqrt{1-\rho_i^2}} exp\left(-\frac{1}{2}\frac{\left(y_{i1} - \gamma t_{ij} - \phi y_{i(j-1)} - \boldsymbol{X}'_{ij}\boldsymbol{\beta}\right)^2}{\sigma^2(1-\rho_i^2)}\right) \times \right.$$

$$\left. exp\left(\alpha + \delta_i y_{i(j-1)}\right) \times exp\left(-e^{\alpha + \delta_i y_{i(j-1)}} t_{ij}\right) \right\}, \qquad (19)$$

where:

$\boldsymbol{y}_i = \left(y_{i1}, \cdots, y_{in_i}\right)$ is $n_i \times 1$ a vector that includes i^{th} subject measurements at times $\boldsymbol{t}_i = \left(t_{i1}, \cdots, t_{in_i}\right)$,

y_{ij} is the outcome for the i^{th} subject measured at the j^{th} time point.

$\boldsymbol{\beta}$ is the effect of the independent variables on outcomes,

$f(t_{i1})$ is the initial time point for the ith subject,

ϕ is the effect of the previous outcome on the mean response of the current outcomes,

γ is the effect of current time on the mean response,

α is the constant parameter for the time process,

δ is the effect of the previous outcome on the mean time,

$\boldsymbol{X}'_{i1}$ is the initial observations of k independent variables,

$\boldsymbol{X}'_{ij}$ is $n \times (k+1)$ design matrix contains the observations of k independent variables, where n is the number of subjects,

σ^2 is the variance of the distribution, and

ρ_i^2 is the relationship between two outcome variables at two-time points.

This Bronsert's Gaussian-Exponential model was further modified and adjusted by Lin (2011), who eliminated the term ρ_i^2 in his model. Lin (2011) pointed out that there was already a term in the model, ϕ, that took care of relationships between two responses and that therefore there was no need to include ρ_i^2. Lin's modified model was:

$$f_{\Theta}(\boldsymbol{y}_i, \boldsymbol{t}_i) = \frac{1}{\sqrt{2\pi(\sigma^2)}} exp\left(-\frac{1}{2}\frac{(y_{i1} - \boldsymbol{X}'_{i1}\boldsymbol{\beta})^2}{\sigma^2}\right) \times f(t_{i1})$$

$$\times \prod_{j=2}^{n_i} \left\{ \frac{1}{\sqrt{2\pi(\sigma^2)}} exp\left(-\frac{1}{2}\frac{\left(y_{i1} - \gamma t_{ij} - \phi y_{i(j-1)} - \boldsymbol{X}'_{ij}\boldsymbol{\beta}\right)^2}{\sigma^2}\right) \times \right.$$

$$\left. exp\big(\alpha + \delta_i y_{i(j-1)}\big) \times exp\big(-e^{\alpha + \delta_i y_{i(j-1)}} t_{ij}\big) \right\}. \quad (20)$$

Later, this modified model was then employed by Seo (2015). In his study showing that the joint models (Gaussian exponential model) of Bronsert's (2009) and Lin's (2011) could be extended to handle longitudinal outcomes distribution from a member of the exponential family of distributions with the informative time that followed an exponential distribution.

However, our studies will be adopted from the joint model of Lin (2011) with informative time with longitudinal responses. Lin (2011) and Seo (2015) used the classical method (maximum likelihood estimation) to compute the parameter estimates, while I will use the Bayesian Approach to estimate the model's parameters.

Bayesian Approach

In general, there are two common approaches in statistical inference which are used to estimate a parameter, address hypothesis testing, and predict a new observation (Bolstad & Curran, 2016). These two approaches are referred to as:

1. Classical approach, which is based on all information from the random sample to model the likelihood of the observed data and make inference on the parameters of interests. This approach associates probability with long-run frequency (Fisher, 1925; Neyman & Pearson, 1933).
2. Bayesian approach, which is a combination of the information from the random sample (likelihood function) and the information of previous research (prior distribution), to produce the posterior distribution that is used to make all inferences about the parameters of interests. This approach interprets probability as the subjective experience of uncertainty (De Finetti, 2017).

The difference between Bayesian statistical and classical (frequentist) statistical methods is how we should deal with the nature of the unknown parameters. In the classical approach, all parameters of interest are assumed to be unknown but fixed. The Bayesian inference assumes that the data are fixed, and all unknown parameters are considered as random variables and treated as uncertain, and therefore should be described by a probability distribution. Bayesian inference has been referred to as the process to fit a probability model to a set of data and summarize the result by a probability distribution on the parameters of the model (Gelman et al., 2014).

In the 21st century, the Bayesian approach has played a major role in statistical analysis and become more popular and widely used in many applied and fundamental research such as educational, economic, and medical research (Poirier, 2006). In addition, Bayesian applications have been increasing and taking place in developmental researches because the background knowledge is incorporated into the analyses and integrated into the statistical model. Also, due to the availability of Bayesian computational methods in popular software packages such as WinBUGS and OpenBUGS (Lunn, Thomas, Best, &

Spiegelhalter, 2000), and a large number of packages within the R statistical computing environment (Albert, 2009).

Why Would Researcher Use Bayesian Statistics?

There are many advantages for using Bayesian methods in comparisons to frequentist statistics, and I will discuss some of them (Kruschke, 2011; Kaplan & Depaoli, 2012, 2013):

Bayesian approach does not depend on large samples. In the classical methods, when the sample size is small, and parameters are not normally distributed, it is often hard to formulate reasonable parameter estimates and attain statistical significant or meaningful results (Button et al., 2013; Scheines, Hoijtink, & Boomsma, 1999). However, the Bayesian method does not assume or require normal distributions underlying the parameters of a model, and the large sample size to make the statistical inference (Gelman et al., 2014). In addition, Bayesian methods may provide more accurate results as they can deal with small data set and asymmetric distributions and/or the whole distribution (Gill, 2008; Muthén & Asparouhov, 2012; Scheines et al., 1999). However, if the sample size is large enough, the maximum likelihood estimation provides all information about parameter estimates from the data. This operation performs the same statistical inferences, in which both methods produce very similar results, as using asymptotic estimation theory (normality assumptions are satisfied) (Gelman et al., 2014).

Common-sense interpretation of statistical conclusions. In both approaches, the interpretation of the results is very different. For any given sample, we can obtain the sample mean and compute the confidence interval. The right frequentist interpretation is that (1-α) % of these confidence intervals capture the true parameter under the null

hypothesis. Unfortunately, the results of the frequentist paradigm are often misunderstood (Gigerenzer, 2004). The Bayesian peer of the frequentist confidence interval is the credibility interval, which is a 95% probability that the population parameter lies between the two values. However, the credibility interval and the confidence interval may numerically be similar and might serve related inferential goals, but they are not mathematically equivalent and conceptually quite different. (Gill, 2008; Muthén & Asparouhov, 2012).

Background knowledge can be incorporated into the analyses. Updating knowledge is another important argument for using the Bayesian approach, which allows researchers to incorporate background knowledge into their analyses instead of testing the same null hypothesis repeatedly. The Bayesian approach could incorporate uncertainty about unknown parameters, which are treated as random variables rather than fixed and assumed to have their own distributions. The distribution is referred to as a prior distribution. Bayesian statistical models incorporate both the sample information and prior information on the parameter estimates (Jiang & Mahadevan, 2009; Lynch, 2007). In conclusion, the higher the precision, the more influence the prior specification has on the posterior results.

Allow coping with complex models. A complex model is a model with large numbers of parameters or multiple dimensions, which sometimes cannot be estimated using conventional methods. It cannot use a natural approach such as numerical integration, which is often required to estimates parameters based on maximum likelihood estimation, and this method is intractable due to the high dimensional integration needed to estimate the maximum likelihood (Dunson, 2000; Gill, 2008; Lynch, 2007; Muthén & Asparouhov, 2012; Scheines et al., 1999). Therefore, Bayesian analysis is considered to be a

computational tool for the study of complicated statistical models with complex data structures.

The Components of Bayesian Statistics

Prior Distribution

The most important aspect of the Bayesian approach is to set up a proper prior to including in the model. Prior Distribution is the background knowledge on the parameters of the model being tested before seeing the data, such as previous studies on similar data (O'Hagan, 2006). Perhaps we can include even more information in our prior distribution to increase precision and therefore contribute to more accurate estimates. The prior reflects our knowledge about the parameters before observing the current data. Usually, knowledge of the prior will be available and included in the analysis as informative prior. Knowledge of the prior may be obtained from opinions of experts or because it reuses the posterior distribution from previous research, but if the knowledge of the prior is uncertain, lost, or ignored, non-informative prior is used. These different priors are called:

Noninformative priors. From a Bayesian point of view, this lack of information is still important to consider and incorporate into our statistical specifications (Ibrahim & Laud, 1991). Using the noninformative (objective) prior can still benefit from using Bayesian statistics, as will be explained throughout the study.

Jeffreys' prior is kind of non-informative prior. Jeffreys' prior is the square root of Fisher information written by $\pi(\theta) = |J(\theta)|^{\frac{1}{2}}$ with $J(\theta)$ is Fisher's information, which is written as

$$J(\theta) = -E_{\theta}\left(\frac{\partial^2 \, logf(Y|\theta)}{\partial \, \theta^2}\right). \quad (21)$$

If $\theta = \left(\theta_1, \theta_2, \ldots, \theta_p\right)^T$ is a vector, it is used $\pi(\theta) = [\det \boldsymbol{J}(\boldsymbol{\theta})]^{\frac{1}{2}}$ with $\boldsymbol{J}(\boldsymbol{\theta})$ is Fisher information matrix. The Fisher information is

$$J_{ij}(\theta) = -E_\theta \left(\frac{\partial^2 \, logf(Y|\theta)}{\partial\theta_i \, \partial\theta_j}\right), \tag{22}$$

with $i = 1, 2, \cdots, p$ and $j = 1, 2, \cdots, p$.

Informative priors. For large samples, if a low-informative prior is stated, the results are hardly influenced by the type of the prior. The more prior information is added, the more subjective it becomes. Informative prior is beneficial because:

(a) findings from previous research can be incorporated into the analyses.

(b) Bayesian credible intervals will be smaller.

Note that a prior distribution needs to be specified for each and every parameter in the model. It is assumed that a distribution for each and every parameter, including covariances for using Bayesian statistics.

The conjugate prior is informative prior, which is an initial probability assumption expressed in the same distribution type (parameterization) as the posterior probability or likelihood function; for example, the beta distribution is a conjugate prior for the binomial distribution. This means the posterior is also beta. In addition, if the likelihood function is normal with known variance, then a normal prior gives a normal posterior. This means that the normal distribution is its own conjugate prior.

Generally, selecting prior is based on the parameter information available or not. If the parameter information is available, we use informative prior. Informative prior has a significant effect on the posterior distribution and more subjective (Gelman et al., 2014). If the parameter information is not available, we use non-informative prior, which is more objective than most classical analyzes.

Likelihood Function

The second element of Bayesian analysis is the information in the data themselves. It is the observed evidence for our parameters in the data. This information is obtained by the likelihood function containing the information about the parameters given the data set. The likelihood function is defined as a function of the parameter θ equal to (or sometimes proportional to) the density of the observed data with respect to a reference measure. The likelihood is a tool for summarizing the data's evidence about unknown parameters. The likelihood function is one of the most fundamental concepts of modern statistics. It is an important component of both frequentist and Bayesian analyses, where it is also obtained when non-Bayesian studies are conducted using ML estimation. Note that the likelihood function is not a probability density function.

In defining likelihood functions in terms of probability density functions, we can suppose the joint probability density function of your sample, $Y = (y_1, y_2, \cdots, y_n)$, is $f(Y|\theta)$, where y_i is independent, and θ is a parameter. $Y = y$ is an observed sample point, then the likelihood function defined as a function of θ:

$$L(\theta|Y) = \prod_{i=1}^{n} f(y_i|\theta). \quad (23)$$

Posterior Distribution

The third component is the posterior distribution, which is based on combining the first two components (the prior knowledge and the observed evidence) via Bayes' theorem. The posterior distribution reflects updating knowledge, balancing background knowledge (the prior) with observed data (the likelihood). The posterior distribution represents knowledge after taking the data into account (Bernardo & Smith, 2009). The posterior results are affected by the prior specification. The higher the prior precession, the smaller the posterior variance.

With a non-informative prior, the posterior estimate may not be influenced by choice of the prior much at all. In most cases, obtaining the posterior distribution is done by simulation, using the Markov Chain Monte Carlo (MCMC) methods (Gelman et al., 2014).

Bayes' Rule for Posterior:

The Bayesian approach is of interest in computing the posterior distribution of the unknown parameter θ given the observed data Y, assuming the data are fixed and all unknown parameters as random variables. This is obtained by:

$$
\begin{aligned}
P(\theta|y) &= \frac{P(y,\theta)}{P(y)} && [Definition\ of\ conditional] \\
&= \frac{P(y|\theta)P(\theta)}{P(y)} && [Bays\ rule] \\
&= \frac{P(y|\theta)P(\theta)}{\int_{\Theta} P(y,\theta)\, d\theta} && [Low\ of\ total\ probability] \\
&= \frac{P(y|\theta)P(\theta)}{\int_{\Theta} P(y|\theta)P(\theta)\, d\theta}. && [Bays\ rule] \qquad (24)
\end{aligned}
$$

The final result depends only on sampling distribution (Likelihood) $P(y|\theta)$, and the prior distribution $P(\theta)$. Because the denominator $P(y) = \int_{\Theta} P(y|\theta)P(\theta)\, d\theta$ does not depend on θ, then we can consider it as a constant:

$$P(\theta|y) \propto P(y|\theta)P(\theta), \qquad (25)$$

where.

$P(\theta)$ Prior distribution function

$P(Y|\theta)$ Likelihood function, which is the joint probability function from Y as the random sample, if θ is known.

$P(\theta|Y)$ Posterior distribution function, which is essentially a combination of the evidence provided by the observed data and prior relevant data from past research evidence.

The proportion of the product of the prior distribution of the parameters and the likelihood function of the sample data provides the posterior distribution, which could be used to obtain parameter estimates through Bayesian inference (Bernardo & Smith, 2009).

Markov Chain Monte Carlo

For estimating posterior distributions in Bayesian inference, Markov Chain Monte–Carlo (MCMC) is a popular method for obtaining information about posterior distributions (Gilks, Richardson, & Spiegelhaler, 1996; Kruschke, 2014; Lee, 2012). When focusing upon posterior distributions using analytic examination, which are often difficult to work with, MCMC is the practical method which has led to an explosion of computational algorithms and created a real revolution in the implementation of Bayesian methods (Ravenzwaaij, Cassey, & Brown, 2018). If posterior distributions are characterized by closed forms (normal, gamma, beta, Poisson, etc.), it is possible to conduct simulations directly by using computer programming routines. In cases where the posterior distributions have complicated or unusual or high dimensional models, the achievement of an approximation of the posterior distribution can be obtained by aligning differing algorithms used to construct and sample arbitrary posterior distributions. In this case, the complex nature of posterior, using MCMC permits the approximation of posterior distributions that cannot be directly calculated (Gamerman & Lopes, 2006; Gilks et al., 1996).

Historically, the algorithms of the MCMC method have been used for close to 60 years. They have served as a reference method to analyze Bayesian complex models,

particularly in the early 1990s. Gelfand and Smith (1990) and Gelman et al. (2014) noted that a particularly strong attribute of MCMC is its use in drawing samples from distributions even where that is known about the distribution is a method for calculating the density for different samples.

Inspecting the name MCMC reveals the combination of two properties: Monte Carlo and Markov Chain. Monte Carlo is the practice of estimating the properties of distribution by examining random samples from the distribution. A Monte Carlo approach would draw a large number of random samples from the distribution, and calculate the properties of distribution (mean, variance, etc.) It can be much easier than calculating the properties of distribution directly from the distribution's equations. This advantage is apparent in cases where it is easy to draw random samples, and the equations of the distribution are challenging to work with (Gilks et al., 1996).

Markov Chain's aspect of MCMC rests upon the construct that the random samples are obtained through a particular sequential process, where each random sample becomes a steppingstone in order to get the next random sample, hence the inclusion of the 'chain.' It is noteworthy to mention that in this process, each new sample is dependent upon the preceding sample, but any new sample is not dependent on any sample before the preceding one, which is called the Markov property (Ravenzwaaij et al., 2018).

As Lee (2012) stated, MCMC methods provide an excellent approach for parameter estimation in a Bayesian framework. In addition, he pointed out that a key feature of MCMC approaches regards inferences about an analytically intractable posterior, often in high dimensions, and how they are conducted by generating a Markov chain converging to a chain of drawing taken from the posterior distribution. In summary, MCMC methods are used to draw samples from some target densities, which are mostly non-standard or

complex forms of distributions, where the target density in Bayesian applications is the joint posterior, or the posterior density of the model parameters (Ravenzwaaij et al., 2018).

By using MCMC, it is possible to analyze all of the parameters or functions of parameters through a posterior distribution via Bayesian applications (O'Neill, 2002). Mathematically, the posterior summaries for individual parameters or joint distributions of parameters such as means, medians, variances. The Markov Chain Monte Carlo generates a sequence of $\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(t)}$ random variables of some set *T*, $(t \geq 0)$, the next state $\theta^{(t+1)}$ is sampled from the conditional distribution of $\theta^{(t+1)}$ given $\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(t)}$ depending only on the current state of the chain, $\theta^{(t)}$ (Gelman et al., 2014; Gilks et al., 1996). Therefore, the unknown parameter at time *t*+1, $\theta^{(t+1)}$, is conditionally independent of the previous values:

$$P\left(\theta^{(t+1)} | \theta^{(0)}, \theta^{(1)}, \cdots, \theta^{(t)}\right) = P\left(\theta^{(t+1)} | \theta^{(t)}\right), \tag{26}$$

where the random variable at time t+1, $\theta^{(t+1)}$, does not depend further on $\theta^{(0)}, \theta^{(1)}, \cdots,$ $\theta^{(t-1)}$. (Gilks et al., 1996; Sitthisan, 2016).

There are two primary sampling methods of constructing the chains within MCMC, including Metropolis-Hastings Algorithms, which was studied by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) and Hastings (1970) and Gibbs sampling, which was initially introduced by Geman and Geman (1984).

The Metropolis-Hastings Algorithms

Another sampling method that also functions as a basis for all other sampling methods was proposed by Metropolis et al. (1953), which used a modified Monte Carlo scenario. Later, the Metropolis et al. (1953) algorithm was adapted by Hastings (1970) in order to relax the assumption of asymmetric proposal distribution, and this adaptation has become known as the Metropolis-Hastings algorithm. This Metropolis-Hastings algorithm

is a form of the MCMC method often utilized in sampling from a specified target distribution that is itself a posterior distribution within Bayesian analysis (Gelman et al., 2014). It is often employed in solving numerical problems related to statistical analysis as it is both simple but also powerful for a variety of issues (O'Neill, 2002).

The Metropolis-Hasting algorithm is an MCMC method that can be used for sampling from the specified target distribution, which is a posterior distribution in Bayesian analysis (Gelman et al., 2014). Each iteration of the Metropolis-Hastings algorithm is divided into three steps:

(a) Generate a line, which is starting point $\theta^{(0)}$ from a starting distribution $\pi_0(\theta)$, by sampling from a candidate, proposal, or a jumping distribution $q\left(\theta^*|\theta^{(t-1)}\right)$.

(b) Propose a new state through the line, and evaluate

$$\alpha(\phi|\theta) = \min\left[\frac{\pi(\phi)q(\theta|\phi)}{\pi(\theta)q(\phi|\theta)}, 1\right], \tag{27}$$

where,

$\pi(\theta)$ is a distribution π with respect to a sequence of random variables $\theta = \theta^{(1)}, \theta^{(2)}, \cdots$, drawn via Markov chain,

$q(\phi|\theta)$ is a transition kernel or a transition probability, which is constructed from the current state $\theta^{(t-1)} = \theta$ to the next realized state $\theta^* = \phi$,

$\alpha(\phi|\theta)$ is the probability of moving, which is introduced to reduce the number of moves from θ to ϕ,

(Chib & Greenberg, 1995; Gamerman & Lopes, 2006; Mengersen & Tweedie, 1996; Roberts & Smith, 1994; Sitthisan, 2016; Tierney, 1994).

(c) Accept or reject the proposed state according to the Metropolis-Hastings probability; or, keep the current state.

The decision to move the state can be made referring to the probability of the move, $\alpha(\phi|\theta)$. If the chain is at a point θ, then it generates a candidate value ϕ for the next step. If the candidate point is accepted, the next state becomes ϕ, so the probability of going from state θ (i. e, θ^{t-1}) to state $\phi(\theta^*)$ is shown as:

$$p^*(\phi|\theta) = \begin{cases} q(\phi|\theta)\alpha(\phi|\theta) & \text{if } \theta \neq \phi \\ 0 & \text{if } \theta = \phi \end{cases}, \tag{28}$$

which is also defined as the off-diagonal density of a Metropolis kernel (Lee, 2012; Tierney, 1994). If the candidate point is rejected, the chain remains in the present state θ. The probability when the algorithm remains at θ is set as

$$r(\theta) = 1 - \sum_{\phi} q(\phi|\theta)\alpha(\phi|\theta). \tag{29}$$

The simulation of a draw from a target (posterior) distribution can be summarized as:

1. Draw a starting point $\theta^{(0)}$ from a starting distribution $p_0(\theta)$.
2. (a) Sample a proposal or a candidate point θ^* from a proposal distribution $q(\theta^*|\theta^{t-1})$ at time *t*.

   (b) Calculate the ratio $\alpha(\theta^*|\theta^{t-1}) = \frac{\pi(\theta^*|y)/q(\theta^*|\theta^{t-1})}{\pi(\theta^{t-1}|y)/q(\theta^{t-1}|\theta^*)}$.
3. Generate U from an independent Uniform distribution on (0, 1).
4. Compare U with $\alpha(\theta^*|\theta^{t-1})$,

   if U $\leq \alpha(\theta^*|\theta^{t-1})$ the move is accepted and define $\theta^t = \theta^*$,

   if U $> \alpha(\theta^*|\theta^{t-1})$ the move is accepted and define $\theta^t = \theta^{t-1}$,
5. Change the time *t* to *t*+1 and return to step 2 to get the sequence of random variable $\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(t)}$,

(Chib & Greenberg, 1995; Gamerman & Lopes, 2006; Lee, 2012; Sitthisan, 2016).

Gibbs Sampling

In Geman and Geman's study (1984), the focus was upon image-processing models (Casella & George, 1992). This original work yielded a widely accepted principle theoretical contribution for investigating the Markov random field in the sampling and computation of the mode of the posterior distribution. The applications of the Gibbs sampling have been used in sampling complicated models (those with various unknown parameters or high dimensional integration) by Smith and Robert (1993), Zhang, Hamagami, Wang, Nesselroade, and Grimm (2007), and Lu, Zhang, and Lubke (2011). At this point, the Gibbs sampling has become the default algorithm in most software, where it uses an iterative process when all the parameters of the model (e.g., means, variances, regression parameters, etc.) are estimated repeatedly. These repeated estimations are able to be summarized by creating plot diagrams of the results from each iteration. Then, this distribution can be used in computing means or confidence intervals, allowing for multiple chains to be specified and sampling from a greater range of locations that are within the posterior distribution. Theoretically, the results of sampling multiple chains will, after many iterations, converge to reflect the same marginal distribution of the model parameters (Casella & George, 1992; Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2019).

An integral characteristic of the Gibbs sampling technique is reflected by the drawing of samples from the full conditional distributions (Smith & Robert, 1993), which are distributions of the parameter of focus that is predicated on the known information available from all the other parameters. Gamerman and Lopes (2006) rephrased this key feature by stating that the Gibbs sampling approach relies upon the full conditional distributions. Let $\pi(\boldsymbol{\theta})$ be the density function of interest with q unknown parameters ($\boldsymbol{\theta} = \theta_1, \theta_2, \cdots, \theta_q$). Each component θ_i can be considered as a scalar, a vector, or a matrix

(Gamerman & Lopes, 2006). Let $\pi(\boldsymbol{\theta}) = \pi\left(\theta_1, \theta_2, \cdots, \theta_q\right)$ denote the joint density function, then $\pi_i(\boldsymbol{\theta_i}) = \pi\left(\theta_i|\theta_1, \theta_2, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_q\right) = \pi(\theta_i|\theta_{-i}),\ i = 1, 2, \cdots, q$ denote the full conditional densities for each of the components θ_i, given all the components of θ, except for θ_i at the current values. Gibbs sampling provides an alternative scheme to draw samples directly from a known marginal distribution when the full conditional densities are known. This technique samples one parameter at a time. For each iteration of the Gibbs sampler, the value of each component cycles through the subvectors of θ. At iteration *t*, each subset θ_i^t is sampled individually from the conditional distribution given all the other components of θ, $\pi(\theta_i^t|\theta_{-i}^{t-1})$. The Gibbs sampling algorithm is defined by the following iterations:

1. Choose a starting values of the chain, $j = 0$; $\boldsymbol{\theta}^{(0)} = \theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_q^{(0)}$;
2. At time t, starting at $j = 1$, obtain the single cycle by drawing a new values $\boldsymbol{\theta}^{(t)} = \theta_1^{(t)}, \theta_2^{(t)}, \cdots, \theta_q^{(t)}$ from successive random drawings from the full conditional distributions $\pi(\theta_i^t|\theta_{-i}^{t-1}, x), i = 1, \cdots, k$ as follows:

   sample $\theta_1^t \sim \vartheta_1|\theta_2^{t-1}, \theta_3^{t-1}, \cdots, \theta_q^{t-1}$

   sample $\theta_2^t \sim \vartheta_2|\theta_1^{t-1}, \theta_3^{t-1}, \cdots, \theta_q^{t-1}$

   sample $\theta_3^t \sim \vartheta_3|\theta_1^{t-1}, \theta_2^{t-1}, \theta_4^{t-1}, \cdots, \theta_q^{t-1}$

   ⋮

   sample $\theta_{q-1}^t \sim \vartheta_{q-1}|\theta_1^{t-1}, \theta_2^{t-1}, \theta_3^{t-1}, \cdots, \theta_{q-2}^{t-1}, \theta_q^{t-1}$

   sample $\theta_q^t \sim \vartheta_q|\theta_1^{t-1}, \theta_2^{t-1}, \theta_3^{t-1}, \cdots, \theta_{q-1}^{t-1}$
3. Increment *t* and repeat until the chain convergence criterion is satisfied (noting that convergence is to stationarity rather than to a point, as it would be for iteratively calculated randomization-based estimators).

When convergence is reached, this means $\theta^{(t)}$ is sampled from π (Gamerman & Lopes, 2006; Sitthisan, 2016).

Theoretically, both The Gibbs sampling method and the Metropolis-Hastings algorithm are simple and straightforward, utilized in simulations of a posterior distribution on spaces of fixed dimension (Richardson & Green, 1997). In addition, both of them are designed to ensure the final convergence to the stationary distribution. When the process stops, a Monte Carlo standard error indicates how close the last values are likely to be to the actual ML estimates (Agresti, 2015).

Convergence of the Markov Chain Monte Carlo Algorithms

Markov Chain Monte Carlo (MCMC) algorithms have been used frequently as a way to fit complicated statistical models when it is challenging to apply traditional estimation techniques. The concept of an MCMC algorithm is to develop a process that has a stationary distribution that matches a posterior distribution of interest. One problem in using an MCMC algorithm is in the determination of the convergence of the algorithm. Convergence technically happens when the Markov chain generated converges in distribution to be a posterior distribution of interest. The convergence is in distribution (not to point), and the generated values will vary even after convergence. As a way to calculate generally some form of statistical analysis to assess convergence of the MCMC algorithms, various convergence diagnostics have been suggested.

It is important to raise a question that is related to applying the convergence diagnostics to a practical problem: “How many parameters to monitor?” Gelman and Rubin (1992) suggest monitoring the convergence of all the model’s parameters. Monitoring convergence only for the parameters of interest, particularly when the problem contains

high dimensional parameters, may lead to a mistake of diagnosing convergence too early (Carlin & Louis, 1996).

With respect to another question that is frequently asked, “How many chains to employ,” there is a variety of expert opinions. Geyer (1992) recommends using one very long chain as he claims that will have the best chance of exploring the whole parameter space, particularly for a slowly moving chain. In contrast, Gelman and Rubin (1992) recommend running several long chains. It is important to note that some of the popular MCMC convergence diagnostics work only for multiple chains. Several diagnostic tests can be applied and will be reviewed in this dissertation.

Convergence Diagnostics

There are several numbers of samples tools currently in existence to make the MCMC convergence assessment and provide useful feedback about the convergence of the MCMC (Brooks & Roberts, 1998; Cowles & Carlin, 1996; Roberts & Smith, 1994), including:

Trace plots (Time series plots). The trace plot, which is sometimes referred to as a time-series plot, shows the sampled values of a parameter over time. This is the plots of the iterations versus the generated values and helps in judging how rapidly the MCMC procedure converges in distribution

Autocorrelation functions plot. Autocorrelation is a term that is used about a pattern of serial correlation in the chain, where sequential draws of a parameter, say $\theta^{(t)}$, from the conditional distribution, are correlated. For this, the autocorrelation is computed as

$$cor\left(\theta^{(B)}, \theta^{(B+t)}\right), \tag{30}$$

where B is the burn-in period. Monitoring autocorrelations has a handy feature since low or high values indicate fast or slow convergence, respectively. Since it will take a very long time to explore the entire posterior distribution, this feature is essential. It should be noted that if the level of autocorrelation is high for a parameter of interest, then a trace plot will be a poor diagnostic for convergence.

Posterior variance of the parameter $(\widehat{R})$. For any given parameter, the estimated posterior variance of the parameter, $(\hat{R})$, was used to assess convergence. The estimated posterior variance of the parameter was estimated by

$$\hat{R} = \sqrt{\frac{\widehat{var}\left(\psi/_y\right)}{W}}, \tag{31}$$

where ψ was the simulated value, which was specified as $\psi_{ij},\ (i = 1, \cdots, n\ ;\ j = 1, \cdots, m)$. The subscripts i and j were specified after discarding the warm-up iterations. Then the post-burn-in iterations were split into the first and second half (i.e., m is the number of subgroups and n is the number of lengths of each chain). This posterior estimated variance consists of the between-sequence variances (B) and within-sequence variances (W). B and W can be computed from the following equations:

$$B = \frac{n}{m-1}\sum_{j=1}^{m}\left(\bar{\psi}_{\cdot j} - \bar{\psi}_{\cdot\cdot}\right), \quad \text{where,} \quad \bar{\psi}_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n}\psi_{ij} \quad \text{and} \quad \bar{\psi}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{n}\psi_{\cdot j}, \tag{32}$$

$$W = \frac{m}{m}\sum_{j=1}^{m}s_j^2, \quad \text{where,} \quad s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\psi_{ij} - \psi_{\cdot j}\right)^2, \tag{33}$$

$$\widehat{var}\left(\psi/_y\right) = \frac{n-1}{n}W + \frac{1}{n}B, \tag{34}$$

where $\psi_{\cdot j}$ is the within-sequence means, $\bar{\psi}$ is the grand mean, and s_j^2 is the variance within the chain (Gelman et al., 2014).

Heidelberger and Welch convergence diagnostic. This diagnostic applies to a single chain. The user pre-specifies ε, which is the desired relative half-width for confidence intervals. The stationarity test of Schruben, Singh, and Tierney (1983) is applied to the chain. If the test rejects the null hypothesis, the first 10% iterations of the chain are discarded. The stationarity test is repeated until 50% of iterations have been dropped. In this latter case, the failure of the chain to pass the stationarity test is an indication of the need to run the MCMC longer. A half-width test is performed on that portion of the chain that does pass the stationarity test for each parameter. Spectral density estimation yields an estimate of the standard error of the mean, leading to an estimated half-width of the confidence interval for the mean. In the case where the latter estimate is less than ε times the sample mean from the retained portion of the chain, the process is stopped, and the sample mean and confidence interval are reported.

The convergence of the simulation is calculated from the unknown parameters to ensure a precise estimation. When all the convergence tests have been passed, the total of parameters will be summarized, based on the four groups for each estimation, as recommended by the Bayesian method.

Bayesian Analysis

While there are many attractive features of a Bayesian approach, Bayesian methods do insist that prior distributions that are not straightforward be specified, especially about variance components (Agresti, 2015). There is also the issue of computation because the commonly used implementation is through a Markov chain Monte Carlo (MCMC), requiring a large computational overhead.

There has been increased interest in the simultaneous analysis of a joint model of longitudinal outcome and informative time in the last few years. It often appears in

practical applications. For example, in clinical trials, the measurements of some biomarkers are collected repeatedly over time for each patient, while some patients may experience death or dropout during the study. This can be seen in clinical trials involving the measurement of biomarkers, which are gathered repeatedly over time for each participant where there may be cases of death or participant dropout during the study. The joint model of longitudinal outcome and informative time is rapidly evolving (Alomair, 2017; Bronsert, 2009; Liang et al., 2009; Lin, 2011; Seo, 2015). At present, longitudinal data analysis does not rest on any assumptions regarding regular times for observation, and due to a large number of parameters in relation to sample size, difficulties can arise in unstructured covariance matrices. This situation has led to the progress of a variety of statistical models and methods that can handle irregular correlated data (Pullenayegum & Lim, 2016). Baghfalaki, Ganjali, and Hashemi (2014) adopted the Bayesian approach using the Markov-chain Monte Carlo method for parameter estimation. Chan and Wan (2011) considered the Bayesian approach via MCMC method for the longitudinal bivariate binary data with informative dropout model.

Employing a Bayesian analysis of generalized linear models necessitates the specification of a proper prior to account for the unknown parameters and can take a variety of forms (Agresti, 2015; Carlin & Louis, 2008; Gelman et al., 2014). In generalized linear models, the posterior distribution normally has no closed-form expression. A presenting difficulty is in the determination of the appropriate constant, which will allow the posterior to integrate to 1 by evaluating the denominator integral that determines the normalizing constant, which is often intractable, or overly complicated. The Markov chain Monte Carlo (MCMC) is commonly employed as the predominant simulation method. Sweeting (1981), for example, used noninformative priors in examining a more general

class of models beyond the GLM's. Extending Sweeting's findings to GLMs using informative normal prior to regression coefficients were West, Harrison, and Migon (1985) and Albert (1988). Ibrahim and Laud (1991) observed that uniform priors, Jeffrey's prior or diffuse priors, serve very frequently as conventional noninformative priors. It is only with great difficulty, however, that priors are elicited directly on regression parameters in a GLM, with the possibility of obtaining improper posterior distributions, resulting in an undesirable uniform prior (O'Hagan, Woodward, & Moodaley, 1990). A more accessible alternative is to obtain prior to canonical parameters, as shown by Das and Dey (2007), in their study that obtained prior to canonical parameters. They demonstrated that with a full rank assumption of the design matrix, a proper prior to the regression parameters could be induced by the various elicited prior to canonical parameters. After that, the usual Bayesian analysis based on the induced prior can be carried out.

Now that there is a user-friendly software program, WinBUGS and/or OpenBUGS, for use with Bayesian analysis that employs MCMC techniques, estimating parameters has become much more attainable for nonexperts (Spiegelhalter, Myles, Jones, & Abrams, 2000). It is possible to extend Lin's (2011) model in Bayesian inference utilizing the MCMC algorithm for the joint model of longitudinal outcome with informative time due to the computational ease, flexibility regarding model extension, and the good knowledge, where can be summarized into a prior distribution for some of the parameters of the model, but where knowledge on the rest of the parameters lacks or unavailable.

Conclusion

Longitudinal designs serve an important function in many areas of research that enhance our understanding of research objectives unattainable in other analytical approaches. The literature is replete with a large number of different longitudinal designs

and approaches with their model assumptions as an attempt to accommodate varying response data, types, and design issues.

The key characteristic of longitudinal designs is their ability to measure the change in outcomes and/or predictors at an individual level over time. Although there have been a variety of methods developed and presented to address the many different outcomes, research design issues, nearly all these methods rest on assumptions regarding the time intervals that they are fixed and/or predetermined. In reality, however, there are occasions where time points need to be based on prior outcomes, resulting in an individual measurement of response at different sets of time points. For this particular kind of research design including irregular measurement, traditional methods are not appropriate for longitudinal design with informative time data, given their assumption of a fixed time. Giving rise to newer approaches, joint model, to better answer research objectives when time points are not fixed or predetermined.

Thus, in the current study, I will use the Bayesian Approach (noninformative, informative, and semi-informative prior) to estimate the parameters of the joint model of Lin (2011). His model was designed to handle longitudinal outcomes that distributed to be a normal distribution with an informative time that followed an exponential distribution. Another aim of this dissertation is to present the Bayesian analysis and its terminology in an easier to comprehend manner without the use of large numbers of formulae. I will illustrate the attractions of the Bayesian approach and present an explanation on how to estimate a model developed within a Bayesian perspective that relies on background knowledge in the actual data analysis, followed by an interpretation of the results. Bayesian computation and an introduction to the Markov Chain Monte Carlo method will also be presented.

CHAPTER III

METHODOLOGY

The joint model by Bronsert (2009) and extended by Lin (2011) was developed under the assumption that outcomes follow a normal distribution, and time follows an exponential distribution. Later, Seo (2015) adapted and modified Bronsert and Lin's joint model to show the parameter estimates of the extended joint models satisfy the normality assumption when the distribution of outcomes is a member of the exponential family of distributions. The purpose of this study is to find Bayesian estimates for the parameters of the joint model longitudinal outcomes and informative time, assuming that the outcomes will follow a normal distribution, whereas time will follow an exponential distribution.

In order to employ the Bayesian model in estimating the value of unknown parameters, it is necessary to place previous knowledge about the parameters upon the model parameters in terms of the distribution. The distributions of these parameters are called prior distribution, which can take on different levels of information. In the review of the literature, I used in deriving the previous information for the unknown parameters, including both informative and noninformative priors. In the current study, I focus on three types of prior information (informative, noninformative, semi-informative priors).

Using simulation for data sets via Markov chain Monto Carlo simulation, the Bayesian estimation approach was implemented to investigate the performance resulting to provide an estimate of the parameters of the joint model of longitudinal outcomes and

informative time. Finally, the statistical inference was conducted based on the samples from the posterior distribution created from the generated Markov chain.

Joint Model with the Notation

The outcome for the i^{th} individual measured at the j^{th} time point is given by y_{ij}; so the i^{th} individual has a vector of outcomes $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \cdots, y_{in_i})'$ collected at a vector of time $\boldsymbol{t}_i = \left(t_{i1}, t_{i2}, \cdots, t_{in_i}\right)'$,

where,

the individuals range from $i = (1, 2, \cdots, m)$,

the time range from $j = (1, 2, \cdots, n_i)$,

n_i allows the measured time to vary from one individual to another individual. The joint distribution of outcomes $(\boldsymbol{y}_i)$ and time points $(\boldsymbol{t}_i)$ is in general given by:

$$f_{\boldsymbol{\Theta}}(\boldsymbol{y}_i, \boldsymbol{t}_i) = f_{\boldsymbol{\Theta}}(\boldsymbol{y}_i | \boldsymbol{t}_i) \cdot f_{\boldsymbol{\Theta}}(\boldsymbol{t}_i), \tag{35}$$

where,

$\boldsymbol{\Theta}$ is a vector of unknown parameters.

A general model can be derived by using this joint distribution of $(\boldsymbol{y}_i)$ and $(\boldsymbol{t}_i)$. Therefore, the general model under the assumptions that the current outcome is dependent on the one-step prior outcome (y_{ij-1}), current outcome (y_{ij}), and current time point (t_{ij}) becomes

$$f_{\boldsymbol{\Theta}}(\boldsymbol{y}_i, \boldsymbol{t}_i) = f_{\boldsymbol{\Theta}}(y_{i1} | t_{i1}) \cdot f_{\boldsymbol{\Theta}}(t_{i1}) \cdot \prod_{j=2}^{n_i} f_{\boldsymbol{\Theta}}(y_{ij} | t_{ij}, y_{ij-1}) \cdot f_{\boldsymbol{\Theta}}(t_{ij} | y_{ij-1}). \tag{36}$$

Based on this general model, a joint model was developed for each member of the exponential family of distributions, while assuming time to follow an exponential distribution. The special case will be termed the Gaussian-Exponential model (GE), can be represented as the following:

$$f_{\boldsymbol{\Theta}}(\boldsymbol{y}_i, \boldsymbol{t}_i) = \frac{1}{\sqrt{2\pi(\sigma^2)}} exp\left(-\frac{1}{2}\frac{(y_{i1} - \boldsymbol{X}'_{i1}\boldsymbol{\beta})^2}{\sigma^2}\right) \times f(t_{i1})$$

$$\times \prod_{j=2}^{n_i} \left\{ \frac{1}{\sqrt{2\pi(\sigma^2)}} exp\left(-\frac{1}{2}\frac{\left(y_{ij} - \gamma t_{ij} - \phi y_{i(j-1)} - \boldsymbol{X}'_{ij}\boldsymbol{\beta}\right)^2}{\sigma^2}\right) \right.$$

$$\left. \times exp\left(\alpha + \delta_i y_{i(j-1)}\right) exp\left(-e^{\alpha + \delta_i y_{i(j-1)}} t_{ij}\right) \right\}. \quad (37)$$

It is assumed that $f(t_{i1})$ does not depend on $\boldsymbol{\Theta}$, so for the purpose of the likelihood function, we can ignore it. Furthermore, the resulting function of the initial observation, y_{i1} is conditioned on time of observation, t_{i1}, which is the same approach found in traditional longitudinal models. However, under the Gaussian-Exponential case, the response variable is considered to be conditionally normal given time, while the time of observation is assumed to be distributed exponentially. In addition, it is assumed that the initial observation is a function only of the unknown regression parameters and that the subsequent responses are then conditioned on these unknown parameters as well as the effects of the prior response outcome and observation times. This conditional association on prior response outcomes contributes to this model's ability to analyze informative time data.

Bayesian Estimation for the Joint Model

Although the classical ML has theoretical appeal, its estimation of model parameters is complicated in that numerical methods are necessary to evaluate some complex marginal likelihood functions of the joint model in the equation. The rise of simulation-based Bayesian MCMC methods over the past few years has seen its acceptance as a popular tool for a variety of complicated statistical models. In fact, given a large

sample size and appropriate regularity conditions, the Bayesian estimator is deemed asymptotically equivalent to ML estimators (Ghosal, Lund, Moin, & Akselvoll, 1995).

The Bayesian procedure, by incorporating both data and prior information for parameters, generates the posterior distribution of unknown parameters. With prior information unavailable, noninformative priors with large variance are used. With informative priors as normal for parameters are included, and inverse gamma (1G) priors for positive parameters as in the variance σ^2, are adopted. In order to obtain the desired joint posterior distribution, the MCMC algorithm, with its ability to construct an irreducible and aperiodic Markov chain, is used to generate the equilibrium distribution.

The application of the Gibbs sampler can produce a sequence of samples of one or more variables at a time that is taken from the set of full conditional distributions. To conduct a posterior analysis, outputs are taken from the simulated chain; i. e., parameters are estimated by their posterior means. In situations where the full conditional distributions are not standard, other methods are permissible, such as the Metropolis-Hastings. An advantage of the MCMC algorithm utilizing Gibbs sampler is its ease in implementation using Bayesian software OprnBUGS and R programming.

High posterior correlations are present between some of the parameters because of the complexity of the models, slowing down convergence rates in the Gibbs samplers. As a way to address this predicament, the number of iterations needs to be substantial enough to obtain a stationary sample. Checking for independence and convergence of the sample is accomplished with trajectory plots, autocorrelation plots, of the simulated values. In addition, the convergence test uses the Cramer-von-Mises statistic to test the null hypothesis that the sampled values come from a stationary distribution is applied as well.

The Bayesian hierarchies and joint posterior distributions for all the models are then presented.

Likelihood Functions

An assumption of the model described above is that $f(t_{i1})$ does not depend on $\boldsymbol{\Theta_i}$, and can, therefore, be ignored with respect to the likelihood function. Also, the function thus obtained from the initial observation, y_{i1}, is conditioned on time of observation, t_{i1}, which is the identical approach found in traditional longitudinal models. However, subsequent observations of the response variable, y_{ij}, are no longer exclusively conditioned on time of observation, t_{ij}, alone but are now also conditioned on the most recent previous observation, y_{ij-1}, and time of observation. Thus, the likelihood function for the above joint model is the product of the density functions for *m* individuals, namely:

$$L(\Theta, y_1, y_2, \cdots, y_m) = \prod_{i=1}^{m} f_{\Theta}(\boldsymbol{y}_i, \boldsymbol{t}_i)$$

$$= \prod_{i=1}^{m} \left\{ \frac{1}{\sqrt{2\pi(\sigma^2)}} exp\left(-\frac{1}{2} \frac{\left(y_{i1} - \boldsymbol{X}'_{i1}\boldsymbol{\beta}\right)^2}{\sigma^2} \right) \right.$$

$$\times \prod_{j=2}^{n_i} \left[\left(\frac{1}{\sqrt{2\pi(\sigma^2)}} \cdot exp\left(-\frac{1}{2} \frac{\left(y_{ij} - \gamma t_{ij} - \phi y_{i(j-1)} - \boldsymbol{X}'_{ij}\boldsymbol{\beta}\right)^2}{\sigma^2} \right) \right) \right.$$

$$\left. \left. \times exp\left(\alpha + \delta_i y_{i(j-1)}\right) \cdot exp\left(-e^{\alpha + \delta_i y_{i(j-1)}} t_{ij}\right) \right] \right\}. \quad (38)$$

Prior Distribution

The model parameters that we need to estimate in the current dissertation project by using Bayesian approach, as shown in the likelihood function for the joint model, are a vector of explanatory variables ($\boldsymbol{\beta}$), the variance (σ^2), the coefficient (ϕ), accounts for the

effect of the prior outcome on the mean response, the coefficient (γ), accounts for the effect of the current time on the mean response, parameters associated with modeling time of observation include a constant parameter (α), and a coefficient that maps time of observation (δ).

Requirements for employing a Bayesian method for estimating the parameters in the joint model include the specification of the priors for the model parameters and the utilization of these priors for the calculation of the posterior distribution of each parameter. It is imperative that the priors be carefully selected for accuracy in making inferences for the parameters in the joint model. There can be a high relationship between the posterior distribution for each parameter and its prior chosen, which can extend to the priors selected for the other unknown parameters of the joint models.

Priors can take two forms: informative and noninformative. Even though researchers commonly use a noninformative prior distribution in the Bayesian analysis, using informative priors is preferable (Depaoli, 2014). With this consideration, the current project takes into account both informative and noninformative priors in estimating the unknown parameters in the joint model. The noninformative priors on some of the parameters will be set up as a proper vague prior. In this dissertation, we will adopt three scenarios for the prior distribution:

1. The prior distributions of all unknown parameters as shown in the likelihood function of this current dissertation, are informative prior, which will be set to be specific informative prior; such as normal for unrestricted parameters, N(a,b), where a and b are determined by previous studies; for positive parameters such as the variance σ^2, should be inverse Gamma (IG) priors.

2. The prior distributions of all unknown parameters are noninformative prior, which will be set to be a uniform distribution or vague prior, such as N(0, 1e6).
3. A combination of two above scenarios, so the prior distributions of some unknown parameters are noninformative, and the others are informative.

The first scenario is all unknown parameters are informative prior. In this scenario, we have to determine the hyperparameters from the previous studies. Since some historical datasets of the proposed joint model were available, then they can be used to estimate the hyperparameters and set up informative prior for this dissertation. The Bayesian approach with informative prior is conducted using the same model as the classical approach that used in the previous studies, such as Bronsert (2009), Lin (2011), and Seo (2015).

Therefore, the prior for the vector of explanatory variables ($\boldsymbol{\beta}$) is the multivariate normal distribution denoted as

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, a\boldsymbol{\Sigma}_\beta), \tag{39}$$

where $\boldsymbol{\mu}_\beta$ is the hyperparameter representing the mean vector with the value of $0.4\mathbf{I}_m$, and $\boldsymbol{\Sigma}_\beta$ is the hyperparameter representing the covariance matrix defined as the identity matrix, $\boldsymbol{I}_m$. The dimension of both depends on m individuals and the hyperparameter. *a.* is set to be 4.0. The multivariate normal distribution has the density function

$$p(\beta) = (2\pi)^{\frac{m}{2}}\left|\boldsymbol{\Sigma}_\beta\right|^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(\beta - \boldsymbol{\mu}_\beta)'(\boldsymbol{\Sigma}_\beta)^{-1}(\beta - \boldsymbol{\mu}_\beta)\right). \tag{40}$$

The priors for the amount of variance (σ^2) is Inverse Gamma (IG) distribution denoted as

$$\sigma^2 \sim IG(a_1, a_2), \tag{41}$$

where the hyperparameters a_1 and a_2 are set to be 0.2. The inverse Gamma has density function,

$$p(\sigma^2) = \frac{a_2{}^{a_1}}{\Gamma(a_1)} (\sigma^2)^{-(a_1+1)} e^{-a_2/(\sigma^2)}, \qquad (\sigma^2) > 0. \tag{42}$$

The prior for the coefficient, (ϕ), which is accounts for the effect of the prior outcome on the mean response is normal distribution denoted as

$$\phi \sim N(\mu_\phi, \tau_\phi^2), \tag{43}$$

where μ_ϕ is the hyperparameter, which is known with a mean vector with the value of 0.2, and the hyperparameter τ_ϕ^2 is set to be 0.2. the normal distribution has the density function

$$p(\phi) = \frac{1}{\sqrt{2\pi\,\tau_\phi^2}} \exp\left(-\frac{1}{2}(\phi - \mu_\phi)^2\right). \tag{44}$$

Similarly, the prior for the coefficient (γ), accounts for the effect of the current time on the mean response is normal distribution denoted as

$$\gamma \sim N(\mu_\gamma, \tau_\gamma^2), \tag{45}$$

where μ_γ is the hyperparameter, which is known with a mean vector with the value of 0.5, and the hyperparameter τ_γ^2 is set to be 0.5.

The priors for the parameters associated with modeling time of observation include a constant parameter (α), and a coefficient that maps time of observation (δ) are also the same as normal distribution denoted as

$$\alpha \sim N(\mu_\alpha, \tau_\alpha^2) \qquad \text{and} \qquad \delta \sim N(\mu_\delta, \tau_\delta^2), \tag{46}$$

where μ_α and μ_δ are the hyperparameters, which is known with mean vectors with the value of 2.0 and 0.2 respectively, and the hyperparameter τ_α^2 and τ_δ^2 are set to be 0.2 and 0.1 respectively.

The next step toward a more general joint model is to assume an independent prior distribution for all unknown parameters in the joint model. The subsequent discussion is the joint prior density, which must have the product from all unknown parameters. Based on the probability density function discussed above, the joint prior density for informative priors corresponds to

$$p(\boldsymbol{\beta}, \phi, \gamma, \alpha, \delta, \sigma^2) = p(\boldsymbol{\beta}) \times p(\phi) \times p(\gamma) \times p(\alpha) \times p(\delta) \times p(\sigma^2)$$

$$= (2\pi)^{\frac{m}{2}} \left|\boldsymbol{\Sigma}_\beta\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\beta - \boldsymbol{\mu}_\beta\right)'\left(\boldsymbol{\Sigma}_\beta\right)^{-1}\left(\beta - \boldsymbol{\mu}_\beta\right)\right)$$

$$\times \frac{1}{\sqrt{2\pi\,\tau_\phi^2}} \exp\left(-\frac{1}{2}\left(\phi - \mu_\phi\right)^2\right) \times \frac{1}{\sqrt{2\pi\,\tau_\gamma^2}} \exp\left(-\frac{1}{2}\left(\gamma - \mu_\gamma\right)^2\right)$$

$$\times \frac{1}{\sqrt{2\pi\,\tau_\alpha^2}} \exp\left(-\frac{1}{2}(\alpha - \mu_\alpha)^2\right) \times \frac{1}{\sqrt{2\pi\,\tau_\delta^2}} \exp\left(-\frac{1}{2}(\delta - \mu_\delta)^2\right)$$

$$\times \frac{{a_2}^{a_1}}{\Gamma(a_1)} \left(\sigma^2\right)^{-(a_1+1)} e^{-a_2/(\sigma^2)} \tag{47}$$

The second scenario is all unknown parameters are noninformative prior. Due to difficulties in the availability of scientifically solid prior information about the unknown parameters, most of the studies used by noninformative priors in Bayesian inference. Therefore, in this dissertation, the vague prior for all parameters were used as the following: The vague priors for the vector of explanatory variables ($\boldsymbol{\beta}$) is the multivariate normal distribution, which is denoted as

$$\boldsymbol{\beta} \sim N(\mathbf{0}_m, 10^6 \mathbf{I}_m). \tag{48}$$

The vague prior for the variance (σ^2) is Inverse Gamma (IG) distribution denoted as

$$\sigma^2 \sim IG(10^{-3}, 10^{-3}). \quad (49)$$

The vague prior for the coefficient (ϕ) is normal distribution denoted as

$$\phi \sim N(0, 10^6). \quad (50)$$

Similarly, the vague prior for the coefficient (γ) is normal distribution denoted as

$$\gamma \sim N(0, 10^6). \quad (51)$$

The vague priors for the parameters (α) and (δ) are also the same as normal distribution denoted as

$$\alpha \sim N(0, 10^6) \quad \text{and} \quad \delta \sim N(0, 10^6). \quad (52)$$

The final form for the joint prior density for noninformative prior distribution is:

$$p(\boldsymbol{\beta}, \phi, \gamma, \alpha, \delta, \sigma^2) = p(\boldsymbol{\beta}) \times p(\phi) \times p(\gamma) \times p(\alpha) \times p(\delta) \times p(\sigma^2). \quad (53)$$

The third scenario is some unknown parameters are informative prior and the others are noninformative prior. The informative prior distributions for the vector of explanatory variables ($\boldsymbol{\beta}$) and the amount of variance (σ^2) are adopted, while for the other parameters, I assumed to be noninformative priors. Therefore,

$$\boldsymbol{\beta} \sim N(\mathbf{0.4}_m, 4\boldsymbol{I}_m), \quad (54)$$

$$\sigma^2 \sim IG(0.2, 0.2). \quad (55)$$

The reasonable noninformative prior density for each of them, as in most previous studies, is also vague priors and they are independent of each other (Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), therefore,

$$\phi \sim N(0, 10^6) \quad \text{and} \quad \gamma \sim N(0, 10^6), \quad (56)$$

$$\alpha \sim N(0, 10^6) \quad \text{and} \quad \delta \sim N(0, 10^6). \quad (57)$$

For all unknown parameters in the joint model is to assume to be an independent prior distribution, and based on the probability density function discussed above, the joint prior density for semi-informative prior correspond to

$$p(\boldsymbol{\beta}, \phi, \gamma, \alpha, \delta, \sigma^2) = p(\boldsymbol{\beta}) \times p(\phi) \times p(\gamma) \times p(\alpha) \times p(\delta) \times p(\sigma^2)$$

$$= (2\pi)^{\frac{m}{2}} \left|\boldsymbol{\Sigma}_{\beta}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\beta - \boldsymbol{\mu}_{\beta}\right)'\left(\boldsymbol{\Sigma}_{\beta}\right)^{-1}\left(\beta - \boldsymbol{\mu}_{\beta}\right)\right)$$

$$\times \frac{{a_2}^{a_1}}{\Gamma(a_1)} \left(\sigma^2\right)^{-(a_1+1)} e^{-a_2/(\sigma^2)} \times p(\phi) \times p(\gamma) \times p(\alpha) \times p(\delta). \quad (58)$$

Posterior Distribution

As we have seen, the posterior distribution is proportional to the likelihood function multiplied by the prior distribution:

$$p\left(\boldsymbol{\beta}, \phi, \gamma, \alpha, \delta, \sigma^2 | \boldsymbol{y}_{i'} \boldsymbol{t}_i\right) = L\left(\Theta, y_1, y_2, \cdots, y_m\right) \times p(\boldsymbol{\beta}, \phi, \gamma, \alpha, \delta, \sigma^2). \quad (59)$$

For each scenario, the estimation of unknown parameters is obtained by using Bayesian methods through the priors of unknown parameters in the joint model, and the likelihood function was previously specified. The marginal posterior distribution of the parameter of interest, which is the aim of Bayesian analyses, is hard to obtain because of high-dimensional integration in the model discussed. Therefore, MCMC techniques were used for the posterior computation of the proposed model in the current project. More specifically, when the dimension of the parameter space in Markov chain simulation changes from one iteration to the next iteration, the Metropolis algorithm using the method of reversible jump sampling is suitable to perform (Gelman et al., 2014). Then the posterior distribution of unknown parameters is obtained.

Markov Chain Monte Carlo Simulation

In this dissertation, in order to obtain the posterior distribution of all parameters, the Metropolis-Hastings algorithm is employed as a part of the Markov Chain Monte Carlo simulation. This algorithm creates a new sample for all dimensions at one time by generating a sequence of samples iteratively with the distribution of the next sample where each sample is only dependent on the just prior sample from the current state of the chain (i.e., Markov chain). This allows one parameter to be updated in line with the acceptance probability, a specification within the Metropolis-Hastings formula.

Convergence and Summary Statistics

At this point in the analysis, an assessment is made of convergence, and the summary statistics are summarized. In order to demonstrate the application of the Bayesian approach, there usually needs to be a burn-in, or warm-up, period before to the convergence of the estimated unknown parameters in the distribution to the right posterior (Geyer, 1992). This burn-in period is that iteration in which a run needs to be discarded to prevent autocorrelation of the samples. The convergence of the chain with the stationary distribution takes place with a larger number of iterations within the algorithm (Leiby, Have, Lynch, & Sammel, 2014). As a result, a burn-in period is specified at 10,000 iterations, with the next 10,000 iterations employed as post-burn-in iterations, as described by Depaoli (2014). This yields a total of 20,000 iterations in order to obtain convergence testing and data analysis.

There is no straightforward method to calculate a convergence diagnostic. The general consensus in various proposed methods is to monitor common parameters in each of the models. So, with that in mind, I will assess convergence of the unknown parameters

using the Heidelberger and Welch (1983) diagnostic (HW), which calculates a test statistic to accept or reject the null hypothesis that the Markov chain is from a stationary distribution. The diagnostic consists of two parts.

The first one is called the stationary test, which assesses the null hypothesis that the Markov Chain is in the stationary distribution and produces *p*-values for each estimated parameter. If the null hypothesis is rejected, then discard the first 10% of the chain, and calculate the test statistic to accept or reject the same null hypothesis. If it is rejected, then discard the next 10% and calculate the test statistic. Repeat until the null hypothesis is accepted, or 50% of the chain is discarded. If the test still rejects the null hypothesis, then the chain fails the test and needs to be run longer.

The second part is called the halfwidth test, which indicates whether there are enough iterations to estimate the mean of a marginal posterior distribution with sufficient precision, assuming that the Markov Chain is in the stationary distribution. If the chain passes the first part of the diagnostic, then it takes the part of the chain not discarded from the first part to test the second part. The halfwidth test calculates half the width of the $(1-\alpha)$ % credible interval around the mean. If the ratio of the halfwidth and the mean is lower than some ϵ, then the chain passes the test. Otherwise, the chain must be run out longer.

The HW tests were chosen because it requires only one realization of the MCMC to use. We used the R package coda with the OpenBUGS program to implement the diagnostic via the function heidle.diag(). The software produces a table showing for each parameter if the parameter passed or failed the stationarity test and halfwidth test, the number of iterations retained and discarded, the Cramer-von Mises statistic, the sample mean, and the estimated halfwidth. Default values were for all arguments of the function were used. In addition, each of the trace plots of the unknown parameter is presented to

examine any systematic deviation from the steady-state of a converged MCMC chain as well as the autocorrelation plots for the common parameters. The convergence of the simulation is calculated from the unknown parameters to ensure a precise estimation. When all the convergence tests have been passed, the total of parameters will be summarized as recommended by the Bayesian method.

The Bayesian estimates theoretically are expected to be biased (Gifford & Swaminathan, 1990). In this way, the focus of this dissertation, to estimate the unknown parameters in the joint model simultaneously, will be obtained. Gelman et al. (2014) remarked that it is not possible to obtain an approximately unbiased estimator when several parameters are simultaneously estimated since the information or knowledge of these parameters is of relevance in estimating other parameters. Theoretically, the bias in the Bayesian estimates is expected, and investigations by Gifford and Swaminathan (1990) demonstrated that both joint maximum likelihood and Bayesian techniques contained biases in the estimation. This was confirmed by Ho et al. (2011) regarding bias in the posterior mean of the parameters. With this in mind, the current dissertation will omit the bias diagnosis across all model parameters in the summary statistics part.

Verifying the Validity of a Simulation

In order to gauge the validity of the MCMC method as a way to estimate unknown parameters on joint models, this dissertation will specify different scenarios and different prior by using the Markov chain Monte Carlo sampling in a Bayesian approach. Simplification of the presentation will be accomplished through simulation design, implementation, and data generation.

Analysis of Simulated Data

Inasmuch as this study is an extension of Bronsert (2009), Lin (2011), and Seo (2015), most of the parameter values and simulation conditions were derived from the work of these authors. Table 1 parameter values and Table 2 simulation conditions were based on the work of Lin (2011). As a method to verify the properties of the MLEs, Monte Carlo simulations were used with SAS/IML and/or R in their model. However, in this dissertation, we will apply R with OpenBUGS programs to find Bayesian estimates for unknown parameters. It is assumed that parameters are the same across subjects for model simplicity. It is also assumed that the observations follow a normal distribution but that observations for the time intervals follow an exponential distribution in the joint model used in this study.

When using simulated data, its basic structure comprises two categorical variables with three levels each along with two continuous variables associated with the response variable. To obtain the first outcome, data is gathered from the normal distribution; the next outcome is then calculated on the relationship between the previous outcome along with the previous time in order to predict the average outcome with fixed parameter values. Observation times follow the exponential distribution, adjusting the mean by the previous outcome. It is assumed that all parameter values are equal across subjects to simplify the model form in simulation studies. All terms with fixed parameter values were based on previous studies (see Table 1).

Table 1

Parameter values for simulations

α	β_0	β_1	β_2	β_3	β_4	β_5	β_6	δ	γ	ϕ	σ^2
2	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.01	0.1	0.8	1
1	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.02	0.1	0.8	1
2	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.01	0.1	0.8	2
1	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.02	0.1	0	2
2	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.01	0.1	0	0.5
1	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.02	0.1	0.8	0.5

As seen in Table 2, the five sample sizes are incorporated with four types of design structures with a different number of observations to detect a certain pattern as the number of observations increases. For the replications, some researchers used 500 replications (Liang et al., 2009), some used 1,000 replications (Lipsitz et al., 2002; Qiu et al., 2013; & Seo, 2015), some used 2,000 replication (Alomair, 2017), and some used 5,000 replication (Lin, 2011). In this study, each simulation design will be run 1,000 times. Each prior distribution will have 120 simulation designs.

To sum up, three different prior distributions of the unknown parameters will be applied based on whether the information about the parameters was known or not, along with the six parameter schemes and five different sample sizes, with four different numbers of observations (see Table 2). With all conditions applied, analyses for the total of 360 simulation designs will be conducted on the joint models in the current dissertation.

Table 2

Simulation designs

Scheme Number	***Sample Size***	***Number of Observation***	***Design Structure***	***Total Number of Observation***
1	18	10	Balanced	180
2		5 & 3	Unbalanced	72
3		10 & 5	Unbalanced	135
4		20 &6	Unbalanced	234
5	36	10	Balanced	360
6		5 & 3	Unbalanced	144
7		10 & 5	Unbalanced	180
8		20 &6	Unbalanced	288
9	54	10	Balanced	540
10		5 & 3	Unbalanced	216
11		10 & 5	Unbalanced	405
12		20 &6	Unbalanced	702
13	90	10	Balanced	900
14		5 & 3	Unbalanced	360
15		10 & 5	Unbalanced	675
16		20 &6	Unbalanced	1170
17	180	10	Balanced	1800
18		5 & 3	Unbalanced	720
19		10 & 5	Unbalanced	1350
20		20 &6	Unbalanced	2340

Simulation Process

Using concepts from Bayesian statistics, the fundamental information of population parameters will be exercised by updating the original distribution of the parameters of interest through the process of the conditioning on data via the likelihood function (Gill, 2008). To better understanding the behavior of the statistical estimates through their sampling distributions, samples will be drawn from the target distribution or posterior distribution of all unknown parameters. Utilizing the Markov chain Monte Carlo with the Metropolis-Hastings algorithm, the sampling distributions will thus be derived from the data.

The process of estimating the unknown parameters for the joint model data files comprises five steps: (a) data generation, (b) calculation of the likelihood function, (c) calculation of the joint prior distribution for all unknown parameters, (d) calculation of the joint posterior probability distribution, and (e) completion of the sampling from the joint posterior distribution of parameters and hyperparameters. After the joint posterior distribution is secured, the simulation procedures will be instituted utilizing the Metropolis-Hastings sampling method. The generation of the Metropolis-Hastings algorithm samples from the point proposal distribution comprises the following four steps at each iteration:

1. Generate Markov chains for model parameters via the Metropolis-Hastings algorithm, i.e., drawing samples from the posterior distribution by set the start values. The initial values for model parameters were assigned from other simulation studies with the joint model (Bronsert, 2009; Lin, 2011; & Seo, 2015).

2. Run the Markov chain for 20,000 iterations where the first 10,000 constitute the burn-in period while the last 10,000 constitute the simulation draws, in order to assure the passing of the convergence test with no autocorrelation.
3. Test the convergence of the Markov chains.
4. Calculate the inferential statistics.

As a way to estimate all unknown parameters in the joint model under a Bayesian approach, the Metropolis-Hastings algorithm will be implemented in the R program. The joint model will comprise 20,000 iterations, the first 10,000 for the burn-in, and the remaining 10,000 for the post-burn-in analysis to attain independent sample values. These sample values will then form the data bank for convergence testing and data analysis. The convergence of the simulated sequences will be monitored after the simulation has been running for a period of time.

The Heidelberger and Welch tests of any given parameter will be computed in assessing convergence. The software produces a table showing for each parameter if the parameter passed or failed the stationarity and halfwidth tests. In the case of each simulation condition, the HW tests of each parameter will be computed and assessed convergence. Then. The summary statistics such as mean, standard deviation, and the average upper and lower limits of the 97.5% confidence interval will be presented based on the 20,000 iterations of the sampling run in the R and OpenBUGS programs.

As described above, the Bayesian analysis will be employed for investigating 120 data sets, where each of the data sets comes from the proposed joint model with the six parameter schemes and four different numbers of observations. The sample sizes will be in all cases N = 18, N = 36, N = 54, N = 90, N = 180. Every data set of the simulated joint model data will be calculated from one and/or three chains, each of which contains 20,000

iterations. The samples that will be drawn from a posterior distribution in each chain will be divided into two groups of 10,000 iterations each. The target distribution samples (posterior distribution) will be used in the calculation of inferential statistics. The parameters will then be estimated after fitting all unconditional joint models without covariances. P-value and 97.5% credibility intervals will be generated in the joint model estimates based on Markov chain Monte Carlo. Specification of the different degrees of prior knowledge for the model parameters will be presented, and the simulation outcomes across different prior distributions will be also compared.

Applied Bayesian Modeling Using OpenBUGS via R

A convenient way to fit Bayesian models using OpenBUGS is to use R packages that function as frontends. These packages make it easy to do all Bayesian data analysis in R, including estimate the model using MCMC and process the output of Bayesian models Gelman and Hill (2007). Fitting Bayesian models using OpenBUGS by installing and loading the respective package (*R2OpenBUGS package*) within R or Rstudio.

R offers a variety of solutions to obtain convergence diagnostics. MCMC objects are a separate class of R objects that contain one or multiple Markov Chains and the respective information about iterations etc. that are needed to conduct convergence diagnostics. In R, there are a variety of commands for diagnostics and presentation using the coda package with MCMC, including density plot, trac plot, Autocorrelation plot, Geweke diagnostic, Raftery and Lewis diagnostic, and Heidelberger and Welch diagnostic (Plummer, Best, Cowles, & Vines, 2006).

CHAPTER IV

RESULTS

The primary purpose of this dissertation was to find Bayesian estimates for the unknown parameters in the proposed joint model based on the three kinds of prior distribution, (informative, noninformative, and semi-informative priors), with the assumptions of a normal distribution for the outcome process and an exponential distribution for informative time, by using developed R with OpenBUGS software. Analyses for a total of 360 convergence diagnostics were conducted on the proposed joint model. The 360 diagnostics refer to three different prior distributions of the unknown parameters, six parameter schemes, and five different sample sizes with four different numbers of observations (see Table 2). These diagnostics were run two times with 20,000 iterations for each run, including one or three chains of MCMC.

The convergence on all parameters in the proposed joint model was examined first. The parameter estimates on the joint models were then calculated based on the convergence diagnostic tests. The parameters of the proposed joint model such as the vector of explanatory variables ($\boldsymbol{\beta}$), the variance (σ^2), the coefficient (ϕ), accounts for the effect of the prior outcome on the mean response, the coefficient (γ), accounts for the effect of the current time on the mean response, parameters associated with modeling time of observation include a constant parameter (α), and a coefficient that maps time of observation (δ) were estimated by using Bayesian approach. Moreover, the proposed joint model-generated data sets were fitted and analyzed using R with OpenBUGS. This chapter

presents the simulation results to verify the Bayesian estimates of the proposed joint model. To achieve these goals, the researcher attempted to study the following research questions:

Q1 How will the Bayesian method be designed for estimating the unknown parameters on the proposed joint model constructed by Lin (2011) for a longitudinal response variable with a set of informative time?

Q2 How are these Bayesian estimates of the proposed joint model influenced by a few select variations in subject sample size, types of design structures with a different number of observations for each subject, and the various parameter schemes, with three types of prior distribution on the parameters (noninformative, informative, and semi-informative priors)?

Q3 How will the developed R program work closely together with OpenBUGS for fitting Bayesian models? Could that support researchers obtain the Bayesian estimations for the unknown parameters on the proposed joint model?

The first and second questions were answered by a simulation study that was presented to evaluate the performance of the proposed Bayesian joint models using the Monto Carlo Markov Chain. To specify the performance of the Bayesian approach, sample sizes of 18, 36, 54, 90, and 180 were considered on each six parameter schemes with four different numbers of observations. Moreover, monitoring convergence was assessed first by calculating the Heidelberger and Welch tests of all parameters. The following section shows the estimation of all parameters in the proposed joint model by using R programs using a Bayesian approach. In conclusion, the summary of the results section shows the performance of the Bayesian method using a developed R with OpenBUGS programs for the current dissertation to estimate the parameters of the proposed joint model.

Steps of Simulation

A general description of the simulation procedures is as follows.

Step1: A design matrix related to the outcome was generated with two continuous and two categorical variables with three levels each.

Step 2: A dataset that contains three variables (outcomes, time, and subject) was created based on the relations among previous and current outcomes, and the current time with the fixed parameter values shown in Table 1.

Step 3: Bayesian models fit in OpenBUGS by installing and loading the respective package (*R2OpenBUGS package*) within R or Rstudio. The model for BUGS can be written in the R script, including the prior distribution.

Step 4: The parameters whose posterior distributions the researcher is interested in summarizing later were defined.

Step 5: The function of starting values for BUGS was created. It is a list that contains one element for each parameter.

Step 6: OpenBUGS can be run from R by using *bugs()* function. Additionally, the location of the model file, the data, the parameters, and the initial values must be specified as well as how many chains the researcher wants to fit and how long he wants to run them.

Step 7: After running, the OpenBUGS window will pop up, and R will freeze up. The model will now run in OpenBUGS for a while. When OpenBUGS is done, its window will close, and R will work again.

Step 8: After finishing OpenBUGS running, the resulting data can be read into R by using a coda package for more analyses of the output, graphical summary of inference, and convergence diagnostics (Sturtz, Ligges, & Gelman, 2005).

Step 9: The previous steps were repeated 1,000 times. Note that every single replication, the HW diagnostic was used every time to test the convergence of each parameter in the proposed joint model. Then the parameter estimates and another statistical inference can be calculated.

The test was conducted separately for the parameter estimates of the different prior distribution. Simulation designs are based on five sample sizes, four different observations, and six parameter schemes. Each prior distribution has 120 (5 * 4 * 6) simulation conditions, which makes 360 simulation designs in total for all three prior distributions. This section aims to answer the third research question, where the developed R with OpenBUGS codes for the above steps can handle the Bayesian estimations for the parameters on the joint model of an outcome from Gaussian distribution, and informative time, which follows an exponential distribution. These codes presented in Appendix A. After having all the input to use the R and OpenBUGS software, the codes compute the estimates and different convergence diagnostics.

Model Convergence

There is no straightforward method to calculate a convergence diagnostic. Therefore, one and three generating Markov chains with a total of 20,000 iterations each were run for the convergence testing. With the burn-in period of 10,000 iterations for each chain, the outputs of Heidelberger and Welch diagnostic (HW) tests of all unknown parameters were monitored. This diagnostic employs two tests. (1) The stationary test, which determines whether the trace of simulated values arises from a stationary stochastic process. (2) The halfwidth test, which determines if there are enough iterations to estimate the mean of the process with acceptable precision. The reason for choosing HW tests is because it only requires one realization of the MCMC to use. OpenBUGS with R package coda was used to implement the diagnostic via the function heidle.diag(). Since there is a huge number of test tables for all parameters in the proposed joint model that cannot be placed in this study, some examples are randomly introduced for the output of all

parameters with different prior (noninformative, informative, and semi-informative priors) are summarized in Tables 3 through 11.

Table 3

The Heidelberger & Welch (HW) tests (stationary and halfwidth tests) of all unknown parameters in the joint model for noninformative Prior (Suggested Example 1)

Parameters	Sample Size = 18 Parameters Scheme = 6 Design Structure = 20 & 6 (Unbalanced)			Sample Size = 36 Parameters Scheme = 4 Design Structure = 5 & 3 (Unbalanced)		
	Stationarity Test	Halfwidth Test	p-value	Stationarity Test	Halfwidth Test	p-value
$\boldsymbol{\alpha}$	Passed	Passed	0.1801	Passed	Passed	0.6822
$\boldsymbol{\beta_0}$	Passed	Passed	0.2249	Passed	Passed	0.7589
$\boldsymbol{\beta_1}$	Passed	Passed	0.1750	Passed	Passed	0.3157
$\boldsymbol{\beta_2}$	Passed	Passed	0.1723	Passed	Passed	0.4473
$\boldsymbol{\beta_3}$	Passed	Passed	0.6622	Passed	Passed	0.9609
$\boldsymbol{\beta_4}$	Passed	Passed	0.1261	Passed	Passed	0.7934
$\boldsymbol{\beta_5}$	Passed	Passed	0.6898	Passed	Passed	0.5965
$\boldsymbol{\beta_6}$	Passed	Passed	0.2762	Passed	Passed	0.3973
$\boldsymbol{\delta}$	Passed	Passed	0.4828	Passed	Passed	0.7379
$\boldsymbol{\gamma}$	Passed	Passed	0.6887	Passed	Passed	0.5422
$\boldsymbol{\phi}$	Passed	Passed	0.8056	Passed	Passed	0.8779
$\boldsymbol{\sigma^2}$	Passed	Passed	0.2040	Passed	Passed	0.1233

Table 4

The Heidelberger & Welch (HW) tests (stationary and halfwidth tests) of all unknown parameters in the joint model for noninformative Prior (Suggested Example 2)

Parameters	Sample Size = 54 Parameters Scheme = 2 Design Structure = 10 & 5 (Unbalanced)			Sample Size = 90 Parameters Scheme = 1 Design Structure = 10 & 10 (Balanced)		
	Stationarity Test	Halfwidth Test	p-value	Stationarity Test	Halfwidth Test	p-value
$\boldsymbol{\alpha}$	Passed	Passed	0.1298	Passed	Passed	0.1151
$\boldsymbol{\beta_0}$	Passed	Passed	0.3695	Passed	Passed	0.4498
$\boldsymbol{\beta_1}$	Passed	Passed	0.2621	Passed	Passed	0.2294
$\boldsymbol{\beta_2}$	Passed	Passed	0.1759	Passed	Passed	0.1489
$\boldsymbol{\beta_3}$	Passed	Passed	0.6643	Passed	Passed	0.4682
$\boldsymbol{\beta_4}$	Passed	Passed	0.2112	Passed	Passed	0.0931
$\boldsymbol{\beta_5}$	Passed	Passed	0.6551	Passed	Passed	0.7356
$\boldsymbol{\beta_6}$	Passed	Passed	0.3379	Passed	Passed	0.3337
$\boldsymbol{\delta}$	Passed	Passed	0.1820	Passed	Passed	0.0297
$\boldsymbol{\gamma}$	Passed	Passed	0.4202	Passed	Passed	0.7648
$\boldsymbol{\phi}$	Passed	Passed	0.7140	Passed	Passed	0.5296
$\boldsymbol{\sigma^2}$	Passed	Passed	0.1925	Passed	Passed	0.1595

Table 5

The Heidelberger & Welch (HW) tests (stationary and halfwidth tests) of all unknown parameters in the joint model for noninformative Prior (Suggested Example 3)

Parameters	Sample Size = 18 Parameters Scheme = 5 Design Structure = 10 & 10 (Balanced)			Sample Size = 180 Parameters Scheme = 3 Design Structure = 5 & 3 (Unbalanced)		
	Stationarity Test	Halfwidth Test	p-value	Stationarity Test	Halfwidth Test	p-value
$\boldsymbol{\alpha}$	Passed	Passed	0.1289	Passed	Passed	0.0892
$\boldsymbol{\beta_0}$	Passed	Passed	0.1426	Passed	Passed	0.7001
$\boldsymbol{\beta_1}$	Passed	Passed	0.0902	Passed	Passed	0.5333
$\boldsymbol{\beta_2}$	Passed	Passed	0.0583	Passed	Passed	0.3103
$\boldsymbol{\beta_3}$	Passed	Passed	0.4209	Passed	Passed	0.9965
$\boldsymbol{\beta_4}$	Passed	Passed	0.1360	Passed	Passed	0.2920
$\boldsymbol{\beta_5}$	Passed	Passed	0.1028	Passed	Passed	0.7268
$\boldsymbol{\beta_6}$	Passed	Passed	0.4482	Passed	Passed	0.0932
$\boldsymbol{\delta}$	Passed	Passed	0.1290	Passed	Passed	0.0515
$\boldsymbol{\gamma}$	Passed	Passed	0.9305	Passed	Passed	0.0782
$\boldsymbol{\phi}$	Passed	Passed	0.8706	Passed	Passed	0.3899
$\boldsymbol{\sigma^2}$	Passed	Passed	0.3107	Passed	Passed	0.2624

Table 6

The Heidelberger & Welch (HW) tests (stationary and halfwidth tests) of all unknown parameters in the joint model for informative Prior (Suggested Example 1)

Parameters	Sample Size = 180 Parameters Scheme = 5 Design Structure = 20 & 6 (Unbalanced)			Sample Size = 18 Parameters Scheme = 4 Design Structure = 10 & 10 (Balanced)		
	Stationarity Test	Halfwidth Test	p-value	Stationarity Test	Halfwidth Test	p-value
$\boldsymbol{\alpha}$	Passed	Passed	0.1033	Passed	Passed	0.1447
$\boldsymbol{\beta_0}$	Passed	Passed	0.4149	Passed	Passed	0.2161
$\boldsymbol{\beta_1}$	Passed	Passed	0.2942	Passed	Passed	0.1537
$\boldsymbol{\beta_2}$	Passed	Passed	0.2210	Passed	Passed	0.1145
$\boldsymbol{\beta_3}$	Passed	Passed	0.1220	Passed	Passed	0.7908
$\boldsymbol{\beta_4}$	Passed	Passed	0.0703	Passed	Passed	0.2469
$\boldsymbol{\beta_5}$	Passed	Passed	0.6870	Passed	Passed	0.7112
$\boldsymbol{\beta_6}$	Passed	Passed	0.3907	Passed	Passed	0.4038
$\boldsymbol{\delta}$	Passed	Passed	0.0755	Passed	Passed	0.0445
$\boldsymbol{\gamma}$	Passed	Passed	0.6841	Passed	Passed	0.9441
$\boldsymbol{\phi}$	Passed	Passed	0.6242	Passed	Passed	0.8787
$\boldsymbol{\sigma^2}$	Passed	Passed	0.3170	Passed	Passed	0.2108

Table 7

The Heidelberger & Welch (HW) tests (stationary and halfwidth tests) of all unknown parameters in the joint model for informative Prior (Suggested Example 2)

Parameters	Sample Size = 90 Parameters Scheme = 3 Design Structure = 5 &3 (Unbalanced)			Sample Size = 54 Parameters Scheme = 1 Design Structure = 10 &10 (Balanced)		
	Stationarity Test	Halfwidth Test	p-value	Stationarity Test	Halfwidth Test	p-value
$\boldsymbol{\alpha}$	Passed	Passed	0.1319	Passed	Passed	0.1670
$\boldsymbol{\beta_0}$	Passed	Passed	0.3163	Passed	Passed	0.8088
$\boldsymbol{\beta_1}$	Passed	Passed	0.3508	Passed	Passed	0.4146
$\boldsymbol{\beta_2}$	Passed	Passed	0.2173	Passed	Passed	0.2250
$\boldsymbol{\beta_3}$	Passed	Passed	0.8946	Passed	Passed	0.8910
$\boldsymbol{\beta_4}$	Passed	Passed	0.1627	Passed	Passed	0.1423
$\boldsymbol{\beta_5}$	Passed	Passed	0.5321	Passed	Passed	0.6869
$\boldsymbol{\beta_6}$	Passed	Passed	0.6210	Passed	Passed	0.1763
$\boldsymbol{\delta}$	Passed	Passed	0.0850	Passed	Passed	0.1083
$\boldsymbol{\gamma}$	Passed	Passed	0.7129	Passed	Passed	0.7036
$\boldsymbol{\phi}$	Passed	Passed	0.7771	Passed	Passed	0.3764
$\boldsymbol{\sigma^2}$	Passed	Passed	0.1779	Passed	Passed	0.2146

Table 8

The Heidelberger & Welch (HW) tests (stationary and halfwidth tests) of all unknown parameters in the joint model for informative Prior (Suggested Example 3)

Parameters	Sample Size = 36 Parameters Scheme = 6 Design Structure = 10 & 5 (Unbalanced)			Sample Size = 180 Parameters Scheme = 2 Design Structure = 20 & 6 (Unbalanced)		
	Stationarity Test	Halfwidth Test	p-value	Stationarity Test	Halfwidth Test	p-value
$\boldsymbol{\alpha}$	Passed	Passed	0.9515	Passed	Passed	0.7724
$\boldsymbol{\beta_0}$	Passed	Passed	0.3374	Passed	Passed	0.3877
$\boldsymbol{\beta_1}$	Passed	Passed	0.3492	Passed	Passed	0.2863
$\boldsymbol{\beta_2}$	Passed	Passed	0.2600	Passed	Passed	0.2012
$\boldsymbol{\beta_3}$	Passed	Passed	0.6481	Passed	Passed	0.1170
$\boldsymbol{\beta_4}$	Passed	Passed	0.1616	Passed	Passed	0.0818
$\boldsymbol{\beta_5}$	Passed	Passed	0.9049	Passed	Passed	0.6831
$\boldsymbol{\beta_6}$	Passed	Passed	0.6972	Passed	Passed	0.3890
$\boldsymbol{\delta}$	Passed	Passed	0.9716	Passed	Passed	0.7336
$\boldsymbol{\gamma}$	Passed	Passed	0.6867	Passed	Passed	0.5445
$\boldsymbol{\phi}$	Passed	Passed	0.8452	Passed	Passed	0.4980
$\boldsymbol{\sigma^2}$	Passed	Passed	0.2097	Passed	Passed	0.3126

Table 9

The Heidelberger & Welch (HW) tests (stationary and halfwidth tests) of all unknown parameters in the joint model for semi-informative Prior (Suggested Example 1)

Parameters	Sample Size = 90 Parameters Scheme = 2 Design Structure = 10 & 10 (Balanced)			Sample Size = 36 Parameters Scheme = 5 Design Structure = 20 & 6 (Balanced)		
	Stationarity Test	Halfwidth Test	p-value	Stationarity Test	Halfwidth Test	p-value
$\boldsymbol{\alpha}$	Passed	Passed	0.6421	Passed	Passed	0.0976
$\boldsymbol{\beta_0}$	Passed	Passed	0.4473	Passed	Passed	0.9400
$\boldsymbol{\beta_1}$	Passed	Passed	0.2181	Passed	Passed	0.5867
$\boldsymbol{\beta_2}$	Passed	Passed	0.1192	Passed	Passed	0.3233
$\boldsymbol{\beta_3}$	Passed	Passed	0.5898	Passed	Passed	0.9776
$\boldsymbol{\beta_4}$	Passed	Passed	0.0895	Passed	Passed	0.3212
$\boldsymbol{\beta_5}$	Passed	Passed	0.5517	Passed	Passed	0.8434
$\boldsymbol{\beta_6}$	Passed	Passed	0.2557	Passed	Passed	0.1572
$\boldsymbol{\delta}$	Passed	Passed	0.7519	Passed	Passed	0.0760
$\boldsymbol{\gamma}$	Passed	Passed	0.8316	Passed	Passed	0.7414
$\boldsymbol{\phi}$	Passed	Passed	0.4136	Passed	Passed	0.4563
$\boldsymbol{\sigma^2}$	Passed	Passed	0.1555	Passed	Passed	0.2064

Table 10

The Heidelberger & Welch (HW) tests (stationary and halfwidth tests) of all unknown parameters in the joint model for semi-informative Prior (Suggested Example 2)

Parameters	Sample Size = 180 Parameters Scheme = 3 Design Structure = 5 &3 (Unbalanced)			Sample Size = 18 Parameters Scheme = 6 Design Structure = 10 &5 (Unbalanced)		
	Stationarity Test	Halfwidth Test	p-value	Stationarity Test	Halfwidth Test	p-value
$\boldsymbol{\alpha}$	Passed	Passed	0.0714	Passed	Passed	0.3702
$\boldsymbol{\beta_0}$	Passed	Passed	0.6735	Passed	Passed	0.7476
$\boldsymbol{\beta_1}$	Passed	Passed	0.5523	Passed	Passed	0.3438
$\boldsymbol{\beta_2}$	Passed	Passed	0.2702	Passed	Passed	0.5194
$\boldsymbol{\beta_3}$	Passed	Passed	0.9932	Passed	Passed	0.8153
$\boldsymbol{\beta_4}$	Passed	Passed	0.2637	Passed	Passed	0.8081
$\boldsymbol{\beta_5}$	Passed	Passed	0.6831	Passed	Passed	0.8865
$\boldsymbol{\beta_6}$	Passed	Passed	0.1272	Passed	Passed	0.3858
$\boldsymbol{\delta}$	Passed	Passed	0.0542	Passed	Passed	0.3631
$\boldsymbol{\gamma}$	Passed	Passed	0.6421	Passed	Passed	0.7623
$\boldsymbol{\phi}$	Passed	Passed	0.3302	Passed	Passed	0.4605
$\boldsymbol{\sigma^2}$	Passed	Passed	0.2585	Passed	Passed	0.1392

Table 11

The Heidelberger & Welch (HW) tests (stationary and halfwidth tests) of all unknown parameters in the joint model for semi-informative Prior (Suggested Example 3)

Parameters	Sample Size = 36 Parameters Scheme = 6 Design Structure = 10 & 5 (Unbalanced)			Sample Size = 180 Parameters Scheme = 2 Design Structure = 20 & 6 (Unbalanced)		
	Stationarity Test	Halfwidth Test	p-value	Stationarity Test	Halfwidth Test	p-value
$\boldsymbol{\alpha}$	Passed	Passed	0.0610	Passed	Passed	0.4275
$\boldsymbol{\beta_0}$	Passed	Passed	0.7919	Passed	Passed	0.2811
$\boldsymbol{\beta_1}$	Passed	Passed	0.5591	Passed	Passed	0.3526
$\boldsymbol{\beta_2}$	Passed	Passed	0.2896	Passed	Passed	0.1803
$\boldsymbol{\beta_3}$	Passed	Passed	0.8834	Passed	Passed	0.5279
$\boldsymbol{\beta_4}$	Passed	Passed	0.2128	Passed	Passed	0.0927
$\boldsymbol{\beta_5}$	Passed	Passed	0.9648	Passed	Passed	0.6303
$\boldsymbol{\beta_6}$	Passed	Passed	0.2391	Passed	Passed	0.3327
$\boldsymbol{\delta}$	Passed	Passed	0.0224	Passed	Passed	0.3724
$\boldsymbol{\gamma}$	Passed	Passed	0.7390	Passed	Passed	0.7704
$\boldsymbol{\phi}$	Passed	Passed	0.4479	Passed	Passed	0.4283
$\boldsymbol{\sigma^2}$	Passed	Passed	0.3329	Passed	Passed	0.1448

Tables 3 through 11 suggest examples randomly of HW outputs for all parameters had complete convergence for all joint models with different samples size, different prior, various parameter schemes, and different design structures. Fail indicates a problem with convergence testing that the simulation may need to run longer. Pass indicates that the sequences have been mixed, and the chain consists of a representative subset.

Another way to see if our chain has converged is to realize how well the chain is moving or mixing around the parameter space. If the chain is taking a long time to move around the parameter space, then it will take longer to converge. We can see how well the chain is mixing through visual inspection for every parameter. In addition, the trace plots is presented to examine any systematic deviation from the steady-state of a converged MCMC chain as well as the autocorrelation plots for unknown parameters.

Some convergence diagnostic methods are required to run multiple chains to inspect convergence, such as Gelman and Rubin (1992) Diagnostic. Thus, the proposed joint model was a rerun, but instead of requesting one chain, I requested three chains that are computed at the same time. The trace-plot of the parameters mean, the density plots and the autocorrelations plots are displayed in Figures from 1 to 12. The three lines represent the three chains that ran parallel but are independent. To determine the convergence has occurred for a model, one should check how the Markov chain is moving around the state space, that is, how well it is mixing. Visible trends or changes in the spread of the trace plot suggest that the stationarity has not been reached yet. It is often said that a good trace plot should look like a hairy caterpillar.

Since there is a huge figure to check the convergence for all parameters in each condition with each replication that cannot be placed in this study, as well, therefore, some examples are randomly proposed for all parameters with a different condition. Note that since the chain was run for long periods of time, most figures for each parameter are very similar. These visual inspections are shown in figures 1 through 12.

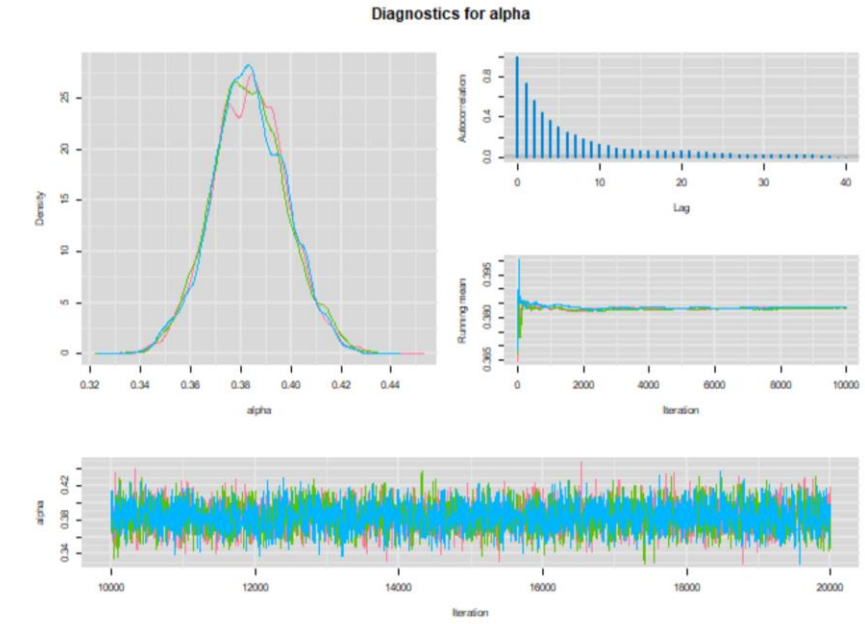

Figure 1. Density, autocorrelation, and trace plots for (α) with three chains of 20000 iterations.

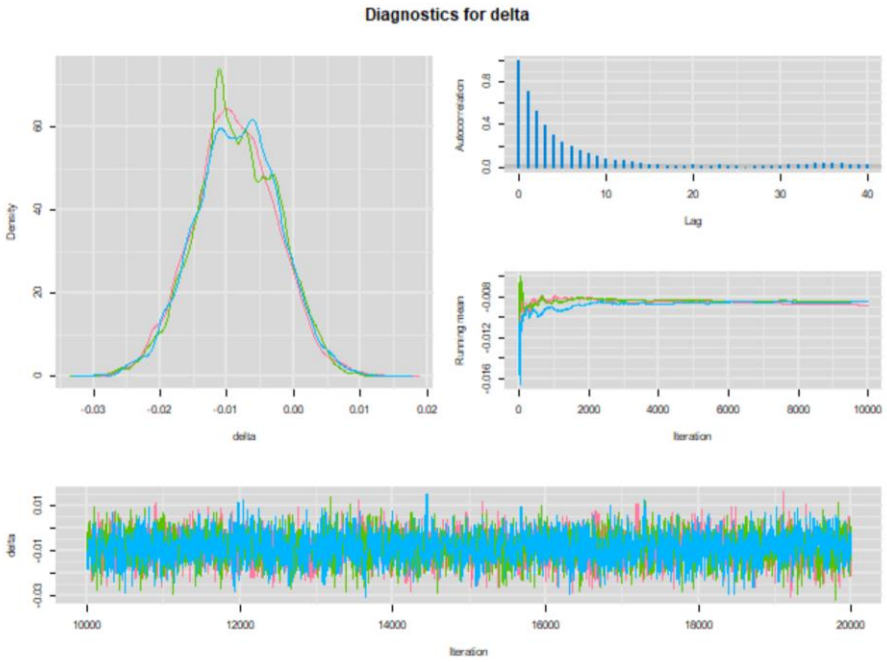

Figure 2. Density, autocorrelation, and trace plots for (δ) with three chains of 20000 iterations.

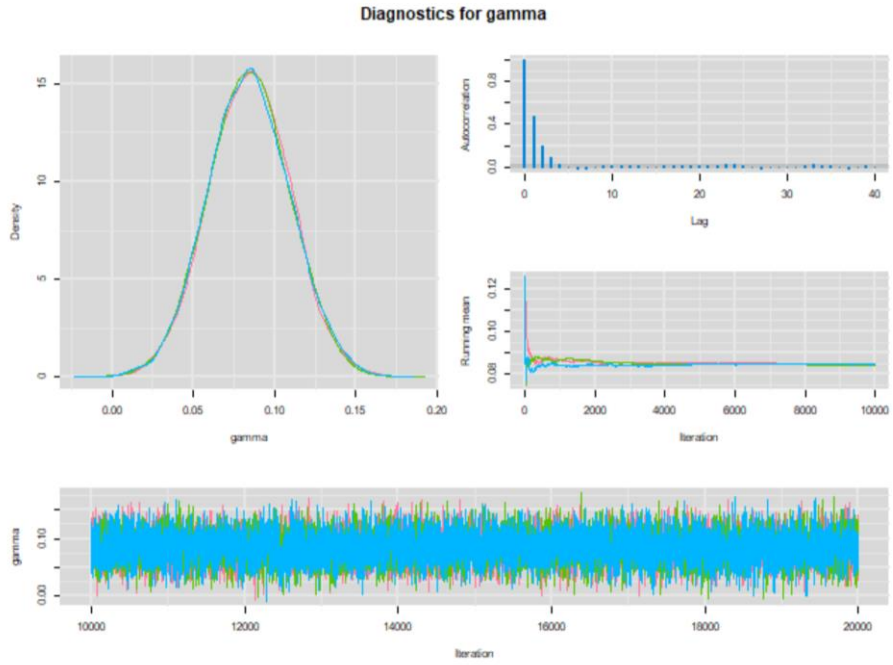

Figure 3. Density, autocorrelation, and trace plots for (γ) with three chains of 20000 iterations.

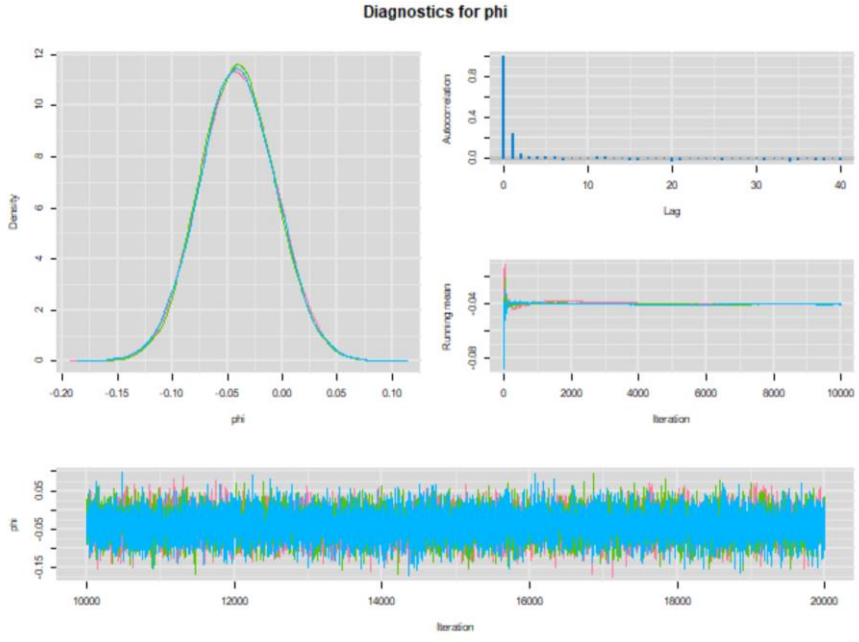

Figure 4. Density, autocorrelation, and trace plots for (ϕ) with three chains of 20000 iterations.

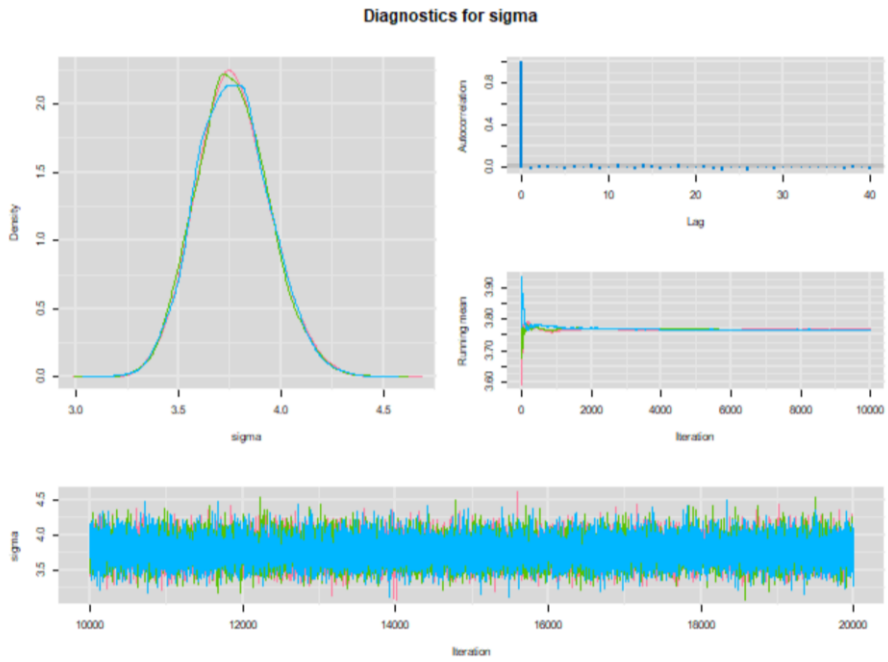

Figure 5. Density, autocorrelation, and trace plots for (σ^2) with three chains of 20000 iterations.

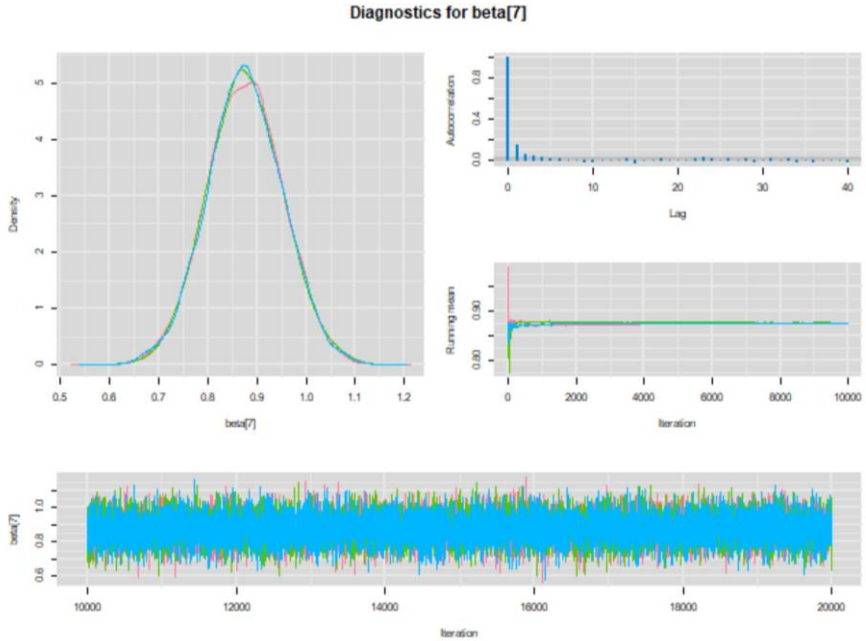

Figure 6. Density, autocorrelation, and trace plots for (β_0) with three chains of 20000 iterations.

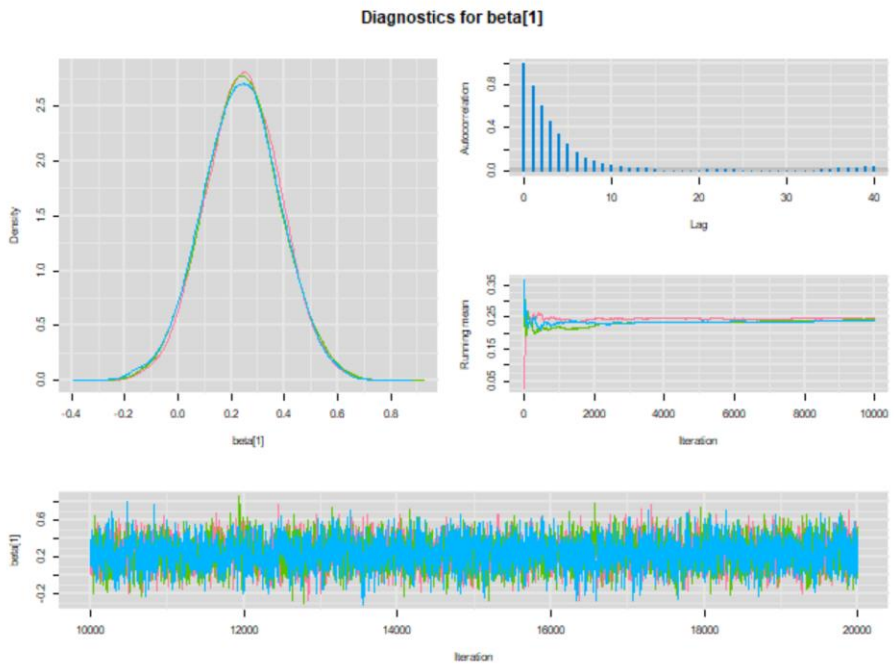

Figure 7. Density, autocorrelation, and trace plots for (β_1) with three chains of 20000 iterations.

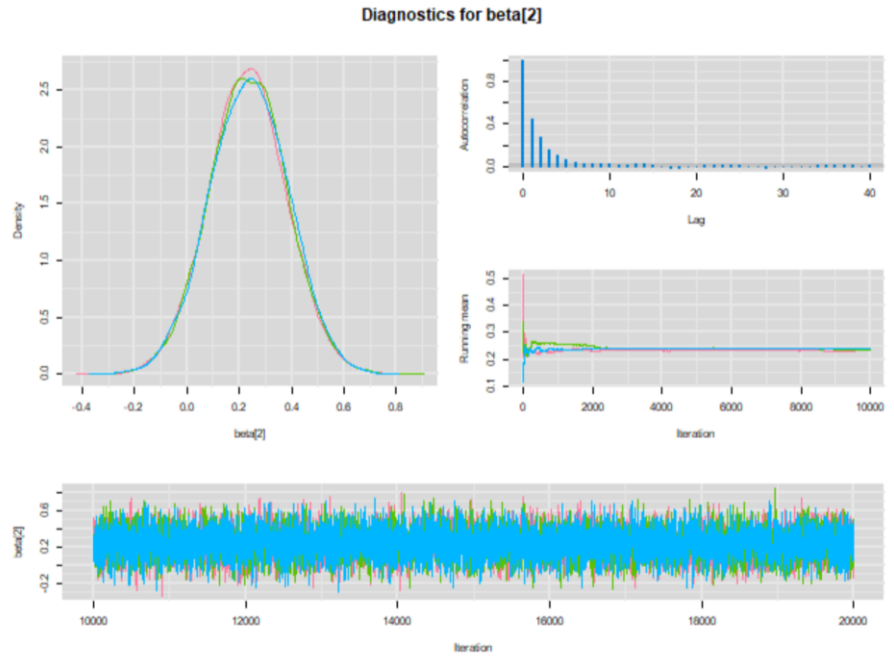

Figure 8. Density, autocorrelation, and trace plots for (β_2) with three chains of 20000 iterations.

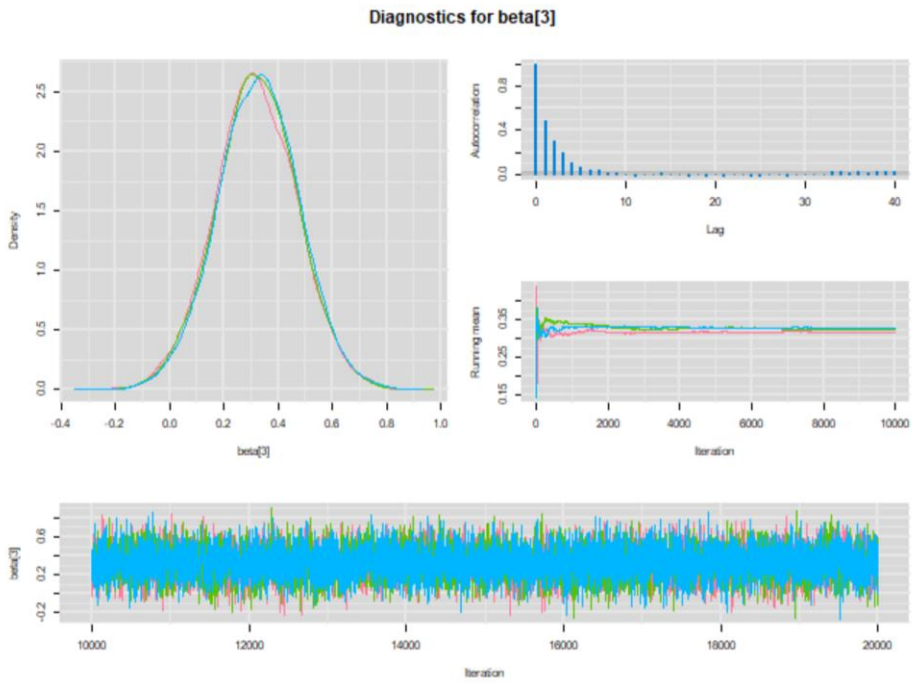

Figure 9. Density, autocorrelation, and trace plots for (β_3) with three chains of 20000 iterations.

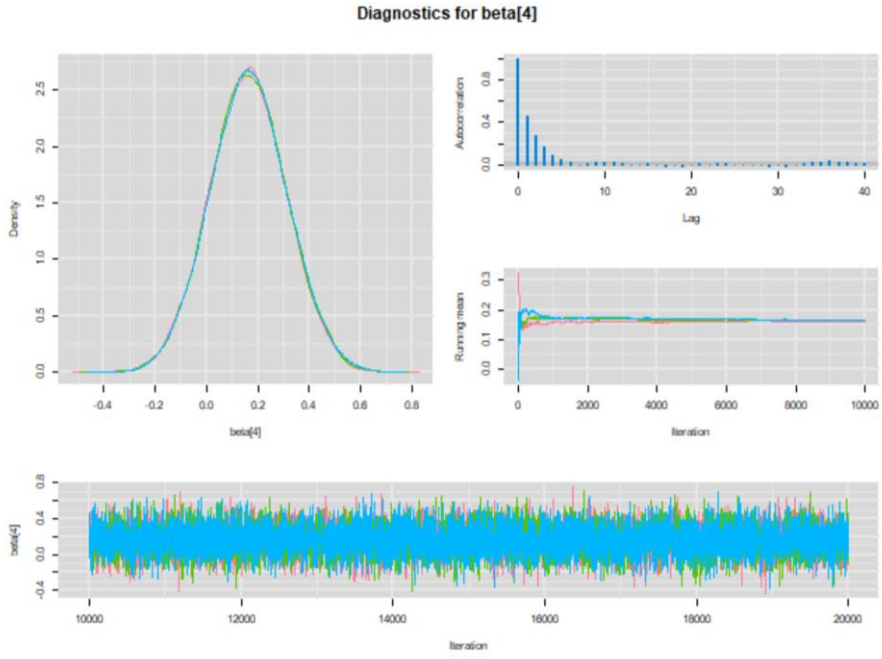

Figure 10. Density, autocorrelation, and trace plots for (β_4) with three chains of 20000 iterations.

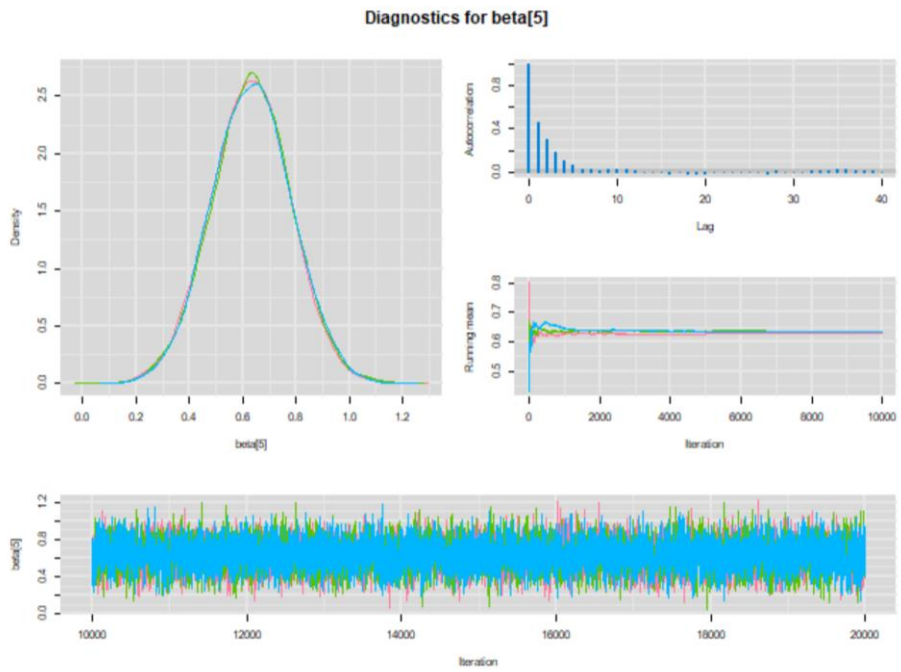

Figure 11. Density, autocorrelation, and trace plots for (β_5) with three chains of 20000 iterations.

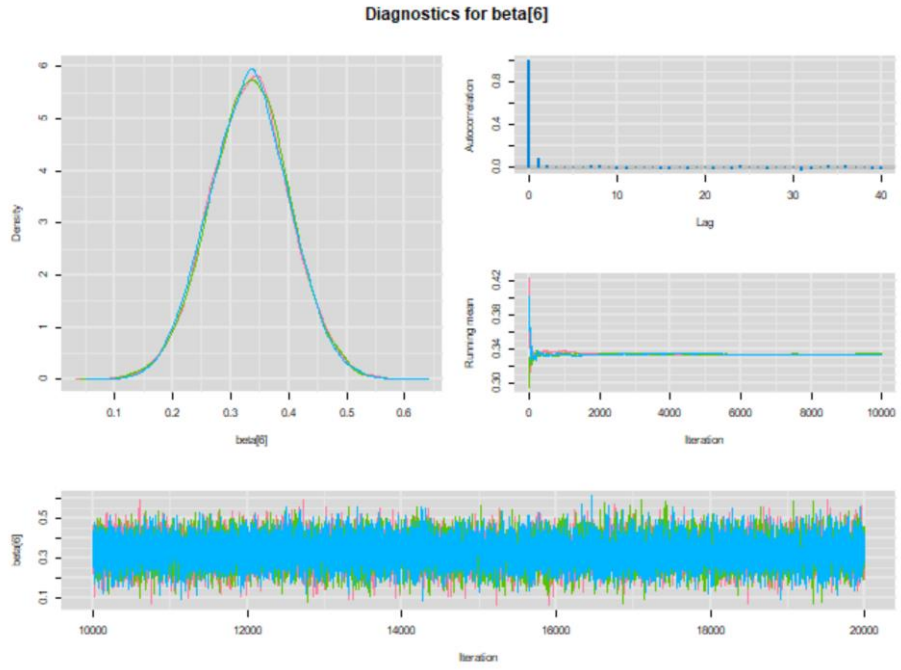

Figure 12. Density, autocorrelation, and trace plots for (β_6) with three chains of 20000 iterations.

We can see a trace plot, which is a plot of the iteration number against the value of the draw of the parameter at each iteration, the density plots, and the autocorrelations plots between the draws of our Markov chain. From the trace plots, we see that the chains mix well. This is also corroborated by the autocorrelation plots. All plots are acceptable and looked satisfactory. However, detecting an unexpected anomaly in MCMC output cannot be noted in all convergence diagnostics. Each diagnostic test is constructed to detect different problems. Hence, some other diagnostics were examined to ensure that convergence has been reached, including the Geweke diagnostic and the Raftery-Lewis diagnostic.

In general, convergence diagnosis for all parameters in the proposed joint model was not a problematic issue in the current project when MCMC was used. All informative, noninformative, and semi-informative priors, based on the design structure, had a similar representative region regarding parameter scheme. Therefore, whether or not the prior information on design structure was known did not affect the HW diagnostic in this case. Also, the sample size did not appear to have an effect on all convergence diagnostics when the simulation had been run long enough.

As can be seen, the more iterations that are used, the higher the accuracy of the histogram, and the better the posterior distribution is approximated. After running enough iterations, MCMC should converge to the posterior distribution of interest. Typically, the more parameters that are needed to be estimated, the more iterations are required (Brooks & Roberts, 1998). Therefore, it would have been necessary for the application of Markov chain simulation to run around 10,000 iterations or more to ensure the accuracy of distributions for the required parameters. In this dissertation, I ran both one and three chains with 20,000 iterations in each chain, which passed the convergence testing.

Since the key purpose of the Markov chain simulation is to create a specific posterior or stationary distribution of the unknown parameters, it was necessary to verify the convergence value of the simulated sequences when the simulation had been applied. The results indicated that the distributions of the current draws were close to the posterior distribution. The convergence, in turn, allowed the inferential statistics to calculate the parameters of interest which are addressed in this chapter. Appendix A shows the details of implementation in the computer language of R.

The Performance of the Estimation

In order to answer the first and second research questions, the proposed joint model, as shown in Equation 37, was applied to simulate data sets to demonstrate the use of a Bayesian method on the estimation for all unknown parameters with three different prior distribution. Conditions set included different levels of sample size, parameter schemes, and different numbers of observations (design structure) (see Table 2). Three statistics, defined earlier, based on converged simulation of posterior sampling data sets, were considered. First, the average estimate was obtained. Second, standard deviations were calculated for each parameter. Third, the average lower and upper limits of the 97.5% confidence intervals were obtained.

Since there are 120 (5 sample size * 4 levels of observation * 6 parameter schemes) simulation conditions, which makes 360 simulation designs in total for all three prior distributions to make statistical inference for 12 unknown parameters, that cannot be placed in this study. Therefore, the selected results for both different design structures (Balanced and Unbalanced) with five different sample sizes are presented in Tables 12 through 41. Each table shows the performance of the Bayesian method of estimation for all unknown parameters based on four selected parameter schemes (1, 3, 4, 6) with three different prior

distribution (noninformative, informative, and semi-informative). Note that the true values of unknown parameters are dependent on the parameter schemes and consistent with previous research (Alomair, 2017; Bronsert, 2009; Lin, 2011; Seo, 2015).

Tables 12 through 17 provide the results for the joint model with four parameters schemes (1, 3, 4, 6) and both different design structures (Balanced and Unbalanced) at sample size 18, including three prior distribution (noninformative, informative, and semi-informative), respectively. Tables 18 through 23 show the results for the joint model with four parameters schemes (1, 3, 4, 6) and both different design structures (Balanced and Unbalanced) at sample size 36, including three prior distribution, respectively. Also, Tables 24 through 29 provide the results for the joint model with the same conditions at sample size 54, and Tables 30 through 35 give the results for the joint model with the same conditions as well at sample size 90, including three prior distribution, respectively. Finally, Tables 36 through 41 provide the results for the joint model with the same conditions at sample size 180, including three prior distribution, respectively. Values in the table mentioned above represent the estimation of the numbers of unknown parameters with 97.5% lower and upper confidence intervals, including the standard deviation for all selected conditions at all sample sizes.

The detailed summary results from the true model with parameter scheme 1 and the true values of parameters ($\alpha = 2,\ \beta_0 = 0.4,\ \beta_1 = 0.2,\ \beta_2 = 0.3,\ \beta_3 = 0.1,\ \beta_4 = 0.3,\ \beta_5 = 0.4,\ \beta_6 = 0.9,\ \delta = 0.01,\ \gamma = 0.1,\ \phi = 0.8,\ \sigma^2 = 1$) for both design structure and sample size of N = 18 was obtained from Tables 12, 14 and 16, N = 36 was obtained from Tables 18, 20 and 22, N = 54 was obtained from Tables 24, 26 and 28, N = 90 was obtained from Tables 30, 32 and 34, and N = 180 was obtained from Tables 36, 38, and 40. All simulation samples passed the convergence as specified earlier. The findings for

parameter scheme 1 varied across all sample sizes with both design structures and prior information toward the unknown parameter α. Also, the unknown parameter δ showed a negative estimate. However, the rest of the parameters were not markedly different. Interestingly, at all sample sizes with both design structures, including all informative, noninformative, semi-informative priors, the true value of α, which is 2, was not contained within the 97.5% confidence interval of the estimation. For the rest of the parameters, the Bayesian method was estimated accurately, or at least, the true values of all parameters except α were contained within the 97.5% confidence interval of the estimation for some conditions. For parameter scheme 1, all parameters except α and δ achieved adequate accuracy in cases when a noninformative prior was applied. For example, with N = 18 based on a noninformative prior as reported in Tables 12, the point estimation for γ was 0.100 (97.5% confidence interval [CI] = 0.080 – 0.121), for ϕ was 0.787 (97.5% confidence interval [CI] = .0737 – 0.837), and for σ^2 was 1.006 (97.5% confidence interval [CI] = 0.815 – 1.241). When N = 90, γ was 0.100 (97.5% confidence interval [CI] = 0.094 – 0.106), for ϕ was 0.800 (97.5% confidence interval [CI] = .0783 – 0.816), and for σ^2 was 1.001 (97.5% confidence interval [CI] = 0.939 – 1.069) (see Tables 23 and 24). the results yielded the narrowest interval. However, when N = 180, the range of the 97.5% confidence interval became wider, which is unexpected.

The results from the true model with parameter scheme 3, which the true values of parameters are $(\alpha = 2,\ \beta_0 = 0.4,\ \beta_1 = 0.2,\ \beta_2 = 0.3,\ \beta_3 = 0.1,\ \beta_4 = 0.3,\ \beta_5 = 0.4,\ \beta_6 = 0.9,\ \delta = 0.01,\ \gamma = 0.1,\ \phi = 0.8,\ \sigma^2 = 2)$ for both design structure and sample size of N = 18 was obtained from Tables 12, 14 and 16, N = 36 was obtained from Tables 18, 20 and 22, N = 54 was obtained from Tables 24, 26 and 28, N = 90 was obtained from Tables 30, 32 and 34, and N = 180 was obtained from Tables 36, 38, and 40. All simulation

samples passed the convergence as specified earlier. For all prior distribution with both design structures, all estimations for all parameters except α and σ^2 fall in the range of the 97.5% confidence interval. When comparing the performance across sample sizes, with a sample size of 180, the results yielded the narrowest interval. In some conditions, the results of the big sample size are narrower range than the small sample size. This occurs throughout the results where parameter scheme 3 is set. The results theoretically confirmed that larger samples tend to give narrower confidence intervals for the estimation of parameters than that of smaller samples, which leads to more precise estimates. However, some results of the simulation conditions did not achieve that claim. This is one advantage of the Bayesian approach, which may provide more accurate results as they can deal with small data set and asymmetric distributions and/or the whole distribution (Gill, 2008; Muthén & Asparouhov, 2012).

As shown in Tables 14 with the true value of all parameters, at sample size of N = 18 for informative prior, for balance design structure (10 & 10) the average estimate for parameters α, for example, was reported to be 0.138 (97.5% CI = 0.109 – 0.171), β_1 = 0.288 (97.5% CI = -0.287 – 0.873), β_6 = 0.825 (97.5% CI = 0.416 – 1.230), γ = 0.100 (97.5% CI = 0.054 – 0.146), ϕ = 0.777 (97.5% CI = 0.657 – 0.898), and σ^2 = 4.017 (97.5% CI = 3.167 – 5.079). Also, the average estimation for unbalance design structure (20 & 6) at the same parameter scheme 3, sample size (N = 18), and prior (informative) was reported for the parameters, α = 0.145 (97.5% CI = 0.115 – 0.182), β_1 = 0.274 (97.5% CI = -0.238 – 0.797), β_6 = 0.888 (97.5% CI = 0.623 – 1.150), γ = 0.100 (97.5% CI = 0.063 – 0.137), ϕ = 0.785 (97.5% CI = 0.689 – 0.882), and σ^2 = 4.004 (97.5% CI = 3.311 – 4.849). As for results of the estimation related to noninformative priors at the same parameter scheme 3, and sample size (N = 18), the average estimate at the balanced design

structure (10 & 10) for parameters α, for example, was reported to be 0.248 (97.5% CI = 0.363 – 0.144), β_1 = 0.226 (97.5% CI = -0.516 – 0.974), β_6 = 0.984 (97.5% CI = 0.531 – 1.431), γ = 0.100 (97.5% CI = 0.060 – 0.141), ϕ = 0.765 (97.5% CI = 0.685 – 0.846), and σ^2 = 4.039 (97.5% CI = 3.270 – 4.985). Also, the average estimation for unbalance design structure (20 & 6) at the same parameter scheme 3, sample size (N = 18), and prior (noninformative) was reported for the parameters, α = 0.153 (97.5% CI = 0.119 – 0.212), β_1 = 0.214 (97.5% CI = -0.296 – 0.727), β_6 = 0.947 (97.5% CI = 0.661 – 1.235), γ = 0.101 (97.5% CI = 0.073 – 0.129), ϕ = 0.784 (97.5% CI = 0.729 – 0.839), and σ^2 = 4.030 (97.5% CI = 3.484 – 4.671). In addition, the estimation related to semi-informative priors at the same parameter scheme 3, and sample size (N = 18), the average estimate at balance design structure (10 & 10) for parameters α was reported to be 0.836 (97.5% CI = 0.771 – 0.910), β_1 = 0.239 (97.5% CI = -0.094 – 0.580), β_6 = 0.894 (97.5% CI = 0.689 – 1.098), γ = 0.100 (97.5% CI = 0.080 – 0.120), ϕ = 0.792 (97.5% CI = 0.750 – 0.835), and σ^2 = 4.007 (97.5% CI = 3.605 – 4.457). Also, the average estimation for unbalance design structure (20 & 6) at the same parameter scheme 3, sample size (N = 18), and prior (noninformative) was reported for the parameters, α = 0.885 (97.5% CI = 0.858 – 0.913), β_1 = 0.208 (97.5% CI = -0.035 – 0.455), β_6 = 0.905 (97.5% CI = 0.789 – 1.021), γ = 0.100 (97.5% CI = 0.086 – 0.115), ϕ = 0.796 (97.5% CI = 0.766 – 0.825), and σ^2 = 3.992 (97.5% CI = 3.707 – 4.707). For all prior distributions, when comparing the performance between both design structures, the results for all parameters except α were similar between them. Overall, the design structures, balanced or unbalanced, do not seem to affect the results of average estimation.

As shown in Tables 18, 20, and 22, the estimation in terms of 97.5% confidence interval for all parameters were reported for all noninformative, informative, and semi-

informative priors at the same parameter scheme 3, with sample size 36. For informative prior, the average estimation of parameters, α = 0.137 (97.5% CI = 0.120 – 0.155), β_1 = 0.213 (97.5% CI = -0.289 – 0.725), β_6 = 0.917 (97.5% CI = 0.697 – 1.137), γ = 0.101 (97.5% CI = 0.071 – 0.130), ϕ = 0.785 (97.5% CI = 0.708 – 0.861), and σ^2 = 4.030 (97.5% CI = 3.483 – 4.670) (see Tables 20). Regarding correct estimation in terms of the 95% confidence interval, the pattern was similar to that for noninformative priors. the average estimation of parameters, α = 0.148 (97.5% CI = 0.119 – 0.197), β_1 = 0.262 (97.5% CI = -0.161 – 0.700), β_6 = 0.885 (97.5% CI = 0.610 – 1.158), γ = 0.100 (97.5% CI = 0.073 – 0.128), ϕ = 0.789 (97.5% CI = 0.736 – 0.843), and σ^2 = 4.017 (97.5% CI = 3.473 – 4.652) (see Tables 18). Regarding correct estimation in terms of the 95% confidence interval except α and σ^2, the pattern was similar to that for semi-informative priors. The average estimation of parameters, α = 0.924 (97.5% CI = 0.902 – 0.946), β_1 = 0.220 (97.5% CI = -0.029 – 0.468), β_6 = 0.902 (97.5% CI = 0.778 – 1.027), γ = 0.100 (97.5% CI = 0.086 – 0.114), ϕ = 0.797 (97.5% CI = 0.769 – 0.826), and σ^2 = 4.009 (97.5% CI = 3.723 – 4.320) (see Tables 22).

Also, for the proposed joint model with the true values of parameter scheme 3 for sample size 54, the pattern of correct estimation for the all parameters except α and σ^2 in terms of 95% confidence interval was similar to the estimation for other sample sizes (see Tables 24, 26, and 28). When the simulation was run based on sample size 54 with the same previously specified conditions (parameter scheme 3), for informative prior, the average estimation of parameters are, α = 0.603 (97.5% CI = 0.481 – 0.738), β_1 = 0.238 (97.5% CI = -0.085 – 0.563), β_6 = 0.901 (97.5% CI = 0.736 – 1.066), γ = 0.100 (97.5% CI = 0.080 – 0.121), ϕ = 0.792 (97.5% CI = 0.7739 – 0.845), and σ^2 = 4.011 (97.5% CI = 3.620 – 4.446), (see Table 26). There were correct estimations on the parameters for

scheme 3 except α and σ^2 in the form of 95% confidence interval for noninformative priors. For example, the average estimation of parameters, α = 0.767 (97.5% CI = 0.693 – 0.842), β_1 = 0.237 (97.5% CI = -0.126 – 0.604), β_6 = 0.901 (97.5% CI = 0.708 – 1.090), γ = 0.101 (97.5% CI = 0.078 – 0.123), ϕ = 0.793 (97.5% CI = 0.751 – 0.836), and σ^2 = 4.019 (97.5% CI = 3.569 – 4.528), (see Table 24). Regarding correct estimation of α and σ^2 in terms of the 95% confidence interval, the pattern was similar to that for semi-informative priors. The average estimations of parameters, α = 0.625 (97.5% CI = 0.515 – 0.745), β_1 = 0.207 (97.5% CI = -0.257 – 0.680), β_6 = 0.935 (97.5% CI = 0.671 – 1.200), γ = 0.101 (97.5% CI = 0.077 – 0.125), ϕ = 0.786 (97.5% CI = 0.740 – 0.832), and σ^2 = 4.030 (97.5% CI = 3.545 – 4.584) (see Table 28).

Additionally, when the simulation was run based on sample size 90 with the same previously specified conditions (parameter scheme 3), for informative prior, the average estimation of parameters, α = 0.138 (97.5% CI = 0.111 – 0.169), β_1 = 0.230 (97.5% CI = -0.457 – 0.914), β_6 = 0.881 (97.5% CI = 0.496 – 1.258), γ = 0.100 (97.5% CI = 0.054 – 0.145), ϕ = 0.776 (97.5% CI = 0.675 – 0.877), and σ^2 = 3.973 (97.5% CI = 3.102 – 5.066), (see Table 32). Also, in terms of the 95% confidence interval, the pattern was similar to that for noninformative priors For example, the average estimation of parameters, α = 0.944 (97.5% CI = 0.930 – 0.957), β_1 = 0.207 (97.5% CI = -0.021 – 0.434), β_6 = 0.909 (97.5% CI = 0.796 – 1.021), γ = 0.100 (97.5% CI = 0.088 – 0.112), ϕ = 0.798 (97.5% CI = 0.773 – 0.822), and σ^2 = 4.006 (97.5% CI = 3.756 – 4.275), (see Table 30). Regarding correct estimation in terms of the 95% confidence interval, the pattern was similar to that for semi-informative priors. the average estimation of parameters, α = 0.873 (97.5% CI = 0.824 – 0.924), β_1 = 0.241 (97.5% CI = -0.098 – 0.589), β_6 = 0.906 (97.5% CI = 0.691 –

1.122), γ = 0.100 (97.5% CI = 0.080 – 0.119), ϕ = 0.794 (97.5% CI = 0.758 – 0.830), and σ^2 = 4004 (97.5% CI = 3.611 – 4.445), (see Table 34).

Finally, when the simulation was run based on sample size 180 with the same previously specified conditions (parameter scheme 3), for informative prior, the average estimation of parameters, α = 0.157 (97.5% CI = 0.21 – 0.220), β_1 = 0.206 (97.5% CI = -0.273 – 0.685), β_6 = 0.925 (97.5% CI = 0.655 – 1.188), γ = 0.101 (97.5% CI = 0.074 – 0.127), ϕ = 0.787 (97.5% CI = 0.735 – 0.840), and σ^2 = 4.016 (97.5% CI = 3.498 – 4.610), (see Table 38). Also, in terms of the 97.5% confidence interval, the pattern was similar to that for noninformative priors For example, the average estimation of parameters, α = 0.144 (97.5% CI = 0.099 – 0.198), β_1 = 0.259 (97.5% CI = -0.466 – 0.982), β_6 = 0.812 (97.5% CI = 0.297 – 1.311), γ = 0.101 (97.5% CI = 0.034 – 0.168), ϕ = 0.770 (97.5% CI = 0.602 – 0.941), and σ^2 = 4.011 (97.5% CI = 2.842 – 5.618), (see Table 36). Regarding correct estimation except α and σ^2 in terms of the 95% confidence interval, the pattern was similar to that for semi-informative priors. The average estimation of parameters, α = 0.914 (97.5% CI = 0.896 – 0.930), β_1 = 0.224 (97.5% CI = -0.043 – 0.493), β_6 = 0.905 (97.5% CI = 0.763 – 1.046), γ = 0.100 (97.5% CI = 0.085 – 0.115), ϕ = 0.796 (97.5% CI = 0.769 – 0.823), and σ^2 = 4.013 (97.5% CI = 3.706 – 4.348), (see Table 40). In this process, the results of average estimation for unknown parameters is not stable at different level of sample size with different prior distributions. For example, some results of small sample size are narrower in range than the big sample size. Also, in term of the different prior, the semi-informative prior sometimes gives a narrower range than the others, and sometimes noninformative prior gives a narrower range than the others.

In addition to the above, the results from the true model with parameter scheme 4, which the true values of parameters are ($\alpha = 1,\ \beta_0 = 0.4,\ \beta_1 = 0.2,\ \beta_2 = 0.3,\ \beta_3 = 0.1,$

$\beta_4 = 0.3,\ \beta_5 = 0.4,\ \beta_6 = 0.9,\ \delta = 0.01,\ \gamma = 0.1,\ \phi = 0,\ \sigma^2 = 2$) and the results from the true model with parameter scheme 6, which the true values of parameters are ($\alpha = 1,\ \beta_0 = 0.4,\ \beta_1 = 0.2,\ \beta_2 = 0.3,\ \beta_3 = 0.1,\ \beta_4 = 0.3,\ \beta_5 = 0.4,\ \beta_6 = 0.9,\ \delta = 0.02,\ \gamma = 0.1,\ \phi = 0.8,\ \sigma^2 = 0.5$) for both design structure and sample size of N = 18 was obtained from Tables 13, 15 and 17, N = 36 was obtained from Tables 19, 21 and 23, N = 54 was obtained from Tables 25, 27 and 29, N = 90 was obtained from Tables 31, 33 and 35, and N = 180 was obtained from Tables 37, 39, and 41. The results for both parameter scheme 4, 6 with the same preconditions that were previously discussed, were somewhat similar to the previous one. Overall, in terms of the different schemes, some of the schemes give better results than the other, which can indicate that fitting the right parameters can help to have the best results.

The comparison of all five-level of sample size received with different schemes and different prior (noninformative, informative, and semi-informative) produced the following results. The prior information concerning the unknown parameter of interest sometimes affected the estimation in that when informative priors were used; it was more likely to obtain a narrower confidence interval than the other priors. The narrower the interval indicates the more precise estimate for the parameter. Therefore, based on the different parameter schemes, representing the 97.5 % confidence interval included the true value of the parameters for each study conducted, with values falling somewhere between the lower and upper bound of the estimation. The performance of the estimates, when all parameter schemes were used, was also confirmed by the HW diagnostic test, which showed convergence for all conditions on all informative, noninformative, semi-informative priors. The results indicated that the estimation of parameters was accurate except α and σ^2.

Table 12

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 18, Two of Design Structure, and four of Parameter Scheme for noninformative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	SD	Upper	Lower	Est.	SD	Upper	Lower
Parameters Scheme = 1								
α	0.306	0.094	0.452	0.167	0.160	0.035	0.237	0.120
β_0	0.436	0.192	0.822	0.062	0.416	0.135	0.686	0.153
β_1	0.223	0.187	0.595	-0.1465	0.208	0.129	0.465	-0.047
β_2	0.325	0.200	0.711	-0.068	0.317	0.135	0.578	0.053
β_3	0.093	0.202	0.484	-0.300	0.092	0.139	0.367	-0.188
β_4	0.300	0.208	0.696	-0.112	0.297	0.132	0.555	0.036
β_5	0.415	0.091	0.596	0.237	0.406	0.062	0.529	0.286
β_6	0.936	0.120	1.170	0.698	0.916	0.078	1.067	0.762
δ	-0.014	0.008	-0.001	-0.026	-0.003	0.003	0.002	-0.010
γ	0.100	0.010	0.121	0.080	0.100	0.007	0.114	0.086
ϕ	0.787	0.025	0.837	0.737	0.794	0.018	0.829	0.759
σ^2	1.006	0.109	1.241	0.815	1.011	0.076	1.171	0.874
Parameters Scheme = 3								
α	0.248	0.071	0.363	0.144	0.153	0.027	0.212	0.119
β_0	0.502	0.379	1.265	-0.237	0.437	0.264	0.964	-0.078
β_1	0.226	0.377	0.974	-0.516	0.214	0.259	0.727	-0.296
β_2	0.339	0.403	1.115	-0.449	0.321	0.269	0.841	-0.205
β_3	0.091	0.407	0.878	-0.698	0.100	0.278	0.650	-0.458
β_4	0.313	0.416	1.108	-0.513	0.308	0.263	0.823	-0.214
β_5	0.436	0.180	0.793	0.083	0.416	0.122	0.657	0.181
β_6	0.984	0.229	1.431	0.531	0.947	0.146	1.235	0.661
δ	-0.008	0.006	0.001	-0.018	-0.002	0.003	0.002	-0.007
γ	0.100	0.020	0.141	0.060	0.101	0.014	0.129	0.073
ϕ	0.765	0.041	0.846	0.685	0.784	0.028	0.839	0.729
σ^2	4.039	0.437	4.985	3.270	4.030	0.302	4.671	3.484

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 13

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 18, Two of Design Structure, and four of Parameter Scheme for noninformative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	SD	Upper	Lower	Est.	SD	Upper	Lower
Parameters Scheme = 4								
α	0.375	0.033	0.442	0.313	0.374	0.023	0.420	0.330
β_0	0.433	0.368	1.174	-0.283	0.403	0.256	0.910	-0.095
β_1	0.227	0.374	0.968	-0.508	0.224	0.258	0.734	-0.283
β_2	0.316	0.399	1.084	-0.465	0.324	0.267	0.841	-0.200
β_3	0.096	0.404	0.881	-0.685	0.079	0.277	0.627	-0.476
β_4	0.285	0.413	1.074	-0.534	0.291	0.262	0.801	-0.228
β_5	0.413	0.176	0.761	0.068	0.411	0.119	0.646	0.182
β_6	0.926	0.214	1.348	0.504	0.919	0.136	1.188	0.653
δ	-0.006	0.013	0.018	-0.031	-0.007	0.009	0.010	-0.025
γ	0.097	0.056	0.207	-0.014	0.100	0.039	0.178	0.023
ϕ	-0.037	0.080	0.121	-0.194	-0.016	0.056	0.093	-0.125
σ^2	4.035	0.436	4.981	3.265	4.024	0.302	4.664	3.477
Parameters Scheme = 6								
α	0.372	0.040	0.452	0.298	0.371	0.027	0.426	0.320
β_0	0.411	0.095	0.600	0.227	0.404	0.066	0.536	0.275
β_1	0.205	0.094	0.392	0.020	0.207	0.065	0.335	0.079
β_2	0.306	0.101	0.501	0.108	0.306	0.067	0.436	0.173
β_3	0.102	0.101	0.299	-0.095	0.100	0.069	0.237	-0.039
β_4	0.305	0.104	0.504	0.098	0.301	0.066	0.431	0.170
β_5	0.409	0.048	0.504	0.315	0.405	0.033	0.470	0.343
β_6	0.919	0.068	1.052	0.783	0.909	0.045	0.995	0.821
δ	-0.006	0.008	0.009	-0.021	-0.007	0.005	0.003	-0.017
γ	0.100	0.014	0.127	0.073	0.101	0.010	0.119	0.082
ϕ	0.793	0.018	0.829	0.757	0.796	0.012	0.821	0.772
σ^2	0.253	0.027	0.312	0.205	0.252	0.019	0.292	0.218

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 14

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 18, Two of Design Structure, and four of Parameter Scheme for informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.137	0.017	0.171	0.106	0.146	0.021	0.187	0.113
β_0	0.359	0.176	0.703	0.017	0.383	0.149	0.678	0.091
β_1	0.245	0.183	0.607	-0.108	0.225	0.155	0.531	-0.075
β_2	0.339	0.195	0.718	-0.042	0.313	0.157	0.614	0.002
β_3	0.158	0.191	0.535	-0.219	0.140	0.155	0.445	-0.169
β_4	0.342	0.187	0.708	-0.025	0.319	0.156	0.622	0.015
β_5	0.406	0.086	0.576	0.240	0.397	0.073	0.542	0.253
β_6	0.889	0.118	1.118	0.658	0.899	0.075	1.046	0.751
δ	0.000	0.004	0.009	-0.009	-0.001	0.004	0.006	-0.009
γ	0.101	0.012	0.124	0.077	0.100	0.010	0.119	0.080
ϕ	0.787	0.041	0.868	0.707	0.793	0.033	0.857	0.728
σ^2	1.006	0.122	1.274	0.792	1.005	0.099	1.218	0.831
Parameters Scheme = 3								
α	0.138	0.016	0.171	0.109	0.145	0.019	0.182	0.115
β_0	0.321	0.272	0.854	-0.210	0.367	0.244	0.845	-0.112
β_1	0.288	0.296	0.873	-0.287	0.274	0.264	0.797	-0.238
β_2	0.404	0.310	1.016	-0.207	0.334	0.267	0.853	-0.198
β_3	0.246	0.308	0.854	-0.360	0.185	0.264	0.706	-0.335
β_4	0.347	0.303	0.939	-0.248	0.358	0.265	0.874	-0.161
β_5	0.410	0.162	0.730	0.098	0.402	0.140	0.680	0.127
β_6	0.825	0.206	1.230	0.416	0.888	0.134	1.150	0.623
δ	-0.001	0.004	0.007	-0.007	-0.002	0.003	0.005	-0.008
γ	0.100	0.024	0.146	0.054	0.100	0.019	0.137	0.063
ϕ	0.777	0.061	0.898	0.657	0.785	0.049	0.882	0.689
σ^2	4.017	0.486	5.079	3.167	4.004	0.393	4.849	3.311

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 15

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 18, Two of Design Structure, and four of Parameter Scheme for informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.380	0.040	0.463	0.306	0.375	0.032	0.441	0.315
β_0	0.308	0.270	0.834	-0.220	0.343	0.243	0.818	-0.134
β_1	0.289	0.296	0.872	-0.285	0.272	0.264	0.794	-0.240
β_2	0.389	0.309	0.999	-0.219	0.328	0.267	0.849	-0.204
β_3	0.254	0.307	0.860	-0.350	0.194	0.264	0.715	-0.326
β_4	0.341	0.302	0.932	-0.253	0.360	0.265	0.876	-0.158
β_5	0.409	0.160	0.725	0.102	0.392	0.140	0.671	0.117
β_6	0.805	0.202	1.203	0.404	0.881	0.129	1.134	0.626
δ	-0.007	0.015	0.022	-0.037	-0.006	0.012	0.017	-0.030
γ	0.099	0.063	0.222	-0.024	0.098	0.052	0.200	-0.004
ϕ	-0.027	0.094	0.159	-0.211	-0.019	0.076	0.131	-0.169
σ^2	4.009	0.485	5.067	3.160	4.021	0.395	4.868	3.326
Parameters Scheme = 6								
α	0.373	0.043	0.462	0.294	0.372	0.035	0.444	0.307
β_0	0.384	0.098	0.577	0.193	0.393	0.080	0.553	0.236
β_1	0.217	0.100	0.415	0.023	0.213	0.082	0.376	0.053
β_2	0.313	0.108	0.521	0.102	0.309	0.084	0.469	0.145
β_3	0.119	0.104	0.324	-0.088	0.108	0.082	0.269	-0.056
β_4	0.311	0.101	0.510	0.109	0.305	0.083	0.466	0.143
β_5	0.401	0.045	0.490	0.313	0.401	0.038	0.476	0.327
β_6	0.898	0.066	1.026	0.768	0.904	0.043	0.988	0.819
δ	-0.003	0.013	0.024	-0.028	-0.005	0.011	0.016	-0.026
γ	0.100	0.017	0.132	0.068	0.101	0.013	0.127	0.075
ϕ	0.795	0.028	0.852	0.739	0.796	0.023	0.841	0.751
σ^2	0.255	0.031	0.322	0.201	0.254	0.025	0.307	0.209

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 16

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 18, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.874	0.039	0.936	0.819	0.866	0.015	0.889	0.843
β_0	0.398	0.092	0.580	0.219	0.396	0.066	0.526	0.270
β_1	0.211	0.092	0.394	0.031	0.212	0.065	0.341	0.085
β_2	0.311	0.093	0.492	0.130	0.309	0.066	0.437	0.181
β_3	0.109	0.092	0.291	-0.072	0.102	0.065	0.230	-0.027
β_4	0.306	0.093	0.488	0.122	0.308	0.066	0.438	0.179
β_5	0.402	0.041	0.482	0.324	0.402	0.032	0.465	0.340
β_6	0.898	0.056	1.008	0.788	0.902	0.032	0.965	0.838
δ	-0.053	0.002	-0.050	-0.057	-0.050	0.001	-0.049	-0.052
γ	0.100	0.005	0.110	0.090	0.100	0.004	0.107	0.093
ϕ	0.797	0.014	0.824	0.768	0.798	0.010	0.818	0.779
σ^2	1.003	0.055	1.116	0.903	1.001	0.038	1.078	0.929
Parameters Scheme = 3								
α	0.836	0.046	0.910	0.771	0.885	0.018	0.913	0.858
β_0	0.379	0.166	0.705	0.058	0.394	0.123	0.637	0.155
β_1	0.239	0.172	0.580	-0.094	0.208	0.124	0.455	-0.035
β_2	0.337	0.174	0.675	-0.003	0.317	0.127	0.563	0.071
β_3	0.136	0.172	0.478	-0.198	0.113	0.125	0.361	-0.132
β_4	0.324	0.174	0.666	-0.021	0.317	0.127	0.567	0.067
β_5	0.405	0.077	0.558	0.254	0.403	0.063	0.526	0.282
β_6	0.894	0.104	1.098	0.689	0.905	0.059	1.021	0.789
δ	-0.044	0.003	-0.041	-0.048	-0.045	0.001	-0.044	-0.046
γ	0.100	0.010	0.120	0.080	0.100	0.007	0.115	0.086
ϕ	0.792	0.022	0.835	0.750	0.796	0.015	0.825	0.766
σ^2	4.007	0.218	4.457	3.605	3.992	0.153	4.301	3.707

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 17

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 18, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.371	0.016	0.404	0.339	0.370	0.011	0.393	0.348
β_0	0.354	0.162	0.671	0.040	0.383	0.121	0.621	0.149
β_1	0.246	0.171	0.585	-0.086	0.210	0.124	0.457	-0.033
β_2	0.344	0.173	0.681	0.005	0.305	0.126	0.550	0.059
β_3	0.135	0.171	0.475	-0.199	0.121	0.125	0.369	-0.125
β_4	0.317	0.174	0.658	-0.026	0.315	0.127	0.564	0.065
β_5	0.407	0.075	0.556	0.262	0.399	0.062	0.520	0.280
β_6	0.882	0.100	1.078	0.684	0.897	0.055	1.006	0.788
δ	-0.007	0.007	0.006	-0.020	-0.007	0.005	0.002	-0.016
γ	0.098	0.028	0.154	0.043	0.101	0.020	0.141	0.062
ϕ	-0.006	0.041	0.076	-0.086	-0.004	0.029	0.053	-0.061
σ^2	4.000	0.218	4.450	3.599	3.997	0.153	4.307	3.711
Parameters Scheme = 6								
α	0.369	0.019	0.406	0.333	0.370	0.013	0.395	0.346
β_0	0.397	0.047	0.490	0.307	0.399	0.033	0.464	0.336
β_1	0.203	0.047	0.297	0.111	0.202	0.033	0.268	0.138
β_2	0.304	0.048	0.397	0.211	0.303	0.034	0.368	0.238
β_3	0.104	0.047	0.197	0.011	0.102	0.033	0.167	0.037
β_4	0.305	0.048	0.398	0.211	0.301	0.034	0.366	0.235
β_5	0.401	0.022	0.444	0.359	0.401	0.017	0.435	0.369
β_6	0.902	0.032	0.964	0.838	0.902	0.019	0.939	0.864
δ	-0.007	0.005	0.002	-0.015	-0.007	0.003	-0.001	-0.013
γ	0.100	0.007	0.114	0.087	0.100	0.005	0.110	0.090
ϕ	0.798	0.011	0.819	0.778	0.799	0.007	0.813	0.786
σ^2	0.251	0.014	0.280	0.226	0.251	0.010	0.270	0.233

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 18

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 36, Two of Design Structure, and four of Parameter Scheme for noninformative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.153	0.031	0.223	0.119	0.725	0.052	0.806	0.645
β_0	0.391	0.126	0.644	0.144	0.392	0.104	0.598	0.191
β_1	0.219	0.123	0.465	-0.020	0.217	0.103	0.423	0.015
β_2	0.313	0.124	0.553	0.072	0.312	0.104	0.510	0.110
β_3	0.127	0.125	0.371	-0.120	0.116	0.107	0.323	-0.096
β_4	0.324	0.129	0.580	0.069	0.311	0.104	0.516	0.106
β_5	0.407	0.057	0.519	0.297	0.405	0.057	0.516	0.294
β_6	0.899	0.078	1.050	0.746	0.902	0.057	1.013	0.788
δ	-0.002	0.003	0.002	-0.009	-0.043	0.003	-0.038	-0.049
γ	0.100	0.007	0.114	0.086	0.100	0.006	0.112	0.089
ϕ	0.796	0.018	0.831	0.762	0.797	0.015	0.826	0.768
σ^2	1.005	0.075	1.164	0.869	1.000	0.061	1.127	0.888
Parameters Scheme = 3								
α	0.148	0.022	0.197	0.119	0.681	0.062	0.780	0.590
β_0	0.371	0.214	0.793	-0.049	0.374	0.184	0.734	0.017
β_1	0.262	0.220	0.700	-0.161	0.244	0.190	0.624	-0.128
β_2	0.344	0.222	0.777	-0.091	0.333	0.192	0.705	-0.041
β_3	0.169	0.223	0.609	-0.270	0.157	0.196	0.540	-0.234
β_4	0.365	0.230	0.818	-0.084	0.327	0.191	0.703	-0.051
β_5	0.426	0.109	0.640	0.215	0.415	0.110	0.630	0.200
β_6	0.885	0.140	1.158	0.610	0.903	0.103	1.105	0.699
δ	-0.002	0.002	0.002	-0.007	-0.034	0.004	-0.029	-0.040
γ	0.100	0.014	0.128	0.073	0.100	0.011	0.123	0.077
ϕ	0.789	0.027	0.843	0.736	0.794	0.023	0.838	0.749
σ^2	4.017	0.301	4.652	3.473	4.023	0.245	4.532	3.573

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 19

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 36, Two of Design Structure, and four of Parameter Scheme for noninformative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.373	0.023	0.419	0.329	0.371	0.018	0.408	0.336
β_0	0.352	0.207	0.761	-0.052	0.365	0.178	0.715	0.017
β_1	0.257	0.219	0.695	-0.165	0.233	0.189	0.612	-0.137
β_2	0.334	0.221	0.765	-0.100	0.325	0.191	0.696	-0.048
β_3	0.158	0.222	0.597	-0.281	0.155	0.196	0.537	-0.234
β_4	0.349	0.229	0.799	-0.097	0.332	0.190	0.706	-0.044
β_5	0.420	0.107	0.629	0.214	0.405	0.109	0.618	0.194
β_6	0.870	0.132	1.126	0.611	0.890	0.094	1.077	0.704
δ	-0.007	0.009	0.010	-0.024	-0.007	0.007	0.007	-0.022
γ	0.099	0.039	0.176	0.023	0.099	0.032	0.162	0.037
ϕ	-0.013	0.055	0.096	-0.121	-0.011	0.045	0.078	-0.100
σ^2	4.028	0.301	4.664	3.481	4.029	0.246	4.539	3.577
Parameters Scheme = 6								
α	0.372	0.026	0.426	0.323	0.370	0.022	0.414	0.329
β_0	0.394	0.065	0.524	0.267	0.401	0.053	0.506	0.297
β_1	0.211	0.064	0.338	0.085	0.203	0.053	0.309	0.099
β_2	0.311	0.065	0.436	0.185	0.305	0.054	0.407	0.201
β_3	0.108	0.065	0.235	-0.020	0.105	0.055	0.211	-0.005
β_4	0.309	0.067	0.443	0.176	0.304	0.054	0.410	0.198
β_5	0.405	0.030	0.464	0.346	0.402	0.030	0.460	0.343
β_6	0.902	0.045	0.990	0.813	0.903	0.035	0.970	0.833
δ	-0.007	0.005	0.003	-0.017	-0.007	0.005	0.002	-0.015
γ	0.100	0.010	0.119	0.081	0.100	0.008	0.115	0.084
ϕ	0.798	0.013	0.823	0.773	0.798	0.011	0.820	0.777
σ^2	0.252	0.019	0.291	0.218	0.252	0.015	0.284	0.224

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 20

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 36, Two of Design Structure, and four of Parameter Scheme for informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.137	0.010	0.156	0.119	0.696	0.057	0.794	0.609
β_0	0.404	0.126	0.656	0.158	0.404	0.088	0.579	0.233
β_1	0.215	0.128	0.470	-0.037	0.200	0.090	0.378	0.023
β_2	0.310	0.129	0.558	0.059	0.298	0.091	0.474	0.120
β_3	0.090	0.132	0.348	-0.171	0.103	0.091	0.282	-0.078
β_4	0.300	0.132	0.561	0.039	0.307	0.092	0.486	0.126
β_5	0.402	0.057	0.514	0.292	0.400	0.040	0.480	0.321
β_6	0.908	0.060	1.025	0.791	0.903	0.046	0.994	0.812
δ	-0.001	0.003	0.005	-0.006	-0.056	0.005	-0.048	-0.064
γ	0.100	0.008	0.115	0.085	0.100	0.005	0.110	0.090
ϕ	0.793	0.026	0.845	0.742	0.797	0.019	0.834	0.760
σ^2	1.008	0.076	1.167	0.871	1.005	0.053	1.115	0.907
Parameters Scheme = 3								
α	0.137	0.009	0.155	0.120	0.568	0.090	0.717	0.436
β_0	0.408	0.252	0.907	-0.083	0.406	0.174	0.752	0.068
β_1	0.213	0.256	0.725	-0.289	0.198	0.179	0.553	-0.155
β_2	0.316	0.257	0.810	-0.187	0.310	0.182	0.662	-0.044
β_3	0.094	0.265	0.612	-0.428	0.092	0.181	0.448	-0.267
β_4	0.302	0.262	0.821	-0.217	0.300	0.183	0.657	-0.060
β_5	0.410	0.112	0.630	0.192	0.402	0.080	0.561	0.247
β_6	0.917	0.112	1.137	0.697	0.906	0.086	1.075	0.737
δ	-0.001	0.002	0.004	-0.005	-0.037	0.007	-0.026	-0.049
γ	0.101	0.015	0.130	0.071	0.101	0.011	0.122	0.081
ϕ	0.785	0.039	0.861	0.708	0.792	0.028	0.847	0.738
σ^2	4.030	0.302	4.670	3.483	3.997	0.211	4.433	3.607

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 21

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 36, Two of Design Structure, and four of Parameter Scheme for informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.375	0.025	0.425	0.328	0.370	0.017	0.405	0.337
β_0	0.395	0.250	0.889	-0.093	0.400	0.173	0.743	0.063
β_1	0.211	0.255	0.720	-0.290	0.213	0.179	0.569	-0.140
β_2	0.330	0.256	0.822	-0.170	0.301	0.182	0.652	-0.053
β_3	0.100	0.264	0.616	-0.419	0.101	0.181	0.457	-0.259
β_4	0.312	0.260	0.826	-0.202	0.298	0.183	0.655	-0.062
β_5	0.403	0.111	0.619	0.186	0.406	0.079	0.565	0.251
β_6	0.907	0.109	1.121	0.693	0.906	0.083	1.071	0.742
δ	-0.007	0.010	0.012	-0.026	-0.007	0.007	0.007	-0.021
γ	0.100	0.041	0.179	0.020	0.099	0.028	0.155	0.044
ϕ	-0.015	0.060	0.102	-0.132	-0.010	0.042	0.073	-0.093
σ^2	4.002	0.300	4.636	3.458	4.007	0.212	4.444	3.615
Parameters Scheme = 6								
α	0.371	0.026	0.424	0.322	0.369	0.019	0.407	0.332
β_0	0.404	0.063	0.528	0.281	0.402	0.044	0.489	0.317
β_1	0.203	0.064	0.331	0.077	0.201	0.045	0.290	0.113
β_2	0.302	0.064	0.426	0.177	0.303	0.046	0.391	0.214
β_3	0.099	0.066	0.228	-0.031	0.099	0.045	0.188	0.009
β_4	0.301	0.066	0.433	0.170	0.301	0.046	0.391	0.210
β_5	0.402	0.029	0.459	0.345	0.400	0.021	0.441	0.360
β_6	0.902	0.034	0.966	0.836	0.902	0.026	0.952	0.851
δ	-0.006	0.008	0.011	-0.022	-0.006	0.007	0.007	-0.020
γ	0.100	0.010	0.120	0.080	0.100	0.007	0.114	0.086
ϕ	0.797	0.018	0.833	0.762	0.798	0.014	0.825	0.771
σ^2	0.252	0.019	0.291	0.217	0.250	0.013	0.277	0.226

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 22

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 36, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.903	0.012	0.922	0.883	0.589	0.120	0.761	0.404
β_0	0.400	0.066	0.531	0.274	0.376	0.163	0.702	0.059
β_1	0.203	0.065	0.332	0.074	0.248	0.161	0.571	-0.069
β_2	0.303	0.066	0.431	0.175	0.331	0.157	0.634	0.022
β_3	0.103	0.066	0.233	-0.027	0.120	0.149	0.415	-0.175
β_4	0.304	0.067	0.434	0.173	0.328	0.159	0.639	0.012
β_5	0.400	0.028	0.456	0.346	0.424	0.073	0.566	0.282
β_6	0.901	0.035	0.969	0.832	0.888	0.109	1.101	0.674
δ	-0.049	0.001	-0.047	-0.050	-0.027	0.007	-0.016	-0.037
γ	0.100	0.004	0.107	0.093	0.100	0.009	0.117	0.083
ϕ	0.799	0.009	0.817	0.781	0.794	0.018	0.830	0.759
σ^2	1.000	0.038	1.078	0.929	1.003	0.095	1.206	0.836
Parameters Scheme = 3								
α	0.924	0.014	0.946	0.902	0.711	0.110	0.874	0.544
β_0	0.394	0.124	0.639	0.153	0.352	0.261	0.865	-0.160
β_1	0.220	0.126	0.468	-0.029	0.315	0.269	0.851	-0.205
β_2	0.319	0.128	0.566	0.071	0.352	0.265	0.863	-0.169
β_3	0.120	0.127	0.370	-0.129	0.154	0.258	0.665	-0.350
β_4	0.313	0.128	0.562	0.061	0.395	0.272	0.926	-0.136
β_5	0.402	0.055	0.512	0.297	0.457	0.133	0.717	0.198
β_6	0.902	0.063	1.027	0.778	0.845	0.195	1.225	0.462
δ	-0.043	0.001	-0.042	-0.044	-0.030	0.006	-0.020	-0.038
γ	0.100	0.007	0.114	0.086	0.101	0.017	0.134	0.067
ϕ	0.797	0.015	0.826	0.769	0.785	0.030	0.843	0.727
σ^2	4.009	0.153	4.320	3.723	4.007	0.377	4.811	3.340

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 23

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 36, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.369	0.012	0.393	0.346	0.374	0.029	0.434	0.319
β_0	0.383	0.122	0.622	0.148	0.323	0.252	0.819	-0.169
β_1	0.216	0.126	0.463	-0.031	0.302	0.266	0.833	-0.214
β_2	0.314	0.127	0.560	0.067	0.343	0.263	0.851	-0.175
β_3	0.118	0.127	0.367	-0.132	0.172	0.257	0.683	-0.329
β_4	0.317	0.128	0.565	0.066	0.379	0.269	0.905	-0.148
β_5	0.403	0.055	0.512	0.298	0.425	0.123	0.663	0.187
β_6	0.900	0.059	1.015	0.784	0.828	0.190	1.200	0.457
δ	-0.007	0.005	0.002	-0.016	-0.007	0.011	0.013	-0.028
γ	0.099	0.020	0.139	0.060	0.097	0.048	0.191	0.004
ϕ	-0.003	0.029	0.053	-0.060	-0.020	0.068	0.115	-0.152
σ^2	4.010	0.153	4.321	3.723	4.006	0.377	4.809	3.340
Parameters Scheme = 6								
α	0.369	0.013	0.395	0.344	0.364	0.034	0.433	0.301
β_0	0.400	0.033	0.466	0.337	0.396	0.087	0.571	0.228
β_1	0.201	0.033	0.267	0.136	0.214	0.087	0.387	0.042
β_2	0.301	0.034	0.366	0.236	0.313	0.084	0.475	0.148
β_3	0.100	0.033	0.166	0.034	0.106	0.079	0.261	-0.050
β_4	0.300	0.034	0.367	0.234	0.310	0.085	0.475	0.140
β_5	0.400	0.015	0.429	0.372	0.409	0.040	0.487	0.331
β_6	0.900	0.020	0.940	0.860	0.901	0.061	1.020	0.781
δ	-0.007	0.003	-0.001	-0.012	-0.005	0.005	0.004	-0.015
γ	0.100	0.005	0.110	0.090	0.100	0.011	0.122	0.077
ϕ	0.800	0.007	0.813	0.787	0.797	0.013	0.822	0.772
σ^2	0.251	0.010	0.270	0.233	0.254	0.024	0.305	0.211

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 24

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 54, Two of Design Structure, and four of Parameter Scheme for noninformative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.760	0.046	0.831	0.687	0.915	0.019	0.945	0.884
β_0	0.391	0.103	0.595	0.192	0.393	0.082	0.556	0.234
β_1	0.216	0.101	0.417	0.019	0.209	0.080	0.367	0.051
β_2	0.312	0.103	0.511	0.111	0.315	0.083	0.474	0.154
β_3	0.119	0.107	0.325	-0.091	0.107	0.079	0.262	-0.050
β_4	0.312	0.104	0.516	0.106	0.302	0.083	0.464	0.138
β_5	0.402	0.047	0.493	0.311	0.404	0.038	0.479	0.330
β_6	0.902	0.053	1.005	0.797	0.902	0.051	1.001	0.801
δ	-0.039	0.003	-0.035	-0.043	-0.054	0.001	-0.052	-0.056
γ	0.100	0.006	0.111	0.089	0.100	0.005	0.109	0.091
ϕ	0.797	0.013	0.824	0.771	0.798	0.012	0.821	0.775
σ^2	1.003	0.061	1.130	0.890	1.004	0.047	1.100	0.917
Parameters Scheme = 3								
α	0.767	0.048	0.842	0.693	0.893	0.024	0.929	0.855
β_0	0.372	0.183	0.732	0.018	0.362	0.150	0.658	0.073
β_1	0.237	0.186	0.604	-0.126	0.231	0.152	0.528	-0.067
β_2	0.327	0.191	0.697	-0.047	0.323	0.156	0.624	0.018
β_3	0.153	0.197	0.534	-0.234	0.134	0.150	0.429	-0.162
β_4	0.344	0.192	0.721	-0.035	0.338	0.157	0.644	0.029
β_5	0.416	0.091	0.593	0.240	0.402	0.073	0.547	0.261
β_6	0.901	0.097	1.090	0.708	0.900	0.093	1.080	0.716
δ	-0.036	0.002	-0.032	-0.040	-0.046	0.001	-0.044	-0.048
γ	0.101	0.011	0.123	0.078	0.100	0.009	0.117	0.083
ϕ	0.793	0.021	0.836	0.751	0.796	0.018	0.831	0.760
σ^2	4.019	0.245	4.528	3.569	4.006	0.188	4.388	3.657

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 25

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 54, Two of Design Structure, and four of Parameter Scheme for noninformative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.372	0.019	0.411	0.336	0.370	0.014	0.398	0.342
β_0	0.362	0.179	0.712	0.014	0.378	0.145	0.663	0.100
β_1	0.240	0.186	0.606	-0.122	0.221	0.152	0.516	-0.077
β_2	0.333	0.190	0.702	-0.041	0.317	0.156	0.617	0.011
β_3	0.148	0.197	0.527	-0.239	0.130	0.150	0.426	-0.165
β_4	0.328	0.191	0.703	-0.050	0.326	0.157	0.630	0.019
β_5	0.403	0.089	0.577	0.231	0.404	0.072	0.547	0.265
β_6	0.895	0.089	1.070	0.717	0.890	0.087	1.059	0.718
δ	-0.007	0.007	0.007	-0.021	-0.007	0.006	0.004	-0.018
γ	0.100	0.032	0.162	0.038	0.099	0.024	0.148	0.052
ϕ	-0.009	0.045	0.079	-0.097	-0.006	0.035	0.064	-0.075
σ^2	4.025	0.246	4.534	3.573	4.011	0.189	4.395	3.662
Parameters Scheme = 6								
α	0.367	0.022	0.412	0.325	0.368	0.017	0.401	0.337
β_0	0.399	0.052	0.503	0.298	0.399	0.041	0.481	0.319
β_1	0.205	0.052	0.308	0.103	0.207	0.041	0.288	0.126
β_2	0.305	0.053	0.406	0.202	0.304	0.042	0.386	0.223
β_3	0.102	0.055	0.208	-0.006	0.101	0.040	0.180	0.022
β_4	0.301	0.053	0.406	0.195	0.301	0.043	0.384	0.217
β_5	0.401	0.025	0.449	0.353	0.401	0.021	0.441	0.361
β_6	0.901	0.031	0.962	0.840	0.902	0.030	0.962	0.842
δ	-0.006	0.004	0.001	-0.014	-0.006	0.004	0.001	-0.014
γ	0.100	0.008	0.115	0.085	0.100	0.006	0.112	0.088
ϕ	0.799	0.009	0.818	0.781	0.798	0.009	0.816	0.781
σ^2	0.252	0.015	0.283	0.223	0.251	0.012	0.276	0.230

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 26

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 54, Two of Design Structure, and four of Parameter Scheme for informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.705	0.058	0.803	0.616	0.140	0.018	0.177	0.108
β_0	0.391	0.085	0.561	0.226	0.364	0.214	0.785	-0.053
β_1	0.207	0.088	0.380	0.035	0.230	0.193	0.612	-0.149
β_2	0.313	0.089	0.484	0.140	0.359	0.209	0.762	-0.047
β_3	0.108	0.089	0.282	-0.069	0.147	0.207	0.554	-0.260
β_4	0.309	0.091	0.488	0.130	0.357	0.209	0.766	-0.056
β_5	0.403	0.039	0.482	0.327	0.429	0.124	0.673	0.186
β_6	0.903	0.045	0.991	0.815	0.898	0.158	1.203	0.586
δ	-0.056	0.005	-0.048	-0.065	-0.001	0.003	0.006	-0.007
γ	0.100	0.005	0.111	0.090	0.101	0.012	0.124	0.077
ϕ	0.796	0.018	0.832	0.761	0.787	0.033	0.851	0.723
σ^2	1.002	0.053	1.111	0.904	1.003	0.126	1.278	0.782
Parameters Scheme = 3								
α	0.603	0.082	0.738	0.481	0.141	0.019	0.179	0.110
β_0	0.366	0.157	0.674	0.064	0.314	0.307	0.916	-0.289
β_1	0.238	0.165	0.563	-0.085	0.271	0.311	0.882	-0.339
β_2	0.335	0.167	0.661	0.007	0.415	0.320	1.044	-0.210
β_3	0.147	0.168	0.476	-0.183	0.192	0.327	0.841	-0.447
β_4	0.328	0.170	0.660	-0.007	0.424	0.324	1.058	-0.213
β_5	0.406	0.077	0.558	0.258	0.441	0.212	0.859	0.028
β_6	0.901	0.084	1.066	0.736	0.857	0.250	1.343	0.353
δ	-0.039	0.006	-0.029	-0.049	-0.001	0.003	0.004	-0.006
γ	0.100	0.010	0.121	0.080	0.100	0.023	0.145	0.054
ϕ	0.792	0.027	0.845	0.739	0.776	0.049	0.873	0.682
σ^2	4.011	0.212	4.446	3.620	4.020	0.504	5.125	3.139

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 27

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 54, Two of Design Structure, and four of Parameter Scheme for informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.370	0.017	0.405	0.337	0.381	0.037	0.457	0.311
β_0	0.373	0.156	0.678	0.071	0.290	0.299	0.880	-0.290
β_1	0.233	0.164	0.558	-0.090	0.252	0.310	0.860	-0.355
β_2	0.324	0.167	0.649	-0.004	0.396	0.317	1.020	-0.224
β_3	0.122	0.168	0.451	-0.208	0.182	0.326	0.828	-0.454
β_4	0.325	0.170	0.656	-0.009	0.389	0.320	1.018	-0.241
β_5	0.404	0.076	0.554	0.258	0.427	0.206	0.831	0.022
β_6	0.895	0.082	1.057	0.733	0.811	0.238	1.281	0.336
δ	-0.007	0.007	0.007	-0.020	-0.006	0.015	0.023	-0.035
γ	0.099	0.028	0.155	0.044	0.097	0.064	0.221	-0.029
ϕ	-0.007	0.042	0.076	-0.089	-0.028	0.092	0.153	-0.209
σ^2	4.005	0.211	4.439	3.614	4.014	0.502	5.113	3.134
Parameters Scheme = 6								
α	0.370	0.019	0.408	0.333	0.375	0.040	0.457	0.299
β_0	0.398	0.044	0.485	0.313	0.389	0.123	0.633	0.150
β_1	0.203	0.045	0.292	0.115	0.214	0.106	0.424	0.007
β_2	0.304	0.045	0.391	0.216	0.322	0.118	0.552	0.091
β_3	0.102	0.046	0.191	0.012	0.116	0.114	0.338	-0.109
β_4	0.302	0.046	0.393	0.210	0.325	0.118	0.557	0.090
β_5	0.400	0.020	0.440	0.361	0.414	0.068	0.547	0.281
β_6	0.901	0.025	0.949	0.852	0.906	0.094	1.089	0.721
δ	-0.007	0.006	0.006	-0.019	-0.005	0.009	0.014	-0.023
γ	0.100	0.007	0.114	0.086	0.100	0.016	0.132	0.068
ϕ	0.799	0.013	0.824	0.774	0.792	0.024	0.840	0.745
σ^2	0.251	0.013	0.279	0.227	0.255	0.032	0.325	0.199

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 28

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 54, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.627	0.079	0.752	0.515	0.893	0.028	0.936	0.851
β_0	0.419	0.120	0.661	0.187	0.400	0.100	0.598	0.205
β_1	0.201	0.118	0.437	-0.031	0.204	0.095	0.392	0.017
β_2	0.302	0.114	0.521	0.082	0.300	0.094	0.482	0.116
β_3	0.093	0.114	0.318	-0.134	0.111	0.097	0.301	-0.081
β_4	0.306	0.118	0.539	0.073	0.305	0.097	0.493	0.113
β_5	0.407	0.051	0.506	0.310	0.399	0.045	0.488	0.313
β_6	0.915	0.072	1.055	0.774	0.907	0.050	1.005	0.808
δ	-0.037	0.006	-0.029	-0.045	-0.047	0.002	-0.044	-0.049
γ	0.100	0.006	0.113	0.088	0.101	0.005	0.110	0.091
ϕ	0.795	0.015	0.824	0.766	0.798	0.011	0.820	0.776
σ^2	1.008	0.066	1.146	0.887	1.002	0.054	1.113	0.903
Parameters Scheme = 3								
α	0.625	0.075	0.745	0.515	0.889	0.028	0.935	0.846
β_0	0.434	0.234	0.902	-0.017	0.435	0.196	0.826	0.056
β_1	0.207	0.237	0.680	-0.257	0.209	0.190	0.583	-0.166
β_2	0.310	0.228	0.748	-0.132	0.312	0.187	0.677	-0.055
β_3	0.116	0.228	0.567	-0.338	0.083	0.193	0.463	-0.301
β_4	0.340	0.234	0.802	-0.122	0.291	0.193	0.666	-0.092
β_5	0.416	0.100	0.611	0.223	0.414	0.088	0.590	0.244
β_6	0.935	0.135	1.200	0.671	0.920	0.092	1.099	0.736
δ	-0.030	0.005	-0.024	-0.038	-0.041	0.001	-0.039	-0.043
γ	0.101	0.012	0.125	0.077	0.100	0.010	0.120	0.081
ϕ	0.786	0.023	0.832	0.740	0.791	0.018	0.827	0.755
σ^2	4.030	0.265	4.584	3.545	4.015	0.215	4.458	3.620

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 29

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 54, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.371	0.019	0.410	0.335	0.369	0.016	0.401	0.338
β_0	0.414	0.224	0.860	-0.023	0.410	0.190	0.788	0.042
β_1	0.210	0.236	0.683	-0.253	0.208	0.189	0.580	-0.163
β_2	0.307	0.227	0.743	-0.135	0.318	0.186	0.679	-0.047
β_3	0.090	0.228	0.538	-0.365	0.087	0.193	0.465	-0.295
β_4	0.297	0.230	0.750	-0.157	0.298	0.191	0.669	-0.082
β_5	0.408	0.098	0.601	0.216	0.403	0.086	0.575	0.236
β_6	0.908	0.125	1.154	0.662	0.905	0.082	1.066	0.744
δ	-0.007	0.008	0.008	-0.022	-0.007	0.006	0.005	-0.019
γ	0.100	0.034	0.168	0.034	0.100	0.028	0.154	0.046
ϕ	-0.015	0.049	0.081	-0.109	-0.012	0.039	0.066	-0.089
σ^2	4.035	0.265	4.589	3.546	4.004	0.214	4.448	3.609
Parameters Scheme = 6								
α	0.370	0.022	0.415	0.328	0.368	0.019	0.405	0.332
β_0	0.405	0.059	0.521	0.291	0.402	0.049	0.499	0.307
β_1	0.205	0.059	0.323	0.089	0.205	0.048	0.300	0.110
β_2	0.306	0.057	0.415	0.195	0.303	0.047	0.395	0.210
β_3	0.099	0.057	0.211	-0.014	0.100	0.048	0.195	0.004
β_4	0.302	0.060	0.420	0.183	0.301	0.049	0.395	0.204
β_5	0.405	0.027	0.458	0.353	0.402	0.023	0.448	0.357
β_6	0.907	0.042	0.988	0.824	0.904	0.029	0.961	0.845
δ	-0.006	0.004	0.002	-0.015	-0.006	0.003	0.000	-0.012
γ	0.100	0.008	0.116	0.084	0.100	0.007	0.113	0.087
ϕ	0.797	0.011	0.819	0.776	0.799	0.008	0.815	0.783
σ^2	0.251	0.017	0.285	0.221	0.250	0.013	0.278	0.226

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 30

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 90, Two of Design Structure, and four of Parameter Scheme for noninformative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.951	0.007	0.962	0.939	0.143	0.021	0.187	0.105
β_0	0.400	0.060	0.519	0.283	0.462	0.301	1.071	-0.130
β_1	0.201	0.058	0.315	0.087	0.208	0.294	0.786	-0.372
β_2	0.300	0.058	0.412	0.187	0.333	0.301	0.917	-0.261
β_3	0.099	0.057	0.211	-0.014	0.084	0.301	0.679	-0.513
β_4	0.298	0.059	0.414	0.184	0.286	0.296	0.869	-0.299
β_5	0.400	0.025	0.449	0.350	0.415	0.158	0.728	0.106
β_6	0.901	0.032	0.963	0.839	0.964	0.150	1.254	0.664
δ	-0.049	0.000	-0.048	-0.050	0.000	0.007	0.015	-0.013
γ	0.100	0.003	0.106	0.094	0.101	0.018	0.137	0.065
ϕ	0.800	0.008	0.816	0.783	0.753	0.073	0.899	0.611
σ^2	1.001	0.033	1.069	0.939	1.028	0.187	1.460	0.720
Parameters Scheme = 3								
α	0.944	0.008	0.957	0.930	0.144	0.021	0.187	0.105
β_0	0.397	0.117	0.629	0.170	0.541	0.594	1.745	-0.624
β_1	0.207	0.115	0.434	-0.021	0.218	0.588	1.374	-0.942
β_2	0.312	0.116	0.535	0.086	0.353	0.598	1.514	-0.828
β_3	0.103	0.115	0.328	-0.122	0.056	0.605	1.250	-1.144
β_4	0.305	0.118	0.535	0.077	0.298	0.595	1.466	-0.878
β_5	0.406	0.049	0.503	0.310	0.409	0.316	1.032	-0.210
β_6	0.909	0.057	1.021	0.796	1.021	0.272	1.546	0.480
δ	-0.043	0.000	-0.042	-0.043	-0.001	0.005	0.010	-0.011
γ	0.100	0.006	0.112	0.088	0.103	0.036	0.174	0.031
ϕ	0.798	0.013	0.822	0.773	0.698	0.104	0.903	0.494
σ^2	4.006	0.133	4.275	3.756	4.082	0.743	5.793	2.860

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 31

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 90, Two of Design Structure, and four of Parameter Scheme for noninformative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.369	0.010	0.389	0.351	0.395	0.055	0.510	0.294
β_0	0.402	0.113	0.627	0.181	0.423	0.581	1.599	-0.711
β_1	0.202	0.115	0.428	-0.025	0.235	0.581	1.377	-0.911
β_2	0.301	0.116	0.522	0.076	0.350	0.590	1.492	-0.815
β_3	0.105	0.115	0.328	-0.121	0.098	0.598	1.277	-1.088
β_4	0.304	0.117	0.534	0.077	0.330	0.587	1.485	-0.832
β_5	0.402	0.048	0.497	0.309	0.449	0.315	1.069	-0.167
β_6	0.905	0.052	1.008	0.804	0.973	0.258	1.470	0.461
δ	-0.007	0.004	0.001	-0.015	-0.005	0.022	0.038	-0.048
γ	0.100	0.017	0.134	0.066	0.101	0.098	0.294	-0.094
ϕ	-0.004	0.025	0.044	-0.053	-0.100	0.142	0.178	-0.379
σ^2	4.010	0.133	4.279	3.759	4.046	0.736	5.739	2.835
Parameters Scheme = 6								
α	0.368	0.011	0.389	0.347	0.390	0.055	0.505	0.289
β_0	0.400	0.030	0.458	0.342	0.424	0.151	0.728	0.127
β_1	0.200	0.029	0.258	0.143	0.208	0.147	0.497	-0.081
β_2	0.301	0.029	0.357	0.244	0.317	0.152	0.611	0.018
β_3	0.100	0.029	0.156	0.043	0.089	0.150	0.385	-0.210
β_4	0.301	0.030	0.359	0.243	0.299	0.148	0.590	0.007
β_5	0.401	0.014	0.427	0.374	0.407	0.080	0.564	0.251
β_6	0.901	0.019	0.938	0.864	0.928	0.088	1.096	0.753
δ	-0.007	0.002	-0.002	-0.011	-0.002	0.022	0.042	-0.044
γ	0.100	0.004	0.108	0.092	0.100	0.025	0.149	0.052
ϕ	0.800	0.006	0.811	0.788	0.780	0.052	0.884	0.679
σ^2	0.250	0.008	0.267	0.235	0.258	0.047	0.367	0.181

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 32

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 90, Two of Design Structure, and four of Parameter Scheme for informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.140	0.016	0.172	0.111	0.136	0.011	0.158	0.116
β_0	0.385	0.192	0.762	0.007	0.376	0.147	0.669	0.089
β_1	0.212	0.232	0.665	-0.244	0.222	0.140	0.500	-0.051
β_2	0.347	0.196	0.728	-0.036	0.335	0.144	0.616	0.051
β_3	0.162	0.202	0.560	-0.234	0.141	0.143	0.423	-0.145
β_4	0.326	0.199	0.719	-0.070	0.332	0.147	0.623	0.041
β_5	0.410	0.170	0.741	0.078	0.405	0.066	0.535	0.277
β_6	0.912	0.117	1.141	0.677	0.904	0.078	1.054	0.749
δ	-0.001	0.003	0.005	-0.008	-0.001	0.002	0.004	-0.005
γ	0.100	0.012	0.124	0.077	0.100	0.008	0.117	0.084
ϕ	0.785	0.035	0.854	0.717	0.792	0.024	0.840	0.744
σ^2	0.997	0.125	1.273	0.777	1.007	0.088	1.194	0.850
Parameters Scheme = 3								
α	0.138	0.015	0.169	0.111	0.137	0.011	0.159	0.118
β_0	0.343	0.291	0.912	-0.227	0.337	0.240	0.809	-0.132
β_1	0.230	0.348	0.914	-0.457	0.274	0.243	0.758	-0.200
β_2	0.369	0.313	0.979	-0.243	0.370	0.249	0.858	-0.121
β_3	0.242	0.318	0.869	-0.383	0.181	0.247	0.668	-0.307
β_4	0.349	0.317	0.965	-0.273	0.374	0.252	0.868	-0.123
β_5	0.429	0.286	0.985	-0.129	0.401	0.125	0.650	0.158
β_6	0.881	0.193	1.258	0.496	0.894	0.138	1.161	0.622
δ	-0.001	0.003	0.004	-0.006	-0.001	0.002	0.003	-0.005
γ	0.100	0.023	0.145	0.054	0.100	0.016	0.132	0.068
ϕ	0.776	0.051	0.877	0.675	0.783	0.036	0.854	0.713
σ^2	3.973	0.498	5.066	3.102	4.001	0.349	4.743	3.378

Note. P = Parameters of interest, Par. Sch. = Parameters Scheme.

Table 33

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 90, Two of Design Structure, and four of Parameter Scheme for informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.382	0.036	0.455	0.314	0.375	0.026	0.427	0.327
β_0	0.317	0.284	0.873	-0.237	0.330	0.237	0.789	-0.132
β_1	0.234	0.348	0.918	-0.455	0.253	0.242	0.733	-0.218
β_2	0.346	0.310	0.950	-0.257	0.345	0.247	0.828	-0.142
β_3	0.246	0.317	0.870	-0.377	0.179	0.246	0.667	-0.308
β_4	0.345	0.316	0.959	-0.275	0.368	0.250	0.856	-0.123
β_5	0.418	0.286	0.974	-0.140	0.398	0.124	0.645	0.158
β_6	0.862	0.185	1.222	0.496	0.871	0.132	1.126	0.611
δ	-0.007	0.015	0.022	-0.035	-0.007	0.011	0.014	-0.028
γ	0.096	0.064	0.221	-0.030	0.099	0.045	0.188	0.011
ϕ	-0.026	0.092	0.153	-0.207	-0.018	0.065	0.110	-0.145
σ^2	4.015	0.503	5.121	3.135	4.009	0.350	4.753	3.384
Parameters Scheme = 6								
α	0.381	0.037	0.458	0.311	0.374	0.027	0.429	0.323
β_0	0.396	0.107	0.607	0.186	0.390	0.078	0.544	0.239
β_1	0.212	0.131	0.468	-0.045	0.214	0.074	0.360	0.070
β_2	0.316	0.109	0.526	0.102	0.311	0.077	0.459	0.161
β_3	0.113	0.112	0.332	-0.109	0.110	0.075	0.257	-0.041
β_4	0.309	0.110	0.528	0.090	0.313	0.078	0.468	0.159
β_5	0.408	0.093	0.590	0.227	0.403	0.035	0.472	0.336
β_6	0.914	0.072	1.052	0.769	0.906	0.047	0.997	0.811
δ	-0.008	0.010	0.011	-0.028	-0.006	0.007	0.008	-0.021
γ	0.100	0.017	0.133	0.068	0.101	0.011	0.123	0.078
ϕ	0.792	0.025	0.842	0.744	0.795	0.018	0.831	0.761
σ^2	0.257	0.032	0.328	0.200	0.252	0.022	0.298	0.212

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 34

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 90, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.881	0.029	0.928	0.837	0.933	0.012	0.953	0.913
β_0	0.391	0.094	0.576	0.209	0.400	0.074	0.547	0.255
β_1	0.213	0.093	0.399	0.032	0.208	0.069	0.344	0.072
β_2	0.305	0.092	0.481	0.128	0.310	0.072	0.451	0.169
β_3	0.112	0.090	0.287	-0.065	0.106	0.071	0.244	-0.034
β_4	0.316	0.092	0.496	0.136	0.304	0.072	0.446	0.163
β_5	0.406	0.049	0.501	0.312	0.403	0.031	0.464	0.343
β_6	0.902	0.060	1.020	0.783	0.903	0.041	0.983	0.822
δ	-0.046	0.002	-0.043	-0.048	-0.045	0.001	-0.044	-0.046
γ	0.100	0.005	0.110	0.090	0.100	0.004	0.108	0.093
ϕ	0.798	0.012	0.821	0.775	0.798	0.009	0.816	0.780
σ^2	1.000	0.054	1.111	0.902	0.999	0.041	1.082	0.922
Parameters Scheme = 3								
α	0.873	0.032	0.924	0.824	0.903	0.014	0.925	0.881
β_0	0.370	0.168	0.700	0.046	0.402	0.137	0.672	0.137
β_1	0.241	0.173	0.589	-0.098	0.228	0.133	0.490	-0.033
β_2	0.320	0.172	0.652	-0.014	0.325	0.138	0.594	0.055
β_3	0.144	0.168	0.476	-0.186	0.117	0.136	0.384	-0.151
β_4	0.336	0.172	0.672	-0.005	0.314	0.139	0.587	0.043
β_5	0.422	0.094	0.606	0.240	0.402	0.060	0.519	0.286
β_6	0.906	0.109	1.122	0.691	0.903	0.073	1.047	0.757
δ	-0.041	0.002	-0.039	-0.043	-0.039	0.001	-0.038	-0.040
γ	0.100	0.010	0.119	0.080	0.100	0.008	0.115	0.085
ϕ	0.794	0.018	0.830	0.758	0.796	0.014	0.824	0.768
σ^2	4.004	0.214	4.445	3.611	3.997	0.164	4.332	3.691

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 35

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 90, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.371	0.016	0.403	0.340	0.369	0.012	0.393	0.346
β_0	0.377	0.161	0.693	0.065	0.373	0.131	0.630	0.121
β_1	0.232	0.173	0.580	-0.105	0.223	0.133	0.484	-0.037
β_2	0.320	0.171	0.649	-0.014	0.322	0.137	0.589	0.052
β_3	0.128	0.168	0.460	-0.201	0.129	0.136	0.397	-0.139
β_4	0.317	0.171	0.652	-0.022	0.317	0.138	0.588	0.047
β_5	0.407	0.092	0.589	0.228	0.403	0.058	0.517	0.292
β_6	0.888	0.100	1.087	0.692	0.899	0.066	1.030	0.771
δ	-0.007	0.006	0.005	-0.020	-0.007	0.005	0.002	-0.017
γ	0.100	0.027	0.155	0.047	0.099	0.021	0.141	0.058
ϕ	-0.007	0.039	0.071	-0.084	-0.002	0.030	0.058	-0.062
σ^2	4.008	0.214	4.450	3.613	4.008	0.165	4.343	3.701
Parameters Scheme = 6								
α	0.369	0.018	0.405	0.335	0.369	0.013	0.394	0.343
β_0	0.401	0.047	0.493	0.310	0.399	0.037	0.472	0.327
β_1	0.203	0.047	0.298	0.110	0.202	0.035	0.272	0.133
β_2	0.305	0.047	0.395	0.213	0.304	0.037	0.376	0.232
β_3	0.104	0.046	0.194	0.013	0.101	0.036	0.172	0.031
β_4	0.304	0.047	0.396	0.212	0.300	0.037	0.373	0.228
β_5	0.403	0.026	0.454	0.353	0.402	0.017	0.435	0.369
β_6	0.905	0.035	0.973	0.835	0.901	0.025	0.950	0.852
δ	-0.007	0.003	0.000	-0.013	-0.007	0.002	-0.002	-0.011
γ	0.100	0.007	0.113	0.087	0.100	0.005	0.110	0.090
ϕ	0.798	0.008	0.815	0.782	0.799	0.007	0.813	0.786
σ^2	0.252	0.014	0.279	0.227	0.251	0.010	0.272	0.232

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 36

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 180, Two of Design Structure, and four of Parameter Scheme for noninformative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.142	0.028	0.200	0.093	0.138	0.016	0.170	0.109
β_0	0.365	0.229	0.820	-0.078	0.373	0.176	0.716	0.029
β_1	0.236	0.262	0.748	-0.280	0.241	0.187	0.611	-0.119
β_2	0.347	0.243	0.821	-0.126	0.328	0.191	0.702	-0.049
β_3	0.200	0.258	0.712	-0.312	0.135	0.187	0.502	-0.232
β_4	0.368	0.254	0.864	-0.134	0.339	0.186	0.701	-0.026
β_5	0.415	0.102	0.615	0.217	0.412	0.102	0.609	0.209
β_6	0.882	0.161	1.193	0.561	0.903	0.099	1.097	0.706
δ	0.001	0.007	0.015	-0.012	0.000	0.005	0.011	-0.011
γ	0.101	0.018	0.136	0.066	0.100	0.012	0.123	0.076
ϕ	0.778	0.058	0.895	0.665	0.784	0.048	0.878	0.690
σ^2	1.030	0.184	1.449	0.727	1.005	0.122	1.271	0.793
Parameters Scheme = 3								
α	0.144	0.026	0.198	0.099	0.138	0.015	0.169	0.111
β_0	0.341	0.324	0.978	-0.286	0.331	0.273	0.867	-0.206
β_1	0.259	0.369	0.982	-0.466	0.307	0.302	0.904	-0.282
β_2	0.373	0.359	1.081	-0.324	0.360	0.309	0.960	-0.252
β_3	0.297	0.369	1.022	-0.427	0.192	0.302	0.793	-0.398
β_4	0.416	0.364	1.119	-0.301	0.356	0.301	0.943	-0.234
β_5	0.422	0.186	0.790	0.057	0.411	0.187	0.776	0.041
β_6	0.812	0.259	1.311	0.297	0.876	0.176	1.218	0.525
δ	0.000	0.005	0.010	-0.011	-0.001	0.004	0.007	-0.008
γ	0.101	0.034	0.168	0.034	0.100	0.024	0.146	0.053
ϕ	0.770	0.086	0.941	0.602	0.771	0.066	0.902	0.642
σ^2	4.011	0.710	5.618	2.842	4.017	0.486	5.077	3.169

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 37

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 180, Two of Design Structure, and four of Parameter Scheme for noninformative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.396	0.060	0.521	0.288	0.382	0.038	0.460	0.310
β_0	0.348	0.320	0.976	-0.271	0.323	0.271	0.855	-0.210
β_1	0.249	0.369	0.970	-0.474	0.289	0.301	0.884	-0.298
β_2	0.389	0.359	1.096	-0.307	0.347	0.308	0.945	-0.263
β_3	0.283	0.368	1.007	-0.438	0.191	0.302	0.792	-0.400
β_4	0.408	0.363	1.108	-0.306	0.359	0.301	0.945	-0.229
β_5	0.418	0.186	0.785	0.054	0.400	0.186	0.762	0.032
β_6	0.801	0.254	1.291	0.296	0.869	0.174	1.208	0.523
δ	-0.006	0.022	0.037	-0.048	-0.006	0.016	0.024	-0.037
γ	0.096	0.092	0.278	-0.083	0.098	0.064	0.223	-0.028
ϕ	-0.046	0.134	0.217	-0.308	-0.030	0.095	0.156	-0.218
σ^2	4.022	0.711	5.631	2.851	4.022	0.486	5.083	3.172
Parameters Scheme = 6								
α	0.382	0.068	0.523	0.258	0.383	0.041	0.468	0.307
β_0	0.389	0.136	0.663	0.122	0.388	0.097	0.579	0.198
β_1	0.209	0.156	0.511	-0.098	0.217	0.102	0.419	0.020
β_2	0.308	0.140	0.580	0.035	0.315	0.104	0.518	0.111
β_3	0.137	0.153	0.441	-0.168	0.110	0.101	0.308	-0.089
β_4	0.331	0.150	0.628	0.034	0.307	0.101	0.506	0.110
β_5	0.409	0.054	0.515	0.305	0.404	0.054	0.509	0.298
β_6	0.902	0.094	1.084	0.715	0.908	0.058	1.022	0.791
δ	0.001	0.022	0.047	-0.040	-0.007	0.018	0.028	-0.042
γ	0.102	0.025	0.151	0.053	0.100	0.017	0.133	0.068
ϕ	0.788	0.041	0.870	0.708	0.790	0.036	0.861	0.721
σ^2	0.263	0.047	0.372	0.185	0.254	0.031	0.321	0.200

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 38

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 180, Two of Design Structure, and four of Parameter Scheme for informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.172	0.046	0.273	0.122	0.902	0.034	0.957	0.854
β_0	0.413	0.122	0.659	0.175	0.410	0.095	0.599	0.226
β_1	0.203	0.121	0.442	-0.037	0.197	0.095	0.384	0.010
β_2	0.306	0.123	0.543	0.065	0.298	0.094	0.482	0.114
β_3	0.091	0.121	0.330	-0.150	0.100	0.093	0.282	-0.083
β_4	0.300	0.132	0.561	0.040	0.304	0.096	0.490	0.115
β_5	0.405	0.053	0.510	0.303	0.401	0.040	0.481	0.324
β_6	0.906	0.071	1.042	0.767	0.908	0.050	1.006	0.808
δ	-0.004	0.004	0.002	-0.012	-0.052	0.002	-0.049	-0.055
γ	0.100	0.007	0.113	0.087	0.100	0.005	0.110	0.090
ϕ	0.796	0.016	0.829	0.764	0.797	0.013	0.823	0.771
σ^2	1.006	0.071	1.155	0.876	1.003	0.055	1.116	0.902
Parameters Scheme = 3								
α	0.157	0.029	0.220	0.121	0.875	0.039	0.937	0.819
β_0	0.426	0.242	0.911	-0.045	0.408	0.187	0.780	0.047
β_1	0.206	0.242	0.685	-0.273	0.217	0.190	0.592	-0.157
β_2	0.327	0.245	0.799	-0.156	0.320	0.189	0.688	-0.050
β_3	0.086	0.243	0.563	-0.397	0.107	0.186	0.470	-0.261
β_4	0.303	0.263	0.824	-0.214	0.304	0.192	0.676	-0.075
β_5	0.414	0.105	0.622	0.214	0.407	0.078	0.563	0.256
β_6	0.925	0.137	1.188	0.655	0.907	0.095	1.092	0.719
δ	-0.002	0.003	0.002	-0.008	-0.045	0.002	-0.042	-0.048
γ	0.101	0.013	0.127	0.074	0.100	0.010	0.121	0.080
ϕ	0.787	0.027	0.840	0.735	0.793	0.021	0.834	0.751
σ^2	4.016	0.284	4.610	3.498	4.014	0.219	4.466	3.609

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 39

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 180, Two of Design Structure, and four of Parameter Scheme for informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.372	0.022	0.418	0.330	0.371	0.017	0.405	0.338
β_0	0.414	0.238	0.889	-0.048	0.407	0.183	0.771	0.053
β_1	0.193	0.241	0.669	-0.285	0.198	0.190	0.571	-0.174
β_2	0.304	0.244	0.772	-0.175	0.308	0.188	0.673	-0.061
β_3	0.095	0.243	0.572	-0.388	0.091	0.186	0.453	-0.276
β_4	0.299	0.262	0.816	-0.217	0.307	0.191	0.678	-0.070
β_5	0.406	0.103	0.611	0.210	0.405	0.076	0.559	0.257
β_6	0.914	0.134	1.173	0.650	0.901	0.089	1.076	0.724
δ	-0.007	0.009	0.009	-0.024	-0.007	0.007	0.006	-0.020
γ	0.101	0.037	0.173	0.028	0.099	0.029	0.156	0.043
ϕ	-0.018	0.053	0.087	-0.121	-0.009	0.041	0.072	-0.090
σ^2	4.015	0.284	4.609	3.497	3.996	0.218	4.448	3.593
Parameters Scheme = 6								
α	0.369	0.026	0.422	0.320	0.368	0.019	0.407	0.331
β_0	0.403	0.060	0.524	0.286	0.402	0.047	0.496	0.311
β_1	0.203	0.061	0.323	0.083	0.203	0.048	0.297	0.109
β_2	0.304	0.062	0.423	0.182	0.305	0.047	0.398	0.213
β_3	0.100	0.061	0.220	-0.020	0.099	0.046	0.190	0.007
β_4	0.300	0.066	0.431	0.170	0.298	0.048	0.391	0.203
β_5	0.403	0.028	0.458	0.350	0.403	0.021	0.444	0.363
β_6	0.905	0.038	0.979	0.830	0.903	0.029	0.959	0.846
δ	-0.006	0.005	0.004	-0.016	-0.006	0.004	0.002	-0.014
γ	0.100	0.009	0.118	0.083	0.100	0.007	0.114	0.087
ϕ	0.798	0.011	0.820	0.775	0.798	0.009	0.817	0.780
σ^2	0.251	0.018	0.288	0.219	0.251	0.014	0.279	0.226

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 40

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 180, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 1								
α	0.930	0.009	0.945	0.914	0.939	0.003	0.943	0.933
β_0	0.392	0.074	0.540	0.247	0.397	0.053	0.500	0.293
β_1	0.210	0.071	0.350	0.071	0.205	0.050	0.302	0.107
β_2	0.311	0.071	0.448	0.173	0.305	0.050	0.401	0.206
β_3	0.104	0.070	0.241	-0.034	0.104	0.050	0.202	0.007
β_4	0.305	0.071	0.445	0.165	0.304	0.051	0.404	0.205
β_5	0.402	0.033	0.468	0.337	0.401	0.022	0.445	0.358
β_6	0.901	0.039	0.977	0.824	0.902	0.028	0.955	0.846
δ	-0.041	0.000	-0.040	-0.041	-0.039	0.000	-0.039	-0.039
γ	0.100	0.004	0.108	0.093	0.100	0.003	0.105	0.095
ϕ	0.799	0.008	0.815	0.783	0.800	0.006	0.812	0.787
σ^2	0.998	0.041	1.082	0.922	1.000	0.029	1.059	0.945
Parameters Scheme = 3								
α	0.914	0.010	0.930	0.896	0.960	0.003	0.965	0.955
β_0	0.390	0.138	0.662	0.123	0.398	0.100	0.595	0.203
β_1	0.224	0.136	0.493	-0.043	0.208	0.097	0.399	0.016
β_2	0.322	0.136	0.586	0.057	0.310	0.098	0.499	0.118
β_3	0.124	0.134	0.389	-0.139	0.113	0.097	0.306	-0.079
β_4	0.317	0.137	0.586	0.048	0.308	0.100	0.504	0.113
β_5	0.407	0.065	0.536	0.280	0.403	0.044	0.489	0.318
β_6	0.905	0.072	1.046	0.763	0.902	0.050	0.998	0.802
δ	-0.036	0.000	-0.035	-0.037	-0.036	0.000	-0.036	-0.037
γ	0.100	0.008	0.115	0.085	0.100	0.005	0.111	0.089
ϕ	0.796	0.014	0.823	0.769	0.798	0.010	0.818	0.779
σ^2	4.013	0.165	4.348	3.706	4.008	0.117	4.244	3.787

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

Table 41

Parameters Estimates, 95% Confidence Intervals, and Empirical Standard Errors, for Parameters of Joint Model with N = 180, Two of Design Structure, and four of Parameter Scheme for semi-informative Prior (continued)

	Design Structure = 10 &10 (Balanced)				Design Structure = 20 & 6 (Unbalanced)			
			97.5% CI				97.5% CI	
P	Est.	Emp.	Upper	Lower	Est.	Emp.	Upper	Lower
Parameters Scheme = 4								
α	0.369	0.013	0.394	0.345	0.369	0.009	0.386	0.353
β_0	0.379	0.132	0.639	0.126	0.384	0.095	0.570	0.199
β_1	0.224	0.135	0.493	-0.042	0.215	0.097	0.406	0.024
β_2	0.318	0.135	0.581	0.053	0.314	0.098	0.503	0.122
β_3	0.126	0.134	0.392	-0.137	0.114	0.097	0.307	-0.077
β_4	0.320	0.136	0.587	0.052	0.312	0.100	0.507	0.117
β_5	0.401	0.064	0.527	0.278	0.401	0.043	0.485	0.319
β_6	0.892	0.064	1.019	0.765	0.898	0.045	0.986	0.810
δ	-0.007	0.005	0.002	-0.017	-0.007	0.003	-0.001	-0.014
γ	0.099	0.021	0.141	0.057	0.100	0.015	0.130	0.071
ϕ	-0.003	0.030	0.056	-0.063	-0.003	0.021	0.039	-0.045
σ^2	4.018	0.165	4.353	3.710	4.003	0.116	4.238	3.782
Parameters Scheme = 6								
α	0.367	0.015	0.396	0.339	0.368	0.009	0.386	0.350
β_0	0.401	0.037	0.475	0.329	0.399	0.026	0.450	0.348
β_1	0.202	0.036	0.273	0.131	0.202	0.025	0.251	0.153
β_2	0.301	0.036	0.371	0.231	0.302	0.025	0.351	0.252
β_3	0.102	0.035	0.171	0.032	0.102	0.025	0.151	0.053
β_4	0.300	0.037	0.372	0.229	0.300	0.026	0.351	0.249
β_5	0.400	0.018	0.435	0.366	0.401	0.012	0.424	0.377
β_6	0.901	0.023	0.946	0.856	0.900	0.017	0.933	0.866
δ	-0.006	0.002	-0.002	-0.011	-0.007	0.002	-0.003	-0.010
γ	0.100	0.005	0.110	0.090	0.100	0.004	0.107	0.093
ϕ	0.799	0.006	0.811	0.788	0.800	0.005	0.809	0.791
σ^2	0.252	0.010	0.273	0.232	0.251	0.007	0.265	0.237

Note. P = Parameters, Est. = Parameter Estimate, SD = Standard Deviation.

The point estimates, confidence intervals, and standard deviations provide information to assess the performance of the Bayesian method and to estimate unknown parameters of the proposed joint model ($\alpha,\ \beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4,\ \beta_5,\ \beta_6,\ \delta,\ \gamma,\ \phi,\ \sigma^2$). Looking across all parameter schemes considered in this study, there are a few general trends worth noting. First, the parameter schemes influence heavily the estimation of unknown parameters of the joint model except α and σ^2. For any of the parameter schemes, Bayesian estimation seems to be the appropriate method for identifying all unknown parameters except α and σ^2 on the joint model. The estimates are significantly higher or lower than the true values of ($\alpha = 1\ or\ 2, and\ \sigma^2 = 0.5,\ 1,\ or\ 2$), depending on the parameter schemes, which are severely misleading even though the model has converged. Further, the Bayesian method correctly identified the other remaining parameters ($\beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4,\ \beta_5,\ \beta_6,\ \delta,\ \gamma,\ \phi$) (all true values were contained in the 95% CI) on the proposed joint model for all sample sizes and prior distributions.

Second, the design structures (balanced or unbalanced) had relatively little influence on parameters estimation. Third, it is interesting to observe that all parameter schemes correctly specified all model parameters except α and σ^2, as shown by falling in the range of the bounds. Depending on the narrowest range of 97.5% confidence intervals, the estimation of model parameters that used informative priors produced sometimes more accuracy in the estimation than the others (noninformative, semi-informative). On the other hand, sometime noninformative priors produced more precision in the estimates than the others (semi-informative and informative). It is surprising that the results of average estimation for unknown parameters were not stable at a different kinds of the prior distributions, which was disappointing.

Specifying informative priors on unknown parameters indicates that the researcher has knowledge about the unknown parameters. Previous information of all unknown parameters was added. It was of interest whether or not these priors were appropriate and would be able to identify the correct extraction of model parameters with ignorance about the other conditions of the proposed joint model. Fourth, the minor difference between large and small sample sizes in all conditions not only demonstrates that the Bayesian method used in the study can estimate the parameters of the proposed joint model quite adequately, even with small sample sizes but also eliminated the problem of failing to converge.

In general, the Bayesian method performed well under conditions of parameter schemes regardless of the size of the observations and the design structures of data collection. As can be seen from the findings of this dissertation, the estimation for the model parameters showed a minor difference between informative, noninformative, and semi-informative priors in each case.

Summary of the Estimation

The summary results are concerned with the point estimation of the parameters of interest, $(\alpha,\ \beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4,\ \beta_5,\ \beta_6,\ \delta,\ \gamma,\ \phi,\ \sigma^2)$ of each condition. The Bayesian method using the MCMC algorithm with and without information set on the unknown parameters estimated the model parameters accurately. The accuracies were seen across a parameter's schemes regardless of the information of prior and sample size. As a result, the estimation was most likely to identify all model parameters except α and σ^2 for all levels of information of prior distribution. The reason that the estimate of both parameters (α and σ^2) are severely misleading may be that they were placed in incorrect prior distributions. The Bayesian method, however, was accurate in certain settings.

CHAPTER V

CONCLUSION AND DISCUSSIONS

The purpose of this dissertation is to develop an R program with OpenBUGS software to obtain Bayesian estimates for the unknown parameters of the proposed joint (Gaussian-Exponential) model (Bronsert, 2009; Lin, 2011). In this dissertation, a joint model was developed to model a longitudinal outcome with an informative time jointly. The outcome distribution considered in this study was the Gaussian distribution. Also, it is assumed that time and covariates are independent of each other, and that time should be informative and exponentially distributed.

The simulation studies were conducted with six parameter schemes to observe how the results change with different parameter values, two different design structures (balanced and unbalanced), and five sample sizes, to evaluate the performance of a program by using Bayesian approach of estimation in the proposed joint model analysis with different level of information concerning the parameters of interest. Furthermore, the computing package using R with OpenBUGS was developed to handle and fit the proposed joint model in order to obtain parameter estimates to ensure the accuracy of the R package estimation for applied researchers conducting the joint model analysis. The joint models presented in this study rely on the relation among the one-step prior outcome, current time, and potential covariates. If any of these assumptions are not satisfied, the joint models proposed in this study should be considered with caution. In this chapter, I summarize and

discuss the prior distribution, convergence, and parameter estimates. Then, recommendations for future research are presented.

Prior Distribution

The main concern in applying Bayesian estimation using the MCMC algorithm is the prior distribution of unknown parameters, which was of interest in the current dissertation. As can be seen, in this dissertation, the prior distributions were placed on all unknown parameters in the proposed joint model, as specified in Chapter 3. However, determining three forms of prior information (informative, noninformative, and semi-informative) about the parameters of interest, ($\alpha,\ \beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4,\ \beta_5,\ \beta_6,\ \delta,\ \gamma,\ \phi,\ \sigma^2$), were applied through an MCMC algorithm.

Since convergence was not an issue, it was concluded that 20,000 iterations of the simulation are enough to yield the proper posterior distribution of all unknown parameters. However, the current investigation showed in the results that there was variation in the sample of the unknown parameters in each iteration of the simulation method. Thus, in calculating the posterior mean of the unknown parameters from this posterior distribution, it was found that some parameters were not accurate in some conditions. As a result, the prior distributions on these parameters should be changed; otherwise, the MCMC algorithm can be considered an inappropriate method for posterior distribution sampling.

Some results that showed up in this dissertation concerned the effect of noninformative, informative, and semi-informative priors, including:

1. The estimation performance of the unknown parameters showed a lack of difference in identifying the correct parameters for all types of priors while holding the sample size constant. It was known that informative prior provides

influence positively on the validity and accuracy of parameter estimates; it was also found that influence when noninformative prior was used.

2. It is surprising that the impact of the degree of informativeness in the current project was not stable when the sample size was constant. For example, in terms of the 97.5% confidence interval, semi-informative priors sometimes give the narrowest range than the others, and sometimes noninformative priors gave the narrowest range as compared as the others when the sample size was constant.

Some unknown parameters were more sensitive to small samples than others were. In other words, with small sample sizes, the value of the estimated parameter increased when the value of the hyperparameter on the prior distribution increased in the analysis of the model (Berger & Bernado, 1992; Lambert, Sutton, Burton, Abrams, & Jones, 2005). When applying the Bayesian approach for the unknown parameters in the joint model, it is important to consider prior distributions for two main reasons:

1. When the research was conducted with small amounts of data, the estimation of some unknown parameters becomes sensitive to the hyperparameter specified in informative priors (Gifford & Swaminathan, 1990). This suggests that researchers desiring to uphold the use of the MCMC method to assess these parameters on the proposed joint model should change the type of prior distribution on these parameters. For example, t distribution, Cauchy distribution, or any different distribution other than normal distribution. Alternatively, researchers could maintain the same type of those prior distributions but instead change the hyperparameter for each one.
2. Researchers need to take into account the effect of the priors' input on the other parameters of the joint model. It is well to consider that prior information

regarding these parameters may have an important influence when estimating other parameters in the same model.

Convergence

In order to express the representative subset of the parameter space, convergence diagnostics were performed for all unknown parameters. Within the variety of ways for testing convergence of iterative simulations or joint posterior density such as Heidelberger and Welch diagnostic (HW) tests and trace plot, for this study, the monitoring of the output of the HW test was selected. There are 12 unknown parameters in the proposed joint model that needed to be estimated, so all of them were diagnosed for the convergence because of the parameters' characteristics (Gelman et al., 2014). This particular diagnostic consists of two tests:

1. The stationary test, which determines whether the trace of simulated values arises from a stationary stochastic process.
2. The halfwidth test which, determines if there are enough iterations to estimate the mean of the process with acceptable precision.

In this dissertation, there were no problems in convergence overall with the proposed joint model data generated. Convergence results across all parameters in the model with various sample sizes were almost identical and were 1 for both tests. The HW tests equal one means passed test, and the sequences of samples have mixed, showing a good indication of representativeness of the sample in the simulation. Another way to see if the chain has converged is to see how well the chain is moving or mixing around the parameter space through visual inspection, for example, the trace-plot of the parameters mean, the density plots, and the autocorrelations plots, that are displayed and monitored for each parameter as well. In this dissertation, 20,000 iterations could represent an adequate

number of iterations for running Markov Chain Monte Carlo simulations for the estimation of unknown parameters in the proposed joint model regardless of sample sizes. The finding in the current dissertation confirms that the sequences are mixed and suggests that it is not necessary to run any more simulations.

Parameter Estimates

The R program developed for the current dissertation was tested using 120 data sets with six parameters schemes, five sample sizes, and four different observations on the proposed joint model, including three types of prior distributions, resulting in 360 simulations. Each simulation was run one and three times with 20,000 iterations each to provide one and three chains of posterior distributions of each parameter, respectively. For Bayesian inference, four important values to assess the performance of MCMC applied in this dissertation are the estimated parameters, 97.5% confidence intervals on the posterior inference for a parameter, and standard deviations. The properties of the simulation seem acceptable. That is, the sequences of the values in the chain were mixed, which means that there was no autocorrelation of simulations, indicating it was not necessary to run any more simulations.

However, when 97.5% posterior inference for a parameter of interest was introduced across sample size, parameter schemes, and prior information on unknown parameters, the true values for α and σ^2 in the proposed joint model were out of bounds for most situations. In general, for the proposed joint model, all true values of parameters except α and σ^2 were contained in the 97.5% confidence interval. The true values of unknown parameters are dependent on the parameter schemes and consistent with previous research (Alomair, 2017; Bronsert, 2009; Lin, 2011; Seo, 2015).

As expected, there was a higher level of accuracy seen in the larger sample size. However, unexpectedly, there was a higher level in the accuracy of estimations in the smaller samples as well. Inspecting the 97.5% posterior inference confidence intervals very carefully sometimes did not indicate that the narrowest interval occurred with a large sample size as was expected for some conditions regardless of the number of observations or parameter schemes. The narrowest interval sometimes happened with a small sample size. However, the posterior mean of all unknown parameters was estimated. Theoretically, the sample mean of all Markov chain Monte Carlo samples should be a reasonably good estimate whenever the mean is calculated from large sample sizes. But it was not always true for Bayesian estimation in the current dissertation. Nevertheless, it should be noted that MCMC under the Bayesian framework permits a very large amount of model flexibility when evaluation of high-dimensional integration around the unknown parameters (Gelman et al., 2014).

The other finding in the current dissertation is a demonstration of the advantage of the Bayesian approach in comparison with the frequentist approach as it treats unknown parameters as random variables. It can be seen how important the role of the posterior distribution is, providing researchers with a way of obtaining the inferential statistics in which they are interested.

Recommendation for Future Research

Carrying out further simulation studies and/or real longitudinal data investigation may shed additional light in determining the helpfulness of the Bayesian approach for the evaluation of the appropriateness of constraints that are imposed on the estimation and might clarify the comprehensive development of a statistical software program for estimating the unknown parameters within the proposed joint model in Bayesian

framework. One particular need is the continuing exploration of the impact of different informative prior distributions on the parameters.

It is essential that researchers who conduct joint model analysis employing the MCMC algorithm sampling method keep in mind the characteristics of the parameters and the distribution that matches them, in line with previous studies. For example, researchers who are interested in adapting the MCMC algorithm sampling technique should need the recommended to conduct joint model analysis that includes covariates with both small and large sample sizes for comparison purposes of their performance. Additionally, research is also indicated in the development of methods using the Bayesian approach for simulating posterior distributions.

Furthermore, the joint models carry multiple assumptions, limiting the use of these joint models. When those assumptions are relaxed, it then is possible to expand the joint models to be more flexible. Next, the researchers can attempt to find parameters estimation for the extended model with the new assumption by applying the Bayesian approach. For example:

1. The response in the proposed joint model in this study is assumed to be dependent upon the one-step prior outcome. However, the current response may depend on the two-step prior outcome or even three-step prior outcome. Next, the Bayesian approach can be applied to find parameters estimation for the extended model with the new assumption
2. Time is assumed to follow an exponential distribution. The distribution of time can be different based on the research design. If that is the situation, the appropriate distribution can be applied to the time process; then, the Bayesian

parameter estimates from the time process can be obtained. Also, this study considers the outcome variable to follow a normal distribution.

However, Seo (2015) developed and extended the current model to handle longitudinal outcomes distribution from a member of the exponential family of distributions such as Bernoulli, Poisson, and Gamma. Next, the Bayesian parameter estimates from generalized linear models can be obtained.

3. In the current model, time and covariates are assumed to be independent of each other. If they are related, another term can be added to define the relations between them in the models. However, Alomair (2017) adapted the current model to be able to incorporate informative time and time-dependent covariates with a longitudinal response. Next, the Bayesian parameter estimates from depended covariate models can be obtained.

All of the alternative assumptions mentioned above are technically possible; and can be further explored by a researcher in order to improve the joint models, and then the Bayesian approach can be applied to the improved model. Finally, Bayesian predictions are recommended as it is an important objective of the joint model besides the estimation and the testing. Bayesian predictions are outcome values simulated from the posterior predictive distribution, which is the distribution of the unobserved (future) data given the observed data.

Conclusion

This dissertation provides a demonstration into the extent to which an alternative method under the Bayesian framework was able to estimate all unknown parameters in the proposed joint model. There is a little previous research to estimate parameters of joint model in Bayesian approach which has manipulated a variety of fit indexes and tests for

determining the correct parameter estimation. In this dissertation, OpenBUGS with R code was developed to both generate data sets for the proposed joint model and analyze the data drawn using the MCMC algorithm as a sampling method to estimate the unknown parameters in joint model with different prior information. It is particularly striking that there was correct identification of the hyperparameters of prior distribution in at least some cases. This method of estimation, however, performs accurately in most conditions, and it can be considered as the pioneer of using the theory associated with a Bayesian approach on joint models in estimating the unknown parameters. Further validation employing joint models or adopting more advanced methods in both the MCMC technique and Bayesian estimations will clarify and expand the usefulness of this approach.

REFERENCES

Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.

Albert, J. (2009). *Bayesian computation with R*. Springer.

Albert, J. H. (1988). Computational methods using a Bayesian hierarchical generalized linear model. *Journal of the American Statistical Association, 83*(404), 1037–1044.

Alomair, M. A. (2017). A Joint Model of Longitudinal Data and Informative Time with Time-Dependent Covariate (Unpublished doctoral dissertation). University of Northern Colorado.

Azevedo, C. L., Fox, J. P., & Andrade, D. F. (2016). Bayesian longitudinal item response modeling with restricted covariance pattern structures. *Statistics and computing*, *26*(1–2), 443–460.

Baghfalaki, T., Ganjali, M., & Hashemi, R. (2014). Bayesian Joint Modeling of Longitudinal Measurements and Time-to-Event Data Using Robust Distributions. *Journal of Biopharmaceutical Statistics*, *24*(4), 834–855.

Bain, L. J., & Engelhardt, M. (1992). *Introduction to probability and mathematical statistics* (2nd ed.). CA: Thomson Learning Academic Resource Center.

Berger, J. O., & Bernado, J. M. (1992). On the development of reference priors. *Bayesian Statistics*, 4, 35–60.

Bernardo, J. M., & Smith, A. F. (2009). *Bayesian theory* (Vol. 405). John Wiley & Sons.

Bolstad, W. M., & Curran, J. M. (2016). *Introduction to Bayesian statistics*. John Wiley & Sons.

Bronsert, M. R. (2009). A joint model of a longitudinal process and informative time schedule data (Unpublished doctoral dissertation). University of Northern Colorado.

Brooks, S. P., & Roberts, G. O. (1998). Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing*, *8*(4), 319–335.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Carlin, B. P., & Louis, T. A. (1996). Identifying prior distributions that produce specific decisions, with application to monitoring clinical trials. *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, 493–503.

Carlin, B. P., & Louis, T. A. (2008). *Bayesian methods for data analysis*. third edition. Chapman and Hall/CRC.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, *46*(3), 167–174.

Chan, J., & Wan, Y. (2011). Bayesian approach to analysing longitudinal bivariate binary data with informative dropout. *Computational Statistics, 26,* 121-144.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *The American Statistician, 49*(4), 327–335.

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis* (No. 519.535 C6).

Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, *91*(434), 883–904.

Das, S., & Dey, D. (2006). On Bayesian analysis of generalized linear models using the Jacobian technique. *The American Statistician, 60*(3), 264–268.

Das, S., & Dey, D. (2007). *On Bayesian analysis of generalized linear models: A new perspective*. University of Connecticut, Department of Statistics.

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements* (2nd ed.). Verlag, NY: Springer.

De Finetti, B. (2017). *Theory of probability: A critical introductory treatment* (Vol. 6). John Wiley & Sons.

Depaoli, S. (2014). The impact of inaccurate informative priors for growth parameters in Bayesian growth mixture modeling. *Structural Equation Modeling A Multidisciplinary Journal, 21,* 239–252.

Diggle, P. J., Heagerty, P., Liang, K. Y., & Zeger, S. L. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford, UK: Oxford University Press.

Dobson, A. J. (2001). *An introduction to generalized linear models.* London: Chapman and Hall/CRC Press.

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(2), 355–366.

Everitt, B. S. (2006). *The Cambridge dictionary of statistics* (3rd ed.). Cambridge, UK: Oxford University Press.

Fisher, R. A. (1925). Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 22, No. 5, pp. 700–725). Cambridge University Press.

Fitzmaurice, G. M. (2008). *Longitudinal data analysis*. Boca Raton, FL: Chapman and Hall/CRC.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (Vol. 998). John Wiley & Sons.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*(2), 389–402.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain monte carlo stochastic simulation for Bayesian inference* (2nd ed.). FL: Chapman and Hall.

Gelfand, A., & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, *85*(410), 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel hierarchical models* (Vol. 1). New York, NY, USA: Cambridge University Press.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457-472.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741.

Geyer, C. J. (1992). Practical Markova Chain Monte Carlo. *Statistical Science*, 4, 473–511.

Ghosal, S., Lund, T. S., Moin, P., & Akselvoll, K. (1995). A dynamic localization model for large-eddy simulation of turbulent flows. *Journal of fluid mechanics*, *286*, 229–255.

Gifford, J. A., & Swaminathan, H. (1990). Bias and the effect of priors in Bayesian estimation of parameters of item response models. *Applied Psychological Measurement, 14*(1), 33–43.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33*(5), 587–606.

Gilks, W. R., Richardson, S., & Spiegelhaler, D. J. (1996). *Markov chain monte Carlo in practice.* NY: Chapman and Hall.

Gill, J. (2008). *Bayesian methods: A social and behavioral sciences approach* (2nd ed.). FL Chapman and Hall CRC: Boca Raton.

Hastings, W. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika, 57*(1), 97-109.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis.* Hoboken, NJ: John Wiley & Sons, Inc.

Heidelberger, P., & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research, 31*(6), 1109–1144.

Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics, 1*, 465–480.

Ho, S. Y. W., Lanfear, R., Philips, M. J., Barnes, I., Thomas, J. A., Kolokotronis, S. O., & Shapiro, B. (2011). Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Systematic Biology, 60*(3), 366–375.

Hox, J. (2002). *Quantitative methodology series. Multilevel analysis techniques and applications.* Lawrence Erlbaum Associates Publishers.

Huang, C. Y., Wang, M. C., & Zhang, Y. (2006). Analysing panel count data with informative observation times. *Biometrika, 93*(4), 763–775.

Ibrahim, J. G., & Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior. *Journal of the American Statistical Association*, *86*(416), 981–986.

Jiang, X., & Mahadevan, S. (2009). Bayesian structural equation modeling method for hierarchical model validation. *Reliability Engineering and System Safety, 94*, 796–809.

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis*. NJ: Pearson Prentice Hall.

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*, 650–673, The Guilford Press.

Kaplan, D., & Depaoli, S. (2013). Bayesian statistical methods. *Oxford handbook of quantitative methods*, 407–437.

Kim, S., Zeng, D., Chambless, L., & Li, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in biosciences*, *4*(2), 262–281.

Kruschke, J. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*(3), 299–312.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., & Jones, D. R. (2005). How vague is vague a simulation study of the impact of the use of vague prior distributions in mcmc using winbugs. *Statistics in Medicine*, *24*(5), 2401–2428.

Lee, P. M. (2012). *Bayesian statistics: An introduction* (4th ed.). United Kingdom: John Wiley and Sons.

Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society*, *58*, 619–678.

Leeuw, J. D., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, *11*, 57–85.

Leiby, B. E., Have, T. R., Lynch, K. G., & Sammel, M. D. (2014). Bayesian multivariate growth curve latent class models for mixed outcomes. *Statistics in medicine*, *33*(20), 3434–3452.

Liang, K., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22.

Liang, Y., Lu, W., & Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, *65*, 377–384.

Lin, H., Scharfstein, D. O., & Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society*, *66*, 791–813.

Lin, Y. (2011). Hypothesis testing for the Gaussian-Exponential longitudinal model (Unpublished doctoral dissertation). University of Northern Colorado.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Gelber, R., & Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics, 58*, 621–630.

Liu, C., Cripe, T. P., & Kim, M. O. (2010). Statistical issues in longitudinal data analysis for treatment efficacy studies in the biomedical sciences. *Molecular Therapy*, *18*(9), 1724–1730.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, *74*(4), 817–827.

Lu, Z. L., Zhang, Z., & Lubke, G. (2011). Bayesian inference for growth mixture models with latent class dependent missing data. *Multivariate Behavioral Research*, *46*, 567–597.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, *10*(4), 325–337.

Lynch, S. M. (2007). *Introduction to applied bayesian statistics and estimation for social scientists.* NY: Springer Science & Business Media.

Ma, Y., Mazumdar, M., & Memtsoudis, S. G. (2012). Beyond repeated-measures analysis of variance: advanced statistical methods for the analysis of longitudinal data in anesthesia research. *Regional Anesthesia & Pain Medicine*, *37*(1), 99-105.

Mengersen, K. L., & Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithm. *The Annals of Statistics*, *24*(1), 101–121.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*, 1087–1092.

Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling a more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335.

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users. *Journal of Statistical Software*, 43, 1–14.

Nelder, J. A., & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, *135*, 370–384.

Neyman, J., & Pearson, E. S. (1933). The testing of statistical hypotheses in relation to probabilities a priori. *Mathematical Proceedings of the Cambridge Philosophical Society*, *29*(4), 492–510.

Ntzoufras, I. (2009). *Bayesian Modeling Using Winbugs*. John Wiley& Sons. *Inc, Canada*.

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, *91*(10–11), 1290–1300.

O'Hagan, A., Woodward, E. G., & Moodaley, L. C. (1990). Practical Bayesian analysis of a simple logistic regression: predicting corneal transplants. *Statistics in Medicine*, *9*(9), 1091–1101.

O'Neill, P. D. (2002). A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov Chain Monte Carlo methods. *Mathematical Biosciences*, *180*, 103–114.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*, 545–554.

Pepe, M. S., & Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics*, *23*, 939–951.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-PLUS*. Verlag, New York: Springer.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R news*, *6*(1), 7–11.

Poirier, D. J. (2006). The growth of bayesian methods in statistics and economics since 1970. *Bayesian Analysis*, *1*, 969–980.

Pullenayegum, E. M., & Lim, L. S. (2016). Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. *Statistical Methods in Medical Research*, *25*(6), 2992–3014.

Qiu, F., Stein, C. M., & Elston, R. C. (2013). Joint modeling of longitudinal data and discrete-time survival outcome. *Statistical Methods in Medical Research*, 1–15.

Quintana, F. A., Johnson, W. O., Waetjen, L. E., & Gold, E. B. (2016). Bayesian nonparametric longitudinal data analysis. *Journal of the American Statistical Association*, *111*(515), 1168–1181.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Ravenzwaaij, D. V., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte Carlo sampling. *Psychonomic Bulletin & Review, 25*(1), 143–154.

Richardson, S., & Green, P. J. (1997). 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.

Roberts, G. O., & Smith, A. F. M. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications, 49*, 207–216.

Ryu, D., Sinha, D., Mallick, D., Lipsitz, S., & Lipshultz, S. (2007). Longitudinal studies with outcome-dependent follow-up: Models and Bayesian regression. *Journal of the American Statistical Association, 102*, 952–961.

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika, 64*, 37–52.

Schober, P., & Vetter, T. R. (2018). Repeated measures design and analysis of longitudinal data: if at first you do not succeed—try, try again. *Anesthesia and Analgesia, 127*(2), 569.

Schruben, L., Singh, H., & Tierney, L. (1983). Optimal tests for initialization bias in simulation output. *Operations Research, 31*(6), 1167-1178.

Seo, J. (2015). Joint models of longitudinal outcomes and informative time (Unpublished doctoral dissertation). University of Northern Colorado.

Sitthisan, C. (2016). The Bayesian Method of Estimation for the Number of Latent Classes in Growth Mixture Models (Unpublished doctoral dissertation). University of Northern Colorado.

Smith, A. F., & Robert, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society, 55*(1), 3–23.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks: Sage Publications.

Song, X., Mu, X., & Sun, L. (2012), Regression Analysis of Longitudinal Data with Time-Dependent Covariates and Informative Observation Times. *Scandinavian Journal of Statistics, 39*, 248–258.

Spiegelhalter, D. J., Myles, J. P., Jones, D. R., & Abrams, K. R. (2000). Bayesian methods in health technology assessment: a review. *Health Technology Assessment, 4*(38), 1–130.

Stevens, G. V. G. (1998). On the Inverse of the Covariance Matrix in Portfolio Analysis. *The Journal of Finance*, *53*, 1821–1827.

Student. (1908). The probable error of a mean. *Biometrika*, 1–25.

Sturtz, S., Ligges, U., & Gelman, A. E. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software, 12*(3), 1–16.

Sun, J., Park, D. H., Sun, L., & Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association*, *100*, 882–889.

Sweeting, T. (1981). Scale Parameters: A Bayesian Treatment. *Journal of the Royal Statistical Society. Series B (Methodological), 43*(3), 333-338.

Tabachnick, B. G. (2007). *Using Multivariate Statistics* (5th ed.). MA: Pearson Education.

Tabachnick, B. G., & Fidell, L. S. (2012). Using Multivariate Statistics. 6h Edition, Person Education, Boston.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, *22*(4), 1701–1728.

Twisk, J. W. (2013). *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide*. Cambridge University Press.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2019). Rank-normalization, folding, and localization: An improved $\widehat {R} $ for assessing convergence of MCMC.

Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics, Springer-Verlag, New-York, 568 pages. ISBN 0–387–95027–3 (hard cover), first printing in 2000, second printing in 2001.

West, M., Harrison, P. J., & Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, *80*(389), 73–83.

Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in Statistics-Simulation and Computation*, *22*(4), 1079–1106.

Wu, L., Liu, W., Yi, G. Y., & Huang, Y. (2012). Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 1–17.

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 1049–1060.

Zhang, Z., Hamagami, F., Wang, L., Nesselroade, J. R., & Grimm, K. J. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, *31*(4), 374–383.

APPENDIX A

R CODES

```
#######################################################################
# Packages #
#######################################################################
install.packages('R2OpenBUGS')          # provide rep(dat,times)
install.packages('MASS')                # provide rep(dat,times)
install.packages('coda')                # provide rep(dat,times)
install.packages('maxLik')              # maxLik
install.packages('AlgDesign')           # gen.factorial
install.packages('mefa')                # provide rep(dat,times)
install.packages('doParallel')          # provide rep(dat,times)
install.packages('parallel')            # provide rep(dat,times)

library(R2OpenBUGS)
library(MASS)
library(coda)
library(maxLik)
library(AlgDesign)
library(mefa)
library(doParallel)
library(parallel)

#######################################################################
# Parameter Setting (Pscheme: 1 to 6) #
#######################################################################
parameter = matrix(c(1,1,2,2,0.5,0.5, #1:sigma
                     0.4,0.4,0.4,0.4,0.4,0.4, #2:beta0
                     0.2,0.2,0.2,0.2,0.2,0.2, #3:beta1
                     0.3,0.3,0.3,0.3,0.3,0.3, #4:beta2
                     0.1,0.1,0.1,0.1,0.1,0.1, #5:beta3
                     0.3,0.3,0.3,0.3,0.3,0.3, #6:beta4
                     0.4,0.4,0.4,0.4,0.4,0.4, #7:beta5
                     0.9,0.9,0.9,0.9,0.9,0.9, #8:beta6
                     0.8,0.8,0.8,0.0,0.0,0.8, #9:phi
                     0.1,0.1,0.1,0.1,0.1,0.1, #10:gamma
                     2,1,2,1,2,1, #11:alpha
                     0.01,0.02,0.01,0.02,0.01,0.02),#12:delta
                   nrow=6)

#####################################################################
# create design matrix (X) with two cat & two cont vars #
#####################################################################
design=function(level=c(3,3),m=18,c=2){
  catg=gen.factorial(levels=level,center=FALSE,factors='all')
  ext=rep(catg,m/(prod(level)))
  des=model.matrix(~.,data=ext) #'~.' is supported by {AlgDesign}
  cont=data.frame(matrix(NA,nrow=m,ncol=c))
  for (i in 1:c){
```

```
   cont[i]=rnorm(m)
  }
  xmatrix=as.matrix(cbind(des,cont))
  xmatrix
}

#####################################################################
# Create Data: c('outcome','time','subject') #
#####################################################################
outcome<- function(m=m,num=num,parm=parm){
  if (num == 1) {n1 = 10; n2=10}
  if (num == 2) {n1 = 5; n2=3}
  if (num == 3) {n1 = 10; n2=5}
  if (num == 4) {n1 = 20; n2=6}
  ndesign = matrix(c(rep(n1,m/2),rep(n2,m/2)),byrow=T)
  nn=cumsum(c(1,ndesign[-length(ndesign)]))
  raw = matrix(NA,sum(ndesign),3) #Null matrix
  mu = xmatrix %*% parm[2:8] # mu is matrix
  raw[nn,1]= mu + rnorm(m)*parm[1]
  raw[nn,2] = rexp(m)
  for (i in 1:m){
    for (j in 2:ndesign[i]){
      yjmin1 = raw[nn[i] - 1 + j - 1,1]
      raw[nn[i] - 1 + j,2] = rexp(1)*
        exp(parm[11] +parm[12] * yjmin1)
      raw[nn[i] - 1 + j,1] =mu[i] + yjmin1 * parm[9] +
        raw[nn[i]-1+j,2]*parm[10]+rnorm(1)*parm[1]
      raw[nn[i],3]=i
      raw[nn[i]-1+j,3]=i
    } #j
  }#i
  result=list(raw=raw,nn=nn,ndesign=ndesign)
  result
} #outcome

#############################################################
# The Bayesian Model   #   First scenarios
#############################################################
## Instead of writing the model in a text editor, we can enter it in R script:

sink("bayesmod1.txt")
cat("
model{
  for (i in 1: m){
    y[nn[i]] ~ dnorm(mu[nn[i]], tau)    #initial obs for each subjects
      mu[nn[i]] <- inprod(xmatrix[i,], beta[])
```

```
    for (j in 2:ndesign[i]){
      y[nn[i]+(j-1)] ~ dnorm(mun[nn[i]+(j-1)], tau)
      t[nn[i]+(j-1)] ~ dexp(theta[nn[i]+(j-1)])
      mun[nn[i]+(j-1)] <- gamma * t[nn[i]+(j-1)] + phi * y[nn[i]+(j-2)] + mu[nn[i]]
      theta[nn[i]+(j-1)] <- alpha + delta * y[nn[i]+(j-2)]

    }
  }
    ####    Prior Distribution     ####
    ## First scenario: Informative Prior ##

    #Priors on the coefficients of covariates
    for (k in 1: p+1){
       beta[k] ~ dnorm(0.4, 4.0)
    }
    gamma ~ dnorm (0.5, 0.5)
    phi ~ dnorm (0.2, 0.2)
    alpha ~ dnorm (2.0, 0.2)
    delta ~ dnorm (0.2, 0.1)
    tau ~ dgamma(0.2, 0.2)
    sigma <- 1/tau    # sigma: variance of the normal distribution
}

", fill=TRUE)
sink()

#################################################################
# The Bayesian Model   #   Second scenarios
#################################################################
## Instead of writing the model in a text editor, we can enter it in R script:

sink("bayesmod2.txt")
cat("
model{
  for (i in 1: m){
    y[nn[i]] ~ dnorm(mu[nn[i]], tau)    #initial obs for each subjects
       mu[nn[i]] <- inprod(xmatrix[i,], beta[])

    for (j in 2:ndesign[i]){
      y[nn[i]+(j-1)] ~ dnorm(mun[nn[i]+(j-1)], tau)
      t[nn[i]+(j-1)] ~ dexp(theta[nn[i]+(j-1)])
      mun[nn[i]+(j-1)] <- gamma * t[nn[i]+(j-1)] + phi * y[nn[i]+(j-2)] + mu[nn[i]]
      theta[nn[i]+(j-1)] <- alpha + delta * y[nn[i]+(j-2)]

    }
  }
    ####    Prior Distribution     ####
```

```
   ## Second scenario: Noninformative Prior ##

   #Priors on the coefficients of covariates
   for (k in 1: p+1){
      beta[k] ~ dnorm(0.0, 1.0E-6)
   }
   gamma ~ dnorm (0.0, 1.0E-6)
   phi ~ dnorm (0.0, 1.0E-6)
   alpha ~ dnorm (0.0, 1.0E-6)
   delta ~ dnorm (0.0, 1.0E-6)
   tau ~ dgamma(0.001, 0.001)
   sigma <- 1/tau    # sigma: variance of the normal distribution
}

", fill=TRUE)
sink()

############################################################
# The Bayesian Model   #   Third scenarios
############################################################
## Instead of writing the model in a text editor, we can enter it in R script:

sink("bayesmod3.txt")
cat("
model{
  for (i in 1: m){
    y[nn[i]] ~ dnorm(mu[nn[i]], tau)    #initial obs for each subjects
       mu[nn[i]] <- inprod(xmatrix[i,], beta[])

    for (j in 2:ndesign[i]){
      y[nn[i]+(j-1)] ~ dnorm(mun[nn[i]+(j-1)], tau)
      t[nn[i]+(j-1)] ~ dexp(theta[nn[i]+(j-1)])
      mun[nn[i]+(j-1)] <- gamma * t[nn[i]+(j-1)] + phi * y[nn[i]+(j-2)] + mu[nn[i]]
      theta[nn[i]+(j-1)] <- alpha + delta * y[nn[i]+(j-2)]

    }
  }
    ####    Prior Distribution     ####
    ## Third scenario: Semi-informative Prior ##

    #Priors on the coefficients of covariates
    for (k in 1: p+1){
       beta[k] ~ dnorm(0.4, 4.0)
    }
    gamma ~ dnorm (0.0, 1.0E-6)
    phi ~ dnorm (0.0, 1.0E-6)
    alpha ~ dnorm (0.0, 1.0E-6)
```

```
   delta ~ dnorm (0.0, 1.0E-6)
    tau ~ dgamma(0.2, 0.2)
   sigma <- 1/tau    # sigma: variance of the normal distribution

}
", fill=TRUE)
sink()

############################################################
# Simulation #
############################################################
#Pschem = r  # parameter schemes, 1 to 6

m = 18      # sample sizes, 18, 36, 54, 90, 180
num = 4    # design structure 1(10,10), 2(5,3),3(10,5),4(20,6)
p = 6         # the explanatory variables.

xmatrix=design(level=c(3,3),m=m,c=2)

fsim<-function(){
  out = array(NA,c(6,ncol(parameter), 4))
  for (r in 1:6){
    parm = parameter[r,]

   # compute some info to be used in optimization
    result=outcome(m=m,num=num,parm=parm)
    y=c(result$raw[,1])
    t=c(result$raw[,2])
    nn=c(result$nn)
    ndesign=c(result$ndesign)
    p = 6  # the explanatory variables.

    # Read in the data frame for BUGS:
    sim.dat.bugs <- list( "y" , "m" , "p" , "t", "nn", "ndesign", "xmatrix")

    ## Define the parameters whose posterior distributions we are
    ## interested in summarizing:
    bayes.mod.params <- c("sigma", "beta", "phi", "gamma", "alpha", "delta", "tau")

    ## Define the starting values for BUGS.
    bayes.mod.inits <- function(){
      list("tau" = parm[1], "beta" = parm[2:8], "phi" = parm[9],
           "gamma" = parm[10], "alpha" = parm[11], "delta" = parm[12])
    }

    ## Now, we are ready to use the bugs() function, which calls OpenBUGS.
    ## We have to specify the location of the model file, the data, the parameters,
```

```
   ## the initial values, as well as how many chains we want to fit and how long
   ## we want to run them.

   bayes.mod.fit.R2OpenBUGS <- bugs( data = sim.dat.bugs,
                              model.file = "bayesmod1.txt",
                              parameters.to.save = bayes.mod.params,
                              inits = bayes.mod.inits,
                              n.chains = 1,
                              n.iter = 20000,
                              n.burnin = 10000,
                              n.thin = 1,
                              debug=FALSE,
                              codaPkg=TRUE)

   code.object <- read.bugs(bayes.mod.fit.R2OpenBUGS)
   ## Heidelberger and Welch Convergence Diagnostic
   hw <- heidel.diag(code.object)[[1]][,1][-c(10,14)]

   hw1[hw==""]<-0
   hw1[is.na(hw)] <- 0

   ## 1 = pass the stationarity test and halfwidth test
   ## 0 = failure of the chain to pass
   if (sum(hw)==12){
     Bayes.Est1 <- summary(code.object)$statistics[,c("Mean","SD")][-c(10,14),]
     Bayes.Est2 <- summary(code.object)$quantiles[,c("2.5%","97.5%")][-c(10,14),]
     Bayes.Est <- as.matrix(cbind(Bayes.Est1,Bayes.Est2))
   } else {
     Bayes.Est <- NA
   }
   out[r,,]=Bayes.Est
  }
  return(out)
}

cl <- makeCluster(32)                  ##########################################
registerDoParallel(cl)          ## TO MAKE SIMULATION FASTER THAN USUAL ##
pack<-c("R2OpenBUGS","coda")    ##########################################
rep=1000

system.time({ results<-foreach(r=1:rep, .packages=pack) %dopar% fsim() })

results
```