Hokkaido University

Doctoral Thesis

Bayesian Analysis on Ordered Probit Models with Individual Heterogeneity

Author:
Lei Shi

Supervisor:
Hikaru Hasegawa

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Economics

in the

Graduate School of Economics and Business

December 24, 2019

Abstract

Bayesian Analysis on Ordered Probit Models with Individual Heterogeneity

by Lei Shi

In their seminal study of the Bayesian ordered-probit model, Albert and Chib (1993) use latent variable representation to estimate the posterior distribution of explanatory variables. Cowles (1996), Nandram and Chen (1996), and Chen and Dey (2000) offer several suggestions on the sampling of cutpoints. Albert and Chib (2001) consider an algorithm to transform cutpoints into an unconstrained variable and remove their ordering constraint using a one-to-one map. Chen and Dey (2000) propose a sampling algorithm identified by fixing two cutpoints, leaving for estimation the variance in error term. Moreover, Müller and Czado (2005) present a Bayesian auto regressive ordered probit model based on the Liu and Sabatti (2000) method. Further, Hasegawa (2009) extends these algorithms to the Bayesian dynamic ordered probit model.

In an analysis of subjective well-being, the backgrounds and preferences of respondents are complicated, and individuals may interpret concepts in different ways (Brady, 1985). Many studies have considered ordered choice models, incorporating heterogeneity among individuals (Farewell (1982); Terza (1985); Greene (2007); Eluru et al. (2008); Greene and Hensher (2010b); King et al. (2004); Kapteyn et al. (2007)). However, to the best of our knowledge, no study has so far accommodated individual heterogeneity in the cutpoints of ordered probit models.

Therefore, we first present a sampling algorithm for the univariate ordered probit model with individual heterogeneity and apply the model to happiness data. Our findings indicate that the model with individual heterogeneity act better than that without heterogeneity do. In some instances, we need to model several related ordered responses from the same individuals. Since these ordered responses are correlated, we need to use a multivariate model, rather than several separate univariate models. Therefore, we propose a Bayesian multivariate ordered probit model with individual heterogeneity to meet the requirements. In addition, to deal with panel data, we conduct a new Bayesian algorithm in a dynamic ordered probit model.

KEY WORDS: Bayesian analysis, Markov chain Monte Carlo (MCMC), Individual heterogeneity, Ordered probit model.

Acknowledgements

First of all, I would like to express my great appreciation to Prof. Hasegawa, Prof. Takagi and Prof. Kakizawa for their patient guidance and excellent advice during my Phd career. Prof. Hasegawa has spent much time reading through each chapter and given me a lot of suggestions for revising this thesis. Without his support from various aspects, I could not complete this thesis. I also thank the China Scholarship Council for support.

I also would like to thank my beloved family for their support and encouragement. I thank my parents for giving me the chance to study in Japan and thank my husband for understanding me and giving me so much love.

Finally, I would like to thank all my friends in Hokkaido. Thank you for giving me so much fun in my life.

Contents

Abstract **i**

Acknowledgements **iii**

List of Tables **viii**

1 Introduction **1**

1.1 Background 1

1.2 Thesis Outline 6

2 Bayesian Univariate Ordered Probit Model with Individual Heterogeneity* **8**

2.1 Introduction 8

2.2 Literature Review 10

2.3 Model Specification 13

2.3.1 Univariate Ordered Probit Model 13

2.3.2 Accommodating Individual Heterogeneity 16

2.3.3 Univariate Ordered Probit Model with Individual Heterogeneity 18

2.4 An Empirical Study 20

2.4.1 Data 20

2.4.2 The Posterior Results 22
2.5 Summary 34

3 Bayesian Multivariate Ordered Probit Model with Individual Heterogeneity 36

3.1 Introduction 36
3.2 Model Specification 39
3.2.1 Multivariate Ordered Probit Model 39
3.2.2 Multivariate Ordered Probit Model with Individual Heterogeneity 43
3.3 An Empirical Study 46
3.3.1 Data 46
3.3.2 Posterior Results 48
3.4 Summary 60

4 Bayesian Dynamic Panel-ordered Probit Model with Individual Heterogeneity 61

4.1 Introduction 61
4.2 Model Specification 63
4.2.1 Dynamic Panel-ordered Probit Model 63
4.2.2 Dynamic Panel-ordered Probit Model with Individual Heterogeneity 66
4.3 An Empirical Study 68
4.3.1 Data 68
4.3.2 Posterior Results 69
4.4 Summary 73

5 Conclusion **75**

5.1 Summary 75

5.2 Remaining Issues and Future Works 77

A Derivation in Chapter 2 **87**

A.1 Derivation of Algorithm 2 87

A.1.1 Full Conditional Distribution of $\boldsymbol{\beta}$ 87

A.1.2 Full Conditional Distribution of $\boldsymbol{y}^*$ 88

A.1.3 Full Conditional Distribution of $\boldsymbol{\gamma}$ 88

A.2 Marginal Likelihood 90

A.3 Marginal Effects 94

B Derivation in Chapter 3 **95**

B.1 Derivation of Algorithm 1 95

B.1.1 Full Conditional Distribution of $\boldsymbol{\beta}$ 95

B.1.2 Full Conditional Distribution of $\boldsymbol{\Sigma}^{-1}$ 97

B.1.3 Full Conditional Distribution of $\boldsymbol{y}^*$ 97

B.1.4 Full Conditional Distribution of $\boldsymbol{\gamma}$ 98

B.2 Derivation of Algorithm 2 99

B.2.1 Full Conditional Distribution of $\boldsymbol{\beta}$ 99

B.2.2 Full Conditional Distribution of $\boldsymbol{\Sigma}^{-1}$ 101

B.2.3 Full Conditional Distribution of $\boldsymbol{y}^*$ 101

B.2.4 Full Conditional Distribution of $\boldsymbol{\gamma}$ 102

B.3 Derivation of the marginal likelihood on Algorithm 1 103

B.4 Derivation of the marginal likelihood on Algorithm 2 111

C Derivation in Chapter 4 **120**

C.1 Derivation of Algorithm 1 . . . 120

C.1.1 Full Conditional Distribution of ϕ . . . 120

C.1.2 Full Conditional Distribution of $\boldsymbol{\beta}$ . . . 121

C.1.3 Full Conditional Distribution of $\boldsymbol{\delta}$ . . . 122

C.1.4 Full Conditional Distribution of $\boldsymbol{\beta}_0$ . . . 123

C.1.5 Full Conditional Distribution of $\boldsymbol{\delta}_0$ . . . 123

C.1.6 Full Conditional Distribution of σ^2 . . . 124

C.1.7 Full Conditional Distribution of $\boldsymbol{y}_0^*$ . . . 125

C.1.8 Full Conditional Distribution of $\boldsymbol{y}^*$ . . . 126

C.1.9 Full Conditional Distribution of $\boldsymbol{\zeta}$ . . . 127

C.2 Derivation of Algorithm 2 . . . 128

C.2.1 Full Conditional Distribution of ϕ . . . 128

C.2.2 Full Conditional Distribution of $\boldsymbol{\beta}$ . . . 129

C.2.3 Full Conditional Distribution of $\boldsymbol{\delta}$ . . . 130

C.2.4 Full Conditional Distribution of $\boldsymbol{\beta}_0$ . . . 131

C.2.5 Full Conditional Distribution of $\boldsymbol{\delta}_0$ . . . 131

C.2.6 Full Conditional Distribution of σ^2 . . . 132

C.2.7 Full Conditional Distribution of $\boldsymbol{y}_0^*$ . . . 133

C.2.8 Full Conditional Distribution of $\boldsymbol{y}^*$ . . . 134

C.2.9 Full Conditional Distribution of $\boldsymbol{\zeta}$ . . . 135

C.3 The Partial Effects of the Explanatory Variables on the Response Probabilities . . . 136

C.3.1 The Partial Effects of Algorithm 2 . . . 136

List of Tables

2.1 The happiness level in the United States (frequency) 23

2.2 Descriptive statistics of data in the United States 23

2.3 The happiness level in Canada (frequency) 24

2.4 Descriptive statistics of data in Canada 24

2.5 The happiness level in Australia (frequency) 24

2.6 Descriptive statistics of data in Australia 25

2.7 Summary of marginal likelihood 27

2.8 Summary of Bayes factor 28

2.9 The posterior results of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for the United States, with individual heterogeneity, and combining father and mother immigration as parent immigration 29

2.10 The posterior results of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for Canada, with individual heterogeneity, and combining father and mother immigration as parent immigration 30

2.11 The marginal effects for the United States, with individual heterogeneity, and combining father and mother immigration as parent immigration 31

2.12 The marginal effects for Canada, with individual heterogeneity, and combining father and mother immigration as parent immigration 31
2.13 The posterior results of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for Australia, with individual heterogeneity, and dividing immigration into father and mother immigration 32
2.14 The marginal effects for Australia, with individual heterogeneity, and dividing immigration into father and mother immigration 33

3.1 Descriptive statistics of explanatory variables in the United States 49
3.2 Descriptive statistics of explained variables in the United States 50
3.3 Descriptive statistics of explanatory variables in Canada 51
3.4 Descriptive statistics of explained variables in Canada 52
3.5 Descriptive statistics of explanatory variables in Australia . . . 53
3.6 Descriptive statistics of explained variables in Australia 54
3.7 Summary of Bayes factor 54
3.8 The posterior results with individual heterogeneity in Australia 56
3.9 The posterior results with individual heterogeneity in Canada . 57
3.10 The posterior results with individual heterogeneity in the United States 58

4.1 Summary statistics 70
4.2 The posterior results for parameters (CFPS data) 72
4.3 The posterior results for partial effects on response probabilities (CFPS data) 74

Chapter 1

Introduction

1.1 Background

In the research field concerning subjective well-being, outcomes are often arranged ordinally. The ordered probit model is a useful approach to estimate models with this type of data. According to frequentists, an ordered probit model can be estimated using the maximum likelihood method. However, with the development of computer science and the Markov chain Monte Carlo (MCMC) method, Bayesian estimation of the ordered probit model has also become popular.

Compared with the non-Bayesian approach to the ordered probit model, the Bayesian approach has several merits. First, as described in Hasegawa and Ueda (2016, p.359), while the Bayesian method allows the estimation of the multivariate ordered probit model with correlation between explained variables, it is difficult to apply the non-Bayesian approach when the number of explained variables is greater than two.[1] Second, the representation of latent

[1] For example, if we use STATA for such an estimation, it is possible to estimate a bivariate ordered probit model using BIOPROBIT, which computes maximum likelihood estimates, but there is no function to estimate ordered probit models with three or more variables. See Hasegawa and Ueda (2016, p.359, footnote).

variables is particularly useful in the estimation of ordered probit models. The values of the latent variables can easily be obtained from the posterior distribution of the Bayesian sampling algorithm.

Following Jeliazkov et al. (2008), there are three main algorithms for the Bayesian analysis of the ordered probit model. The seminal study is Albert and Chib (1993), where the authors proposed a Bayesian sampling algorithm in the univariate ordered probit model. They employ the latent variable representation to estimate the posterior results of the explanatory variables. One shortcoming of their algorithm is that the sampling of the cutpoints is not conditioned on the latent data. Later, Cowles (1996), Nandram and Chen (1996), and Chen and Dey (2000) offered several suggestions on the sampling of cutpoints. Further to these studies, Albert and Chib (2001) considered an algorithm to transform cutpoints into an unconstrained variable and remove their ordering constraint using a one-to-one map. Thus, the proposal density for a new variable can be acquired by unconstrained optimization, with the prior of the variable set as unrestricted. Subsequently, Chen and Dey (2000) proposed a sampling algorithm identified by fixing two cutpoints, leaving for estimation the variance in error term.

Besides the main algorithms mentioned above, Müller and Czado (2005) present a Bayesian auto regressive ordered probit model based on the Liu and Sabatti (2000) method. Further, Hasegawa (2009) extends these algorithms to the Bayesian dynamic ordered probit model. From the result obtained in Hasegawa (2009) using simulated data, the algorithms following Liu and Sabatti (2000) and Chen and Dey (2000) perform well.

In an analysis of subjective well-being, the backgrounds and preferences of respondents are complicated, and individuals may interpret concepts in different ways (Brady, 1985). Many studies have considered ordered choice models, incorporating heterogeneity among individuals. The authors of Farewell (1982) proposed an ordered Weibull model, and Terza (1985) considered an alternative generalized model, in which the explanatory variables in the cutpoints regression are considered to vary across outcomes. Later, in his study on three outcomes, Terza (1985) assumed that the explanatory variables in the cutpoints regression are the same with the explanatory variables in the generalized model. However, a linear specification of thresholds in Terza (1985) model cannot ensure the cutpoints always take positive values. Subsequently, Greene (2007), Eluru et al. (2008), and Greene and Hensher (2010b) proposed a hierarchical ordered probit model, where the nonlinear specification of thresholds in the model ensures the cutpoints always take positive values. The authors of King et al. (2004) and Kapteyn et al. (2007) suggested that the response scales differ systematically between groups. Therefore, they employed the anchoring vignettes method to adjust for these differences.

Although the method of King et al. (2004) cannot be applied to the survey data from World Values Survey and many other data, their idea lead us to consider that individuals with different backgrounds may self-assess happiness using different criteria. Thus, individuals in different groups should not be treated using the same thresholds in the ordered probit model. However, to the best of our knowledge, no study has so far accommodated individual heterogeneity in the cutpoints of ordered probit models. Therefore, we first present a sampling algorithm for the univariate ordered probit model with individual heterogeneity and apply the model to happiness data. Our findings indicate that the model with individual heterogeneity act better than that without heterogeneity do. In some instances, we need to model several related ordered responses from the same individuals, such as responses related to happiness and self-assessed health. Since these ordered responses are correlated, we need to use a multivariate model, rather than several separate univariate models. Therefore, we propose a Bayesian multivariate ordered probit model with individual heterogeneity to meet the requirements. In addition, to deal with panel data, we conduct a new Bayesian algorithm in a dynamic ordered probit model.

We apply the new proposed algorithms to real data. In the first empirical study, we analyze happiness levels in the United States, Canada, and Australia using data from the World Values Survey (WVS), Wave 5, and accommodate immigration and religion as individual heterogeneity in the threshold model. We calculate the marginal likelihood and Bayes factor. The results indicate that the Bayesian univariate ordered probit models that include individual heterogeneity perform better than those without individual heterogeneity do. The posterior results for the explanatory variables are the same as those of other empirical studies; that is, being married, earning a high income, and having a high education level have positive effects on happiness, while being male has a negative effect on happiness in Canada. Furthermore, we find that the effects of heterogeneity vary between these three countries. Belonging to a religion has an evident impact on the threshold in the United States, but has no evident effects in Canada or Australia. In addition, parents' immigration status can affect the threshold in Australia, but shows no evident effects in the United States or Canada.

Subsequently, we apply the algorithm from Chen and Dey (2000) and the new proposed Bayesian multivariate ordered probit model with individual heterogeneity to the WVS data that is same with our first empirical study. We choose three immigration countries, namely Australia, Canada, the United States, and study individuals' life satisfaction, state of health, interest in politics, and freedom of choice as the correlated responses. Furthermore, we accommodate parent immigration status and religious status as individual heterogeneity factors in the thresholds in the new algorithm. The results for the Bayes factor indicate that the models with individual heterogeneity outperforms that without heterogeneity. Moreover, the effects of heterogeneity are different among the three countries. Belonging to a religion has an evident impact on the state of health and freedom of choice in Australia, while it has significant effects on the state of health and life satisfaction in the United States. However, our findings also suggest that parent immigration status cannot affect the thresholds in all three countries.

Finally, we apply the algorithm with individual heterogeneity in Bayesian dynamic panel ordered probit model to subjective well-being in China. We employ panel data taken from the China Family Panel Studies (CFPS), which launched the baseline survey in 2010 and three waves of their follow-up surveys in 2012, 2014, and 2016. We select marriage, income, and work as the time-variant explanatory variables, with age, gender, and education year as the time-invariant variables. Furthermore, we incorporate age, gender, and education year as individual heterogeneity factors in the cutpoints function. From the posterior results, the lagged latent variable increases life satisfaction; marriage, work, and age have positive effects on life satisfaction for all periods; and education year increases life satisfaction initially. Also, age has evident impacts on the thresholds, but education year and gender have no impact.

1.2 Thesis Outline

In the forthcoming chapters we propose three new Bayesian algorithms with individual heterogeneity in the univariate, multivariate and dynamic panel-ordered probit model. Each chapter begins by introducing the standard Bayesian ordered probit model and moves to present a new Bayesian method. Furthermore, We evaluate the performance of new algorithms using real data to demonstrate the practical use of the new algorithms. The focus and contributions of each chapter of this thesis are:

Chapter 2. We introduce the Bayesian approach to the ordered probit model from Albert and Chib (2001). We clarify the necessity of accommodating individual heterogeneity in subjective well-being and present a new sampling algorithm for the univariate ordered probit model. Then, we apply the new Bayesian univariate ordered probit model to happiness data on Australia, Canada, and the United States, with immigration status and religion status reflecting individual heterogeneity in the threshold model. We present our findings for the United States, Canada, and Australia using the WVS data. Lastly, we summary this chapter.

Chapter 3. Building on the model presented in chapter 2, we propose a new algorithm that includes individual heterogeneity in the cutpoints function of Bayesian multivariate ordered probit model based on Chen and Dey (2000). We present the comparison of the algorithm from Chen and Dey (2000) and the new proposed algorithm using real data from World Values Survey wave 5. Importantly, We conclude from the empirical results that the model with individual heterogeneity outperforms that without heterogeneity.

Chapter 4. Following the Bayesian dynamic panel-ordered probit model described in Hasegawa (2009), we develop an algorithm with individual heterogeneity in the cutpoints function. We conduct an empirical study using panel data of China. Finally, we summarise this chapter's concluding remarks.

Chapter 5. This concluding chapter summarises the contributions of this thesis, and discuss potential paths for future research.

Chapter 2

Bayesian Univariate Ordered Probit Model with Individual Heterogeneity*

2.1 Introduction

A wide range of factors can affect happiness, including marital status, education, age, income, gender, employment, number of children, personality, self-reported health status, and social relationships (Argyle, 2001; Diener et al., 1999, 2009). In an empirical study, Frey and Stutzer (2002) establish a link between happiness and economics, showing that economic conditions such as unemployment and inflation decrease individual happiness. Moreover, recent studies have shown that religion and immigration are relevant to happiness. Bartram (2011) finds that immigrants are less happy than natives, after controlling for other variables. Safi (2010) obtains the same finding as Bartram (2011), and also examines the generation effect; that is, second generations

[0]This chapter is based on Shi and Hasegawa (2019).

whose parents are immigrants seem to be at least as dissatisfied with their life as the first-generation members were. Koenig et al. (2001) examine the statistical association between religion and life satisfaction, and find that 79% of studies show a positive association between the two. Furthermore, the results of Van Cappellen et al. (2016) indicate that religion improves subjective well-being through self-transcendent positive emotions (awe, gratitude, love, and peace).

Because happiness data are usually ordered, the ordered probit model is a useful approach for statistical analyses of such data. In this study, we conduct empirical analyses on happiness data using the Bayesian ordered probit model, focusing on the following two issues: (1) the impact of marital status, education level, age, income, gender, and labor force participation on happiness; and (2) how father's immigration status, mother's immigration status, and religion affect the thresholds in the ordered probit model.

In our empirical study, we analyze happiness levels in the United States, Canada, and Australia using data from the World Values Survey (WVS), Wave 5, and accommodate immigration and religion as individual heterogeneity in the threshold model. The empirical results indicate that the models that include individual heterogeneity perform better than those without individual heterogeneity do. The results for the explanatory variables are the same as those of other empirical studies; that is, being married, earning a high income, and having a high education level have positive effects on happiness, while being male has a negative effect on happiness in Canada. Furthermore, we find that the effects of heterogeneity vary between these three countries. Belonging to a religion has an evident impact on the threshold in the United States, but has no evident effects in Canada or Australia. In addition, parents' immigration status can affect the threshold in Australia, but shows no evident effects in the United States or Canada.

2.2 Literature Review

Blanchflower and Oswald (2004) study happiness in the United States and Britain, but regard happiness as life satisfaction in the study of Britain. The determinants of individuals' happiness tend to be similar across countries. According to Blanchflower and Oswald (2004), those who are female, married, and employed, as well as those who earn a higher income, have a higher level of education, and are not from a divorced family obtain higher levels of happiness.

Many studies have found that religious behavior is positively related to subjective well-being (Witter et al., 1985; Ellison, 1991; Diener et al., 1999; Koenig et al., 2001; Argyle, 2001; Ellison et al., 2001; Diener et al., 2009; Kortt et al., 2015; Van Cappellen et al., 2016). Although most studies have been carried out in the United States, an increasing number of non-U.S.-based studies report a similar positive association between religious activities and happiness (Inglehart, 2000; Abdel-Khalek and Lester, 2009; Fisher, 2013). Clark and Lelkes (Clark and Lelkes) report that religious behavior is positively correlated with individual life satisfaction, even after controlling for social capital and other personal characteristics. Koenig et al. (2001) examine the statistical association between religion and life satisfaction, and find that 79% of studies show a positive association between the two. Joshanloo and Weijers (2016) indicate that religious belief alleviates the passive effects of income inequality on life satisfaction. Furthermore, Cohen and Johnson (2017) discuss how religion is related to unhappiness, depression, and dissatisfaction with life, and find that religion can ameliorate the pain of loss. A recent paper by Van Cappellen et al. (2016) shows that religion improves subjective well-being through self-transcendent positive emotions (awe, gratitude, love, and peace).

Previous studies also report that immigration is related to individuals' happiness. Werkuyten and Nekuee (1999), Bălţătescu (2007), Safi (2010), Bartram (2011), Gokdemir and Dumludag (2012), Obućina (2013), Hendriks (2015), and Kóczán (2016) find that immigrants do not reach similar levels of happiness to those of natives. Bartram (2011) shows that immigrants are less happy than natives, after controlling for other variables. Safi (2010) examines the generation effect, finding that second generations whose parents are immigrants seem to be at least as dissatisfied with their life as the first-generation members were. Stillman et al. (2012) use a natural experiment to compare successful and unsuccessful applicants to a migration lottery to experimentally estimate the impact of migration on objective and subjective well-being. They find that international migration improves objective well-being in terms of income and expenditure, but decreases individuals' happiness. Hwang et al. (2011) examine the short-term impact of involuntary migration and discover negative changes in migrants' well-being, despite a relative improvement in housing quality. The findings of Gatina (2016) indicate that immigrants to Australia have a lower level of life satisfaction compared to that of the native population. Furthermore, Gatina (2016) shows that an immigrant's cultural background is one of the crucial factors affecting his/her life satisfaction. Several psychology and sociology studies find that racial discrimination has negative impacts on the psychological distress and mental health of immigrants (Vega and Rumbaut, 1991; Finch et al., 2000; Taylor and Turner, 2002; Sellers et al., 2003), which reduces their happiness.

One issue we need to address here is that individuals understand concepts in different ways (Brady, 1985), leading them to feel and self-assess happiness in quite different ways. King et al. (2004) and Kapteyn et al. (2007) consider that the response scales differ systematically between groups. As a result, their studies use the anchoring vignettes method to adjust for these differences. Although the method of King et al. (2004) cannot be applied to the WVS data, their idea leads us to consider that individuals with different backgrounds may self-assess happiness using different criteria. That is, individuals in different groups should not be treated using the same thresholds in the ordered probit model.

Several studies show that religiosity affects happiness through positive emotions (Fredrickson, 2002; Van Cappellen et al., 2016; Cohen and Johnson, 2017), and that religion can help individuals ameliorate the pain of loss (Cohen and Johnson, 2017). Some researchers discuss the potential of positive emotions to increase subjective well-being, but explain that empirical support for this claim is indirect only (Fredrickson et al., 2008; Park and Slattery, 2012). Furthermore, Inkeles (1993) suggests that individuals in certain cultural groups are socialized to have a particular perspective, which leads them to "see things in a positive or negative light." Gatina (2016) emphasizes the important role of an immigrant's cultural background as a crucial factor affecting his/her life satisfaction. Therefore, we consider to group individuals with different cultural backgrounds, religion and social values, because they may feel and self-assess happiness using different criteria, even under the same conditions.

2.3 Model Specification

2.3.1 Univariate Ordered Probit Model

Let y_i denote the ordinal discrete response of individual i, for $i = 1, \ldots, n$, that is, $y_i = j$ for $j = 1, \ldots, J$, where J is the number of ordinal categories. In addition, let y_i^* denote the latent variable of individual i, such that

$$y_i = j \ \text{ if } \ \mu_{j-1} < y_i^* \leq \mu_j, \ j = 1, \ldots, J, \tag{2.1}$$

where μ_j is a cutpoint of an ordinal response. The unknown cutpoint parameters satisfy the following condition:

$$-\infty = \mu_0 < \mu_1 = 0 < \mu_2 < \cdots < \mu_{J-1} < \mu_J = \infty,$$

where $\mu_1 = 0$ is required to establish the identifiability of the cutoff parameters.[1] The latent variable y_i^* is assumed to depend on a k-dimensional vector of covariates $\boldsymbol{x}_i$ through the following linear model:

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \ \epsilon_i \sim \mathrm{N}(0, 1), \ i = 1, \ldots, n, \tag{2.2}$$

[1]See, for example, Albert and Chib (1993, p.673).

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters, and $\boldsymbol{x}_i$ is independent of ϵ_i. Then, the probability of observing $y_i = j$ can be written as

$$\begin{aligned}\Pr(y_i = j|\boldsymbol{\beta}, \boldsymbol{\mu}) &= \Pr(\mu_{j-1} < \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i \leq \mu_j) \\ &= \Phi(\mu_j - \boldsymbol{x}_i'\boldsymbol{\beta}) - \Phi(\mu_{j-1} - \boldsymbol{x}_i'\boldsymbol{\beta}),\end{aligned} \tag{2.3}$$

where $\boldsymbol{\mu} = (\mu_2, \dots, \mu_{J-1})'$, and $\Phi(\cdot)$ is the distribution function of the standard normal distribution. From (2.3), we obtain the following likelihood function for the ordered probit model:

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\mu}) = \prod_{i=1}^{n}\prod_{j=1}^{J}[\Phi(\mu_j - \boldsymbol{x}_i'\boldsymbol{\beta}) - \Phi(\mu_{j-1} - \boldsymbol{x}_i'\boldsymbol{\beta})]^{1(y_i=j)}, \tag{2.4}$$

where $\boldsymbol{y} = (y_1, \dots, y_n)'$ and $1(y_i = j)$ is an indicator function. The indicator function takes the value one if the event $y_i = j$ is true, and zero otherwise.

The seminal study of Bayesian analyses using an ordered probit model is that of Albert and Chib (1993). Although their algorithm is convenient for sampling the full conditional distributions of y_i^* and $\boldsymbol{\beta}$, the sampling of the cutpoints μ_j is not conditioned on the latent data. In this case, Albert and Chib (2001) indicate that we can sample the cutpoints $\boldsymbol{\mu}$ by transforming them into $\gamma_j = \log(\mu_j - \mu_{j-1})$, $(2 \leq j \leq J-1)$ in order to remove the ordering constraints using a one-to-one mapping.

Under the prior distributions $\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)$ and $\boldsymbol{\gamma} \sim \mathrm{N}(\boldsymbol{\gamma}_0, \boldsymbol{G}_0)$, we obtain the following joint posterior distribution:

$$\begin{aligned} p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{y}^*|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{y}^*)p(\boldsymbol{y}^*|\boldsymbol{\beta}, \boldsymbol{\gamma})p(\boldsymbol{\beta})p(\boldsymbol{\gamma}) \\ &= \left[\prod_{i=1}^{n} p(y_i|\boldsymbol{\beta}, \boldsymbol{\gamma}, y_i^*)p(y_i^*|\boldsymbol{\beta}, \boldsymbol{\gamma})\right] p(\boldsymbol{\beta})p(\boldsymbol{\gamma}) \\ &= \left\{\prod_{i=1}^{n} 1(\mu_{j-1} < y_i^* \le \mu_j)p(y_i^*|\boldsymbol{\beta}, \boldsymbol{\gamma})\right\} p(\boldsymbol{\beta})p(\boldsymbol{\gamma}), \end{aligned} \tag{2.5}$$

where $\boldsymbol{\gamma} = (\gamma_2, \ldots, \gamma_{J-1})'$.

Algorithm 1. Sampling in a univariate ordered probit model — Albert and Chib (2001).

---

1. Draw $\tilde{\boldsymbol{\gamma}} \sim q(\boldsymbol{\gamma}|\cdots)$ from a multivariate t proposal density $q(\boldsymbol{\gamma}|\cdots) = f_T(\boldsymbol{\gamma}|\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{G}}, \boldsymbol{\nu})$, where "$|\cdots$" denotes the conditioning of the other unspecified variables in the model, ν is the degrees of freedom, and

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \log p(\boldsymbol{\gamma}|\cdots), \; \hat{\boldsymbol{G}} = \left\{\left[-\frac{\partial^2 \log p(\boldsymbol{\gamma}|\cdots)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right]_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}\right\}^{-1}.$$

2. Compute the acceptance ratio

$$r = \frac{p(\tilde{\boldsymbol{\gamma}}|\cdots)}{p(\boldsymbol{\gamma}|\cdots)} \frac{f_T(\boldsymbol{\gamma}|\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{G}}, \boldsymbol{\nu})}{f_T(\tilde{\boldsymbol{\gamma}}|\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{G}}, \boldsymbol{\nu})}.$$

3. Return $\tilde{\boldsymbol{\gamma}}$, with probability $\min\{1, r\}$; otherwise, repeat using the previous value of $\boldsymbol{\gamma}$.

4. Sample $y_i^*|\cdots \sim \mathrm{TN}_{(\mu_{j-1}, \mu_j]}(\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$, where $\mathrm{TN}(\cdot)$ denotes the truncated

normal distribution.

5. Sample $\boldsymbol{\beta}|\cdots \sim \mathrm{N}(\boldsymbol{\beta}^*, \boldsymbol{B}^*)$, where $\boldsymbol{B}^* = (\boldsymbol{B}_0^{-1}+\boldsymbol{X}'\boldsymbol{X})^{-1}$, $\boldsymbol{\beta}^* = \boldsymbol{B}^*(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0+\boldsymbol{X}'\boldsymbol{y}^*)$.

The merit of sampling $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_{J-1})'$ rather than $\boldsymbol{\mu} = (\mu_2, \ldots, \mu_{J-1})'$ is that, by transforming the cutpoints $\boldsymbol{\mu}$ using $\gamma_j = \log(\mu_j - \mu_{j-1})$, $2 \leq j \leq J-1$, the proposal density of $\boldsymbol{\gamma}$ can be determined using unconstrained optimization and the prior distribution of $\boldsymbol{\gamma}$ can be set as unrestricted.

2.3.2 Accommodating Individual Heterogeneity

In biometric studies, such as insecticide experiments, researchers may safely assume that the population is homogeneous enough to be treated with a zero mean, homogeneous variance, and the same thresholds. However, in applications to subjective well-being, the backgrounds and preferences of individuals are more complicated, and individuals understand concepts in different ways (Brady, 1985). King et al. (2004) and Kapteyn et al. (2007) suggest that the response scales differ systematically between groups and, therefore, use the anchoring vignettes method to adjust for these differences. Although the method of King et al. (2004) cannot be applied to data from the WVS, we need to consider that individuals with different backgrounds may self-assess happiness using different criteria. Thus, individuals in different groups should not be treated using the same thresholds in the ordered probit model.

The seminal research on accommodating individual heterogeneity in thresholds is that of Terza (1985), who considers the following cutpoints:

$$\mu_{ij} = \mu_j + \boldsymbol{\gamma}_j' \boldsymbol{z}_i.$$

In this equation, $\boldsymbol{\gamma}_j$ is suggested to vary across outcomes. In his studies on three outcomes, Terza (1985) assumes that the variables in $\boldsymbol{z}_i$ are the same as those in $\boldsymbol{x}_i$ in the model. After fixing cutpoint $\mu_0 = 0$, the model becomes

$$y_i^* = \boldsymbol{x}_i' \boldsymbol{\beta} + \epsilon_i, \;\; \epsilon_i \sim \mathrm{N}(0, 1),$$

and the outcome y_i is

$$\begin{gathered} y_i = 0 \;\; \text{if} \;\; y_i^* \leq 0, \\ y_i = 1 \;\; \text{if} \;\; 0 < y_i^* \leq \mu_1 + \boldsymbol{\gamma}_1' \boldsymbol{x}_i, \\ y_i = 2 \;\; \text{if} \;\; y_i^* > \mu_1 + \boldsymbol{\gamma}_1' \boldsymbol{x}_i. \end{gathered}$$

However, a linear specification of thresholds, such as $\mu_{ij} = \mu_j + \boldsymbol{\gamma}_j' \boldsymbol{z}_i$, cannot ensure that the probabilities always take positive values.

Thus, Greene (2007), Eluru et al. (2008), and Greene and Hensher (2010a) propose the following hierarchical ordered probit model:

$$\begin{gathered} y_i^* = \boldsymbol{x}_i' \boldsymbol{\beta} + \epsilon_i, \\ y_i = j \;\; \text{if} \;\; \mu_{i,j-1} < y_i^* \leq \mu_{ij}, \\ \mu_{i0} = 0, \;\; \mu_{ij} = \exp(\lambda_j + \boldsymbol{\gamma}_j' \boldsymbol{z}_i). \end{gathered}$$

This nonlinear specification of thresholds has the following merits. First, the equation of the cutpoints is mathematically different from the regression model. Second, using an exponential transformation, the cutpoints μ_{ij} always take positive values. Furthermore, we modify the cutpoint function

$$\mu_{i,j} = \mu_{i,j-1} + \exp(\boldsymbol{\gamma}_j' \boldsymbol{z}_i)$$

to ensure the ordering of the cutpoints.

2.3.3 Univariate Ordered Probit Model with Individual Heterogeneity

We assume that a continuous latent variable y_i^* depends on a k-dimensional vector of covariates $\boldsymbol{x}_i$ in the following model:

$$y_i^* = \boldsymbol{x}_i' \boldsymbol{\beta} + \epsilon_i, \ \ \epsilon_i \sim \mathrm{N}(0, 1), \ \ i = 1, \ldots, n. \tag{2.6}$$

Furthermore, the observed outcome y_i results from

$$y_i = j \ \ \text{if} \ \ \mu_{i,j-1} < y_i^* \leq \mu_{ij}, \ \ j = 1, \ldots, J. \tag{2.7}$$

In this model, we use the following cutpoint function:

$$\mu_{ij} = \mu_{i,j-1} + \exp(\boldsymbol{\gamma}_j' \boldsymbol{z}_i), \ \ j = 2, \ldots, J-1. \tag{2.8}$$

Here, we fix the cutpoints as $\mu_{i0} = -\infty$, $\mu_{i1} = 0$, and $\mu_{iJ} = \infty$.

Then, the probability of observing $y_i = j$ is given by

$$\begin{aligned}\Pr(y_i = j|\boldsymbol{\beta}, \boldsymbol{\mu}) &= \Pr(\mu_{i,j-1} < \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i \leq \mu_{ij}) \\ &= \Phi(\mu_{ij} - \boldsymbol{x}_i'\boldsymbol{\beta}) - \Phi(\mu_{i,j-1} - \boldsymbol{x}_i'\boldsymbol{\beta}).\end{aligned} \tag{2.9}$$

The likelihood function for this ordered probit model can be written as follows:

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^{n}\prod_{j=1}^{J}\left[\Phi(\mu_{ij} - \boldsymbol{x}_i'\boldsymbol{\beta}) - \Phi(\mu_{i,j-1} - \boldsymbol{x}_i'\boldsymbol{\beta})\right]^{1(y_i=j)}, \tag{2.10}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_2', \dots, \boldsymbol{\gamma}_{J-1}')'$. According to Bayes' theorem, the joint posterior distribution of the model can be written as

$$\begin{aligned}p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{y}^*|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{y}^*)p(\boldsymbol{y}^*|\boldsymbol{\beta}, \boldsymbol{\gamma})p(\boldsymbol{\beta})p(\boldsymbol{\gamma}) \\ &= \left[\prod_{i=1}^{n} p(y_i|\boldsymbol{\beta}, \boldsymbol{\gamma}, y_i^*)p(y_i^*|\boldsymbol{\beta}, \boldsymbol{\gamma})\right] p(\boldsymbol{\beta}) \left[\prod_{j=2}^{J-1} p(\boldsymbol{\gamma}_j)\right],\end{aligned} \tag{2.11}$$

where the prior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j$ are specified as $\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)$ and $\boldsymbol{\gamma}_j \sim \mathrm{N}(\gamma_0, \boldsymbol{G}_0)$, respectively.

As a result, we have the following algorithm.[2]

Algorithm 2. Sampling in a univariate ordered probit model (Accommodating individual heterogeneity)

---

1. Draw $\tilde{\boldsymbol{\gamma}}_j \sim q(\boldsymbol{\gamma}_j|\cdots)$ from a multivariate t proposal density $q(\boldsymbol{\gamma}_j|\cdots) =$

[2]The derivation of Algorithm 2 is provided in Appendix A.1.

$f_T(\boldsymbol{\gamma}_j|\hat{\boldsymbol{\gamma}}_j, \hat{\boldsymbol{G}}_{J-1}, \boldsymbol{\nu})$, where ν is the degrees of freedom and

$$\hat{\boldsymbol{\gamma}}_j = \arg\max_{\boldsymbol{\gamma}_j} \log p(\boldsymbol{\gamma}_j|\cdots), \;\; \hat{\boldsymbol{G}}_j = \left\{ \left[-\frac{\partial^2 \log p(\boldsymbol{\gamma}_j|\cdots)}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j'} \right]_{\boldsymbol{\gamma}_j=\hat{\boldsymbol{\gamma}}_j} \right\}^{-1}.$$

2. Compute the acceptance ratio

$$r = \frac{p(\tilde{\boldsymbol{\gamma}}_j|\cdots)}{p(\boldsymbol{\gamma}_j|\cdots)} \frac{f_T(\boldsymbol{\gamma}_j|\hat{\boldsymbol{\gamma}}_j, \hat{\boldsymbol{G}}, \boldsymbol{\nu})}{f_T(\tilde{\boldsymbol{\gamma}}_j|\hat{\boldsymbol{\gamma}}_j, \hat{\boldsymbol{G}}, \boldsymbol{\nu})}.$$

3. Return $\tilde{\boldsymbol{\gamma}}_j$, with probability $\min\{1, r\}$; otherwise, repeat using the previous value of $\boldsymbol{\gamma}_j$.

4. Sample $y_i^*|\cdots \sim TN_{(\mu_{j-1},\mu_j]}(\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$.

5. Sample $\boldsymbol{\beta}|\cdots \sim \mathrm{N}(\boldsymbol{\beta}^*, \boldsymbol{B}^*)$, where $\boldsymbol{B}^* = (\boldsymbol{B}_0^{-1}+\boldsymbol{X}'\boldsymbol{X})^{-1}$, $\boldsymbol{\beta}^* = \boldsymbol{B}^*(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0+\boldsymbol{X}'\boldsymbol{y}^*)$.

2.4 An Empirical Study

2.4.1 Data

Studies of happiness can apply a Bayesian analysis of an ordered probit model because happiness data are usually arranged on an ordinal scale. In this section, we discuss the variables, such as education, labor force participation, marital status, and so on, that affect individuals' happiness. The data are taken from the WVS, which launched six waves of surveys between 1981 and 2014. Because we attempt to accommodate immigration and religion status as individual heterogeneity in the thresholds, we choose data on three nations of immigrants, namely, Australia, Canada, and the United States. However, the latest wave of the survey, Wave 6, does not include data on Canada, so we use Wave 5 data, which were collected between 2005 and 2009. In the Wave 5 questionnaire, over 200 questions inquired about respondents' feelings about their life, state of health, economic situation, attitudes toward political action, and so on.

The data on happiness, marital status, education level, age, income, gender, labor force participation, father's immigration status, mother's immigration status, and religion are used in our analysis. There are four choices on the scale of happiness in the data. In our study, we modify the data so that the ratings of happiness range from 1 to 4, where 1 denotes "not at all happy," 2 is "not very happy," 3 is "rather happy," and 4 is "very happy." Then, marital status, gender, labor force participation, father's immigration status, mother's immigration status, and religion are all modified as dummy variables. Here, the value one denotes "married," "male," "having paid employment," "father is an immigrant," "mother is an immigrant," and "belongs to a religious denomination," and zero denotes the opposite result in each case. Furthermore, following Layard et al. (2008), we construct numerical values of income using the midpoint of each band. In addition, if the respondents are in the lowest income band, we set the income value as two-thirds of the upper limit of the band. If the respondents are in the highest income band, we set the income value as 1.5 times the lower income limit of the band.

Moreover, in the Wave 5 survey, the educational level in the United States is classified as less than high school, some high school but no diploma, graduated from high school, some college but no degree, associate degree, bachelor's degree, master's degree, professional degree, and doctorate degree. These are represented by the values 9, 10.5, 12, 13, 14, 16, 18, 18, and 21, respectively, which denote the time required to complete each education option. In Canada, the educational level is classified as no formal education, incomplete primary school, complete primary school, incomplete secondary school: technical/vocational type, complete secondary school: technical/vocational type, incomplete secondary school: university-preparatory type, complete secondary school: university-preparatory type, some university-level education without a degree, and university-level education with a degree. These are represented by the values 0, 3, 6, 10.5, 12, 10.5, 12, 14, and 16, respectively. In Australia, the educational level is classified as no formal education, incomplete primary school, complete primary school, incomplete secondary school: to year 10, complete secondary school: to year 10, incomplete secondary school: to year 12, complete secondary school: to year 12, trade qualification or apprenticeship, certificate or diploma, complete university: bachelor, and complete university: postgraduate. These are represented by the values 0, 3, 6, 8, 10, 11, 12, 14, 14, 16, and 18, respectively. The descriptive statistics of the data are listed in Tables 2.1 to 2.6.

2.4.2 The Posterior Results

There are three models in our analysis. In the first model, we do not accommodate individual heterogeneity. In the second model, we accommodate "father

Table 2.1: The happiness level in the United States (frequency)

	Frequency	%
Not at all happy (happiness=1)	5	0.46
Not very happy (happiness=2)	76	7.01
Rather happy (happiness=3)	623	57.47
Very happy (happiness=4)	380	35.06

Table 2.2: Descriptive statistics of data in the United States

Continuous variables	Min	Median	Max	Mean	SD
Education	9.00	13.00	21.00	13.29	2.47
Age	18.00	48.00	91.00	47.96	17.03
Income	3333.33	44999.50	262500	51456.69	42333.93
Discrete variables				Frequency	%
Marital status	Married (married=1)			639	58.95
	Unmarried (married=0)			445	41.05
Respondent's sex	Male (male=1)			544	50.18
	Female (male=0)			540	49.82
Employment status	Respondent is employed (work=1)			622	57.38
	Respondent is not employed (work=0)			462	42.62
Mother immigration status	Mother is an immigrant (mothim=1)			95	8.76
	Mother is not an immigrant (mothim=0)			989	91.24
Father immigration status	Father is an immigrant (fathim=1)			101	9.32
	Father is not an immigrant (fathim=0)			983	90.68
Religion status	Have a religion(religion=1)			811	74.82
	Not have a religion (religion=0)			273	25.18

immigration," "mother immigration," and "religion status" as individual heterogeneity. In the third model, we combine "father immigration" and "mother immigration" as "parent immigration." The Markov chain Monte Carlo (MCMC) simulations are run for $15,000$ iterations, with the first $5,000$ samples being discarded as a burn-in period. In this chapter, the posterior results of these

Table 2.3: The happiness level in Canada (frequency)

	Frequency	%
Not at all happy (happiness=1)	5	0.48
Not very happy (happiness=2)	58	5.52
Rather happy (happiness=3)	482	45.86
Very happy (happiness=4)	506	48.14

Table 2.4: Descriptive statistics of data in Canada

Continuous variables	Min	Median	Max	Mean	SD
Education	0.00	12.00	16.00	12.58	2.01
Age	16.00	47.00	94.00	47.75	16.83
Income	8333.33	46250.50	225000	57262.18	46200.28
Discrete variables				Frequency	%
Marital status	Married (married=1)			527	50.14
	Unmarried (married=0)			524	49.86
Respondent's sex	Male (male=1)			437	41.58
	Female (male=0)			614	58.42
Employment status	Respondent is employed (work=1)			584	55.57
	Respondent is not employed (work=0)			467	44.43
Mother immigration status	Mother is an immigrant (mothim=1)			179	17.03
	Mother is not an immigrant (mothim=0)			872	82.97
Father immigration status	Father is an immigrant (fathim=1)			189	17.98
	Father is not an immigrant (fathim=0)			862	82.02
Religion status	Have a religion(religion=1)			740	70.41
	Not have a religion (religion=0)			311	29.59

Table 2.5: The happiness level in Australia (frequency)

	Frequency	%
Not at all happy (happiness=1)	8	0.74
Not very happy (happiness=2)	78	7.17
Rather happy (happiness=3)	622	57.16
Very happy (happiness=4)	380	34.93

TABLE 2.6: Descriptive statistics of data in Australia

Continuous variables	Min	Median	Max	Mean	SD
Education	0.00	14.00	18.00	12.94	3.36
Age	18.00	49.00	90.00	48.49	15.73
Income	12000	48750.50	172500	64128.83	47053.72
Discrete variables				Frequency	%
Marital status	Married (married=1)			668	61.40
	Unmarried (married=0)			420	38.60
Respondent's sex	Male (male=1)			508	46.69
	Female (male=0)			580	53.31
Employment status	Respondent is employed (work=1)			727	66.82
	Respondent is not employed (work=0)			361	33.18
Mother immigration status	Mother is an immigrant (mothim=1)			252	23.16
	Mother is not an immigrant (mothim=0)			836	76.84
Father immigration status	Father is an immigrant (fathim=1)			275	25.28
	Father is not an immigrant (fathim=0)			813	74.72
Religion status	Have a religion(religion=1)			659	60.57
	Not have a religion (religion=0)			429	39.43

models are obtained using `R`. The convergence of MCMC chains has been confirmed by checking the effective sample size and the sample path.

In the first model, the ordered probit model is

$$y_i^* = \beta_1 + \beta_2 married_i + \beta_3 male_i + \beta_4 work_i$$

$$+\beta_5 education_i + \beta_6 age_i + \beta_7 income_i + \epsilon_i,\ \epsilon_i \sim \mathrm{N}(0,1)$$

$$y_i = j \ \text{ if } \ \mu_{j-1} < y_i^* \leq \mu_j,\ j = 1, \ldots, 4,\ i = 1, \ldots, n,$$

where *income* is the logarithm of total household income. In this model, we use algorithm 1 (see Section 2) and sample the cutpoints $\boldsymbol{\mu}$ by transforming them into $\mu_j = \mu_{j-1} + \exp(\gamma_j)$. The prior distributions are specified as $\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I})$ and $\boldsymbol{\gamma} \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I})$.

After including "father immigration," "mother immigration," and "religion status" as individual heterogeneity, the ordered probit model in the second model can be written as:

$$
\begin{gathered}
y_i^* = \beta_1 + \beta_2 married_i + \beta_3 male_i + \beta_4 work_i \\
+\beta_5 education_i + \beta_6 age_i + \beta_7 income_i + \epsilon_i,\ \epsilon_i \sim \mathrm{N}(0,1) \\
\mu_{ij} = \mu_{i,j-1} + \exp(\gamma_{j1} + \gamma_{j2} mothim_i + \gamma_{j3} fathim_i + \gamma_{j4} religion_i) \\
y_i = j \ \text{ if } \ \mu_{i,j-1} < y_i^* \leq \mu_{ij},\ j = 1, \ldots, 4,\ i = 1, \ldots, n,
\end{gathered}
$$

where $mothim$, $fathim$, and $religion$ are dummy variables denoting the mother's immigration status, father's immigration status, and religion status, respectively. If a respondent's mother and father are immigrants and the respondent belongs to a religious denomination, then $mothim = 1$, $fathim = 1$, and $religion = 1$. In this model, we use algorithm 2 (see Section 2) and specify the prior distributions as $\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I})$ and $\boldsymbol{\gamma}_j \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I})$. Furthermore, the transformations for the cutpoints are $\mu_{ij} = \mu_{i,j-1} + \exp(\boldsymbol{\gamma}_j' \boldsymbol{z}_i)$. Because $J = 4$ and we fix the cutpoints as $\boldsymbol{\mu}_0 = -\infty\boldsymbol{\iota}$, $\boldsymbol{\mu}_1 = \mathbf{0}$, and $\boldsymbol{\mu}_4 = \infty\boldsymbol{\iota}$, where $\boldsymbol{\iota} = (\underbrace{1, \ldots, 1}_{n})'$, we only need to estimate $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$, which include four parameters each.

In the final model, we use the same sampling algorithm and priors as in model 2, but we combine "father immigration" and "mother immigration" as "parent immigration." Then, we have the following model:

$$y_i^* = \beta_1 + \beta_2 married_i + \beta_3 male_i + \beta_4 work_i$$
$$+\beta_5 education_i + \beta_6 age_i + \beta_7 income_i + \epsilon_i, \;\; \epsilon_i \sim \mathrm{N}(0,1)$$
$$\mu_{ij} = \mu_{i,j-1} + \exp(\gamma_{j1} + \gamma_{j2} parim_i + \gamma_{j3} religion_i)$$
$$y_i = j \;\; \text{if} \;\; \mu_{i,j-1} < y_i^* \leq \mu_{ij}, \;\; j = 1, \ldots, 4, \;\; i = 1, \ldots, n,$$

where *parim* is a vector of dummy variables representing the parents' immigration status; that is, when at least one of the parents is an immigrant, *parim* takes the value one.

The summary results of the marginal likelihood and the Bayes factor are listed in Tables 2.7 and 2.8, respectively.[3]

TABLE 2.7: Summary of marginal likelihood

Country	Model fitted	log(marginal)
Australia	Model 1	−1892.853
	Model 2	−1066.411
	Model 3	−1068.793
Canada	Model 1	−1743.308
	Model 2	−967.824
	Model 3	−964.801
United States	Model 1	−1387.922
	Model 2	−1041.927
	Model 3	−1040.445

The Bayes factor for any two models i and l is calculated as $B_{il} = m(\boldsymbol{y}|M_i)/m(\boldsymbol{y}|M_l)$, where $m(\boldsymbol{y}|M_i)$ and $m(\boldsymbol{y}|M_l)$ are the marginal likelihood functions of M_i and

[3]The derivation of the marginal likelihood is provided in Appendix A.2.

Table 2.8: Summary of Bayes factor

Country	Logarithm of Bayes factor	Results
Australia	$\log_{10} B_{21}$	358.919
	$\log_{10} B_{31}$	357.885
	$\log_{10} B_{23}$	1.034
Canada	$log_{10}B_{21}$	336.788
	$log_{10}B_{31}$	338.101
	$log_{10}B_{32}$	1.313
United States	$\log_{10} B_{21}$	146.355
	$\log_{10} B_{31}$	146.999
	$\log_{10} B_{32}$	0.644

M_l, respectively. In addition, Jeffreys (1961) provides a scale on which to judge the evidence of M_i against M_l.[4] The scale is:

- if $0 < \log_{10}(B_{il}) < 0.5$, the evidence of M_i against M_l is *poor*,
- if $0.5 < \log_{10}(B_{il}) < 1$, the evidence against M_l is *substantial*,
- if $1 < \log_{10}(B_{il}) < 2$, the evidence is *strong*, and
- if $2 < \log_{10}(B_{il})$, the evidence is *decisive*.

The results of the Bayes factor analysis show that the models with individual heterogeneity perform better than those without individual heterogeneity do. Furthermore, the results indicate that model 3 outperforms model 1 and model 2 for both Canada and the United States, while model 2 performs best for Australia. Since the difference between model 2 and model 3 is that model 2 includes more detailed information about father's and mother's immigration status, we can conclude that the more detailed model performs better for Australia, while for both Canada and the United States, more detailed model about immigration status is not necessary.

[4]See Robert (2001, p.228).

TABLE 2.9: The posterior results of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for the United States, with individual heterogeneity, and combining father and mother immigration as parent immigration

United States			
Covariate	Mean	SD	Median
Intercept (β_1)	1.0062	0.4834	1.0067**
Married (β_2)	0.2574	0.0757	0.2578**
Male (β_3)	−0.0421	0.0717	−0.0420
Work (β_4)	−0.0914	0.0818	−0.0923
Education (β_5)	0.0096	0.0160	0.0097
Age (β_6)	0.0020	0.0023	0.0020
Income (β_7)	0.1152	0.0474	0.1153**
Intercept (γ_{21})	0.2685	0.1152	0.2702**
Parim (γ_{22})	−0.0666	0.1824	−0.0546
Religion (γ_{23})	−0.3196	0.1053	−0.3173**
Intercept (γ_{31})	0.5021	0.0667	0.5027**
Parim (γ_{32})	0.0940	0.0959	0.0968
Religion (γ_{33})	0.1483	0.0736	0.1482**

Notes: "$**$" denotes that zero is not included in the 95% credible interval. Mean, Median, and SD are the posterior mean, posterior median, and posterior standard deviation, respectively.

The posterior results for the United States and Canada in model 3 are listed in Tables 2.9 and 2.10, respectively. The results indicate that in both countries, *married* and *income* have positive effects on happiness. In addition,*male* has negative effects in Canada, but no evident impact in the United States. Furthermore, from the posterior results of $\boldsymbol{\gamma}$, we find that, in the United States, *religion* has negative effects on $\boldsymbol{\mu}_2$; that is, *religion* decreases the value of $\boldsymbol{\mu}_2$. However, *religion* still has positive effects on $\boldsymbol{\mu}_3$; that is, *religion* increases the value of $\boldsymbol{\mu}_3$. Thus, in the United States, religion can increase the interval $(\boldsymbol{\mu}_2, \boldsymbol{\mu}_3]$, which corresponds to the choice of $y = 3$ ("rather happy"). In Canada, the three variables *mothim*, *fathim*, and *religion* do not have evident effects

TABLE 2.10: The posterior results of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for Canada, with individual heterogeneity, and combining father and mother immigration as parent immigration

	Canada		
Covariate	Mean	SD	Median
Intercept (β_1)	0.6196	0.5649	0.6216
Married (β_2)	0.3220	0.0804	0.3222**
Male (β_3)	−0.1290	0.0733	−0.1296*
Work (β_4)	0.0499	0.0850	0.0497
Education (β_5)	−0.0130	0.0197	−0.0128
Age (β_6)	0.0020	0.0025	0.0021
Income (β_7)	0.1845	0.0552	0.1848**
Intercept (γ_{21})	0.0052	0.1583	0.0129
Parim (γ_{22})	0.1695	0.1313	0.1742
Religion (γ_{23})	−0.0304	0.1265	−0.0321
Intercept (γ_{31})	0.4941	0.0750	0.4955**
Parim (γ_{32})	−0.0489	0.0931	−0.0470
Religion (γ_{33})	0.0388	0.0836	0.0381

Notes: "∗∗" and "∗" denote that zero is not included in the 95% and 90% credible intervals, respectively. Mean, Median, and SD are the posterior mean, posterior median, and posterior standard deviation, respectively.

on $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$.

The marginal effects in the United States and Canada in model 3 are listed in Tables 2.11 and 2.12, respectively.[5]

We find that, in both countries, the marginal effects of *married* and *income* take negative values when $y = 1$, $y = 2$, and $y = 3$, and take a positive value in the case of $y = 4$. That is, *married* and *income* have negative marginal effects on being "not at all happy," "not very happy," and "rather happy," and have positive marginal effects on being "very happy." Then, *male* in Canada has positive marginal effects on being "not at all happy," "not very happy," and

[5] The derivation of the marginal effects is provided in Appendix A.3.

TABLE 2.11: The marginal effects for the United States, with individual heterogeneity, and combining father and mother immigration as parent immigration

United States				
Covariate	$y = 1$	$y = 2$	$y = 3$	$y = 4$
Married	-0.0050^{**}	-0.0361^{**}	-0.0555^{**}	0.0934^{**}
Male	0.0010	0.0061	0.0092	-0.0154
Work	0.0022	0.0131	0.0203	-0.0336
Education	-0.0002	-0.0014	-0.0021	0.0035
Age	0.0000	-0.0003	-0.0004	0.0007
Income	-0.0027^{**}	-0.0166^{**}	-0.0251^{**}	0.0421^{**}

Notes: "$**$" denotes that zero is not included in the 95% credible interval. The values are posterior means.

TABLE 2.12: The marginal effects for Canada, with individual heterogeneity, and combining father and mother immigration as parent immigration

Canada				
Covariate	$y = 1$	$y = 2$	$y = 3$	$y = 4$
Married	-0.0043^{**}	-0.0294^{**}	-0.0898^{**}	0.1177^{**}
Male	0.0023^{*}	0.0134^{*}	0.0347^{*}	-0.0465^{*}
Work	-0.0008	-0.0049	-0.0135	0.0180
Education	0.0003	0.0017	0.0034	-0.0050
Age	0.0000	-0.0003	-0.0005	0.0008
Income	-0.0043^{**}	-0.0240^{**}	-0.0486^{**}	0.0710^{**}

Notes: "$**$" and "$*$" denote that zero is not included in the 95% and 90% credible intervals, respectively. The values are posterior means.

"rather happy," but has negative marginal effects on being "very happy."

Table 2.13 illustrates the posterior results for Australia obtained from model 2. The results indicate that both *married* and *education* have positive effects on happiness in Australia. Furthermore, *mothim* has negative

TABLE 2.13: The posterior results of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for Australia, with individual heterogeneity, and dividing immigration into father and mother immigration

	Australia		
Covariate	Mean	SD	Median
Intercept (β_1)	1.0182	0.5895	1.0165*
Married (β_2)	0.5226	0.0798	0.5220**
Male (β_3)	−0.0565	0.0730	−0.0557
Work (β_4)	0.0016	0.0923	0.0022
Education (β_5)	0.0272	0.0117	0.0272**
Age (β_6)	0.0002	0.0028	0.0002
Income (β_7)	0.0833	0.0564	0.0833
Intercept (γ_{21})	0.0159	0.1364	0.0194
Mothim (γ_{22})	0.3082	0.1774	0.3190
Fathim (γ_{23})	−0.1172	0.1796	−0.1244
Religion (γ_{24})	0.0097	0.1019	0.0099
Intercept (γ_{31})	0.6811	0.0548	0.6828**
Mothim (γ_{32})	−0.2362	0.1218	−0.2425*
Fathim (γ_{33})	0.1597	0.1147	0.1691
Religion (γ_{34})	−0.0561	0.0638	−0.0557

Notes: "∗∗" and "∗" denote that zero is not included in the 95% and 90% credible intervals, respectively. Mean, Median, and SD are the posterior mean, posterior median, and posterior standard deviation, respectively.

effects on $\boldsymbol{\mu}_2$. That is, if the mother is an immigrant, this decreases the value of $\boldsymbol{\mu}_2$. In conclusion, *mothim* can increase the interval $(\boldsymbol{\mu}_3, \infty\boldsymbol{\iota}]$, which corresponds to the choice of $y = 4$ ("very happy"). From the marginal effects for Australia, listed in Table 2.14, we find that *income* has no evident marginal effects. However, *married* and *education* take negative values when $y = 1$, $y = 2$, and $y = 3$, and take positive values when $y = 4$. These results show that *married* and *education* have negative marginal effects on being "not at all happy," "not very happy," and "rather happy," and have positive effects on being "very happy."

TABLE 2.14: The marginal effects for Australia, with individual heterogeneity, and dividing immigration into father and mother immigration

	Australia			
Covariate	$y = 1$	$y = 2$	$y = 3$	$y = 4$
Married	-0.0138^{**}	-0.0726^{**}	-0.1044^{**}	0.1851^{**}
Male	0.0026	0.0081	0.0113	−0.0207
Work	−0.0001	−0.0004	0.0001	0.0005
Education	-0.0012^{**}	-0.0039^{**}	-0.0055^{**}	0.0100^{**}
Age	0.0000	0.0000	0.0000	0.0001
Income	−0.0038	−0.0120	−0.0168	0.0305

Notes: "'∗∗" denotes that zero is not included in the 95% credible interval. The values are posterior means.

In conclusion, *married* has positive effects on happiness in all three countries, which is consistent with the findings of many previous studies (Diener et al., 1999; Argyle, 2001; Blanchflower and Oswald, 2004; Diener et al., 2009). Then, *income* has positive effects in the United States and Canada, but no evident effect in Australia. In addition, *education* has positive effects in Australia only, which supports the empirical results of Nattavudh et al. (2015) and Tony and Matthew (2014). Then, *male* has negative effects in Canada only. Furthermore, the effects of heterogeneity vary between the three countries. Many studies have found a positive relationship between religious behavior and subjective well-being in the United States (Witter et al., 1985; Ellison, 1991; Diener et al., 1999; Koenig et al., 2001; Argyle, 2001; Ellison et al., 2001; Diener et al., 2009; Kortt et al., 2015; Van Cappellen et al., 2016). Our results confirm that *religion* has an evident impact in the United States, but no effect in Canada or Australia. In addition, parents' immigration status can affect the threshold in Australia. This fact supports the finding of Gatina (2016). Although many studies indicate a negative relationship between immigration and happiness (Werkuyten and Nekuee, 1999; Bălţătescu, 2007; Safi, 2010; Bartram, 2011; Gokdemir and Dumludag, 2012; Obućina, 2013; Hendriks, 2015; Kóczán, 2016), our results show that immigration status has no evident effects on the threshold in the United States or Canada.

2.5 Summary

In this chapter, we analyze happiness levels in Australia, Canada, and the United States using a new Bayesian univariate ordered probit model, and accommodating immigration status and religion status as individual heterogeneity in the cutpoint functions.

Our main findings are as follows:

- The results for the explanatory variables agree with those of other studies; that is, marriage, high income, and a high education level have positive effects on happiness, while being male has negative effects on happiness in Canada.
- The models with individual heterogeneity outperform those without individual heterogeneity.
- The effects of heterogeneity vary between the three countries. Having a religious affiliation has an evident impact on the thresholds in the United States, but shows no evident effects in Canada or Australia. In addition, parents' immigration status can affect the thresholds in Australia, but shows no effects in the United States or Canada.

In some instances, we need to model several related ordered responses from the same individuals, such as responses related to happiness and self-assessed health. Because these ordered responses are correlated, we need to use a multivariate model, rather than several separate univariate models. In next chapter, we conduct a Bayesian analysis using a multivariate ordered probit model, and propose a new algorithm that will allow us to include individual heterogeneity in the model.

Chapter 3

Bayesian Multivariate Ordered Probit Model with Individual Heterogeneity

3.1 Introduction

In the research field concerning subjective well-being, outcomes are often arranged ordinally. The ordered probit model is a useful approach to estimate models with this type of data. According to frequentists, an ordered probit model can be estimated using the maximum likelihood method. However, with the development of computer science and the Markov chain Monte Carlo (MCMC) method, Bayesian estimation of the ordered probit model has also become popular.

Compared with the non-Bayesian approach to the ordered probit model, the Bayesian approach has several merits. First, as described in Hasegawa and Ueda (2016, p.359), while the Bayesian method allows the estimation of the multivariate ordered probit model with correlation between explained variables, it is difficult to apply the non-Bayesian approach when the number of explained variables is greater than two.[1] Second, the representation of latent variables is particularly useful in the estimation of ordered probit models. The values of the latent variables can easily be obtained from the posterior distribution of the Bayesian sampling algorithm.

Following Jeliazkov et al. (2008), there are three main algorithms for the Bayesian analysis of the ordered probit model. The seminal study is Albert and Chib (1993), where the authors proposed a Bayesian sampling algorithm in the univariate ordered probit model. They employ the latent variable representation to estimate the posterior results of the explanatory variables. One shortcoming of their algorithm is that the sampling of the cutpoints is not conditioned on the latent data. Subsequently, Cowles (1996), Nandram and Chen (1996), and Chen and Dey (2000) proposed several approaches to sampling the cutpoints. Based on these ideas, Albert and Chib (2001) considered a new algorithm that transforms the cutoff points into a new unconstrained variable, to remove the ordering constraints using a one-to-one map. Moreover, Chen and Dey (2000) suggested removing the ordering constraints on the cutpoints through a transformation and proposed a sampling algorithm that is identified by fixing two cutpoints, and leaves the variance of the error term to be estimated.

In an analysis of subjective well-being, the backgrounds and preferences of respondents are complicated, and individuals may interpret concepts in different ways (Brady, 1985). The authors of King et al. (2004) and Kapteyn et al.

[1]For example, if we use STATA for such an estimation, it is possible to estimate a bivariate ordered probit model using BIOPROBIT, which computes maximum likelihood estimates, but there is no function to estimate ordered probit models with three or more variables. See Hasegawa and Ueda (2016, p.359, footnote).

(2007) suggested that the response scales differ systematically between groups. Therefore, they employed the anchoring vignettes method to adjust for these differences. Furthermore, many studies have considered ordered choice models, incorporating heterogeneity among individuals. The authors of Farewell (1982) proposed an ordered Weibull model, and Terza (1985) considered an alternative generalized model, in which the explanatory variables in the cutpoints regression are considered to vary across outcomes. Subsequently, Greene (2007), Eluru et al. (2008), and Greene and Hensher (2010b) proposed a hierarchical ordered probit model, where the nonlinear specification of thresholds in the model ensures that the cutpoints always take positive values. Further, Shi and Hasegawa (2019) introduced a Bayesian algorithm, which accommodates individual heterogeneity in the univariate ordered probit model. However, in the case of modeling several related responses, such as self-assessed health and subjective well-being from the same individuals, it is necessary to employ a multivariate ordered probit model. Therefore, we propose a new algorithm that includes individual heterogeneity in the multivariate ordered probit model. To our best knowledge, there are no previous studies that include individual heterogeneity in the multivariate ordered probit model.

In an empirical study, we apply the algorithm from Chen and Dey (2000) and the new proposed algorithm to real data. The data we employ is from the World Values Survey (WVS) wave 5, collected between 2005 and 2009. We choose three immigration countries, namely Australia, Canada, the United States, and study individuals' life satisfaction, state of health, interest in politics, and freedom of choice as the correlated responses. Furthermore, we accommodate parent immigration status and religious status as individual heterogeneity factors in the thresholds in the new algorithm. The results for the Bayes factor indicate that the model with individual heterogeneity outperforms that without heterogeneity. Moreover, the effects of heterogeneity are different among the three countries. Belonging to a religion has an evident impact on the state of health and freedom of choice in Australia, while it has significant effects on the state of health and life satisfaction in the United States. However, our findings also suggest that parent immigration status cannot affect the thresholds in all three countries.

3.2 Model Specification

3.2.1 Multivariate Ordered Probit Model

In this section, we describe the Bayesian multivariate ordered probit model introduced by Chen and Dey (2000). Let y_{ij} denote the ordinal response of an individual i to a question j, where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. That is, $y_{ij} = c_j$ for $c_j = 1, \ldots, C_j$, where C_j denotes the total number of distinct categories.

Further, let $\boldsymbol{y}_i^*$ denote an m-dimensional vector of continuous latent variables, which depends on a matrix of covariates $\boldsymbol{X}_i$ in the following model:

$$\boldsymbol{y}_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i,\ \ \boldsymbol{\epsilon}_i \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{3.1}$$

where

$$\boldsymbol{y}_i^* = \begin{pmatrix} y_{i1}^* \\ y_{i2}^* \\ \vdots \\ y_{im}^* \end{pmatrix}, \ \boldsymbol{x}_{ij} = \begin{pmatrix} x_{ij1} \\ x_{ij2} \\ \vdots \\ x_{ijk_j} \end{pmatrix}, \ \boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_{i1}' & & & \\ & \boldsymbol{x}_{i2}' & & \\ & & \ddots & \\ & & & \boldsymbol{x}_{im}' \end{pmatrix},$$

$$\boldsymbol{\beta}_j = \begin{pmatrix} \boldsymbol{\beta}_{j1} \\ \boldsymbol{\beta}_{j1} \\ \vdots \\ \boldsymbol{\beta}_{jk_j} \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_m \end{pmatrix}, \ \boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{im} \end{pmatrix},$$

and $\boldsymbol{\Sigma}$ is an $m \times m$ positive-definite covariance matrix.

Then, the outcome y_{ij} is determined by

$$y_{ij} = c_j \ \text{ if } \ \gamma_{j(c_j-1)} < y_{ij}^* \leq \gamma_{jc_j}, \tag{3.2}$$

where γ_{jc_j} is a cutpoint for the jth ordinal response. Here, the cutpoints are specified as

$$-\infty = \gamma_{j0} < \gamma_{j1} = 0 < \gamma_{j2} < \cdots < \gamma_{j(C_j-1)} = 1 < \gamma_{jC_j} = \infty,$$

where $\gamma_{j1} = 0$ and $\gamma_{j(C_j-1)} = 1$ are required to establish the identifiability of the cutoff parameters.[2]

[2]See, for example, Chen and Dey (2000, p.136).

In this model, we specify that

$$\mathcal{G}_{ij} = (\gamma_{j(c_j-1)}, \gamma_{jc_j}] \ \text{ if } \ y_{ij} = c_j, c_j = 1, \dots, C_j, j = 1, \dots, m,$$
$$\mathcal{G}_i = \mathcal{G}_{i1} \times \cdots \times \mathcal{G}_{im}, \ i = 1, \dots, n.$$

Then, we have that

$$p(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}_i^*) = 1_{\{\boldsymbol{y}_i^* \in \mathcal{G}_i\}}, \ i = 1, \dots, n, \tag{3.3}$$

where $1_{(\cdot)}$ is an indicator function.

Furthermore, Chen and Dey (2000) introduced the following transformation for γ_{jc_j} to remove the constraints on the cutpoints:

$$\delta_{jc_j} = \log\left(\frac{\gamma_{jc_j} - \gamma_{j(c_j-1)}}{1 - \gamma_{jc_j}}\right), \quad c_j = 2, \dots, C_j - 2, \tag{3.4}$$

where $\boldsymbol{\delta}_j = (\delta_{j2}, \dots, \delta_{j(C_j-2)})'$ $(j = 1, \cdots, m)$ and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \dots, \boldsymbol{\delta}_m')'$. The prior distribution of $\boldsymbol{\gamma}_j$ can be specified as follows:

$$\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) \sim \mathrm{N}(\boldsymbol{\delta}_{j0}, \boldsymbol{D}_{j0}), \ j = 1, \dots, m.$$

Moreover, we introduce the prior distributions $\boldsymbol{\beta}_j \sim \mathrm{N}(\boldsymbol{\beta}_{j0}, \boldsymbol{B}_{j0})$ and $\boldsymbol{\Sigma}^{-1} \sim$

$W(n_0, \boldsymbol{Q}_0^{-1})$, where $W(n_0, \boldsymbol{Q}_0^{-1})$ denotes the Wishart distribution with n_0 degrees of freedom and the scale matrix $\boldsymbol{Q}^{-1}$. Then, the joint posterior distribution of the model can be written as

$$\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}^*|\boldsymbol{y}) &\propto p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}^*)p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}^*) \\
&= p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})p(\boldsymbol{y}^*|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}^*) \\
&= p(\boldsymbol{\beta})p(\boldsymbol{\gamma})p(\boldsymbol{\Sigma}) \left[\prod_{i=1}^{n} p(\boldsymbol{y}_i^*|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})p(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}_i^*)\right] \\
&= \left[\prod_{j=1}^{m} p(\boldsymbol{\beta}_j)\right] \left[\prod_{j=1}^{m} \prod_{c_j=1}^{\boldsymbol{C}_{j-1}} p(\gamma_{jc_j})\right] \\
&\quad \times p(\boldsymbol{\Sigma}) \left\{\prod_{i=1}^{n} 1_{\{\boldsymbol{y}_i^* \in \mathcal{G}_i\}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}_i^* - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i^* - \boldsymbol{X}_i\boldsymbol{\beta})\right]\right\}.
\end{aligned} \tag{3.5}$$

The algorithm of Chen and Dey (2000) is summarized as follows:[3]

Algorithm 1. Sampling in a multivariate ordered probit model — Chen and Dey (2000).

---

1. Let $\boldsymbol{\delta}_j^s$ denote the value of $\boldsymbol{\delta}_j$ at the sth iteration. At the $(s+1)$th iteration, draw $\boldsymbol{\delta}_j^{s+1} \sim q(\boldsymbol{\delta}_j|\cdots)$ from a Student-t proposal density $q(\boldsymbol{\delta}_j|\cdots) = f_T(\boldsymbol{\delta}_j|\hat{\boldsymbol{\delta}}_j, \hat{\boldsymbol{G}}_j, \nu)$, where "$|\cdots$" denotes the conditioning of the other unspecified variables in the model, ν denotes the degrees of freedom, and

$$\hat{\boldsymbol{\delta}}_j = \arg\max \log p(\boldsymbol{\delta}_j|\cdots), \hat{\boldsymbol{G}}_j = \left\{\left[-\frac{\partial^2 \log p(\boldsymbol{\delta}_j|\cdots)}{\partial\boldsymbol{\delta}_j\partial\boldsymbol{\delta}_j'}\right]_{\boldsymbol{\delta}_j=\hat{\boldsymbol{\delta}}_j}\right\}^{-1}.$$

[3] The derivation of algorithm 1 is provided in Appendix B.1.

2. Compute the acceptance ratio

$$r = \frac{p(\boldsymbol{\delta}_j^{s+1}|\cdots)}{p(\boldsymbol{\delta}_j^{s}|\cdots)} \frac{f_T(\boldsymbol{\delta}_j^{s}|\hat{\boldsymbol{\delta}}_j, \hat{\boldsymbol{G}}_j, \nu)}{f_T(\boldsymbol{\delta}_j^{s+1}|\hat{\boldsymbol{\delta}}_j, \hat{\boldsymbol{G}}_j, \nu)}.$$

3. Return $\boldsymbol{\delta}_j^{s+1}$ with probability $\min\{1, r\}$; otherwise, repeat using the previous value of $\boldsymbol{\delta}_j^{s}$.

4. Sample $y_{ij}^*|\cdots \sim \mathrm{TN}_{(\gamma_{j(c_j-1)},\gamma_{jc_j}]}(\tilde{\mu}_{ij}, \tilde{\sigma}_j^2)$, where $\mathrm{TN}(\cdot)$ denotes the truncated normal distribution.

5. Sample $\boldsymbol{\beta}_j|\cdots \sim \mathrm{N}(\tilde{\boldsymbol{\beta}}_j, \tilde{\boldsymbol{B}}_j)$, where $\tilde{\boldsymbol{B}}_j = (\boldsymbol{B}_{j0}^{-1} + \sigma^{jj}\sum_{i=1}^n \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}')^{-1}$,
$\tilde{\boldsymbol{\beta}}_j = \tilde{\boldsymbol{B}}_j \left[\boldsymbol{B}_{j0}^{-1}\boldsymbol{\beta}_{j0} + \sigma^{jj}\sum_{i=1}^n \boldsymbol{x}_{ij}y_{ij}^* + \sum_{i=1}^n \boldsymbol{x}_{ij}\boldsymbol{\sigma}^{(-j)'}(\boldsymbol{y}_{i(-j)}^* - \boldsymbol{X}_{i(-j)}\boldsymbol{\beta}_{(-j)})\right]$.

6. Sample $\boldsymbol{\Sigma}^{-1}|\cdots \sim W\,[n_0 + n, \boldsymbol{Q}_0 + \sum_{i=1}^n (y_i^* - \boldsymbol{X}_i\boldsymbol{\beta})(y_i^* - \boldsymbol{X}_i\boldsymbol{\beta})']$.

3.2.2 Multivariate Ordered Probit Model with Individual Heterogeneity

Following the previous studies concerning the accommodation of individual heterogeneity in thresholds discussed in chapter 2, we consider the following transformation for γ_{ijc_j} to add heterogeneity in the cutpoints:

$$\gamma_{ijc_j} = \frac{\gamma_{ij(c_j-1)} + \exp(\boldsymbol{\zeta}_{jc_j}'\boldsymbol{z}_i)}{1 + \exp(\boldsymbol{\zeta}_{jc_j}'\boldsymbol{z}_i)}. \tag{3.6}$$

In this model, the outcome y_{ij} are specified as

$$y_{ij} = c_j \;\text{ if }\; \gamma_{ij(c_j-1)} < y_{ij}^* \leq \gamma_{ijc_j},\; j = 1, \ldots, m,\; c_j = 1, \ldots, \boldsymbol{C}_j, \tag{3.7}$$

where γ_{ijc_j} is a cutpoint for an individual i to the jth ordinal response, and C_j denotes the total number of categories.

Here, we set $\gamma_{ij1} = 0$ and $\gamma_{ij(C_j-1)} = 1$, to establish the identifiability of the cutoff parameters. Then, the cutpoints are fixed as

$$-\infty = \gamma_{ij0} < \gamma_{ij1} = 0 < \gamma_{ij2} < \cdots < \gamma_{ij(C_j-1)} = 1 < \gamma_{ijC_j} = \infty.$$

Further, we specify that

$$\mathcal{G}_{ij} = (\gamma_{ij(c_j-1)}, \gamma_{ijc_j}] \;\text{ if }\; y_{ij} = c_j, c_j = 1, \ldots, C_j, j = 1, \ldots, m,$$
$$\mathcal{G}_i = \mathcal{G}_{i1} \times \cdots \times \mathcal{G}_{im},\; i = 1, \ldots, n.$$

Then, the density function of $\boldsymbol{y}_i$ is

$$p(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}_i^*) = 1_{\{\boldsymbol{y}_i^* \in \mathcal{G}_i\}},\; i = 1, \ldots, n,$$

where $1_{(\cdot)}$ is an indicator function.

According to Bayes' theorem, the joint posterior distribution of the model can be written as

$$\begin{aligned}
&p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}^* | \boldsymbol{y}) \propto p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}^*) p(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}^*) \\
&\quad = p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) p(\boldsymbol{y}^* | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) p(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}^*) \\
&\quad = p(\boldsymbol{\beta}) p(\boldsymbol{\gamma}) p(\boldsymbol{\Sigma}) \left[\prod_{i=1}^{n} p(\boldsymbol{y}_i^* | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) p(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{y}_i^*) \right] \\
&\quad = \left[\prod_{j=1}^{m} p(\boldsymbol{\beta}_j) \right] \left[\prod_{j=1}^{m} \prod_{c_j=1}^{\boldsymbol{C}_{j-1}} p(\boldsymbol{\gamma}_{jc_j}) \right] \\
&\qquad \times p(\boldsymbol{\Sigma}) \left\{ \prod_{i=1}^{n} 1_{\{\boldsymbol{y}_i^* \in \mathcal{G}_i\}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\boldsymbol{y}_i^* - \boldsymbol{X}_i \boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i^* - \boldsymbol{X}_i \boldsymbol{\beta}) \right] \right\}.
\end{aligned} \tag{3.8}$$

The prior distributions of $\boldsymbol{\beta}_j$ and $\boldsymbol{\Sigma}$ can be specified as $\boldsymbol{\beta}_j \sim \mathrm{N}(\boldsymbol{\beta}_{j0}, \boldsymbol{B}_{j0})$ and $\boldsymbol{\Sigma}^{-1} \sim W(n_0, \boldsymbol{Q}_0^{-1})$, respectively.

Further, we specify the prior distribution of $\boldsymbol{\zeta}_{jc_j}$ as follows:

$$\boldsymbol{\zeta}_{jc_j} \sim \mathrm{N}(\boldsymbol{\zeta}_{j0}, \boldsymbol{G}_{j0}), \;\; j = 1, \ldots, m.$$

Then, we obtain the following algorithm.[4]

Algorithm 2. Sampling in a multivariate ordered probit model with individual heterogeneity.

---

[4]The derivation of algorithm 2 is provided in Appendix B.2.

1. Let $\boldsymbol{\zeta}_{jc_j}^s$ denote the value of $\boldsymbol{\zeta}_{jc_j}$ at the sth iteration. At the $(s+1)$th iteration, drawing $\boldsymbol{\zeta}_{jc_j}^{s+1} \sim q(\boldsymbol{\zeta}_{jc_j}|\cdots)$ from a Student-t proposal density $q(\boldsymbol{\zeta}_{jc_j}|\cdots) = f_T(\boldsymbol{\zeta}_{jc_j}|\hat{\boldsymbol{\zeta}}_{jc_j}, \hat{\boldsymbol{G}}_{jc_j}, \nu)$, where "$|\cdots$" denotes the conditioning of the other unspecified variables in the model and ν is the degrees of freedom,

$$\hat{\boldsymbol{\zeta}}_{jc_j} = \arg\max \log p(\boldsymbol{\zeta}_{jc_j}|\cdots), \hat{\boldsymbol{G}}_{jc_j} = \left\{ \left[-\frac{\partial^2 \log p(\boldsymbol{\zeta}_{jc_j}|\cdots)}{\partial \boldsymbol{\zeta}_{jc_j} \partial \boldsymbol{\zeta}'_{jc_j}} \right]_{\boldsymbol{\zeta}_{jc_j} = \hat{\boldsymbol{\zeta}}_{jc_j}} \right\}^{-1}.$$

2. Compute the acceptance ratio

$$r = \frac{p(\boldsymbol{\zeta}_{jc_j}^{s+1}|\cdots)}{p(\boldsymbol{\zeta}_{jc_j}^{s}|\cdots)} \frac{f_T(\boldsymbol{\zeta}_{jc_j}^{s}|\hat{\boldsymbol{\zeta}}_{jc_j}, \hat{\boldsymbol{G}}_{jc_j}, \nu)}{f_T(\tilde{\boldsymbol{\zeta}}_{jc_j}^{s+1}|\hat{\boldsymbol{\zeta}}_{jc_j}, \hat{\boldsymbol{G}}_{jc_j}, \nu)}.$$

3. Return $\boldsymbol{\zeta}_{jc_j}^{s+1}$ with probability $\min\{1, r\}$; otherwise, repeat using the previous value of $\boldsymbol{\zeta}_{jc_j}^{s}$.

4. Sample $y_{ij}^*|\cdots \sim \mathrm{TN}_{(\gamma_{ij(c_j-1)}, \gamma_{ijc_j}]}(\tilde{\mu}_{ij}, \tilde{\sigma}_j^2)$, where $\mathrm{TN}(\cdot)$ denotes the truncated normal distribution.

5. Sample $\boldsymbol{\beta}_j|\cdots \sim \mathrm{N}(\tilde{\boldsymbol{\beta}}_j, \tilde{\boldsymbol{B}}_j)$, where $\tilde{\boldsymbol{B}}_j = (\boldsymbol{B}_{j0}^{-1} + \sigma^{jj} \sum_{i=1}^n \boldsymbol{x}_{ij}\boldsymbol{x}'_{ij})^{-1}$,
$\tilde{\boldsymbol{\beta}}_j = \tilde{\boldsymbol{B}}_j \left[\boldsymbol{B}_{j0}^{-1}\boldsymbol{\beta}_{j0} + \sigma^{jj} \sum_{i=1}^n \boldsymbol{x}_{ij} y_{ij}^* + \sum_{i=1}^n \boldsymbol{x}_{ij} \boldsymbol{\sigma}^{(-j)\prime} (\boldsymbol{y}_{i(-j)}^* - \boldsymbol{X}_{i(-j)}\boldsymbol{\beta}_{(-j)}) \right]$.

6. Sample $\boldsymbol{\Sigma}^{-1}|\cdots \sim W\left[n_0 + n, \boldsymbol{Q}_0 + \sum_{i=1}^n (y_i^* - \boldsymbol{X}_i\boldsymbol{\beta})(y_i^* - \boldsymbol{X}_i\boldsymbol{\beta})'\right]$.

3.3 An Empirical Study

3.3.1 Data

In the empirical study, we employ the survey data from the WVS, which launched six waves of surveys between 1981 and 2014. In each wave, over

200 questions were proposed concerning respondents' feelings about their lives, state of health, economic situation, attitudes toward political action, and so on. Because we attempt to include immigration and religious statuses as individual heterogeneity in the cutpoint function, we choose three immigration countries, namely Australia, Canada, the United States, and employ the data from wave 5.[5]

We emply the following variables:

- Latent dependent variables: $lifesat$, $health$, *politicsinte*, and $choicefr$.
- Explanatory variables in the main regression: *married*, *male*, *work*, *midedu*, *highedu*, *age*, and *income*.
- Explanatory variables in the cutpoint function: *parim* and *religion*.

Here, $lifesat$ denotes life satisfaction, where the number of responses is $C_1 = 7$;[6] $health$ denotes the state of health, where the number of responses is $C_2 = 4$;[7] *politicsinte* denotes the interest in politics, which ranges from 1 to 4, where 1 denotes "not at all interested," 2 is "not very interested," 3 is "somewhat interested," and 4 is "very interested;"[8] $choicefr$ indicates how much freedom of choice and control the respondents feel, where 1 indicates "no choice at

[5]The latest wave of the survey, wave 6, does not include data on Canada, and so we decided to use wave 5, which was collected between 2005 and 2009.

[6]The ordinal category of life satisfaction is $C_1 = 10$ in wave 5, where $lifesat = 1$ denotes "completely dissatisfied" and $lifesat = 10$ denotes "completely satisfied." However, because the frequencies of $lifesat = 1$, $lifesat = 2$, $lifesat = 3$, and $lifesat = 4$ are small, we combine these four categories into the single category "$lifesat = 1$."

[7]The respondents are required to describe their state of health. $health = 1$ denotes "very good," and $health = 4$ denotes "poor." In our study, we modify the data such that $health = 1$ denotes "poor" and $health = 4$ denotes "very good."

[8]The original ranking is such that $politicsinte = 1$ denotes "very interested" and $politicsinte = 4$ denotes "not at all interested." We modify the data to ensure that $politicsinte = 1$ denotes "poor" and $politicsinte = 4$ denotes "very good."

all" and 7 means "a great deal of choice;"[9] and *married*, *male*, *work*, *parim*, and *religion* are dummy variables, which take the value 1 if the respondent is "married," "male," "having paid employment," "at least one of the parents is an immigrant," and "belonging to a religious denomination." Finally, *midedu* and *highedu* represent the middle and high education levels of respondents;[10] *age* denote the ages of a respondent; and *income* is the logarithm of the total household income.[11] The descriptive statistics of the data are listed in Tables 3.1 to 3.6.

3.3.2 Posterior Results

In this study, we conducted empirical analyses on the survey data from the WVS, focusing on the following research questions: (1) the impact of marital status, gender, labor force participation, education level, age and income on several related responses; and (2) how parent immigration status and religious status affect the thresholds in the multivariate ordered probit model.

There are two models in our empirical analysis. In the first, we employ algorithm 1 from Chen and Dey (2000) (see Section 2). In the second, we

[9] The original responses range from 1 to 10. However, the frequencies of $choicefr = 1$, $choicefr = 2$, $choicefr = 3$, and $choicefr = 4$ are small, so we combine them into "$choicefr = 1$."

[10] We capture education levels by three dummy variables for low (no formal education, incomplete primary school, complete primary school, less than high school, some high school but no diploma, incomplete secondary school: technical/vocational type, incomplete secondary school: to year 10), medium (graduated from high school, complete secondary school: technical/vocational type, incomplete secondary school: university-preparatory type, complete secondary school: university preparatory type) and high education (some college but no degree, trade apprenticeship, associate degree, bachelor's degree, master's degree, professional degree, doctorate degree).

[11] We construct numerical values for the income using the midpoint of each band(Layard et al. (2008)). If the respondents are in the lowest income band, then the income value is two-thirds of the upper limit of the band. If the respondents are in the highest income band, then the income value is 1.5 times the lower income limit of the band.

Table 3.1: Descriptive statistics of explanatory variables in the United States

Continuous variables	Min	Median	Max	Mean	SD
Age	18.00	48.00	91.00	48.02	17.02
Income	3333.33	44999.50	262500	52429.70	43177.45
Discrete variables				Frequency	%
Marital status	Married (*married*=1)			633	58.88
	Unmarried (*married*=0)			442	41.12
Low education	Education level is low (*lowedu*=1)			144	13.40
	Education level is not low (*lowedu*=0)			931	86.60
Median education	Education level is median (*medianedu*=1)			367	34.14
	Education level is not median (*medianedu*=0)			708	65.86
High education	Education level is high (*highedu*=1)			564	52.47
	Education level is not high (*highedu*=0)			511	47.53
Respondent's sex	Male (*male*=1)			540	50.23
	Female (*male*=0)			535	49.77
Employment status	Respondent is employed (*work*=1)			617	57.40
	Respondent is not employed (*work*=0)			458	42.60
Mother immigration status	Mother is an immigrant (*mothim*=1)			94	8.74
	Mother is not an immigrant (*mothim*=0)			981	91.26
Father immigration status	Father is an immigrant (*fathim*=1)			100	9.30
	Father is not an immigrant (*fathim*=0)			975	90.70
Religion status	Have a religion(*religion*=1)			803	74.70
	Not have a religion (*religion*=0)			272	25.30

accommodate heterogeneity in the cutpoints function, and employ the new algorithm 2 (see Section 2). The Markov chain Monte Carlo (MCMC) simulations were run for $11,000$ iterations, with the first $1,000$ samples being discarded as a burn-in period. The posterior results of these models were obtained using R. The convergence of MCMC chains has been confirmed by checking the effective sample size and the sample path.[12]

[12]The supplementary materials provide the tables of effective sample size and the figures of sample path.

Table 3.2: Descriptive statistics of explained variables in the United States

Explanatory variables		Frequency	%
Life satisfaction	(life satisfaction=1)	89	8.28
	(life satisfaction=2)	82	7.63
	(life satisfaction=3)	98	9.12
	(life satisfaction=4)	245	22.79
	(life satisfaction=5)	316	29.39
	(life satisfaction=6)	187	17.40
	(life satisfaction=7)	58	5.39
State of health	(state of health=1)	45	4.19
	(state of health=2)	168	15.63
	(state of health=3)	564	52.46
	(state of health=4)	298	27.72
Interest in politics	(interest in politics=1)	130	12.09
	(interest in politics=2)	275	25.58
	(interest in politics=3)	496	46.14
	(interest in politics=4)	174	16.19
Freedom of choice	(freedom of choice=1)	39	3.63
	(freedom of choice=2)	93	8.65
	(freedom of choice=3)	93	8.65
	(freedom of choice=4)	210	19.53
	(freedom of choice=5)	271	25.21
	(freedom of choice=6)	184	17.12
	(freedom of choice=7)	185	17.21

The linear functions for the latent variables are as follows:

$$\begin{aligned}
lifesat_i^* &= \beta_{10} + \beta_{11} married_i + \beta_{12} midedu_i + \beta_{13} highedu_i \\
&\quad + \beta_{14} age_i + \beta_{15} income_i + \beta_{16} male_i + \beta_{17} work_i + \epsilon_i, \\
health_i^* &= \beta_{20} + \beta_{21} married_i + \beta_{22} midedu_i + \beta_{23} highedu_i \\
&\quad + \beta_{24} age_i + \beta_{25} income_i + \beta_{26} male_i + \beta_{27} work_i + \epsilon_i, \\
politicsinte_i^* &= \beta_{30} + \beta_{31} married_i + \beta_{32} midedu_i + \beta_{33} highedu_i \\
&\quad + \beta_{34} age_i + \beta_{35} income_i + \beta_{36} male_i + \beta_{37} work_i + \epsilon_i, \\
choicefr_i^* &= \beta_{40} + \beta_{41} married_i + \beta_{42} midedu_i + \beta_{43} highedu_i \\
&\quad + \beta_{44} age_i + \beta_{45} income_i + \beta_{46} male_i + \beta_{47} work_i + \epsilon_i, ,
\end{aligned}$$

TABLE 3.3: Descriptive statistics of explanatory variables in Canada

Continuous variables	Min	Median	Max	Mean	SD
Age	16.00	47.00	94.00	47.76	16.83
Income	8333.33	46250.50	225000	57307.65	46285.74
Discrete variables				Frequency	%
Marital status	Married (married=1)			525	50.19
	Unmarried (married=0)			521	49.81
Low education	Education level is low (*lowedu*=1)			213	20.36
	Education level is not low (*lowedu*=0)			833	79.64
Median education	Education level is median (*medianedu*=1)			512	48.95
	Education level is not median (*medianedu*=0)			534	51.05
High education	Education level is high (*highedu*=1)			321	30.69
	Education level is not high (*highedu*=0)			725	69.31
Respondent's sex	Male (male=1)			434	41.49
	Female (male=0)			612	58.51
Employment status	Respondent is employed (work=1)			580	55.45
	Respondent is not employed (work=0)			466	44.55
Mother immigration status	Mother is an immigrant (mothim=1)			179	17.11
	Mother is not an immigrant (mothim=0)			867	82.89
Father immigration status	Father is an immigrant (fathim=1)			188	17.97
	Father is not an immigrant (fathim=0)			858	82.03
Religion status	Have a religion(religion=1)			736	70.36
	Not have a religion (religion=0)			310	29.64

where $\epsilon_i \sim \mathrm{N}(0, \mathbf{\Sigma})$ and $i = 1, \ldots, n$.

In the first model, the outcomes y_{ij} arise according to $y_{ij} = c_j$ if $\gamma_{j(c_j-1)} < y^*_{ij} \leq \gamma_{jc_j}$. We sample the cutpoints γ_{jc_j} by transforming them into $\delta_{jc_j} = \log\left(\frac{\gamma_{jc_j} - \gamma_{j(c_j-1)}}{1-\gamma_{jc_j}}\right)$. The prior distributions are specified as

$$\begin{aligned}
&\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I}), \\
&\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I}), \ j = 1, 2, 3, 4, \\
&\mathbf{\Sigma}^{-1} \sim W(5, 100 \times \boldsymbol{I}).
\end{aligned}$$

TABLE 3.4: Descriptive statistics of explained variables in Canada

Explanatory variables		Frequency	%
Life satisfaction	(life satisfaction=1)	49	4.68
	(life satisfaction=2)	70	6.69
	(life satisfaction=3)	77	7.36
	(life satisfaction=4)	181	17.31
	(life satisfaction=5)	350	33.46
	(life satisfaction=6)	168	16.06
	(life satisfaction=7)	151	14.43
State of health	(state of health=1)	46	4.40
	(state of health=2)	153	14.63
	(state of health=3)	421	40.25
	(state of health=4)	426	40.72
Interest in politics	(interest in politics=1)	188	17.97
	(interest in politics=2)	291	27.82
	(interest in politics=3)	414	39.58
	(interest in politics=4)	153	14.63
Freedom of choice	(freedom of choice=1)	46	4.40
	(freedom of choice=2)	88	8.41
	(freedom of choice=3)	100	9.56
	(freedom of choice=4)	170	16.25
	(freedom of choice=5)	307	29.35
	(freedom of choice=6)	166	15.87
	(freedom of choice=7)	169	16.16

In the second model, we add "parent immigration," and "religion status" to incorporate individual heterogeneity in the cutpoints function in algorithm 2 (see Section 2), which can be written as:

$$\gamma_{ijc_j} = \frac{\gamma_{ij(c_j-1)} + \exp(\zeta'_{jc_j0} + \zeta'_{jc_j1}parim_i + \zeta'_{jc_j2}religion_i)}{1 + \exp(\zeta'_{jc_j0} + \zeta'_{jc_j1}parim_i + \zeta'_{jc_j2}religion_i)},$$

where $parim_i$ and $religion_i$ are dummy variables denoting parents' immigration, and religious statuses, respectively. In the second model, we specify the

TABLE 3.5: Descriptive statistics of explanatory variables in Australia

Continuous variables	Min	Median	Max	Mean	SD
Age	18.00	49.00	90.00	48.43	15.72
Income	12000	48750.50	172500	64209.10	47180.70
Discrete variables				Frequency	%
Marital status	Married (*married*=1)			661	61.32
	Unmarried (*married*=0)			417	38.68
Low education	Education level is low (*lowedu*=1)			135	12.52
	Education level is not low (*lowedu*=0)			943	84.48
Median education	Education level is median (*medianedu*=1)			307	28.48
	Education level is not median (*medianedu*=0)			771	71.52
High education	Education level is high (*highedu*=1)			636	59.00
	Education level is not high (*highedu*=0)			442	41.00
Respondent's sex	Male (*male*=1)			503	46.66
	Female (*male*=0)			575	53.34
Employment status	Respondent is employed (*work*=1)			722	66.98
	Respondent is not employed (*work*=0)			356	33.02
Mother immigration status	Mother is an immigrant (*mothim*=1)			251	23.28
	Mother is not an immigrant (*mothim*=0)			827	76.72
Father immigration status	Father is an immigrant ($fathim$=1)			275	25.51
	Father is not an immigrant ($fathim$=0)			803	74.49
Religion status	Have a religion(*religion*=1)			653	60.58
	Not have a religion (*religion*=0)			425	39.42

prior distributions as

$$\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I}),$$

$$\boldsymbol{\zeta}_{jc_j} \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I}), \; j = 1, 2, 3, 4, \; c_j = 1, \ldots, C_j,$$

$$\boldsymbol{\Sigma}^{-1} \sim W(5, 100 \times \boldsymbol{I}).$$

In the above two models, $C_j = 7$ for both $lifesat_i$ and $choicefr_i$. Here, we fix the cutpoints as $\boldsymbol{\gamma}_{j0} = -\infty\boldsymbol{\iota}$, $\boldsymbol{\gamma}_{j1} = \mathbf{0}$, $\boldsymbol{\gamma}_{j6} = \boldsymbol{\iota}$, and $\boldsymbol{\gamma}_{j7} = \infty\boldsymbol{\iota}$, where $\boldsymbol{\iota} = (\underbrace{1, \ldots, 1}_{n})'$. For $health_i$ and $politicsinte_i$, $C_j = 4$, and we fix the cutpoints

TABLE 3.6: Descriptive statistics of explained variables in Australia

Explanatory variables		Frequency	%
Life satisfaction	(life satisfaction=1)	77	7.14
	(life satisfaction=2)	83	7.70
	(life satisfaction=3)	95	8.81
	(life satisfaction=4)	237	21.99
	(life satisfaction=5)	373	34.60
	(life satisfaction=6)	154	14.29
	(life satisfaction=7)	59	5.47
State of health	(state of health=1)	39	3.62
	(state of health=2)	184	17.07
	(state of health=3)	540	50.09
	(state of health=4)	315	29.22
Interest in politics	(interest in politics=1)	110	12.20
	(interest in politics=2)	311	28.85
	(interest in politics=3)	494	45.83
	(interest in politics=4)	163	15.12
Freedom of choice	(freedom of choice=1)	58	5.38
	(freedom of choice=2)	85	7.88
	(freedom of choice=3)	86	7.98
	(freedom of choice=4)	198	18.37
	(freedom of choice=5)	286	26.53
	(freedom of choice=6)	158	14.66
	(freedom of choice=7)	207	19.20

as $\boldsymbol{\gamma}_{j0} = -\infty\boldsymbol{\iota}$, $\boldsymbol{\gamma}_{j1} = \mathbf{0}$, $\boldsymbol{\gamma}_{j3} = \boldsymbol{\iota}$, and $\boldsymbol{\gamma}_{j4} = \infty\boldsymbol{\iota}$. In this article, the Bayesian

TABLE 3.7: Summary of Bayes factor

Country	Logarithm of Bayes factor	Results
Australia	$\log_{10} B_{21}$	29600.31
Canada	$log_{10}B_{21}$	31006.85
United States	$\log_{10} B_{21}$	28733.88

model selection proceeds by comparing the Bayes factors. The Bayes factor for any two models i and l is calculated as

$$B_{il} = \frac{m(\boldsymbol{y}|M_i)}{m(\boldsymbol{y}|M_l)},$$

where $m(\boldsymbol{y}|M_i)$ and $m(\boldsymbol{y}|M_l)$ are the marginal likelihood functions of M_i and M_l, respectively.

In addition, Jeffreys (1961) provided a scale on which to judge the evidence of M_i against M_l.[13] This scale is as follows:

- if $0 < \log_{10}(B_{il}) < 0.5$, then the evidence of M_i against M_l is *poor*,
- if $0.5 < \log_{10}(B_{il}) < 1$, then the evidence against M_l is *substantial*,
- if $1 < \log_{10}(B_{il}) < 2$, then the evidence is *strong*, and
- if $2 < \log_{10}(B_{il})$, then the evidence is *decisive*.

A summary of the Bayes factors is presented in Table 3.7.[14] The results show that model 2 with individual heterogeneity performs better than model 1 for all three countries.

Table 3.8 presents the results for Australia using model 2. The results indicate that *married* increases the respondents' state of health and level of life satisfaction; *male* has a positive effect on the interest in politics; *work* positively affects the state of health but has a negative effect on interest in politics; Both *midedu* and *highedu* have evident positive effects on state of health and life satisfaction, i.e., better educated respondents have higher degrees of health and life satisfaction; *age* increases interest in politics, but decreases the state of health; and *income* has an evident positive effect on all four response variables. In addition, the posterior results for $\boldsymbol{\zeta}$ show that *religion* increases the value

[13] See Robert (2001, p.228).

[14] The derivation of the marginal likelihood is provided in Appendix B.3 and B.4.

TABLE 3.8: The posterior results with individual heterogeneity in Australia

	State of health		Interest in politics		Life satisfaction		Freedom of choice	
Covariate	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Intercept (β_0)	−0.4169*	0.2297	−0.6906**	0.2321	−0.3851**	0.1568	−0.3976*	0.2101
Married (β_1)	0.0659**	0.0305	−0.0028	0.0307	0.0973**	0.0206	0.0687**	0.0276
Midedu (β_2)	0.1170**	0.0465	0.0662	0.0480	0.0533*	0.0319	0.0632	0.0435
Highedu (β_3)	0.1677**	0.0448	0.1430**	0.0462	0.0806**	0.0313	0.0491	0.0419
Age (β_4)	−0.0021*	0.0011	0.0068**	0.0011	0.0021**	0.0007	0.0025**	0.0010
Income (β_5)	0.0969**	0.0214	0.0757**	0.0217	0.0567**	0.0146	0.0765**	0.0195
Male (β_6)	−0.0032	0.0279	0.0949**	0.0285	−0.0017	0.0187	−0.0137	0.0252
Work (β_7)	0.0980**	0.0356	−0.0711*	0.0365	−0.0183	0.0244	0.0092	0.0325
Intercept (ζ_{20})	−0.5510**	0.1353	−0.1900**	0.1041	−1.9379**	0.2051	−1.5570**	0.2272
Parim (ζ_{21})	0.1049	0.1385	−0.1395	0.1317	0.1506	0.1999	−0.3134	0.2234
Religion (ζ_{22})	0.2909**	0.1387	−0.0853	0.1203	−0.0731	0.1926	0.4372**	0.2154
Intercept (ζ_{30})					−2.0836**	0.1772	−1.5497**	0.1873
Parim (ζ_{31})					−0.0424	0.2141	0.2488	0.2222
Religion (ζ_{32})					0.0892	0.2017	−0.2594	0.2178
Intercept (ζ_{40})					−1.0107**	0.1122	−0.6381**	0.1358
Parim (ζ_{41})					−0.0493	0.1405	−0.1523	0.1717
Religion (ζ_{42})					−0.1008	0.1298	−0.0552	0.1596
Intercept (ζ_{50})					−0.6150**	0.1155	−0.1703	0.1421
Parim (ζ_{51})					−0.0467	0.1304	0.0190	0.1673
Religion (ζ_{52})					−0.0014	0.1251	0.0257	0.1623

Notes: "∗" denotes that zero is not included in the 90% credible interval. "∗∗" denotes that zero is not included in the 95% credible interval. Mean and SD are the posterior mean and posterior standard deviation, respectively.

of $\boldsymbol{\gamma}_2$ on the state of health and freedom of choice. That is, *religon* can expand the interval $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2]$ on these explained variables. Thus, under the same conditions, individuals identifying with a religion report lower degrees of health and less control over choice with the categories $health = 3$ and $choicefr = 3$.

The posterior results for Canada using model 2 are listed in Table 3.9. The results illustrate that married respondents have higher levels of life satisfaction. Furthermore, males experience a lower freedom of choice and are more

Table 3.9: The posterior results with individual heterogeneity in Canada

	State of health		Interest in politics		Life satisfaction		Freedom of choice	
Covariate	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Intercept (β_0)	−0.1834	0.2638	−0.0625	0.2564	−0.0139	0.1754	0.0161	0.1827
Married (β_1)	0.0172	0.0375	0.0443	0.0358	0.1250**	0.0249	−0.0148	0.0260
Midedu (β_2)	0.1346**	0.0470	0.1059**	0.0461	−0.0035	0.0318	0.0328	0.0329
Highedu (β_3)	0.1763**	0.0535	0.3216**	0.0524	−0.0036	0.0360	0.0267	0.0370
Age (β_4)	−0.0022*	0.0012	0.0043**	0.0011	0.0018**	0.0008	0.0015**	0.0008
Income (β_5)	0.0928**	0.0258	0.0106	0.0249	0.0430**	0.0171	0.0476**	0.0177
Male (β_6)	−0.0482	0.0344	0.1446**	0.0336	−0.0360	0.0226	−0.0537**	0.0233
Work (β_7)	0.1238**	0.0391	−0.0389	0.0379	0.0188	0.0262	0.0411	0.0274
Intercept (ζ_{20})	−0.3600**	0.1607	−0.4109**	0.1338	−1.7977**	0.2347	−1.4531**	0.2006
Parim (ζ_{21})	0.2915	0.1764	0.1183	0.1591	−0.0426	0.2389	0.2452	0.1990
Religion (ζ_{22})	−0.0016	0.1728	0.0278	0.1505	0.2342	0.2362	0.0832	0.1985
Intercept (ζ_{30})					−1.7502**	0.2104	−1.5538**	0.1889
Parim (ζ_{31})					−0.4310	0.3011	−0.0336	0.2407
Religion (ζ_{32})					−0.0335	0.2372	0.0190	0.2174
Intercept (ζ_{40})					−0.8620**	0.1405	−1.0551**	0.1577
Parim (ζ_{41})					0.1575	0.1765	−0.2745	0.2006
Religion (ζ_{42})					−0.2448	0.1615	0.1744	0.1786
Intercept (ζ_{50})					−0.2752**	0.1371	−0.2052	0.1460
Parim (ζ_{51})					−0.0143	0.1627	−0.0148	0.1760
Religion (ζ_{52})					0.0574	0.1520	−0.0338	0.1630

Notes: "∗" denotes that zero is not included in the 90% credible interval. "∗∗" denotes that zero is not included in the 95% credible interval. Mean and SD are the posterior mean and posterior standard deviation, respectively.

interested in politics; respondents in paid employment have a higher state of health; and better educated respondents reported having a higher state of health, and are more interested in politics. Furthermore, *age* has positive effects on interest in politics, life satisfaction and freedom of choice; and *income* has a positive effect on state of health, life satisfaction and freedom of choice. In addition, *religion* and *parim* exhibit no evident impact on the response variables.

TABLE 3.10: The posterior results with individual heterogeneity in the United States

	State of health		Interest in politics		Life satisfaction		Freedom of choice	
Covariate	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Intercept (β_0)	−0.1846	0.1844	−0.1147	0.1952	−0.1252	0.1359	0.1200	0.1543
Married (β_1)	0.0035	0.0298	0.0227	0.0310	0.0704**	0.0216	0.0319	0.0242
Midedu (β_2)	−0.0085	0.0441	0.0260	0.0474	−0.0059	0.0324	−0.0271	0.0373
Highedu (β_3)	0.0883**	0.0442	0.2240**	0.0472	0.0362	0.0325	−0.0015	0.0368
Age (β_4)	−0.0029**	0.0009	0.0055**	0.0009	0.0006	0.0006	0.0013*	0.0007
Income (β_5)	0.0900**	0.0182	0.0195	0.0192	0.0457**	0.0134	0.0439**	0.0152
Male (β_6)	0.0034	0.0280	0.0951**	0.0293	0.0026	0.0202	−0.0533**	0.0231
Work (β_7)	0.1262**	0.0322	0.0009	0.0335	−0.0152	0.0234	0.0352	0.0264
Intercept (ζ_{20})	−0.2645**	0.1365	−0.2888**	0.1270	−1.7249**	0.1815	−1.1164**	0.1800
Parim (ζ_{21})	0.0350	0.2235	0.2640	0.1841	−0.1127	0.3419	0.0661	0.2703
Religion (ζ_{22})	−0.3207**	0.1526	−0.2322	0.1419	−0.4139*	0.2067	−0.1683	0.1868
Intercept (ζ_{30})					−1.9193**	0.1860	−1.5648**	0.2010
Parim (ζ_{31})					0.1583	0.2888	0.1733	0.3084
Religion (ζ_{32})					−0.1566	0.2141	−0.0733	0.2296
Intercept (ζ_{40})					−1.0017**	0.1401	−0.7050**	0.1574
Parim (ζ_{41})					−0.1089	0.2027	0.1275	0.2321
Religion (ζ_{42})					0.0193	0.1536	−0.0182	0.1763
Intercept (ζ_{50})					−0.6305**	0.1615	−0.1838**	0.1790
Parim (ζ_{51})					0.0102	0.1970	−0.0348	0.2740
Religion (ζ_{52})					0.0654	0.1717	0.0186	0.1978

Notes: "$*$" denotes that zero is not included in the 90% credible interval. "$**$" denotes that zero is not included in the 95% credible interval. Mean and SD are the posterior mean and posterior standard deviation, respectively.

Table 3.10 presents the posterior results of model 2 for the United States. The results indicate that *married* has a positive effect on life satisfaction but no evident impact on respondents' state of health, interest in politics, or feeling of freedom. Males have a significantly higher interest in politics, but lower feeling of freedom than females. The respondents with paid employment experience a higher state of health. Furthermore, *highedu* has a positive effect on the state of health, interest in politics, but does not significantly impact life satisfaction and freedom of choice; *age* positively affects respondents' interest in politics but negatively affects the state of health; and *income* has a positive effect on the state of health, life satisfaction and freedom of choice. In addition, the posterior results for $\boldsymbol{\zeta}$ imply that *religion* decreases the value of $\boldsymbol{\gamma}_2$ on the state of health and life satisfaction. That is, *religon* can narrow the interval $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2]$ on these explained variables. Thus, under the same conditions, individuals identifying with a religion report a higher state of health and a higher level of life satisfaction in the categories of $health = 2$, $politicsinte = 2$, and $lifesat = 2$.

In summary, *married* can increase the level of life satisfaction in all three countries, but only has a positive effect on the state of health and freedom of choice in Australia. Furthermore, *male* tend to be more interested in politics. *income* have siginificant positive effects on the life satisfaction and freedom of choice. In addition, respondents in paid employment tend to be healthier. Better educated individuals appear to enjoy better health, and are more interested in politics. With aging, respondents become more interested in politics, while feeling less healthy. Moreover, the effects of heterogeneity are different among the three countries. *religion* evidently impacts the state of health and freedom of choice in Australia, and has significant effects on the state of health, interest in politics, and life satisfaction in the United States. Furthermore, our findings suggest that *parim* has no evident impact on the thresholds in any of the three countries.

3.4 Summary

In this chapter, we describe the Bayesian multivariate ordered probit model introduced by Chen and Dey (2000) (algorithm 1) and propose a new algorithm that includes individual heterogeneity in the cutpoints function (algorithm 2). Then, we apply the two algorithms to an analysis of individuals' life satisfaction, state of health, interest in politics, and freedom of choice, using the survey data from WVS 5, collected between 2005 and 2009. The empirical results indicate the following:

- The algorithm with individual heterogeneity performs better than that without individual heterogeneity.
- The posterior results for the explanatory variables indicate that being in paid employment, having a higher income, and having a higher education level have positive effects on the state of health. Being male, being better educated, and being older can increase the interest in politics. Being married, and having a higher income lead to a higher degree of life satisfaction and freedom of choice.
- The effects of heterogeneity are different among the three countries. *religion* has an evident impact on the state of health and freedom of choice in Australia and has a significant effect on the state of health, interest in politics, and life satisfaction in the United States. However, parents' immigration status has no evident impact on any of the thresholds in any of the three countries.

Chapter 4

Bayesian Dynamic Panel-ordered Probit Model with Individual Heterogeneity

4.1 Introduction

In their seminal study of the Bayesian ordered-probit model, Albert and Chib (1993) use latent variable representation to estimate the posterior distribution of explanatory variables. However, the sampling of cutpoints in their model is not conditioned on latent data. Later, Cowles (1996), Nandram and Chen (1996), and Chen and Dey (2000) offered several suggestions on the sampling of cutpoints. Further to these studies, Albert and Chib (2001) considered an algorithm to transform cutpoints into an unconstrained variable and remove their ordering constraint using a one-to-one map. Thus, the proposal density for a new variable can be acquired by unconstrained optimization, with the prior of the variable set as unrestricted. Subsequently, Chen and Dey (2000) proposed a sampling algorithm identified by fixing two cutpoints, leaving for estimation the variance in error term.

Besides the main algorithms mentioned above, Müller and Czado (2005) present a Bayesian auto regressive ordered probit model based on the Liu and Sabatti (2000) method. Further, Hasegawa (2009) extends these algorithms to the Bayesian dynamic ordered probit model. From the result obtained in Hasegawa (2009) using simulated data, the algorithms following Liu and Sabatti (2000) and Chen and Dey (2000) perform well. In this chapter, we extend the Hasegawa (2009) model to a new Bayesian algorithm using Chen and Dey (2000) with individual heterogeneity in the cutpoints function.

In subjective well-being analysis, the backgrounds and preferences of individuals could be complicated, with the respondents understanding the concepts differently (Brady, 1985). Several studies have suggested the accommodation of individual heterogeneity in ordered choice models. The seminal paper Terza (1985) considered a linear specification to ensure that the ordered choice model cutpoints vary across individuals. Subsequently, Greene (2007), Eluru et al. (2008), and Greene and Hensher (2010b) proposed a hierarchical ordered probit model with nonlinear cutpoint specification. Furthermore, King et al. (2004) and Kapteyn et al. (2007) suggest an anchoring vignettes method to adjust for response scale differences between groups. Shi and Hasegawa (2019) introduce a Bayesian univariate ordered probit model incorporating individual heterogeneity in cutpoints function. However, to the best of our knowledge, no study has so far accommodated individual heterogeneity in dynamic panel-ordered probit models. Thus, we consider a new Bayesian algorithm with individual heterogeneity to deal with panel data in a dynamic ordered probit model.

In this empirical study, we apply an algorithm with individual heterogeneity to subjective well-being in China. We employ panel data taken from the China Family Panel Studies (CFPS) baseline survey launched in 2010 and three waves of their follow-up surveys in 2012, 2014, and 2016. We select marriage, income, and work as the time-variant explanatory variables, with age, gender, and education year as the time-invariant variables. Furthermore, we incorporate age, gender, and education year as individual heterogeneity factors in the cutpoints function. From the posterior results, the lagged latent variable increases life satisfaction; marriage, work, and age have positive effects on life satisfaction for all periods; and education year increases life satisfaction initially. Also, age has evident impacts on the thresholds, but education year and gender have no impact.

4.2 Model Specification

4.2.1 Dynamic Panel-ordered Probit Model

In this section, we first describe the Bayesian dynamic panel-ordered probit model based on Hasegawa (2009). Let y_{it} be the ordinal response of individual i at time t for $i = 1, \ldots, n$ and $t = 0, 1, \ldots, T$; that is, $y_{it} = j$ for $j = 1, \ldots, J$, where J is the total number of distinct categories and $t = 0$ denotes the initial period. Let y_i^* denote a latent variable that depends on the linear model:

$$\begin{aligned} y_{i0}^* &= \boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0 + u_{i0},\ t = 0, \qquad (4.1) \\ y_{it}^* &= \phi y_{i(t-1)}^* + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta} + u_{it},\ t = 1, \ldots, T, \end{aligned}$$

where $\boldsymbol{x}_{it} = (x_{it1}, \ldots, x_{itk})'$ is a vector of time-variant explanatory variables, $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{ip})'$ is a vector of cross-sectional time-invariant explanatory variables, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)'$, $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0k})'$, and $\boldsymbol{\delta}_0 = (\delta_{01}, \ldots, \delta_{0p})'$.

Then, the outcome y_{it} can be determined by

$$y_{it} = j \ \text{ if } \ \gamma_{j-1} < y_{it}^* \leq \gamma_j, \ t = 0, \ldots, T, \tag{4.2}$$

where γ_j is a cutpoint for the jth ordinal response. Here, the cutpoints are specified as

$$-\infty = \gamma_0 < \gamma_1 = 0 < \gamma_2 < \cdots < \gamma_{(J-1)} = 1 < \gamma_J = \infty,$$

such that $\gamma_1 = 0$ and $\gamma_{(J-1)} = 1$ are required to establish the identifiability of the cutoff parameters.[1]

In addition, Chen and Dey (2000) considered the following transformation for γ_j in order to remove the constraints on the cutpoints:

$$\nu_j = \log\left(\frac{\gamma_j - \gamma_{j-1}}{1 - \gamma_j}\right), \ j = 2, \ldots, J-2.$$

Given the prior distributions

$$\phi \sim \mathrm{N}(\tilde{\phi}, \tilde{\omega})1_{\{\phi \in (-1,1)\}}, \quad \boldsymbol{\beta} \sim \mathrm{N}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{B}}), \quad \boldsymbol{\delta} \sim \mathrm{N}(\tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{D}}), \quad \boldsymbol{\beta}_0 \sim \mathrm{N}(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{B}}_0),$$
$$\boldsymbol{\delta}_0 \sim \mathrm{N}(\tilde{\boldsymbol{\delta}}_0, \tilde{\boldsymbol{D}}_0), \quad \sigma^{-2} \sim \mathrm{gamma}(\tilde{c}, \tilde{d}), \quad \boldsymbol{\nu}(\boldsymbol{\gamma}) \sim \mathrm{N}(\boldsymbol{\nu}_0, \boldsymbol{V}_0), \quad i = 1, \ldots, n,$$

[1] See Chen and Dey (2000, p.136).

the joint posterior distribution of parameters in the model can be written as

$$
\begin{aligned}
p(\phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{y}^*, \boldsymbol{y}_0^* | \boldsymbol{y}, \boldsymbol{y}_0) &\propto p(\phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{y}^*, \boldsymbol{y}_0^*) \\
&\times p(\boldsymbol{y}, \boldsymbol{y}_0 | \phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{y}^*, \boldsymbol{y}_0^*) \\
&= p(\phi) p(\boldsymbol{\beta}) p(\boldsymbol{\delta}) p(\boldsymbol{\beta}_0) p(\boldsymbol{\delta}_0) p(\sigma^2) \prod_{j=2}^{J-2} p(\boldsymbol{\gamma}_j) \\
\times \prod_{i=1}^{n} [\{ p(\boldsymbol{y}_i^* | \phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \boldsymbol{\gamma}, y_{i0}^*) & p(\boldsymbol{y}_i | \phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{y}^*, y_{i0}^*) \} \\
\times \{ p(y_{i0}^* | \phi, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\gamma}) & p(y_{i0} | \phi, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\gamma}, y_{i0}^*) \}],
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})', \ \boldsymbol{y}_i^* = (y_{i1}^*, y_{i2}^*, \ldots, y_{iT}^*)', \ i = 1, \ldots, n \\
\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_n')', \ \boldsymbol{y}^* = (\boldsymbol{y}_1^{*\prime}, \boldsymbol{y}_2^{*\prime}, \ldots, \boldsymbol{y}_n^{*\prime})' \\
\boldsymbol{y}_0 = (y_{10}, y_{20}, \ldots, y_{n0})', \ \boldsymbol{y}_0^* = (y_{10}^*, y_{20}^*, \ldots, y_{n0}^*)'.
\end{aligned}
$$

Then, the algorithm is summarized as follows:[2]

Algorithm 1. Sampling in a Bayesian dynamic panel ordered probit model.

---

1. Sample $\phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2$ from their full conditional distributions(FCDs).

2. Sample $y_{i0}^* | \cdots \sim \mathrm{TN}_{(\gamma_{i(j-1)}, \gamma_{ij}]}((\hat{y}_{i0}^*, \frac{\sigma^2}{1+\phi^2})$, where $\mathrm{TN}(\cdot)$ denotes the truncated normal distribution.

3. Sample $y_{it}^* | \cdots \sim \mathrm{TN}_{(\gamma_{i(j-1)}, \gamma_{ij}]}(\hat{y}_{it}^*, \frac{\sigma^2}{1+\phi^2})$, where $t = 1, \ldots, T-1$.

4. Sample $y_{iT}^* | \cdots \sim \mathrm{TN}_{(\gamma_{i(j-1)}, \gamma_{ij}]}(\phi y_{i(T-1)}^* + \boldsymbol{x}_{iT}' \boldsymbol{\beta} + \boldsymbol{w}_i' \boldsymbol{\delta}, \sigma^2)$.

[2]The derivation of algorithm 1 is provided in Appendix C.1.

5. Let $\boldsymbol{\nu}_j^s$ denote the value of $\boldsymbol{\nu}_j$ at the sth iteration. At the $(s+1)$th iteration, drawing $\boldsymbol{\nu}_j^{s+1} \sim q(\boldsymbol{\nu}_j|\cdots)$ from a Student-t proposal density $q(\boldsymbol{\nu}_j|\cdots) = f_T(\boldsymbol{\nu}_j|\hat{\boldsymbol{\nu}}_j, \hat{\boldsymbol{G}}_j, \nu)$,

$$\hat{\boldsymbol{\nu}}_j = \arg\max \log p(\boldsymbol{\nu}_j|\cdots), \hat{\boldsymbol{G}}_j = \left\{ \left[-\frac{\partial^2 \log \pi(\boldsymbol{\nu}_j|\cdots)}{\partial \boldsymbol{\nu}_j \partial \boldsymbol{\nu}_j'} \right]_{\boldsymbol{\nu}_j = \hat{\boldsymbol{\nu}}_j} \right\}^{-1}.$$

6. Compute the acceptance ratio

$$r = \frac{p(\boldsymbol{\nu}_j^{s+1}|\cdots)}{p(\boldsymbol{\nu}_j^{s}|\cdots)} \frac{f_T(\boldsymbol{\nu}_j^s|\hat{\boldsymbol{\nu}}_j, \hat{\boldsymbol{G}}_j, \nu)}{f_T(\tilde{\boldsymbol{\nu}}_j^{s+1}|\hat{\boldsymbol{\nu}}_j, \hat{\boldsymbol{G}}_j, \nu)}.$$

7. Return $\boldsymbol{\nu}_j^{s+1}$ with probability $\min\{1, r\}$; otherwise, repeat using the previous value of $\boldsymbol{\nu}_j^s$.

8. Calculate $\gamma_j = \frac{\gamma_{j-1} + \exp(\nu_j)}{1 + \exp(\nu_j)}$.

4.2.2 Dynamic Panel-ordered Probit Model with Individual Heterogeneity

We modify the cutpoints function to accommodate individual heterogeneity as follows:

$$\gamma_{ij} = \frac{\gamma_{i(j-1)} + \exp(\boldsymbol{\zeta}_j' \boldsymbol{z}_i)}{1 + \exp(\boldsymbol{\zeta}_j' \boldsymbol{z}_i)}, \; j = 2, \ldots, J-2, \tag{4.3}$$

where $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im})'$ and $\boldsymbol{\zeta}_j = (\zeta_{j1}, \ldots, \zeta_{jm})'$. Furthermore, we introduce the prior distributions

$$\phi \sim \mathrm{N}(\tilde{\phi}, \tilde{\omega})1_{\{\phi\in(-1,1)\}}, \quad \boldsymbol{\beta} \sim \mathrm{N}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{B}}), \quad \boldsymbol{\delta} \sim \mathrm{N}(\tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{D}}), \quad \boldsymbol{\beta}_0 \sim \mathrm{N}(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{B}}_0),$$
$$\boldsymbol{\delta}_0 \sim \mathrm{N}(\tilde{\boldsymbol{\delta}}_0, \tilde{\boldsymbol{D}}_0), \quad \sigma^{-2} \sim \mathrm{gamma}(\tilde{c}, \tilde{d}), \quad \boldsymbol{\zeta}_j \sim \mathrm{N}(\tilde{\boldsymbol{\zeta}}, \tilde{\boldsymbol{G}}), \quad i = 1, \ldots, n.$$

Then, the joint posterior distribution can be written as

$$\begin{aligned} p(\phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\zeta}, \boldsymbol{y}^*, \boldsymbol{y}_0^* | \boldsymbol{y}, \boldsymbol{y}_0) &\propto p(\phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\zeta}, \boldsymbol{y}^*, \boldsymbol{y}_0^*) \\ &\quad \times p(\boldsymbol{y}, \boldsymbol{y}_0 | \phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\zeta}, \boldsymbol{y}^*, \boldsymbol{y}_0^*) \\ &= p(\phi)p(\boldsymbol{\beta})p(\boldsymbol{\delta})p(\boldsymbol{\beta}_0)p(\boldsymbol{\delta}_0)p(\sigma^2)\prod_{j=2}^{J-2} p(\boldsymbol{\zeta}_j) \\ \times \prod_{i=1}^{n} [\{p(\boldsymbol{y}_i^* | \phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \boldsymbol{\zeta}, y_{i0}^*) & p(\boldsymbol{y}_i | \phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \boldsymbol{\zeta}, \boldsymbol{y}^*, y_{i0}^*)\} \\ &\times \{p(y_{i0}^* | \phi, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\zeta}) p(y_{i0} | \phi, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2, \boldsymbol{\zeta}, y_{i0}^*)\}]. \end{aligned}$$

Now, we obtain the following algorithm.[3]

Algorithm 2. Sampling in a Bayesian dynamic panel ordered probit model with individual heterogeneity.

---

1. Sample $\phi, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0, \sigma^2$ from their FCDs.
2. Sample $y_{i0}^* | \cdots \sim \mathrm{TN}_{(\gamma_{i(j-1)}, \gamma_{ij}]}((\hat{y}_{i0}^*, \frac{\sigma^2}{1+\phi^2})$.
3. Sample $y_{it}^* | \cdots \sim \mathrm{TN}_{(\gamma_{i(j-1)}, \gamma_{ij}]}(\hat{y}_{it}^*, \frac{\sigma^2}{1+\phi^2})$, where $t = 1, \ldots, T-1$.
4. Sample $y_{iT}^* | \cdots \sim \mathrm{TN}_{(\gamma_{i(j-1)}, \gamma_{ij}]}(\phi y_{i(T-1)}^* + \boldsymbol{x}_{iT}' \boldsymbol{\beta} + \boldsymbol{w}_i' \boldsymbol{\delta}, \sigma^2)$.

[3] The derivation of Algorithm 2 is provided in Appendix C.2.

5. Let $\boldsymbol{\zeta}_j^s$ denote the value of $\boldsymbol{\zeta}_j$ at the sth iteration. At the $(s+1)$th iteration, drawing $\boldsymbol{\zeta}_j^{s+1} \sim q(\boldsymbol{\zeta}_j|\cdots)$ from a Student-t proposal density $q(\boldsymbol{\zeta}_j|\cdots) = f_T(\boldsymbol{\zeta}_j|\hat{\boldsymbol{\zeta}}_j, \hat{\boldsymbol{G}}_j, \nu)$,

$$\hat{\boldsymbol{\zeta}}_j = \arg\max \log p(\boldsymbol{\zeta}_j|\cdots), \hat{\boldsymbol{G}}_j = \left\{ \left[-\frac{\partial^2 \log \pi(\boldsymbol{\zeta}_j|\cdots)}{\partial \boldsymbol{\zeta}_j \partial \boldsymbol{\zeta}_j'} \right]_{\boldsymbol{\zeta}_j = \hat{\boldsymbol{\zeta}}_j} \right\}^{-1}.$$

6. Compute the acceptance ratio

$$r = \frac{p(\boldsymbol{\zeta}_j^{s+1}|\cdots)}{p(\boldsymbol{\zeta}_j^{s}|\cdots)} \frac{f_T(\boldsymbol{\zeta}_j^s|\hat{\boldsymbol{\zeta}}_j, \hat{\boldsymbol{G}}_j, \nu)}{f_T(\tilde{\boldsymbol{\zeta}}_j^{s+1}|\hat{\boldsymbol{\zeta}}_j, \hat{\boldsymbol{G}}_j, \nu)}.$$

7. Return $\boldsymbol{\zeta}_j^{s+1}$ with probability $\min\{1, r\}$; otherwise, repeat using the previous value of $\boldsymbol{\zeta}_j^s$.

8. Calculate $\gamma_{ij} = \frac{\gamma_{ij-1} + \exp(\boldsymbol{\zeta}_j' \boldsymbol{z}_i)}{1 + \exp(\boldsymbol{\zeta}_j' \boldsymbol{z}_i)}$.

4.3 An Empirical Study

4.3.1 Data

For our empirical study, we use panel data from the China Family Panel Studies (CFPS), administered by the Institute of Social Science Survey at Peking University. The CFPS is a national annual longitudinal survey of Chinese communities, families, and individuals, launched in 2010 as a baseline survey, with three waves of follow-up surveys in 2012, 2014, and 2016. Each wave covers over 10,000 households, interviewing them on their economic activities, educational outcomes, family relationships, life satisfaction, state of health, and so on. Since we are analyzing the happiness level in China, we choose the following variables:

- Latent dependent variables: $lifesat(\boldsymbol{y})$.
- Explanatory variables in the main regression:
  - Time-variant explanatory variables: $married(\boldsymbol{x}_1)$, $logincome(\boldsymbol{x}_2)$, $inwork(\boldsymbol{x}_3)$.
  - Time-invariant explanatory variables: $age(\boldsymbol{w}_1)$, $male(\boldsymbol{w}_2)$, $eduy(\boldsymbol{w}_3)$.
- Explanatory variables in the cutpoints function: $age(\boldsymbol{z}_1)$, $male(\boldsymbol{z}_2)$, $eduy(\boldsymbol{z}_3)$.

Here, $lifesat$ denotes life satisfaction with five responses; $married$, $inwork$, and $male$ are dummy variables,[4] which take the value 1 if the respondent is "married," "has paid employment," and is "male;" $logincome$ is the logarithm of total annual income;[5] age denotes the respondents' age in Wave 1; and $eduy$ represents the respondents' education years in Wave 1. The descriptive statistics of the data are listed in Table 4.1.

4.3.2 Posterior Results

In this section, we apply the proposed algorithm to real data. We run MCMC simulations for $30,000$ iterations, with the first $10,000$ samples discarded as burn-in period. We then obtain the posterior results of these models using `R`. We confirm the convergence of MCMC by checking the effective sample size and sample path.[6]

[4] We obtain the data of *male* from Wave 1.

[5] In the case of zero income, we set *logincome* equal to zero.

[6] The supplementary materials provide the tables of effective sample size and the figures of sample path.

Table 4.1: Summary statistics

	life satisfaction				
	Strongly dissatisfied	Dissatisfied	Moderately satisfied	Satisfied	Strongly satisfied
Wave	1	2	3	4	5
1	48(4.98)	120(12.45)	343(35.58)	272(28.22)	181(18.78)
2	63(6.54)	120(12.45)	423(43.88)	215(22.30)	143(14.83)
3	25(2.59)	53(5.50)	295(30.60)	332(34.44)	259(26.87)
4	51(5.29)	111(11.51)	353(36.62)	238(24.69)	211(21.89)
	married		*inwork*		
Wave	no(0)	yes(1)	no(0)	yes(1)	
1	90(9.34)	874(90.66)	344(35.68)	620(64.32)	
2	66(6.85)	898(93.15)	119(12.34)	845(87.66)	
3	72(7.47)	892(92.53)	51(5.29)	913(94.71)	
4	71(7.37)	893(92.63)	63(6.54)	901(93.46)	
			log(income)		
Wave	Min	Median	Max	Mean	SD
1	0.0000	8.8537	13.5924	7.0925	3.8231
2	0.0000	8.5172	12.5913	5.6154	4.7624
3	0.0000	9.2103	12.9991	6.2036	4.7304
4	0.0000	9.4727	13.8547	6.6161	4.7327
	Min	Median	Max	Mean	SD
age	16	40	73	40.85	10.43
eduy	0	9	16	7.93	3.94
	no(0)	yes(1)			
male	421(43.67)	543(56.33)			

Notes: Min, Median and Max denote the minimum value, median and maximum value. Mean and SD denote the sample mean and standard deviation. The value in $(\cdot)$ indicates a percentage. The data of *age*, *eduy* and *male* are obtained from Wave 1.

The estimated functions for the latent variables are

$$lifesat^*_{i0} = \beta_{01} married_{i0} + \beta_{02} logincome_{i0} + \beta_{03} inwork_{i0}$$

$$+\delta_{01} age_i + \delta_{02} eduy_i + \delta_{03} male_i + u_{i0},$$

$$lifesat^*_{it} = \phi lifesat^*_{i(t-1)} + \beta_1 married_{it} + \beta_2 logincome_{it} + \beta_3 inwork_{it}$$

$$+\delta_1 age_i + \delta_2 eduy_i + \delta_3 male_i + u_{it},$$

where $u_{it} \sim \mathrm{N}(0, \sigma^2)$, $t = 1, \ldots, 3$ and $i = 1, \ldots, n$. Further, we incorporate "age," "male," and "education year" as individual heterogeneity in the cutpoints function. Then, we have the following function:

$$\gamma_{ij} = \frac{\gamma_{i(j-1)} + \exp(\zeta'_{j0} + \zeta'_{j1} age_i + \zeta'_{j2} male_i + \zeta'_{j3} eduy_i)}{1 + \exp(\zeta'_{j0} + \zeta'_{j1} age_i + \zeta'_{j2} male_i + \zeta'_{j3} eduy_i)}, \quad j = 2, 3.$$

In this model, the prior distributions are specified as

$$\phi \sim \mathrm{N}(0, 100) 1_{\{\phi \in (-1,1)\}}, \quad \boldsymbol{\beta}_0 \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I}), \quad \boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I}),$$
$$\boldsymbol{\delta} \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I}), \quad \boldsymbol{\delta}_0 \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I}), \quad \boldsymbol{\zeta}_j \sim \mathrm{N}(\mathbf{0}, 100 \times \boldsymbol{I}), \ j = 2, 3,$$
$$\sigma^{-2} \sim \mathrm{gamma}(0.01, 0.01).$$

We then fix the cutpoints as $\boldsymbol{\gamma}_0 = -\infty \boldsymbol{\iota}$, $\boldsymbol{\gamma}_1 = \mathbf{0}$, $\boldsymbol{\gamma}_4 = \boldsymbol{\iota}$, and $\boldsymbol{\gamma}_5 = \infty \boldsymbol{\iota}$, where $\boldsymbol{\iota} = (\underbrace{1, \ldots, 1}_{n})'$.

The posterior results for all parameters are presented in Table 4.2. From the results, the lagged latent variable ($y^*_{(-1)}$) has an evident positive effect on life satisfaction; *married* and *work* increase the respondents' life satisfaction for all periods; the time-invariant variable *age* has positive effects on life satisfaction for all periods; education year has positive effects only at the initial period $t = 0$, but no effect for $t \geq 1$; and *income* and *gender* do not have evident effects on life satisfaction for all periods. Additionally, the posterior results for $\boldsymbol{\zeta}$ show that *age* increases γ_3, with no effect on the value of γ_2. Further, *education* and *gender* do not have any evident effects on both γ_2 and γ_3.

Table 4.3 shows the posterior results for the partial effects of the explanatory variables on the response probabilities.[7] The lagged latent variable

[7]The details of partial effects are provided in Appendix C.3.

TABLE 4.2: The posterior results for parameters (CFPS data)

	Mean	SD
$\phi(y^*_{(-1)})$	0.2777^{**}	0.0132
Married (β_{01})	0.1005^{**}	0.0440
Income (β_{02})	0.0009	0.0040
Work (β_{03})	0.0718^{**}	0.0299
Age (δ_{01})	0.0087^{**}	0.0010
Education (δ_{02})	0.0117^{**}	0.0034
Gender (δ_{03})	-0.0090	0.0306
Married (β_1)	0.1769^{**}	0.0275
Income (β_2)	0.0013	0.0017
Work (β_3)	0.0783^{**}	0.0262
Age (δ_1)	0.0053^{**}	0.0006
Education (δ_2)	-0.0011	0.0020
Gender (δ_3)	-0.0067	0.0175
Intercept (ζ_{20})	-1.6976^{**}	0.2586
Age (ζ_{21})	0.0043	0.0052
Education (ζ_{22})	-0.0131	0.0136
Gender (ζ_{23})	0.1011	0.1074
Intercept (ζ_{30})	-0.0362	0.1929
Age (ζ_{31})	0.0063^{*}	0.0038
Education (ζ_{32})	0.0043	0.0102
Gender (ζ_{33})	-0.1228	0.0783

Notes: "$**$" and "$*$" denote that zero is not included in the 95% and 90% credible intervals. Mean and SD denote the posterior mean and posterior standard deviation.

($(y^*_{(-1)})$), *married*, *work*, *age* and *education* have evident partial effects on life satisfaction, while *income* and *gender* do not. The lagged latent variable ($(y^*_{(-1)})$) increases the probabilities of $\Pr(y = 4)$ and $\Pr(y = 5)$ by 39.13% and 17.13% points. *married* increases the probabilities of $\Pr(y = 4)$ and $\Pr(y = 5)$ by 24.63% and 4.90% points. *work* increases the probabilities of $\Pr(y = 4)$ and $\Pr(y = 5)$ by 12.22% and 3.22% points. *age* increases the probabilities of $\Pr(y = 4)$ and $\Pr(y = 5)$ by 2.30% and 0.55% points. *education* increases the probabilities of $\Pr(y = 4)$ by 4.28% points. Therefore, the lagged latent variable (($y^*_{(-1)}$)) has the largest positive partial effects on life satisfaction.

In summary, the lagged latent variable ($y^*_{(-1)}$) increases life satisfaction. Married respondents with paid employment have higher life satisfaction levels for all waves. With aging, respondents become more satisfied with their lives. Furthermore, *age* evidently impacts the threshold, but *eduy* and *gender* do not.

4.4 Summary

In this chapter, we propose a Bayesian algorithm that accommodates individual heterogeneity in a dynamic panel-ordered probit model. We conduct an empirical study using panel data from CFPS, administered by the Institute of Social Science Survey at Peking University. The posterior results for the explanatory variables indicate that the lagged latent variable increases life satisfaction. With aging, married respondents with paid employment attain higher life satisfaction levels. Additionally, *age* evidently impacts the threshold, but *eduy* and *gender* do not.

TABLE 4.3: The posterior results for partial effects on response probabilities (CFPS data)

		Mean	SD
$\phi(y^*_{(-1)})$	$y=1$	-0.0435^{**}	0.0075
	$y=2$	-0.0823^{**}	0.0097
	$y=3$	-0.4487^{**}	0.0220
	$y=4$	0.3913^{**}	0.0231
	$y=5$	0.1713^{**}	0.0187
Married	$y=1$	-0.0482^{**}	0.0133
	$y=2$	-0.0691^{**}	0.0155
	$y=3$	-0.1952^{**}	0.0299
	$y=4$	0.2463^{**}	0.0362
	$y=5$	0.0490^{**}	0.0065
Income	$y=1$	-0.0002	0.0003
	$y=2$	-0.0006	0.0010
	$y=3$	-0.0037	0.0078
	$y=4$	0.0034	0.0074
	$y=5$	0.0010	0.0015
Work	$y=1$	-0.0120^{**}	0.0048
	$y=2$	-0.0255^{**}	0.0088
	$y=3$	-0.1175^{**}	0.0296
	$y=4$	0.1222^{**}	0.0344
	$y=5$	0.0322^{**}	0.0083
Age	$y=1$	-0.0012^{**}	0.0002
	$y=2$	-0.0031^{**}	0.0007
	$y=3$	-0.0241^{**}	0.0043
	$y=4$	0.0230^{**}	0.0044
	$y=5$	0.0055^{**}	0.0008
Education	$y=1$	-0.0003	0.0004
	$y=2$	-0.0033^{**}	0.0016
	$y=3$	-0.0391^{**}	0.0133
	$y=4$	0.0428^{**}	0.0140
	$y=5$	0.0023	0.0017
Gender	$y=1$	0.0009	0.0021
	$y=2$	0.0065	0.0058
	$y=3$	-0.0268	0.0356
	$y=4$	0.0250	0.0351
	$y=5$	-0.0039	0.0088

Notes: "$**$" and "$*$" denote that zero is not included in the 95% and 90% credible intervals. Mean and SD denote the posterior mean and posterior standard deviation.

Chapter 5

Conclusion

5.1 Summary

In an analysis of subjective well-being, the backgrounds and preferences of respondents are complicated, and individuals may interpret concepts in different ways (Brady, 1985). Many studies have considered ordered choice models, incorporating heterogeneity among individuals (Farewell (1982); Terza (1985); Greene (2007); Eluru et al. (2008); Greene and Hensher (2010b); King et al. (2004); Kapteyn et al. (2007)). However, to the best of our knowledge, no study has so far accommodated individual heterogeneity in the cutpoints of ordered probit models.

Therefore, in Chapter 2, we present a sampling algorithm for the univariate ordered probit model with individual heterogeneity. We apply the new proposed algorithms to analyze happiness levels in the United States, Canada, and Australia using data from the World Values Survey (WVS), Wave 5, with immigration and religion as individual heterogeneity in the threshold model. We calculate the marginal likelihood and Bayes factor. The results indicate that the Bayesian univariate ordered probit models that include individual heterogeneity perform better than those without individual heterogeneity do. The posterior results for the explanatory variables are the same as those of other empirical studies; that is, being married, earning a high income, and having a high education level have positive effects on happiness, while being male has a negative effect on happiness in Canada. Furthermore, we find that the effects of heterogeneity vary between these three countries. Belonging to a religion has an evident impact on the threshold in the United States, but has no evident effects in Canada or Australia. In addition, parents' immigration status can affect the threshold in Australia, but shows no evident effects in the United States or Canada.

In some instances, we need to model several related ordered responses from the same individuals, such as responses related to happiness and self-assessed health. Since these ordered responses are correlated, we need to use a multivariate model, rather than several separate univariate models. In Chapter 3, we propose a Bayesian multivariate ordered probit model with individual heterogeneity to meet the requirements. Subsequently, we apply the algorithm from Chen and Dey (2000) and the new proposed Bayesian multivariate ordered probit model with individual heterogeneity to the WVS data that is same with our first empirical study. We choose three immigration countries, namely Australia, Canada, the United States, and study individuals' life satisfaction, state of health, interest in politics, and freedom of choice as the correlated responses. We accommodate parent immigration status and religious status as individual heterogeneity factors in the thresholds in the new algorithm. The results for the Bayes factor indicate that the models with individual heterogeneity outperforms that without heterogeneity. Moreover, the effects of heterogeneity are different among the three countries. Belonging to a religion has an evident impact on the state of health and freedom of choice in Australia, while it has significant effects on the state of health and life satisfaction in the United States. However, our findings also suggest that parent immigration status cannot affect the thresholds in all three countries.

In addition, to deal with panel data, we conduct a new Bayesian algorithm in a dynamic ordered probit model in Chapter 4. we apply the algorithm with individual heterogeneity in Bayesian dynamic panel ordered probit model to subjective well-being in China. We employ panel data taken from the China Family Panel Studies (CFPS), which launched the baseline survey in 2010 and three waves of their follow-up surveys in 2012, 2014, and 2016. We select marriage, income, and work as the time-variant explanatory variables, with age, gender, and education year as the time-invariant variables. Furthermore, we incorporate age, gender, and education year as individual heterogeneity factors in the cutpoints function. From the posterior results, the lagged latent variable increases life satisfaction; marriage, work, and age have positive effects on life satisfaction for all periods; and education year increases life satisfaction initially. Also, age has evident impacts on the thresholds, but education year and gender have no impact.

5.2 Remaining Issues and Future Works

After this dissertation, there are several extensions to the models we would like to tackle.

In Chapter 2, the identifiability of the new proposed algorithm is established by fixing one cutpoint and assuming the error term to be normally distributed with zero mean and unit variance. However, we can also identify the model by fixing two cutpoints. Therefore, in the future work, we could consider an alternative algorithm with two cutpoints fixed, and compare the performance of the alternative algorithm with the proposed algorithm in Chapter 2.

In Chapter 3, we propose a Bayesian multivariate ordered probit model with individual heterogeneity. In the future work, we consider to extend this multivariate model to a dynamic version.

In Chapter 4, the cutpoints in the new proposed algorithm are considered to be time-invariant. However, in some cases, we need to take time into consideration since time-variant variables may also have impacts on thresholds. Thus, we consider to propose an algorithm that could include time-variant variables in the cutpoints functions.

In this study, all the models are applied to the analyses of life satisfaction and happiness. However, ordered probit models can deal with many other kinds of ordinal discrete responses. For example, using an ordered probit model, Hasegawa (2010) analyses the tourists' satisfaction. Hauret and Williams (2017) estimate the self-reported job satisfaction. In the future work, we will conduct more application studies to demonstrate the practical use of the new algorithms.

Bibliography

Abdel-Khalek, A. M. and D. Lester (2009). A significant association between religiosity and happiness in a sample of kuwaiti students. *Psychological Reports 105*, 381–382.

Albert, J. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*, 669–679.

Albert, J. and S. Chib (2001). Sequential ordinal modeling with applications to survival data. *Biometrics 57*, 829–836.

Argyle, M. (2001). *The Psychology of Happiness*. London: Routledge.

Bartram, D. (2011). Economic migration and happiness: Comparing immigrants' and natives' happiness gains from income. *Social Indicators Research 103*, 57–76.

Blanchflower, D. and A. Oswald (2004). Well-being over time in Britain and the USA. *Journal of Public Economics 88*, 1359–1386.

Brady, H. E. (1985). The perils of survey research: Inter-personally incomparable responses. *Political Methodology 11*, 269–290.

Bălţătescu, S. (2007). Central and eastern europeans migration' subjective quality of life. a comparative study. *Journal of Identity and Migration Studies 1*, 67–81.

Chen, M.-H. and D. Dey (2000). Bayesian analysis for correlated ordinal data models. In D. K. Dey, S. K. Ghosh, and B. K. Mallick (Eds.), *Generalized Linear Models: A Bayesian Perspective*, pp. 133–157. New York: Marcel-Dekker.

Chib, S. (1995a). Marginal likelihood from the gibbs outout. *Journal of the American Statistical Association 90*, 1313–1321.

Chib, S. (1995b). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association 90*, 1313–1321.

Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association 96*, 270–281.

Clark, A. E. and O. Lelkes. Let us pray: Religious interactions in life satisfaction. *PSE Working papers*.

Cohen, A. B. and K. A. Johnson (2017). The relation between religion and well-being. *Applied Research in Quality of Life 12*, 533–547.

Cowles, M. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing 6*, 101–111.

Diener, E., P. Kesebir, and W. Tov (2009). *Handbook of individual differences in social behavior*. New York: Guilford.

Diener, E., E. M. Suh, R. E. Lucas, and H. L. Smith (1999). Subjective well-being: Three decades of progress. *Psychology Bulletin 125*, 276–302.

Ellison, C. G. (1991). Religious involvement and subjective well-being. *Journal of Health and Social Behavior 32*, 80–99.

Ellison, C. G., J. D. Boardman, D. R. Williams, and J. S. Jackson (2001). Religious involvement, stress, and mental health: Findings from the 1995 detroit area study. *Social Forces 80*, 215–249.

Eluru, N., C. Bhat, and D. Hensher (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity levels in traffic crashes. *Accident Analysis and Prevention 40*, 1033–1054.

Farewell, V. (1982). A note on regression analysis of ordinal data with variability of classification. *Biometrika*, 533–538.

Finch, B., B. Kolody, and W. Vega (2000). Perceived discrimination and depression among Mexican-origin adults in California. *Journal of Health and Social Behavior 41*, 295–313.

Fisher, J. (2013). Relating with god contributes to variance in happiness, over that from personality and age. *Religions 4*, 313–324.

Fredrickson, B. L. (2002). How does religion benefit health and well-being? are positive emotions active ingredients? *Psychological Inquiry 13*, 209–213.

Fredrickson, B. L., M. A. Cohn, K. A. Coffey, J. Pek, and S. M. Finkel (2008). Open hearts build lives: Positive emotions, induced through loving-kindness meditation, build consequential personal resources. *Journal of Personality and Social Psychology 95*, 1045–1062.

Frey, B. and A. Stutzer (2002). *Happiness and Economics*. Princeton: Princeton University Press.

Gatina, L. (2016). Does money buy happiness? financial and general well-being of immigrants in Australia. *Journal of Behavioral and Experimental Economics 63*, 91–105.

Gokdemir, O. and D. Dumludag (2012). Life satisfaction among turkish and moroccan immigrants in the netherlands: the role of absolute and relative income. *Social Indicators Research 106*, 407–417.

Greene, W. (2007). Limdep version 9.0: Reference guide.

Greene, W. and D. Hensher (2010a). Ordered choices and heterogeneity in attribute processing. *Journal of Transport Economics and Policy 44*, 331–364.

Greene, W. H. and D. A. Hensher (2010b). *Modeling Ordered Choices: A Primer*. Cambridge: Cambridge University Press.

Hasegawa, H. (2009). Bayesian dynamic panel-ordered probit model and its application to subjective well-being. *Communications in Statistics-Simulation and Computation 38*(6), 1321–1347.

Hasegawa, H. (2010). Analyzing tourists' satisfaction: A multivariate ordered probit approach. *Tourism Management 31*(1), 86–97.

Hasegawa, H. and K. Ueda (2016). Multidimensional inequality for current status of japanese private companies' employees. *Metron 74*(3), 357–373.

Hauret, L. and D. R. Williams (2017). Cross-national analysis of gender differences in job satisfaction. *Industrial Relations: A Journal of Economy and Society 56*(2), 203–235.

Hendriks, M. (2015). The happiness of interntional migrants: A review of research findings. *Migration Studies 3*, 343–369.

Hwang, S.-S., Y. Cao, and J. Xi (2011). The short-term impact of involuntary migration in China's three gorges: a prospective study. *Social Indicators Research 101*, 73–92.

Inglehart, R. (2000). Faith and freedom: Traditional and modern ways to happiness. In E. Diener, F. H. John, and K. Daniel (Eds.), *International Differences in Well-being*, pp. 351–397. New York: Oxford University Press.

Inkeles, A. (1993). Industrialization, modernization and the quality of life. *International Journal of Comparative Sociology 34*, 1–23.

Jeffreys, H. (1961). *Theory of Probability.* London: Oxford University Press.

Jeliazkov, I., J. Graves, and M. Kutzbach (2008). Fitting and comparison of models for multivariate ordinal outcomes. In S. Chib, G. Koop, W. Griffiths, and D. Terrell (Eds.), *Advances in Econometrics: Bayesian Econometrics*, Volume 23, pp. 115–156. Bingley: Emerald.

Joshanloo, M. and D. Weijers (2016). Religiosity moderates the relationship between income inequality and life satisfaction across the Globe. *Social Indicators Research 128*, 731–750.

Kapteyn, A., J. Smith, and A. van Soest (2007). Vignettes and self-reports of work disability in the United States and the Netherlands. *American Economic Review 97*, 461–473.

King, G., C. Murray, J. Salomon, and A. Tandon (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review 98*, 191–207.

Koenig, H., M. McCullough, and D. Larson (2001). *Handbook of Religion and Health*. New York: Oxfrod University Press.

Kortt, M., B. Dollery, and B. Grant (2015). Religion and life satisfaction down under. *Journal of Happiness Studies 16*, 277–293.

Kóczán, Z. (2016). (Why) are immigrants unhappy. *IZA Journal of Migration 5*, 1–25.

Layard, R., G. Mayraz, and S. Nickell (2008). The marginal utility of income. *Journal of Public Economics 92*, 1846–1857.

Liu, J. S. and C. Sabatti (2000). Generalised gibbs sampler and multigrid monte carlo for bayesian computation. *Biometrika 87*(2), 353–369.

Müller, G. and C. Czado (2005). An autoregressive ordered probit model with application to high-frequency financial data. *Journal of Computational and Graphical Statistics 14*(2), 320–338.

Nandram, B. and M.-H. Chen (1996). Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistics Computation and Simulation 54*, 129–144.

Nattavudh, P., N. L. Warn, and W. Mark (2015). What's the good of education on our overall quality of life? a simultaneous equation model of education and life satisfaction for Australia. *Journal of Behavioral and Experimental Economics 54*, 10–21.

Obućina, O. (2013). The patterns of satisfaction among immigrants in germany. *Social Indicators Research 113*, 1105–1127.

Park, C. L. and J. M. Slattery (2012). Spirituality, emotions, and physical health. In L. J. Miller (Ed.), *The Oxford handbook of psychology and spirituality*, pp. 379–387. New York: Oxford University Press.

Robert, C. (2001). *The Bayesian Choice.* New York: Springer.

Safi, M. (2010). Immigrants' life satisfaction in Europe: Between assimilation and discrimination. *European Sociological Review 26*, 159–171.

Sellers, R., C. Caldwell, K. Schmeelk-Cone, and M. Zimmerman (2003). Racial identity, racial discrimination, perceived stress, and psychological distress among African American young adults. *Journal of Health and Social Behavior 44*, 302–317.

Shi, L. and H. Hasegawa (2019). Bayesian analysis of happiness with individual heterogeneity. *Behaviormetrika 46*, 23–48.

Stillman, S., J. Gibson, D. McKenzie, and H. Rohorua (2012). Miserable migrants? natural experiment evidence on international migration and objective and subjective well-being. *IZA Discussion Paper 6871.*

Taylor, J. and R. Turner (2002). Perceived discrimination, social stress, and depression in the transition to adulthood: racial contrasts. *Social Psychology Quarterly 65*, 213–225.

Terza, J. (1985). Ordered probit: A generalization. *Communications in Statistics - Theory and Methods 14*, 1–11.

Tony, V. and E. Matthew (2014). The social dimensions of happiness and life satisfaction of Australians: Evidence from the World Values Survey. *International Journal of Social Welfare 23*, 240–253.

Van Cappellen, P., M. Toth-Gauthier, V. Saroglou, and B. Fredrickson (2016). Religion and well-being: The mediating role of positive emotions. *Journal of Happiness Studies 17*, 485–505.

Vega, W. and R. Rumbaut (1991). Ethnic minorities and mental health. *Annual Review of Sociology 17*, 351–383.

Werkuyten, M. and S. Nekuee (1999). Subjective well-being, discrimination and cultural conflict: Iranians living in the netherlands. *Social Indicators Research 47*, 281–306.

Witter, R. A., W. A. Stock, M. A. Okun, and M. J. Haring (1985). Religion and subjective well-being in adulthood: A quantitative synthesis. *Review of Religious Research 26*, 332–342.

Appendix A

Derivation in Chapter 2

A.1 Derivation of Algorithm 2

A.1.1 Full Conditional Distribution of $\boldsymbol{\beta}$

From equation (2.11), the full conditional distribution (FCD) of $\boldsymbol{\beta}$ depends only on the distribution of $\boldsymbol{y}^*$ and on the prior distribution $p(\boldsymbol{\beta})$. Because $y_i^*|\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$ and $\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)$, the FCD of $\boldsymbol{\beta}$ is

$$\begin{aligned} p(\boldsymbol{\beta}|\cdots) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{y}^* - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}^* - \boldsymbol{X}\boldsymbol{\beta})\right] \times \exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\boldsymbol{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right] \\ &\propto \exp\left[-\frac{1}{2}\boldsymbol{\beta}'(\boldsymbol{B}_0^{-1} + \boldsymbol{X}'\boldsymbol{X})\boldsymbol{\beta} + \boldsymbol{\beta}'(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \boldsymbol{X}'\boldsymbol{y}^*)\right]. \end{aligned}$$

Thus, the FCD of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}|\cdots \sim \mathrm{N}(\boldsymbol{\beta}^*, \boldsymbol{B}^*),$$

where

$$\boldsymbol{B}^* = (\boldsymbol{B}_0^{-1} + \boldsymbol{X}'\boldsymbol{X})^{-1},\ \boldsymbol{\beta}^* = \boldsymbol{B}^*(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \boldsymbol{X}'\boldsymbol{y}^*).$$

A.1.2 Full Conditional Distribution of $\boldsymbol{y}^*$

Because the model of $\boldsymbol{y}^*$ is $y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i$ and we have assumed $\epsilon_i \sim \mathrm{N}(0,1)$, the conditional distribution of $\boldsymbol{y}^*$ is given by

$$p(\boldsymbol{y}^*|\cdots) \propto \prod_{i=1}^{n} 1(\mu_{i,j-1} < y_i^* \leq \mu_{ij})\phi(\boldsymbol{x}_i'\boldsymbol{\beta}, 1),$$

where $\phi(\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$ is the probability density function of the normal distribution $\mathrm{N}(\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$. Thus, the FCD of y_i^* is the truncated normal distribution, which means y_i^* must belong to the interval $(\mu_{i,j-1}, \mu_{ij}]$, for given $y_i = j$.

A.1.3 Full Conditional Distribution of $\boldsymbol{\gamma}$

For $i = 1, \ldots, n$, the cutoff points are specified as follows:

$$\mu_{i0} = -\infty,\ \mu_{i1} = 0,\ \mu_{i2} = \exp(\boldsymbol{\gamma}_2'\boldsymbol{z}_i)$$

$$\mu_{i3} = \mu_{i2} + \exp(\boldsymbol{\gamma}_3'\boldsymbol{z}_i) = \exp(\boldsymbol{\gamma}_2'\boldsymbol{z}_i) + \exp(\boldsymbol{\gamma}_3'\boldsymbol{z}_i)$$

$$\mu_{ij} = \mu_{i,j-1} + \exp(\boldsymbol{\gamma}_j'\boldsymbol{z}_i) = \exp(\boldsymbol{\gamma}_2'\boldsymbol{z}_i) + \exp(\boldsymbol{\gamma}_3'\boldsymbol{z}_i) + \cdots + \exp(\boldsymbol{\gamma}_j'\boldsymbol{z}_i),\ \mu_{iJ} = \infty.$$

Because $\boldsymbol{\gamma}_j \sim \mathrm{N}(\boldsymbol{\gamma}_0, \boldsymbol{G}_0)$, we have

$$p(\boldsymbol{\gamma}_j) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_0)'\boldsymbol{G}_0^{-1}(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_0)\right].$$

Then, we have

$$
\begin{aligned}
p(\boldsymbol{\gamma}_j|\cdots) &\propto p(\boldsymbol{\gamma}_j)p(\boldsymbol{y}|\boldsymbol{\beta},\gamma) \\
&\propto \exp\left[-\frac{1}{2}(\boldsymbol{\gamma}_j-\boldsymbol{\gamma}_0)'\boldsymbol{G}_0^{-1}(\boldsymbol{\gamma}_j-\boldsymbol{\gamma}_0)\right] \\
&\times \prod_{i=1}^{n}\Bigg\{[\Phi(\mu_{ij}-\boldsymbol{x}_i'\boldsymbol{\beta})-\Phi(\mu_{i,j-1}-\boldsymbol{x}_i'\boldsymbol{\beta})]^{1(y_i=j)} \\
\times\cdots\times [\Phi(\mu_{i,J-1}-\boldsymbol{x}_i'\boldsymbol{\beta})-\Phi(\mu_{i,J-2}-\boldsymbol{x}_i'\boldsymbol{\beta})]^{1(y_i=J-1)}&[1-\Phi(\mu_{i,J-1}-\boldsymbol{x}_i'\boldsymbol{\beta})]^{1(y_i=J)}\Bigg\}.
\end{aligned}
$$

In general, we have

$$
\begin{aligned}
\log p(\boldsymbol{\gamma}_j|\cdots) &= \text{const.} - \frac{1}{2}(\boldsymbol{\gamma}_j-\boldsymbol{\gamma}_0)'\boldsymbol{G}_0^{-1}(\boldsymbol{\gamma}_j-\boldsymbol{\gamma}_0) \\
&+ \sum_{i:y_i=j}\log[\Phi(\mu_{ij}-\boldsymbol{x}_i'\boldsymbol{\beta})-\Phi(\mu_{i,j-1}-\boldsymbol{x}_i'\boldsymbol{\beta})] \\
&+\cdots+\sum_{i:y_i=J-1}\log[\Phi(\mu_{i,J-1}-\boldsymbol{x}_i'\boldsymbol{\beta})-\Phi(\mu_{i,J-2}-\boldsymbol{x}_i'\boldsymbol{\beta})] \\
&+\sum_{i:y_i=J}\log[1-\Phi(\mu_{i,J-1}-\boldsymbol{x}_i'\boldsymbol{\beta})].
\end{aligned}
$$

Using the Metropolis-Hastings (M-H) algorithm, we can sample $\tilde{\boldsymbol{\gamma}}_j$ by drawing $\tilde{\boldsymbol{\gamma}}_j \sim q(\boldsymbol{\gamma}_j|\cdots)$ from a multivariate t proposal density $q(\boldsymbol{\gamma}_j|\cdots) = f_T(\boldsymbol{\gamma}_j|\hat{\boldsymbol{\gamma}}_j, \hat{\boldsymbol{G}}_{J-1}, \boldsymbol{\nu})$, where ν is the degrees of freedom, and

$$
\hat{\boldsymbol{\gamma}}_j = \arg\max_{\boldsymbol{\gamma}_j}\log p(\boldsymbol{\gamma}_j|\cdots),\ \hat{\boldsymbol{G}}_j = \left\{\left[-\frac{\partial^2\log p(\boldsymbol{\gamma}_j|\cdots)}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j'}\right]_{\boldsymbol{\gamma}_j=\hat{\boldsymbol{\gamma}}_j}\right\}^{-1}.
$$

A.2 Marginal Likelihood

According to Bayes' theorem, the marginal likelihood can be written as

$$m(\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\boldsymbol{y})},$$

where $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function, $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\theta}|\boldsymbol{y})$ are the prior and posterior densities of $\boldsymbol{\theta}$, respectively, and $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3\}$.

Following Chib (1995b), the proposed estimate of the marginal likelihood on the logarithm scale is

$$\log \hat{m}(\boldsymbol{y}) = \log p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}) + \log p(\hat{\boldsymbol{\theta}}) - \log \hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y}), \tag{A.1}$$

where $\hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y})$ is the posterior density estimate at $\hat{\boldsymbol{\theta}}$. Suppose that $\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2, \hat{\boldsymbol{\gamma}}_3\}$ is the selected point and the Gibbs sampler is applied to the complete conditional densities

$$p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{y}^*),\ \ p(\boldsymbol{y}^*|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3).$$

Then, the posterior density $p(\hat{\boldsymbol{\theta}}|\boldsymbol{y})$ at $\hat{\boldsymbol{\theta}}$ can be expressed as

$$p(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = p(\hat{\boldsymbol{\beta}}|\boldsymbol{y}) \times p(\hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}) \times p(\hat{\boldsymbol{\gamma}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2),$$

where

$$p(\hat{\boldsymbol{\beta}}|\boldsymbol{y}) = \int p(\hat{\boldsymbol{\beta}}|\boldsymbol{y}, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{y}^*) p(\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{y}^*|\boldsymbol{y}) d\boldsymbol{\gamma}_2 d\boldsymbol{\gamma}_3 d\boldsymbol{y}^*.$$

Let the main output be given by $\{\boldsymbol{\beta}^{(s)}, \boldsymbol{\gamma}_2^{(s)}, \boldsymbol{\gamma}_3^{(s)}, \boldsymbol{y}^{*(s)}\}_{s=1}^{S}$. Then, the appropriate Monte Carlo estimate of $p(\hat{\boldsymbol{\beta}}|\boldsymbol{y})$ is

$$\hat{p}(\hat{\boldsymbol{\beta}}|\boldsymbol{y}) = S^{-1}\sum_{s=1}^{S} p(\hat{\boldsymbol{\beta}}|\boldsymbol{y}, \boldsymbol{\gamma}_2^{(s)}, \boldsymbol{\gamma}_3^{(s)}, \boldsymbol{y}^{*(s)}).$$

However, we cannot use the same method to calculate $\hat{p}(\hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}})$ and $\hat{p}(\hat{\boldsymbol{\gamma}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2)$, because $\boldsymbol{\gamma}_2$ and $\boldsymbol{\gamma}_3$ are sampled using the M-H algorithm.

Chib and Jeliazkov (2001) provide an approach to estimate the marginal likelihood from the M-H output. Let $q(\boldsymbol{\gamma}_2, \tilde{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)$ denote the proposal density of the M-H chain for the transition from the current $\boldsymbol{\gamma}_2$ to a proposed value $\tilde{\boldsymbol{\gamma}}_2$, and let

$$\alpha(\boldsymbol{\gamma}_2, \tilde{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*) = \frac{p(\boldsymbol{y}|\boldsymbol{\beta}, \tilde{\boldsymbol{\gamma}}_2, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)p(\boldsymbol{\beta}, \tilde{\boldsymbol{\gamma}}_2, \boldsymbol{\gamma}_3)}{p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)p(\boldsymbol{\beta}, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)}\frac{q(\tilde{\boldsymbol{\gamma}}_2, \boldsymbol{\gamma}_2|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}_3, y^*)}{q(\boldsymbol{\gamma}_2, \tilde{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}_3, y^*)}$$

denote the acceptance ratio of moving from $\boldsymbol{\gamma}_2$ to $\tilde{\boldsymbol{\gamma}}_2$. In addition, let

$$p(\boldsymbol{\gamma}_2, \hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*) = \alpha(\boldsymbol{\gamma}_2, \hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)q(\boldsymbol{\gamma}_2, \hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*) \tag{A.2}$$

denote the subkernel of the M-H algorithm. Then, according to the reversibility of this subkernel, we obtain

$$p(\boldsymbol{\gamma}_2, \hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)p(\boldsymbol{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*) = p(\hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)p(\hat{\boldsymbol{\gamma}}_2, \boldsymbol{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*). \tag{A.3}$$

After multiplying both sides of (A.3) by $p(\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*|\boldsymbol{y})$ and integrating out $\hat{\boldsymbol{\beta}}$, $\boldsymbol{\gamma}_2$, $\boldsymbol{\gamma}_3$, and $\boldsymbol{y}^*$, we have

$$\begin{aligned}&\int p(\boldsymbol{\gamma}_2, \hat{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)p(\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{y}^*|\boldsymbol{y})d\hat{\boldsymbol{\beta}}d\boldsymbol{\gamma}_2 d\boldsymbol{\gamma}_3 d\boldsymbol{y}^* \\ = p(\hat{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}) &\int p(\hat{\gamma}_2, \boldsymbol{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)p(\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\gamma}_2)d\hat{\boldsymbol{\beta}}d\boldsymbol{\gamma}_2 d\boldsymbol{\gamma}_3 d\boldsymbol{y}^*. \quad \text{(A.4)}\end{aligned}$$

From (A.4) and the definition of $p(\boldsymbol{\gamma}_2, \hat{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)$ in (A.2), we obtain

$$p(\hat{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}) = \frac{E_1\{\alpha(\boldsymbol{\gamma}_2, \hat{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)q(\boldsymbol{\gamma}_2, \hat{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)\}}{E_2\{\alpha(\hat{\gamma}_2, \boldsymbol{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)\}},$$

where E_1 is the expectation with respect to the conditional posterior $p(\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{y}^*|\boldsymbol{y})$, and E_2 is the expectation with respect to $p(\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\gamma}_2)\times q(\hat{\gamma}_2, \boldsymbol{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)$.

We estimate $E_1\{\alpha(\boldsymbol{\gamma}_2, \hat{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)q(\boldsymbol{\gamma}_2, \hat{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)\}$ using the posterior draws of $\boldsymbol{\gamma}_2^{(s)}$, $\boldsymbol{\gamma}_3^{(s)}$, and $\boldsymbol{y}^{*(s)}$ from the main MCMC output, and then average the quantity. However, the draws of $\boldsymbol{\gamma}_2$, $\boldsymbol{\gamma}_3$, and $\boldsymbol{y}^*$ from the main MCMC output are from the distributions $p(\boldsymbol{\gamma}_2|\boldsymbol{y}, \boldsymbol{\gamma}_3, \boldsymbol{\beta}, \boldsymbol{y}^*)$, $p(\boldsymbol{\gamma}_3|\boldsymbol{y}, \boldsymbol{\gamma}_2, \boldsymbol{\beta}, \boldsymbol{y}^*)$, and $p(\boldsymbol{y}^*|\boldsymbol{y}, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{\beta})$, respectively, and are not conditional on $\hat{\gamma}_2$. Thus, we need to sample an additional G iterations for the estimate of $E_2\{\alpha(\hat{\gamma}_2, \boldsymbol{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3, \boldsymbol{y}^*)\}$. We sample $\boldsymbol{\gamma}_2^{(g)}$, $\boldsymbol{\gamma}_3^{(g)}$, and $\boldsymbol{y}^{*(g)}$ with the full conditional densities

$$p(\boldsymbol{\gamma}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\gamma}_2, \boldsymbol{y}^*),\ p(\boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\gamma}_2, \boldsymbol{\gamma}_3).$$

Furthermore, we generate $\boldsymbol{\gamma}_2^{(g)} \sim q(\hat{\boldsymbol{\gamma}}_2, \boldsymbol{\gamma}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3^{(g)}, \boldsymbol{y}^{*(g)})$ at each sampling step. Then, the estimate of $p(\hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}})$ can be written as

$$\hat{p}(\hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}) = \frac{S^{-1}\sum_{s=1}^{S} \alpha(\boldsymbol{\gamma}_2^{(s)}, \hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3^{(s)}, \boldsymbol{y}^{*(s)}) q(\boldsymbol{\gamma}_2^{(s)}, \hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3^{(s)}, \boldsymbol{y}^{*(s)})}{G^{-1}\sum_{g=1}^{G} \alpha(\hat{\boldsymbol{\gamma}}_2, \boldsymbol{\gamma}_2^{(g)}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3^{(g)}, \boldsymbol{y}^{*(g)})}.$$

Using a the similar calculation, we have

$$\hat{p}(\hat{\boldsymbol{\gamma}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2) = \frac{S^{-1}\sum_{s=1}^{S} \alpha(\boldsymbol{\gamma}_3^{(s)}, \hat{\boldsymbol{\gamma}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2, \boldsymbol{y}^{*(s)}) q(\boldsymbol{\gamma}_3^{(s)}, \hat{\boldsymbol{\gamma}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2, \boldsymbol{y}^{*(s)})}{G^{-1}\sum_{g=1}^{G} \alpha(\hat{\boldsymbol{\gamma}}_3, \boldsymbol{\gamma}_3^{(g)}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2, \boldsymbol{y}^{*(g)})}.$$

Substituting the estimates of $p(\hat{\boldsymbol{\theta}}|\boldsymbol{y})$ into equation (A.1), we obtain the following estimate of the marginal likelihood:

$$\begin{aligned}
\log \hat{m}(\boldsymbol{y}) &= \log p(\boldsymbol{y}|\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2, \hat{\boldsymbol{\gamma}}_3) + \log p(\hat{\boldsymbol{\beta}}) + \log p(\hat{\boldsymbol{\gamma}}_2) + \log p(\hat{\boldsymbol{\gamma}}_3) \\
&\quad - \log \left[S^{-1} \sum_{s=1}^{S} p(\hat{\boldsymbol{\beta}}|\boldsymbol{y}, \boldsymbol{\gamma}_2^{(s)}, \boldsymbol{\gamma}_3^{(s)}, \boldsymbol{y}^{*(s)}) \right] \\
&\quad - \log \left[\frac{S^{-1}\sum_{s=1}^{S} \alpha(\boldsymbol{\gamma}_2^{(s)}, \hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3^{(s)}, \boldsymbol{y}^{*(s)}) q(\boldsymbol{\gamma}_2^{(s)}, \hat{\boldsymbol{\gamma}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3^{(s)}, \boldsymbol{y}^{*(s)})}{G^{-1}\sum_{g=1}^{G} \alpha(\hat{\boldsymbol{\gamma}}_2, \boldsymbol{\gamma}_2^{(g)}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}_3^{(g)}, \boldsymbol{y}^{*(g)})} \right] \\
&\quad - \log \left[\frac{S^{-1}\sum_{s=1}^{S} \alpha(\boldsymbol{\gamma}_3^{(s)}, \hat{\boldsymbol{\gamma}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2, \boldsymbol{y}^{*(s)}) q(\boldsymbol{\gamma}_3^{(s)}, \hat{\boldsymbol{\gamma}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2, \boldsymbol{y}^{*(s)})}{G^{-1}\sum_{g=1}^{G} \alpha(\hat{\boldsymbol{\gamma}}_3, \boldsymbol{\gamma}_3^{(g)}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2, \boldsymbol{y}^{*(g)})} \right],
\end{aligned}$$

where

$$\begin{aligned}
\log p(\boldsymbol{y}|\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}_2, \hat{\boldsymbol{\gamma}}_3) &= \sum_{i:y_i=1} \log[\Phi(\hat{\mu}_{ij} - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}) - \Phi(\hat{\mu}_{i,j-1} - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}})] \\
&\quad + \cdots + \sum_{i:y_i=J-1} \log[\Phi(\hat{\mu}_{i,J-1} - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}) - \Phi(\hat{\mu}_{i,J-2} - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}})] \\
&\quad + \sum_{i:y_i=J} \log[\Phi(\hat{\mu}_{i,J} - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}) - \Phi(\hat{\mu}_{i,J-1} - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}})].
\end{aligned}$$

A.3 Marginal Effects

From Section 2.1.3, the probability of observing $y_i = j$ can be written as

$$\Pr(y_i = j|\boldsymbol{\beta}, \boldsymbol{\mu}) = \Phi(\mu_{ij} - \boldsymbol{x}_i'\boldsymbol{\beta}) - \Phi(\mu_{i,j-1} - \boldsymbol{x}_i'\boldsymbol{\beta}).$$

Furthermore, for a continuous variable x_l, we can obtain the marginal effect from the partial derivative:

$$\frac{\partial \Pr(y_i = j|\boldsymbol{\beta}, \boldsymbol{\mu})}{\partial x_l} = [\phi(\mu_{ij} - \boldsymbol{x}_i'\boldsymbol{\beta}) - \phi(\mu_{i,j-1} - \boldsymbol{x}_i'\boldsymbol{\beta})]\beta_l,$$

If x_l is a discrete variable, we cannot differentiate. In this case, the marginal effect of x_l going from c_l to $c_l + 1$ should be calculated as follows:

$$\begin{aligned} &\Phi[\mu_{ij} - (\beta_0 + x_1\beta_1 + \cdots + (c_l + 1)\beta_l + \cdots + x_k\beta_k)] \\ &\quad - \Phi[\mu_{ij} - (\beta_0 + x_1\beta_1 + \cdots + c_l\beta_l + \cdots + x_k\beta_k)]. \end{aligned}$$

Appendix B

Derivation in Chapter 3

B.1 Derivation of Algorithm 1

B.1.1 Full Conditional Distribution of $\boldsymbol{\beta}$

First, we replace the jth element as the first element, that is,

$$\boldsymbol{y}_i^* = \begin{pmatrix} y_{ij}^* \\ \boldsymbol{y}_{i(-j)}^* \end{pmatrix}, \ \boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_{ij}' & \mathbf{0}' \\ \mathbf{O} & \boldsymbol{X}_{i(-j)} \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_j \\ \boldsymbol{\beta}_{(-j)} \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{jj} & \boldsymbol{\sigma}_{(-j)}' \\ \boldsymbol{\sigma}_{(-j)} & \boldsymbol{\Sigma}_{(-j)} \end{pmatrix}, \ \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma^{jj} & \boldsymbol{\sigma}^{(-j)\prime} \\ \boldsymbol{\sigma}^{(-j)} & \boldsymbol{\Sigma}^{(-j)} \end{pmatrix}.$$

The FCD of $\boldsymbol{\beta}_j$ is

$$\begin{aligned} p(\boldsymbol{\beta}_j|\dots) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j0})'\boldsymbol{B}_{j0}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j0})\right] \\ &\times \exp\left[\sum_{i=1}^{n}(\boldsymbol{y}_i^* - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i^* - \boldsymbol{X}_i\boldsymbol{\beta})\right]. \end{aligned}$$

After some manipulation, we have the following FCD of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}_j | \cdots \sim \mathrm{N}(\tilde{\boldsymbol{\beta}}_j, \tilde{\boldsymbol{B}}_j),$$

where

$$\begin{aligned}
\tilde{\boldsymbol{B}}_j &= (\boldsymbol{B}_{j0}^{-1} + \sigma^{jj} \sum_{i=1}^{n} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}')^{-1}, \\
\tilde{\boldsymbol{\beta}}_j &= \tilde{\boldsymbol{B}}_j \left[\boldsymbol{B}_{j0}^{-1} \boldsymbol{\beta}_{j0} + \sigma^{jj} \sum_{i=1}^{n} \boldsymbol{x}_{ij} y_{ij}^* + \sum_{i=1}^{n} \boldsymbol{x}_{ij} \boldsymbol{\sigma}^{(-j)\prime} (\boldsymbol{y}_{i(-j)}^* - \boldsymbol{X}_{i(-j)} \boldsymbol{\beta}_{(-j)}) \right].
\end{aligned}$$

If $\boldsymbol{X}_i = \boldsymbol{I}_m \otimes \boldsymbol{x}_i'$, that is, m equations have the same explanatory variables, we have

$$\begin{aligned}
&\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' = \boldsymbol{X}'\boldsymbol{X}, \ \sum_{i=1}^{n} \boldsymbol{x}_i y_{ij}^* = \boldsymbol{X}' \tilde{\boldsymbol{y}}_j^* \\
&\sum_{i=1}^{n} \boldsymbol{x}_{ij} \boldsymbol{\sigma}^{(-j)\prime} \boldsymbol{y}_{i(-j)}^* = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_{i(-j)}^{*\prime} \boldsymbol{\sigma}^{(-j)} = \boldsymbol{X}' \boldsymbol{Y}_{(-j)}^* \boldsymbol{\sigma}^{(-j)} \\
&\sum_{i=1}^{n} \boldsymbol{x}_{ij} \boldsymbol{\sigma}^{(-j)\prime} \boldsymbol{X}_{i(-j)} \boldsymbol{\beta}_{(-j)} = \sum_{i=1}^{n} \boldsymbol{x}_i (\boldsymbol{x}_i' \boldsymbol{\beta}_1, \ldots, \boldsymbol{x}_i' \boldsymbol{\beta}_{j-1}, \boldsymbol{x}_i' \boldsymbol{\beta}_{j+1}, \ldots, \boldsymbol{x}_i' \boldsymbol{\beta}_m) \boldsymbol{\sigma}^{(-j)} \\
&\quad = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \boldsymbol{B}_{(-j)} \boldsymbol{\sigma}^{(-j)} = \boldsymbol{X}' \boldsymbol{X} \boldsymbol{B}_{(-j)} \boldsymbol{\sigma}^{(-j)},
\end{aligned}$$

where

$$\begin{aligned}
&\boldsymbol{Y}^* = \begin{pmatrix} \boldsymbol{y}_1^{*\prime} \\ \vdots \\ \boldsymbol{y}_n^{*\prime} \end{pmatrix} = (\tilde{\boldsymbol{y}}_1^*, \ldots, \tilde{\boldsymbol{y}}_m^*), \ \boldsymbol{Y}_{(-j)}^* = \begin{pmatrix} \boldsymbol{y}_{1(-j)}^{*\prime} \\ \vdots \\ \boldsymbol{y}_{n(-j)}^{*\prime} \end{pmatrix}, \ \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix} \\
&\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m), \ \boldsymbol{B}_{(-j)} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1}, \boldsymbol{\beta}_{j+1}, \ldots, \boldsymbol{\beta}_m).
\end{aligned}$$

B.1.2 Full Conditional Distribution of Σ^{-1}

The FCD of $\boldsymbol{\Sigma}^{-1}$ is

$$\boldsymbol{\Sigma}^{-1}|\cdots \sim W(\tilde{n}, \tilde{\boldsymbol{Q}}^{-1}),$$

where

$$\tilde{n} = n_0 + n,\ \tilde{\boldsymbol{Q}} = \boldsymbol{Q}_0 + \sum_{i=1}^{n}(y_i^* - \boldsymbol{X}_i\boldsymbol{\beta})(y_i^* - \boldsymbol{X}_i\boldsymbol{\beta})'.$$

If $\boldsymbol{X}_i = \boldsymbol{I}_m \otimes \boldsymbol{x}_i'$, that is,

$$\sum_{i=1}^{n}(y_i^* - \boldsymbol{X}_i\boldsymbol{\beta})(y_i^* - \boldsymbol{X}_i\boldsymbol{\beta})' = (\boldsymbol{Y}^* - \boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y}^* - \boldsymbol{X}\boldsymbol{B}).$$

B.1.3 Full Conditional Distribution of $\boldsymbol{y}^*$

Let $\boldsymbol{y}_{(j)}^* = (y_{1j}^*, y_{2j}^*, y_{3j}^*, \ldots, y_{nj}^*)'$ denote the vector of the jth element y_{ij}^* from $\boldsymbol{y}_i^*(i = 1, \ldots, n)$. Further, let $\boldsymbol{y}_{(-j)}^*$ denote the vector obtained by removing $\boldsymbol{y}_{(j)}^*$ from $\boldsymbol{y}^*$, and let $\boldsymbol{y}_{i(-j)}^*$ denote the vector obtained by removing $\boldsymbol{y}_{ij}^*$ from $\boldsymbol{y}_i^*$. We define $\boldsymbol{\gamma}_j = (\boldsymbol{\gamma}_{j2}', \cdots, \boldsymbol{\gamma}_{j(C_j-2)}')'$. Then, we can obtain the joint conditional distribution of $\boldsymbol{\gamma}_j$ and $\boldsymbol{y}_j^*$, that is,

$$p(\boldsymbol{\gamma}_j, \boldsymbol{y}_j^*|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}_{(-j)}^*, \boldsymbol{y}) = p(\boldsymbol{\gamma}_j|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}_{(-j)}^*, \boldsymbol{y})p(\boldsymbol{y}_j^*|\boldsymbol{\gamma}_j, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}_{(-j)}^*, \boldsymbol{y}).$$

Since $\boldsymbol{y}_i^* \sim \mathrm{N}(\boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\Sigma})$, we have

$$y_{ij}^*|\boldsymbol{\gamma}_j, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}_{(-j)}^*, \boldsymbol{y} \sim \mathrm{N}(\tilde{\mu}_{ij}, \tilde{\sigma}_j^2)1\left\{y_{ij}^* \in \mathcal{G}_{ij}\right\},$$

where

$$\tilde{\mu}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_j + \boldsymbol{\sigma}'_{(-j)}\boldsymbol{\Sigma}^{-1}_{(-j)}(\boldsymbol{y}^*_{i(-j)} - \boldsymbol{X}_{i(-j)}\boldsymbol{\beta}_{(-j)})$$
$$\tilde{\sigma}^2_j = \sigma_{jj} - \boldsymbol{\sigma}'_{(-j)}\boldsymbol{\Sigma}^{-1}_{(-j)}\boldsymbol{\sigma}_{(-j)}.$$

B.1.4 Full Conditional Distribution of $\boldsymbol{\gamma}$

Since $\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) \sim \mathrm{N}(\boldsymbol{\delta}_{j0}, \boldsymbol{D}_{j0})$, we have

$$p(\boldsymbol{\gamma}_j) = p(\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j)) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) - \boldsymbol{\delta}_{j0})'\boldsymbol{D}^{-1}_{j0}(\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) - \boldsymbol{\delta}_{j0})\right].$$

Furthermore, we can have the following conditional distribution of $\boldsymbol{\gamma}_j$:

$$\begin{aligned}
p(\boldsymbol{\gamma}_j|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}^*_{(-j)}, \boldsymbol{y}) &\propto p(\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j))f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}^*_{(-j)}) \\
&\propto \exp\left[-\frac{1}{2}(\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) - \boldsymbol{\delta}_{j0})'\boldsymbol{D}^{-1}_{j0}(\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) - \boldsymbol{\delta}_{j0})\right] \\
&\quad \times \prod_{i:y_{ij}=2}\left[\Phi\left(\frac{\gamma_{j2} - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right) - \Phi\left(-\frac{\tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right)\right] \\
&\quad \times \prod_{i:y_{ij}=3}\left[\Phi\left(\frac{\gamma_{j3} - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right) - \Phi\left(\frac{\gamma_{j2} - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right)\right] \\
&\quad \times \cdots \times \prod_{i:y_{ij}=C_j-1}\left[\Phi\left(\frac{1 - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right) - \Phi\left(\frac{\gamma_{j(C_j-2)} - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right)\right],
\end{aligned}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. Therefore, the conditional distribution of $\boldsymbol{\delta}_j$ is

$$p(\boldsymbol{\delta}_j|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}^*_{(-j)}, \boldsymbol{y}) \propto p(\boldsymbol{\gamma}_j|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}^*_{(-j)}, \boldsymbol{y}) \prod_{c_j=2}^{C_j-2} \frac{(1 - \gamma_{j(c_j-1)})\exp(\delta_{jc_j})}{(1 + \exp(\delta_{jc_j}))^2}.$$

Let $\boldsymbol{\delta}_j^s$ denote the value of $\boldsymbol{\delta}_j$ at the sth iteration. Using the Metropolis-Hastings algorithm, we can sample $\boldsymbol{\delta}_j$ by drawing $\boldsymbol{\delta}_j^{(s+1)} \sim q(\boldsymbol{\delta}_j|\cdots)$ from a Student-t proposal density $q(\boldsymbol{\delta}_j|\cdots) = f_T(\boldsymbol{\delta}_j|\hat{\boldsymbol{\delta}}_j, \hat{\boldsymbol{G}}_j, \boldsymbol{\nu})$, where ν is the degrees of freedom, and

$$\begin{aligned}
\hat{\boldsymbol{\delta}}_j &= \arg\max \log p(\boldsymbol{\delta}_j|\cdots) \\
\hat{\boldsymbol{G}}_j &= \left\{ \left[-\frac{\partial^2 \log \pi(\boldsymbol{\delta}_j|\cdots)}{\partial \boldsymbol{\delta}_j \partial \boldsymbol{\delta}_j'} \right]_{\boldsymbol{\delta}_j = \hat{\boldsymbol{\delta}}_j} \right\}^{-1}.
\end{aligned}$$

Return $\boldsymbol{\delta}_j^{s+1}$, with probability of $\min\left\{1, \dfrac{p(\boldsymbol{\delta}_j^{s+1}|\cdots)}{p(\boldsymbol{\delta}_j^{s}|\cdots)} \dfrac{f_T(\boldsymbol{\delta}_j^{s}|\hat{\boldsymbol{\delta}}_j, \hat{\boldsymbol{G}}_j, \nu)}{f_T(\boldsymbol{\delta}_j^{s+1}|\hat{\boldsymbol{\delta}}_j, \hat{\boldsymbol{G}}_j, \nu)}\right\}$; otherwise, repeat using the previous value of $\boldsymbol{\delta}_j^s$.

Then, γ_{jc_j} can be calculated using the following equation:

$$\gamma_{jc_j} = \frac{\gamma_{j(c_j-1)} + \exp(\delta_{jc_j})}{1 + \exp(\delta_{jc_j})}.$$

B.2 Derivation of Algorithm 2

B.2.1 Full Conditional Distribution of $\boldsymbol{\beta}$

Replace the jth element as the first element, that is,

$$\begin{aligned}
\boldsymbol{y}_i^* &= \begin{pmatrix} y_{ij}^* \\ \boldsymbol{y}_{i(-j)}^* \end{pmatrix}, \; \boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_{ij}' & \mathbf{0}' \\ \mathbf{O} & \boldsymbol{X}_{i(-j)} \end{pmatrix}, \; \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_j \\ \boldsymbol{\beta}_{(-j)} \end{pmatrix}, \\
\boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_{jj} & \boldsymbol{\sigma}_{(-j)}' \\ \boldsymbol{\sigma}_{(-j)} & \boldsymbol{\Sigma}_{(-j)} \end{pmatrix}, \; \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma^{jj} & \boldsymbol{\sigma}^{(-j)\prime} \\ \boldsymbol{\sigma}^{(-j)} & \boldsymbol{\Sigma}^{(-j)} \end{pmatrix}.
\end{aligned}$$

The FCD of $\boldsymbol{\beta}_j$ is

$$p(\boldsymbol{\beta}_j|\dots) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j0})'\boldsymbol{B}_{j0}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j0})\right]$$
$$\times \exp\left[\sum_{i=1}^{n}(\boldsymbol{y}_i^* - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i^* - \boldsymbol{X}_i\boldsymbol{\beta})\right].$$

After some manipulation, we have the following FCD of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}_j|\dots \sim \mathrm{N}(\tilde{\boldsymbol{\beta}}_j, \tilde{\boldsymbol{B}}_j),$$

where

$$\tilde{\boldsymbol{B}}_j = (\boldsymbol{B}_{j0}^{-1} + \sigma^{jj}\sum_{i=1}^{n}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}')^{-1},$$
$$\tilde{\boldsymbol{\beta}}_j = \tilde{\boldsymbol{B}}_j\left[\boldsymbol{B}_{j0}^{-1}\boldsymbol{\beta}_{j0} + \sigma^{jj}\sum_{i=1}^{n}\boldsymbol{x}_{ij}y_{ij}^* + \sum_{i=1}^{n}\boldsymbol{x}_{ij}\boldsymbol{\sigma}^{(-j)'}(\boldsymbol{y}_{i(-j)}^* - \boldsymbol{X}_{i(-j)}\boldsymbol{\beta}_{(-j)})\right].$$

If $\boldsymbol{X}_i = \boldsymbol{I}_m \otimes \boldsymbol{x}_i'$, that is, m equations have the same explanatory variables, we have

$$\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' = \boldsymbol{X}'\boldsymbol{X},\ \sum_{i=1}^{n}\boldsymbol{x}_i y_{ij}^* = \boldsymbol{X}'\tilde{\boldsymbol{y}}_j^*$$
$$\sum_{i=1}^{n}\boldsymbol{x}_{ij}\boldsymbol{\sigma}^{(-j)'}\boldsymbol{y}_{i(-j)}^* = \sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{y}_{i(-j)}^{*\prime}\boldsymbol{\sigma}^{(-j)} = \boldsymbol{X}'\boldsymbol{Y}_{(-j)}^*\boldsymbol{\sigma}^{(-j)}$$
$$\sum_{i=1}^{n}\boldsymbol{x}_{ij}\boldsymbol{\sigma}^{(-j)'}\boldsymbol{X}_{i(-j)}\boldsymbol{\beta}_{(-j)} = \sum_{i=1}^{n}\boldsymbol{x}_i(\boldsymbol{x}_i'\boldsymbol{\beta}_1,\dots,\boldsymbol{x}_i'\boldsymbol{\beta}_{j-1},\boldsymbol{x}_i'\boldsymbol{\beta}_{j+1},\dots,\boldsymbol{x}_i'\boldsymbol{\beta}_m)\boldsymbol{\sigma}^{(-j)}$$
$$= \sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{B}_{(-j)}\boldsymbol{\sigma}^{(-j)} = \boldsymbol{X}'\boldsymbol{X}\boldsymbol{B}_{(-j)}\boldsymbol{\sigma}^{(-j)},$$

where

$$\boldsymbol{Y}^* = \begin{pmatrix} \boldsymbol{y}_1^{*\prime} \\ \vdots \\ \boldsymbol{y}_n^{*\prime} \end{pmatrix} = (\tilde{\boldsymbol{y}}_1^*, \ldots, \tilde{\boldsymbol{y}}_m^*),\ \boldsymbol{Y}^*_{(-j)} = \begin{pmatrix} \boldsymbol{y}_{1(-j)}^{*\prime} \\ \vdots \\ \boldsymbol{y}_{n(-j)}^{*\prime} \end{pmatrix},\ \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix}$$

$$\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m),\ \boldsymbol{B}_{(-j)} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1}, \boldsymbol{\beta}_{j+1}, \ldots, \boldsymbol{\beta}_m).$$

B.2.2 Full Conditional Distribution of $\boldsymbol{\Sigma}^{-1}$

The FCD of $\boldsymbol{\Sigma}^{-1}$ is

$$\boldsymbol{\Sigma}^{-1} | \cdots \sim W(\tilde{n}, \tilde{\boldsymbol{Q}}^{-1}),$$

where

$$\tilde{n} = n_0 + n,\ \tilde{\boldsymbol{Q}} = \boldsymbol{Q}_0 + \sum_{i=1}^{n} (y_i^* - \boldsymbol{X}_i \boldsymbol{\beta})(y_i^* - \boldsymbol{X}_i \boldsymbol{\beta})'.$$

If $\boldsymbol{X}_i = \boldsymbol{I}_m \otimes \boldsymbol{x}_i'$, that is,

$$\sum_{i=1}^{n} (y_i^* - \boldsymbol{X}_i \boldsymbol{\beta})(y_i^* - \boldsymbol{X}_i \boldsymbol{\beta})' = (\boldsymbol{Y}^* - \boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y}^* - \boldsymbol{X}\boldsymbol{B}).$$

B.2.3 Full Conditional Distribution of $\boldsymbol{y}^*$

Let $\boldsymbol{y}^*_{(j)} = (y^*_{1j}, y^*_{2j}, y^*_{3j}, \ldots, y^*_{nj})'$ denote the vector of the jth element y^*_{ij} from $\boldsymbol{y}^*_i (i = 1, \ldots, n)$. Further, let $\boldsymbol{y}^*_{(-j)}$ denote the vector obtained by removing $\boldsymbol{y}^*_{(j)}$ from $\boldsymbol{y}^*$, and let $\boldsymbol{y}^*_{i(-j)}$ denote the vector obtained by removing $\boldsymbol{y}^*_{ij}$ from $\boldsymbol{y}^*_i$. We define $\boldsymbol{\gamma}_j = (\boldsymbol{\gamma}'_{j2}, \cdots, \boldsymbol{\gamma}'_{j(C_j-2)})'$. Then, we can obtain the joint conditional distribution of $\boldsymbol{\gamma}_j$ and $\boldsymbol{y}^*_j$, that is,

$$p(\boldsymbol{\gamma}_j, \boldsymbol{y}^*_j | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}^*_{(-j)}, \boldsymbol{y}) = p(\boldsymbol{\gamma}_j | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}^*_{(-j)}, \boldsymbol{y}) p(\boldsymbol{y}^*_j | \gamma_j, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}^*_{(-j)}, \boldsymbol{y}).$$

Since $\boldsymbol{y}_i^* \sim \mathrm{N}(\boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\Sigma})$, we have

$$y_{ij}^*|\boldsymbol{\gamma}_j, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}_{(-j)}^*, \boldsymbol{y} \sim \mathrm{N}(\tilde{\mu}_{ij}, \tilde{\sigma}_j^2) 1\left\{y_{ij}^* \in \mathcal{G}_{ij}\right\},$$

where

$$\begin{aligned}
\tilde{\mu}_{ij} &= \boldsymbol{x}_{ij}'\boldsymbol{\beta}_j + \boldsymbol{\sigma}_{(-j)}'\boldsymbol{\Sigma}_{(-j)}^{-1}(\boldsymbol{y}_{i(-j)}^* - \boldsymbol{X}_{i(-j)}\boldsymbol{\beta}_{(-j)}) \\
\tilde{\sigma}_j^2 &= \sigma_{jj} - \boldsymbol{\sigma}_{(-j)}'\boldsymbol{\Sigma}_{(-j)}^{-1}\boldsymbol{\sigma}_{(-j)}.
\end{aligned}$$

B.2.4 Full Conditional Distribution of $\boldsymbol{\gamma}$

Since $\boldsymbol{\zeta}_{jc_j} \sim \mathrm{N}(\boldsymbol{\zeta}_{j0}, \boldsymbol{G}_{j0})$, we have

$$p(\boldsymbol{\zeta}_{jc_j}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\zeta}_{jc_j} - \boldsymbol{\zeta}_{j0})'\boldsymbol{G}_{j0}^{-1}(\boldsymbol{\zeta}_{jc_j} - \boldsymbol{\zeta}_{j0})\right].$$

Then, the following conditional distribution of $\boldsymbol{\gamma}_{jc_j}$ can be wrriten as:

$$\begin{aligned}
p(\boldsymbol{\zeta}_{jc_j}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}_{(-j)}^*, \boldsymbol{y}) &\propto p(\boldsymbol{\zeta}_{jc_j}) f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{y}_{(-j)}^*) \\
&\propto \exp\left[-\frac{1}{2}(\boldsymbol{\zeta}_{jc_j} - \boldsymbol{\zeta}_{j0})'\boldsymbol{G}_{j0}^{-1}(\boldsymbol{\zeta}_{jc_j} - \boldsymbol{\zeta}_{j0})\right] \\
&\quad \times \prod_{i:y_{ij}=2}\left[\Phi\left(\frac{\gamma_{ij2} - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right) - \Phi\left(-\frac{\tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right)\right] \\
&\quad \times \prod_{i:y_{ij}=3}\left[\Phi\left(\frac{\gamma_{ij3} - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right) - \Phi\left(\frac{\gamma_{ij2} - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right)\right] \\
&\quad \times \cdots \times \prod_{i:y_{ij}=C_j-1}\left[\Phi\left(\frac{1 - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right) - \Phi\left(\frac{\gamma_{ij(C_j-2)} - \tilde{\mu}_{ij}}{\tilde{\sigma}_j}\right)\right],
\end{aligned}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution.

Let $\boldsymbol{\zeta}_{jc_j}^s$ denote the value of $\boldsymbol{\zeta}_{jc_j}$ at the sth iteration. Using the Metropolis-Hastings algorithm, we can sample $\boldsymbol{\zeta}_{jc_j}$ by drawing $\boldsymbol{\zeta}_{jc_j}^{s+1} \sim q(\boldsymbol{\zeta}_{jc_j}|\cdots)$ from a

Student-t proposal density $q(\boldsymbol{\zeta}_{jc_j}|\cdots) = f_T(\boldsymbol{\zeta}_{jc_j}|\hat{\boldsymbol{\zeta}}_{jc_j}, \hat{\boldsymbol{G}}_{jc_j}, \boldsymbol{\nu})$, where ν is the degrees of freedom, and

$$\hat{\boldsymbol{\zeta}}_{jc_j} = \arg\max_{\boldsymbol{\zeta}_{jc_j}} \log p(\boldsymbol{\zeta}_{jc_j}|\cdots)$$
$$\hat{\boldsymbol{G}}_{jc_j} = \left\{ \left[-\frac{\partial^2 \log \pi(\boldsymbol{\zeta}_{jc_j}|\cdots)}{\partial \boldsymbol{\zeta}_{jc_j} \partial \boldsymbol{\zeta}'_{jc_j}} \right]_{\boldsymbol{\zeta}_{jc_j} = \hat{\boldsymbol{\zeta}}_{jc_j}} \right\}^{-1}.$$

Return $\boldsymbol{\zeta}^{s+1}_{jc_j}$, with probability of $\min \left\{ 1, \dfrac{p(\boldsymbol{\zeta}^{s+1}_{jc_j}|\cdots)}{p(\boldsymbol{\zeta}^{s}_{jc_j}|\cdots)} \dfrac{f_T(\boldsymbol{\zeta}^{s}_{jc_j}|\hat{\boldsymbol{\zeta}}_{jc_j}, \hat{\boldsymbol{G}}_{jc_j}, \nu)}{f_T(\tilde{\boldsymbol{\zeta}}^{s+1}_{jc_j}|\hat{\boldsymbol{\zeta}}_{jc_j}, \hat{\boldsymbol{G}}_{jc_j}, \nu)} \right\}$; otherwise, repeat using the previous value of $\boldsymbol{\zeta}^{s}_{jc_j}$.

Then, γ_{ijc_j} can be calculated using the following equation:

$$\gamma_{ijc_j} = \frac{\gamma_{ij(c_j-1)} + \exp(\boldsymbol{\zeta}'_{jc_j} \boldsymbol{z}_i)}{1 + \exp(\boldsymbol{\zeta}'_{jc_j} \boldsymbol{z}_i)}.$$

B.3 Derivation of the marginal likelihood on Algorithm 1

Let $\boldsymbol{\theta}$ denote a vector of parameters included in the model. Then, according to the Bayes' Theorem, the marginal likelihood for this model can be written as

$$m(\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\boldsymbol{y})},$$

where $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function, $p(\boldsymbol{\theta})$ is the prior, $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior density of $\boldsymbol{\theta}$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5)$.

Using Chib's (1995a) method, the log marginal likelihood at a selected points, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \hat{\boldsymbol{\delta}}_3, \hat{\boldsymbol{\delta}}_4, \hat{\boldsymbol{\delta}}_5)$, is then

$$\log \hat{m}(\boldsymbol{y}) = \log p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}) + \log p(\hat{\boldsymbol{\theta}}) - \log \hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y}).$$

Under this model, we have

$$\begin{aligned}
\log p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}) &= \sum_{j=1}^{4} \log p(\boldsymbol{y}_j|\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \hat{\boldsymbol{\delta}}_3, \hat{\boldsymbol{\delta}}_4, \hat{\boldsymbol{\delta}}_5), \\
\log p(\hat{\boldsymbol{\theta}}) &= \log p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) + \log p(\hat{\boldsymbol{\Sigma}^{-1}}) + \log p(\hat{\boldsymbol{\delta}}_2, \hat{\boldsymbol{\delta}}_3, \hat{\boldsymbol{\delta}}_4, \hat{\boldsymbol{\delta}}_5) \\
\log \hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) &= \log \hat{p}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \hat{\boldsymbol{\delta}}_3, \hat{\boldsymbol{\delta}}_4, \hat{\boldsymbol{\delta}}_5|\boldsymbol{y})
\end{aligned}$$

Suppose the Gibbs sampler is applied to the complete conditional densities

$$\begin{aligned}
&p(\boldsymbol{\beta}_1|\boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{\beta}_2|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{\beta}_3|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{\beta}_4|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{\Sigma}^{-1}|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{y}_j^*|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*_{(-j)}), \;\; j = 1 \cdots 4.
\end{aligned}$$

The posterior density $\hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y})$ at $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \hat{\boldsymbol{\delta}}_3, \hat{\boldsymbol{\delta}}_4, \hat{\boldsymbol{\delta}}_5)$ can be expressed as,

$$
\begin{aligned}
\hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) &= \hat{p}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}) \times \hat{p}(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1) \times \hat{p}(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) \times \hat{p}(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) \\
&\times \hat{p}(\widehat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) \times \hat{p}(\hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}) \\
&\times \hat{p}(\hat{\boldsymbol{\delta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2) \times \hat{p}(\hat{\boldsymbol{\delta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \hat{\boldsymbol{\delta}}_3) \\
&\times \hat{p}(\hat{\boldsymbol{\delta}}_5|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \hat{\boldsymbol{\delta}}_3, \hat{\boldsymbol{\delta}}_4),
\end{aligned}
$$

where

$$
\begin{aligned}
p(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}) = \int & p(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) \\
&\times p(\boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y}) d\boldsymbol{\beta}_2 d\boldsymbol{\beta}_3 d\boldsymbol{\beta}_4 d\boldsymbol{\Sigma}^{-1} d\boldsymbol{\delta}_2 d\boldsymbol{\delta}_3 d\boldsymbol{\delta}_4 d\boldsymbol{\delta}_5 d\boldsymbol{y}^*,
\end{aligned}
$$

$$
\begin{aligned}
p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1) = \int & p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) \\
&\times p(\boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1) d\boldsymbol{\beta}_3 d\boldsymbol{\beta}_4 d\boldsymbol{\Sigma}^{-1} d\boldsymbol{\delta}_2 d\boldsymbol{\delta}_3 d\boldsymbol{\delta}_4 d\boldsymbol{\delta}_5 d\boldsymbol{y}^*,
\end{aligned}
$$

$$
\begin{aligned}
p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = \int & p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) \\
&\times p(\boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) d\boldsymbol{\beta}_4 d\boldsymbol{\Sigma}^{-1} d\boldsymbol{\delta}_2 d\boldsymbol{\delta}_3 d\boldsymbol{\delta}_4 d\boldsymbol{\delta}_5 d\boldsymbol{y}^*,
\end{aligned}
$$

$$
\begin{aligned}
p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) = \int & p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) \\
&\times p(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) d\boldsymbol{\Sigma}^{-1} d\boldsymbol{\delta}_2 d\boldsymbol{\delta}_3 d\boldsymbol{\delta}_4 d\boldsymbol{\delta}_5 d\boldsymbol{y}^*,
\end{aligned}
$$

and

$$p(\hat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) = \int p(\hat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)$$
$$\times\, p(\boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) d\boldsymbol{\delta}_2 d\boldsymbol{\delta}_3 d\boldsymbol{\delta}_4 d\boldsymbol{\delta}_5 d\boldsymbol{y}^*.$$

Let the main output from the algorithm be given by $\{\boldsymbol{\beta}_1^{(s)}, \boldsymbol{\beta}_2^{(s)}, \boldsymbol{\beta}_3^{(s)}, \boldsymbol{\beta}_4^{(s)}, \boldsymbol{\Sigma}^{-1(s)},$ $\boldsymbol{\delta}_2^{(s)}, \boldsymbol{\delta}_3^{(s)}, \boldsymbol{\delta}_4^{(s)}, \boldsymbol{\delta}_5^{(s)}, \boldsymbol{y}^{*(s)}\}_{s=1}^{S}$. Then, the appropriate Monte Carlo estimate of $p(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y})$ is

$$\hat{p}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}) = S^{-1}\sum_{s=1}^{S} p(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}, \boldsymbol{\beta}_2^{(s)}, \boldsymbol{\beta}_3^{(s)}, \boldsymbol{\beta}_4^{(s)}, \boldsymbol{\Sigma}^{-1(s)}, \boldsymbol{\delta}_2^{(s)}, \boldsymbol{\delta}_3^{(s)}, \boldsymbol{\delta}_4^{(s)}, \boldsymbol{\delta}_5^{(s)}, \boldsymbol{y}^{*(s)}).$$

However, the conditional density of $p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1), p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2), p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3),$ $p(\hat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4)$ can not be averaged directly, because the main outputs of $\boldsymbol{\beta}_2, \boldsymbol{\beta}_3,$
$\boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5$ and $\boldsymbol{y}^*$ are conditioned on $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4$ not on $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4$. For instance, to estimate $p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1)$, an additional G_1 interation need to be

computed with the conditional densities

$$p(\boldsymbol{\beta}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*);$$

$$p(\boldsymbol{\beta}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*);$$

$$p(\boldsymbol{\Sigma}^{-1}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*);$$

$$p(\boldsymbol{\delta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*);$$

$$p(\boldsymbol{\delta}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*);$$

$$p(\boldsymbol{\delta}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_5, \boldsymbol{y}^*);$$

$$p(\boldsymbol{\delta}_5|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{y}^*);$$

$$p(\boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*).$$

Consequently, the calculation of $\hat{p}(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}_j, \hat{\boldsymbol{\beta}}_1)$ is

$$\hat{p}(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1) = G_1^{-1} \sum_{g_1=1}^{G_1} p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_3^{(g_1)}, \boldsymbol{\beta}_4^{(g_1)}, \boldsymbol{\Sigma}^{-1(g_1)}, \boldsymbol{\delta}_2^{(g_1)}, \boldsymbol{\delta}_3^{(g_1)}, \boldsymbol{\delta}_4^{(g_1)}, \boldsymbol{\delta}_5^{(g_1)}, \boldsymbol{y}^{*(g_1)}).$$

Using a similar technique, the estimate of $p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2), p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3)$, $p(\hat{\boldsymbol{\Sigma}}^{-1}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4)$ are

$$\hat{p}(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = G_2^{-1} \sum_{g_2=1}^{G_2} p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \boldsymbol{\beta}_4^{(g_2)}, \boldsymbol{\Sigma}^{-1(g_1)}, \boldsymbol{\delta}_2^{(g_2)}, \boldsymbol{\delta}_3^{(g_2)}, \boldsymbol{\delta}_4^{(g_2)}, \boldsymbol{\delta}_5^{(g_2)}, \boldsymbol{y}^{*(g_2)}),$$

$$\hat{p}(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) = G_3^{-1} \sum_{g_3=1}^{G_3} p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \boldsymbol{\Sigma}^{-1(g_1)}, \boldsymbol{\delta}_2^{(g_3)}, \boldsymbol{\delta}_3^{(g_3)}, \boldsymbol{\delta}_4^{(g_3)}, \boldsymbol{\delta}_5^{(g_3)}, \boldsymbol{y}^{*(g_3)}),$$

$$\hat{p}(\hat{\mathbf{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) = G_4^{-1} \sum_{g_4=1}^{G_4} p(\hat{\mathbf{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \boldsymbol{\delta}_2^{(g_4)}, \boldsymbol{\delta}_3^{(g_4)}, \boldsymbol{\delta}_4^{(g_4)}, \boldsymbol{\delta}_5^{(g_4)}, \boldsymbol{y}^{*(g_4)}),$$

The same method can not be used to calculate $\hat{p}(\hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\mathbf{\Sigma}^{-1}})$, $\hat{p}(\hat{\boldsymbol{\delta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2,$
$\hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\mathbf{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2), \hat{p}(\hat{\boldsymbol{\delta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\mathbf{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \hat{\boldsymbol{\delta}}_3)$ and $\hat{p}(\hat{\boldsymbol{\delta}}_5|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\mathbf{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \hat{\boldsymbol{\delta}}_3, \hat{\boldsymbol{\delta}}_4)$, because $\boldsymbol{\delta}_2$, $\boldsymbol{\delta}_3$, $\boldsymbol{\delta}_4$ and $\boldsymbol{\delta}_5$ are sampled using the M-H algorithm.

Accordingly, Chib and Jeliazkov (2001) provide an approach to estimate the marginal likelihood from the M-H output. Let $q(\boldsymbol{\delta}_2, \tilde{\boldsymbol{\delta}}_2|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \mathbf{\Sigma}^{-1}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)$ denote the proposal generating density of the M-H chain for the transition from the current $\boldsymbol{\delta}_2$ to a proposed value $\tilde{\boldsymbol{\delta}}_2$, and let

$$\begin{aligned}\alpha(\boldsymbol{\delta}_2, \tilde{\boldsymbol{\delta}}_2|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \mathbf{\Sigma}^{-1}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) &= \frac{p(\boldsymbol{y}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \mathbf{\Sigma}^{-1}, \tilde{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)}{p(\boldsymbol{y}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \mathbf{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)} \\ \times \frac{p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \mathbf{\Sigma}^{-1}, \tilde{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5) q(\tilde{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_2|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \mathbf{\Sigma}^{-1}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)}{p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \mathbf{\Sigma}^{-1}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5) q(\boldsymbol{\delta}_2, \tilde{\boldsymbol{\delta}}_2|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \mathbf{\Sigma}^{-1}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)}\end{aligned}$$

denote the acceptance ratio of moving from $\boldsymbol{\delta}_2$ to $\tilde{\boldsymbol{\delta}}_2$. In addition, let

$$\begin{aligned}p(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\mathbf{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) &= \alpha(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\mathbf{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) \\ \times q(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\mathbf{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)\end{aligned} \tag{B.1}$$

denote the subkernel of the M-H algorithm. Then, according to the reversibility of this subkernel, we obtain

$$\begin{aligned}
&p(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)p(\boldsymbol{\delta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) \\
&\quad = p(\hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)p(\hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*).
\end{aligned} \tag{B.2}$$

After multiplying both sides of (B.2) by $p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y})$ and integrating out $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5$, and $\boldsymbol{y}^*$, we have

$$\begin{aligned}
&\int p(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y}) \\
&\qquad\qquad\qquad d\hat{\boldsymbol{\beta}}_1 d\hat{\boldsymbol{\beta}}_2 d\hat{\boldsymbol{\beta}}_3 d\hat{\boldsymbol{\beta}}_4 d\hat{\boldsymbol{\Sigma}^{-1}} d\boldsymbol{\delta}_2 d\boldsymbol{\delta}_3 d\boldsymbol{\delta}_4 d\boldsymbol{\delta}_5 d\boldsymbol{y}^* \\
&\quad = p(\hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}) \int p(\hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) \\
&\qquad \times p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\delta}}_2) d\boldsymbol{\delta}_2 d\boldsymbol{\delta}_3 d\boldsymbol{\delta}_4 d\boldsymbol{\delta}_5 d\boldsymbol{y}^*.
\end{aligned} \tag{B.3}$$

From (B.3) and the definition of $p(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)$ in (B.1), we can obtain that

$$\begin{aligned}
&p(\hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}) \\
&= \frac{E_1\{\alpha(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)q(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)\}}{E_2\{\alpha(\hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)\}},
\end{aligned}$$

where E_1 is the expectation with respect to the conditional posterior $p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y})$, and the denominator expectation E_2 is with respect to $p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\delta}}_2) \times q(\hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)$.

We estimate $E_1\{\alpha(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*) \times q(\boldsymbol{\delta}_2, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)\}$ using the posterior draws of $\boldsymbol{\delta}_2^{(s)}$, $\boldsymbol{\delta}_3^{(s)}$, $\boldsymbol{\delta}_4^{(s)}$, $\boldsymbol{\delta}_5^{(s)}$ and $\boldsymbol{y}^{*(s)}$ from the main MCMC output, and then average the quantity. However, the draws of $\boldsymbol{\delta}_2$, $\boldsymbol{\delta}_3$, $\boldsymbol{\delta}_4$, $\boldsymbol{\delta}_5$ and $\boldsymbol{y}^*$ from the main MCMC output are not conditional on $\hat{\boldsymbol{\delta}}_2$, so we need to sample an additional G_5 iterations for the estimate of $E_2\{\alpha(\hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*)\}$. We sample $\boldsymbol{\delta}_3^{(g_5)}$, $\boldsymbol{\delta}_4^{(g_5)}$, $\boldsymbol{\delta}_5^{(g_5)}$ and $\boldsymbol{y}^{*(g_5)}$ with the full conditional densities

$$
\begin{aligned}
&p(\boldsymbol{\delta}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5, \boldsymbol{y}^*);\\
&p(\boldsymbol{\delta}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_5, \boldsymbol{y}^*);\\
&p(\boldsymbol{\delta}_5|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{y}^*);\\
&p(\boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_3, \boldsymbol{\delta}_4, \boldsymbol{\delta}_5,).
\end{aligned}
$$

Furthermore, we generate $\boldsymbol{\delta}_2^{(g_5)} \sim q(\hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3^{(g_5)}, \boldsymbol{\delta}_4^{(g_5)}, \boldsymbol{\delta}_5^{(g_5)}, \boldsymbol{y}^{*(g_5)})$ at each sampling step. Then, the estimate of $p(\hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}})$ can be written as

$$
\begin{aligned}
&\hat{p}(\hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}})\\
&= \frac{S^{-1}\sum_{s=1}^{S}\{\alpha(\boldsymbol{\delta}_2^{(s)}, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3^{(s)}, \boldsymbol{\delta}_4^{(s)}, \boldsymbol{\delta}_5^{(s)}, \boldsymbol{y}^{*(s)}) \times q(\boldsymbol{\delta}_2^{(s)}, \hat{\boldsymbol{\delta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3^{(s)}, \boldsymbol{\delta}_4^{(s)}, \boldsymbol{\delta}_5^{(s)}, \boldsymbol{y}^{*(s)})\}}{G_5^{-1}\sum_{g_5=1}^{G_5}\alpha(\hat{\boldsymbol{\delta}}_2, \boldsymbol{\delta}_2^{(g_5)}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\delta}_3^{(g_5)}, \boldsymbol{\delta}_4^{(g_5)}, \boldsymbol{\delta}_5^{(g_5)}, \boldsymbol{y}^{*(g_5)})}.
\end{aligned}
$$

Using a the similar calculation, we have that

$$\hat{p}(\hat{\boldsymbol{\delta}}_3|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\widehat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\delta}}_2)$$
$$= S^{-1}\sum_{s=1}^{S}\{\alpha(\boldsymbol{\delta}_3^{(s)},\hat{\boldsymbol{\delta}}_3|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\widehat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\delta}}_2,\boldsymbol{\delta}_4^{(s)},\boldsymbol{\delta}_5^{(s)},\boldsymbol{y}^{*(s)})$$
$$\frac{\times q(\boldsymbol{\delta}_3^{(s)},\hat{\boldsymbol{\delta}}_3|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\widehat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\delta}}_2,\boldsymbol{\delta}_4^{(s)},\boldsymbol{\delta}_5^{(s)},\boldsymbol{y}^{*(s)})\}}{G_6^{-1}\sum_{g_6=1}^{G_6}\alpha(\hat{\boldsymbol{\delta}}_3,\boldsymbol{\delta}_3^{(g_6)}|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\delta}}_2,\boldsymbol{\delta}_4^{(g_6)},\boldsymbol{\delta}_5^{(g_6)},\boldsymbol{y}^{*(g_6)})},$$

$$\hat{p}(\hat{\boldsymbol{\delta}}_4|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\widehat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\delta}}_2,\hat{\boldsymbol{\delta}}_3)$$
$$= S^{-1}\sum_{s=1}^{S}\{\alpha(\boldsymbol{\delta}_4^{(s)},\hat{\boldsymbol{\delta}}_4|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\widehat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\delta}}_2,\hat{\boldsymbol{\delta}}_3,\boldsymbol{\delta}_5^{(s)},\boldsymbol{y}^{*(s)})$$
$$\frac{\times q(\boldsymbol{\delta}_4^{(s)},\hat{\boldsymbol{\delta}}_4|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\widehat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\delta}}_2,\hat{\boldsymbol{\delta}}_3,\boldsymbol{\delta}_5^{(s)},\boldsymbol{y}^{*(s)})\}}{G_7^{-1}\sum_{g_7=1}^{G_7}\alpha(\hat{\boldsymbol{\delta}}_4,\boldsymbol{\delta}_4^{(g_7)}|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\delta}}_2,\hat{\boldsymbol{\delta}}_3,\boldsymbol{\delta}_5^{(g_7)},\boldsymbol{y}^{*(g_7)})},$$

$$\hat{p}(\hat{\boldsymbol{\delta}}_5|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\widehat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\delta}}_2,\hat{\boldsymbol{\delta}}_3,\hat{\boldsymbol{\delta}}_4)$$
$$= S^{-1}\sum_{s=1}^{S}\{\alpha(\boldsymbol{\delta}_5^{(s)},\hat{\boldsymbol{\delta}}_5|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\widehat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\delta}}_2,\hat{\boldsymbol{\delta}}_3,\hat{\boldsymbol{\delta}}_4,\boldsymbol{y}^{*(s)})$$
$$\frac{\times q(\boldsymbol{\delta}_5^{(s)},\hat{\boldsymbol{\delta}}_5|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\widehat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\delta}}_2,\hat{\boldsymbol{\delta}}_3,\hat{\boldsymbol{\delta}}_4,\boldsymbol{y}^{*(s)})\}}{G_8^{-1}\sum_{g_8=1}^{G_8}\alpha(\hat{\boldsymbol{\delta}}_5,\boldsymbol{\delta}_5^{(g_8)}|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\delta}}_2,\hat{\boldsymbol{\delta}}_3,\hat{\boldsymbol{\delta}}_4,\boldsymbol{y}^{*(g_8)})}.$$

B.4 Derivation of the marginal likelihood on Algorithm 2

Let $\boldsymbol{\theta}$ denote a vector of parameters included in the model. Then, according to the Bayes' Theorem, the marginal likelihood for this model can be written as

$$m(\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\boldsymbol{y})},$$

where $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function, $p(\boldsymbol{\theta})$ is the prior, $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior density of $\boldsymbol{\theta}$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5)$.

Using Chib's (1995a) method, the log marginal likelihood at a selected points, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \hat{\boldsymbol{\zeta}}_4, \hat{\boldsymbol{\zeta}}_5)$, is then

$$\log \hat{m}(\boldsymbol{y}) = \log p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}) + \log p(\hat{\boldsymbol{\theta}}) - \log \hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y}).$$

Under this model, we have

$$\begin{aligned}
\log p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}) &= \sum_{j=1}^{4} \log p(\boldsymbol{y}_j|\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \hat{\boldsymbol{\zeta}}_4, \hat{\boldsymbol{\zeta}}_5), \\
\log p(\hat{\boldsymbol{\theta}}) &= \log p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) + \log p(\hat{\boldsymbol{\Sigma}^{-1}}) + \log p(\hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \hat{\boldsymbol{\zeta}}_4, \hat{\boldsymbol{\zeta}}_5) \\
\log \hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) &= \log \hat{p}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \hat{\boldsymbol{\zeta}}_4, \hat{\boldsymbol{\zeta}}_5|\boldsymbol{y})
\end{aligned}$$

Suppose the Gibbs sampler is applied to the complete conditional densities

$$\begin{aligned}
&p(\boldsymbol{\beta}_1|\boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{\beta}_2|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{\beta}_3|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{\beta}_4|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{\Sigma}^{-1}|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*); \\
&p(\boldsymbol{y}_j^*|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}_{(-j)}^*),\ j = 1 \cdots 4.
\end{aligned}$$

The posterior density $\hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y})$ at $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \hat{\boldsymbol{\zeta}}_4, \hat{\boldsymbol{\zeta}}_5)$ can be expressed as,

$$\begin{aligned}
\hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) &= \hat{p}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}) \times \hat{p}(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1) \times \hat{p}(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) \times \hat{p}(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) \\
&\times \hat{p}(\hat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) \times \hat{p}(\hat{\boldsymbol{\zeta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}) \\
&\times \hat{p}(\hat{\boldsymbol{\zeta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2) \times \hat{p}(\hat{\boldsymbol{\zeta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3) \\
&\times \hat{p}(\hat{\boldsymbol{\zeta}_5}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}_2}, \hat{\boldsymbol{\zeta}_3}, \hat{\boldsymbol{\zeta}_4}),
\end{aligned}$$

where

$$\begin{aligned}
p(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}) &= \int p(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) \\
&\times p(\boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*|\boldsymbol{y}) d\boldsymbol{\beta}_2 d\boldsymbol{\beta}_3 d\boldsymbol{\beta}_4 d\boldsymbol{\Sigma}^{-1} d\boldsymbol{\zeta}_2 d\boldsymbol{\zeta}_3 d\boldsymbol{\zeta}_4 d\boldsymbol{\zeta}_5 d\boldsymbol{y}^*,
\end{aligned}$$

$$\begin{aligned}
p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1) &= \int p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) \\
&\times p(\boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1) d\boldsymbol{\beta}_3 d\boldsymbol{\beta}_4 d\boldsymbol{\Sigma}^{-1} d\boldsymbol{\zeta}_2 d\boldsymbol{\zeta}_3 d\boldsymbol{\zeta}_4 d\boldsymbol{\zeta}_5 d\boldsymbol{y}^*,
\end{aligned}$$

$$\begin{aligned}
p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) &= \int p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) \\
&\times p(\boldsymbol{\beta}_4, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) d\boldsymbol{\beta}_4 d\boldsymbol{\Sigma}^{-1} d\boldsymbol{\zeta}_2 d\boldsymbol{\zeta}_3 d\boldsymbol{\zeta}_4 d\boldsymbol{\zeta}_5 d\boldsymbol{y}^*,
\end{aligned}$$

$$\begin{aligned}
p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) &= \int p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) \\
&\times p(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) d\boldsymbol{\Sigma}^{-1} d\boldsymbol{\zeta}_2 d\boldsymbol{\zeta}_3 d\boldsymbol{\zeta}_4 d\boldsymbol{\zeta}_5 d\boldsymbol{y}^*,
\end{aligned}$$

and

$$p(\hat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) = \int p(\hat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*)$$
$$\times\, p(\boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) d\boldsymbol{\zeta}_2 d\boldsymbol{\zeta}_3 d\boldsymbol{\zeta}_4 d\boldsymbol{\zeta}_5 d\boldsymbol{y}^*.$$

Let the main output from the algorithm be given by $\{\boldsymbol{\beta}_1^{(s)}, \boldsymbol{\beta}_2^{(s)}, \boldsymbol{\beta}_3^{(s)}, \boldsymbol{\beta}_4^{(s)}, \boldsymbol{\Sigma}^{-1(s)}, \boldsymbol{\zeta}_2^{(s)}, \boldsymbol{\zeta}_3^{(s)}, \boldsymbol{\zeta}_4^{(s)}, \boldsymbol{\zeta}_5^{(s)}, \boldsymbol{y}^{*(s)}\}_{s=1}^S$. Then, the appropriate Monte Carlo estimate of $p(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y})$ is

$$\hat{p}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}) = S^{-1}\sum_{s=1}^{S} p(\hat{\boldsymbol{\beta}}_1|\boldsymbol{y}, \boldsymbol{\beta}_2^{(s)}, \boldsymbol{\beta}_3^{(s)}, \boldsymbol{\beta}_4^{(s)}, \boldsymbol{\Sigma}^{-1(s)}, \boldsymbol{\zeta}_2^{(s)}, \boldsymbol{\zeta}_3^{(s)}, \boldsymbol{\zeta}_4^{(s)}, \boldsymbol{\zeta}_5^{(s)}, \boldsymbol{y}^{*(s)}).$$

However, the conditional density of $p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1), p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2), p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3)$, $p(\hat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4)$ can not be averaged directly, because the main outputs of $\boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4$,
$\boldsymbol{\Sigma}^{-1}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5$ and $\boldsymbol{y}^*$ are conditioned on $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4$ not on $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4$. For instance, to estimate $p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1)$, an additional G_1 interation need to be

computed with the conditional densities

$$p(\boldsymbol{\beta}_3|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*);$$

$$p(\boldsymbol{\beta}_4|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*);$$

$$p(\boldsymbol{\Sigma}^{-1}|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\zeta}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*);$$

$$p(\boldsymbol{\zeta}_2|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*);$$

$$p(\boldsymbol{\zeta}_3|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_2,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*);$$

$$p(\boldsymbol{\zeta}_4|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_5,\boldsymbol{y}^*);$$

$$p(\boldsymbol{\zeta}_5|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{y}^*);$$

$$p(\boldsymbol{y}^*|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*).$$

Consequently, the calculation of $\hat{p}(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y}_j,\hat{\boldsymbol{\beta}}_1)$ is

$$\hat{p}(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1) = G_1^{-1}\sum_{g_1=1}^{G_1} p(\hat{\boldsymbol{\beta}}_2|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\boldsymbol{\beta}_3^{(g_1)},\boldsymbol{\beta}_4^{(g_1)},\boldsymbol{\Sigma}^{-1(g_1)},\boldsymbol{\zeta}_2^{(g_1)},\boldsymbol{\zeta}_3^{(g_1)},\boldsymbol{\zeta}_4^{(g_1)},\boldsymbol{\zeta}_5^{(g_1)},\boldsymbol{y}^{*(g_1)}).$$

A similar technique, the estimate of $p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2)$, $p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3)$, $p(\hat{\boldsymbol{\Sigma}}^{-1}|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4)$ are

$$\hat{p}(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2) = G_2^{-1}\sum_{g_2=1}^{G_2} p(\hat{\boldsymbol{\beta}}_3|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\boldsymbol{\beta}_4^{(g_2)},\boldsymbol{\Sigma}^{-1(g_1)},\boldsymbol{\zeta}_2^{(g_2)},\boldsymbol{\zeta}_3^{(g_2)},\boldsymbol{\zeta}_4^{(g_2)},\boldsymbol{\zeta}_5^{(g_2)},\boldsymbol{y}^{*(g_2)}),$$

$$\hat{p}(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3) = G_3^{-1}\sum_{g_3=1}^{G_3} p(\hat{\boldsymbol{\beta}}_4|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\boldsymbol{\Sigma}^{-1(g_1)},\boldsymbol{\zeta}_2^{(g_3)},\boldsymbol{\zeta}_3^{(g_3)},\boldsymbol{\zeta}_4^{(g_3)},\boldsymbol{\zeta}_5^{(g_3)},\boldsymbol{y}^{*(g_3)}),$$

$$\hat{p}(\hat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4) = G_4^{-1}\sum_{g_4=1}^{G_4} p(\hat{\boldsymbol{\Sigma}^{-1}}|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\boldsymbol{\zeta}_2^{(g_4)},\boldsymbol{\zeta}_3^{(g_4)},\boldsymbol{\zeta}_4^{(g_4)},\boldsymbol{\zeta}_5^{(g_4)},\boldsymbol{y}^{*(g_4)}),$$

The same method can not be used to calculate $\hat{p}(\hat{\boldsymbol{\zeta}}_2|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\Sigma}^{-1}})$, $\hat{p}(\hat{\boldsymbol{\zeta}}_3|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,$ $\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\zeta}}_2)$, $\hat{p}(\hat{\boldsymbol{\zeta}}_4|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\zeta}}_2,\hat{\boldsymbol{\zeta}}_3)$ and $\hat{p}(\hat{\boldsymbol{\zeta}}_5|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\Sigma}^{-1}},\hat{\boldsymbol{\zeta}}_2,\hat{\boldsymbol{\zeta}}_3,\hat{\boldsymbol{\zeta}}_4)$, because $\boldsymbol{\zeta}_2$, $\boldsymbol{\zeta}_3$, $\boldsymbol{\zeta}_4$ and $\boldsymbol{\zeta}_5$ are sampled using the M-H algorithm.

Accordingly, Chib and Jeliazkov (2001) provide an approach to estimate the marginal likelihood from the M-H output. Let $q(\boldsymbol{\zeta}_2,\tilde{\boldsymbol{\zeta}}_2|\boldsymbol{y},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*)$ denote the proposal generating density of the M-H chain for the transition from the current $\boldsymbol{\zeta}_2$ to a proposed value $\tilde{\boldsymbol{\zeta}}_2$, and let

$$\alpha(\boldsymbol{\zeta}_2,\tilde{\boldsymbol{\zeta}}_2|\boldsymbol{y},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*) = \frac{p(\boldsymbol{y}|\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\tilde{\boldsymbol{\zeta}}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*)}{p(\boldsymbol{y}|\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*)}$$
$$\times\frac{p(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\tilde{\boldsymbol{\zeta}}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5)q(\tilde{\boldsymbol{\zeta}}_2,\boldsymbol{\zeta}_2|\boldsymbol{y},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*)}{p(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_2,\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5)q(\boldsymbol{\zeta}_2,\tilde{\boldsymbol{\zeta}}_2|\boldsymbol{y},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4,\boldsymbol{\Sigma}^{-1},\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*)}$$

denote the acceptance ratio of moving from $\boldsymbol{\zeta}_2$ to $\tilde{\boldsymbol{\zeta}}_2$. In addition, let

$$p(\boldsymbol{\zeta}_2,\hat{\boldsymbol{\zeta}}_2|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\Sigma}^{-1}},\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*) = \alpha(\boldsymbol{\zeta}_2,\hat{\boldsymbol{\zeta}}_2|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\Sigma}^{-1}},\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*)$$
$$\times\, q(\boldsymbol{\zeta}_2,\hat{\boldsymbol{\zeta}}_2|\boldsymbol{y},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\beta}}_3,\hat{\boldsymbol{\beta}}_4,\hat{\boldsymbol{\Sigma}^{-1}},\boldsymbol{\zeta}_3,\boldsymbol{\zeta}_4,\boldsymbol{\zeta}_5,\boldsymbol{y}^*) \quad \text{(B.4)}$$

denote the subkernel of the M-H algorithm. Then, according to the reversibility of this subkernel, we obtain

$$
\begin{aligned}
&p(\boldsymbol{\zeta}_2, \hat{\boldsymbol{\zeta}}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) p(\boldsymbol{\zeta}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) \\
&\quad = p(\hat{\boldsymbol{\zeta}}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) p(\hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*).
\end{aligned}
\tag{B.5}
$$

After multiplying both sides of (B.5) by $p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^* | \boldsymbol{y})$ and integrating out $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5$, and $\boldsymbol{y}^*$, we have

$$
\begin{aligned}
&\int p(\boldsymbol{\zeta}_2, \hat{\boldsymbol{\zeta}}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^* | \boldsymbol{y}) \\
&\qquad\qquad\qquad\qquad d\hat{\boldsymbol{\beta}}_1 d\hat{\boldsymbol{\beta}}_2 d\hat{\boldsymbol{\beta}}_3 d\hat{\boldsymbol{\beta}}_4 d\widehat{\boldsymbol{\Sigma}^{-1}} d\boldsymbol{\zeta}_2 d\boldsymbol{\zeta}_3 d\boldsymbol{\zeta}_4 d\boldsymbol{\zeta}_5 d\boldsymbol{y}^* \\
&\quad = p(\hat{\boldsymbol{\zeta}}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}) \int p(\hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) \\
&\qquad \times p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^* | \boldsymbol{y}, \hat{\boldsymbol{\zeta}}_2) d\boldsymbol{\zeta}_2 d\boldsymbol{\zeta}_3 d\boldsymbol{\zeta}_4 d\boldsymbol{\zeta}_5 d\boldsymbol{y}^*.
\end{aligned}
\tag{B.6}
$$

From (B.6) and the definition of $p(\boldsymbol{\zeta}_2, \hat{\boldsymbol{\zeta}}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*)$ in (B.4), we can obtain that

$$
\begin{aligned}
&p(\hat{\boldsymbol{\zeta}}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}) \\
&= \frac{E_1\{\alpha(\boldsymbol{\zeta}_2, \hat{\boldsymbol{\zeta}}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) q(\boldsymbol{\zeta}_2, \hat{\boldsymbol{\zeta}}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*)\}}{E_2\{\alpha(\hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*)\}},
\end{aligned}
$$

where E_1 is the expectation with respect to the conditional posterior $p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^* | \boldsymbol{y})$, and the denominator expectation E_2 is with respect to $p(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^* | \boldsymbol{y}, \hat{\boldsymbol{\zeta}}_2) \times q(\hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_2 | \boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*)$.

We estimate $E_1\{\alpha(\boldsymbol{\zeta}_2, \hat{\boldsymbol{\zeta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*) \times q(\boldsymbol{\zeta}_2, \hat{\boldsymbol{\zeta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*)\}$ using the posterior draws of $\boldsymbol{\zeta}_2^{(s)}$, $\boldsymbol{\zeta}_3^{(s)}$, $\boldsymbol{\zeta}_4^{(s)}$, $\boldsymbol{\zeta}_5^{(s)}$ and $\boldsymbol{y}^{*(s)}$ from the main MCMC output, and then average the quantity. However, the draws of $\boldsymbol{\zeta}_2$, $\boldsymbol{\zeta}_3$, $\boldsymbol{\zeta}_4$, $\boldsymbol{\zeta}_5$ and $\boldsymbol{y}^*$ from the main MCMC output are not conditional on $\hat{\boldsymbol{\zeta}}_2$, so we need to sample an additional G_5 iterations for the estimate of $E_2\{\alpha(\hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*)\}$. We sample $\boldsymbol{\zeta}_3^{(g_5)}$, $\boldsymbol{\zeta}_4^{(g_5)}$, $\boldsymbol{\zeta}_5^{(g_5)}$ and $\boldsymbol{y}^{*(g_5)}$ with the full conditional densities

$$
\begin{aligned}
&p(\boldsymbol{\zeta}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5, \boldsymbol{y}^*);\\
&p(\boldsymbol{\zeta}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_5, \boldsymbol{y}^*);\\
&p(\boldsymbol{\zeta}_5|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{y}^*);\\
&p(\boldsymbol{y}^*|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_3, \boldsymbol{\zeta}_4, \boldsymbol{\zeta}_5,).
\end{aligned}
$$

Furthermore, we generate $\boldsymbol{\zeta}_2^{(g_5)} \sim q(\hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3^{(g_5)}, \boldsymbol{\zeta}_4^{(g_5)}, \boldsymbol{\zeta}_5^{(g_5)}, \boldsymbol{y}^{*(g_5)})$ at each sampling step. Then, the estimate of $p(\hat{\boldsymbol{\zeta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}})$ can be written as

$$
\begin{aligned}
&\hat{p}(\hat{\boldsymbol{\zeta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}})\\
&= S^{-1}\sum_{s=1}^{S}\{\alpha(\boldsymbol{\zeta}_2^{(s)}, \hat{\boldsymbol{\zeta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3^{(s)}, \boldsymbol{\zeta}_4^{(s)}, \boldsymbol{\zeta}_5^{(s)}, \boldsymbol{y}^{*(s)})\\
&\qquad \frac{\times q(\boldsymbol{\zeta}_2^{(s)}, \hat{\boldsymbol{\zeta}}_2|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3^{(s)}, \boldsymbol{\zeta}_4^{(s)}, \boldsymbol{\zeta}_5^{(s)}, \boldsymbol{y}^{*(s)})\}}{G_5^{-1}\sum_{g_5=1}^{G_5}\alpha(\hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_2^{(g_5)}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \widehat{\boldsymbol{\Sigma}^{-1}}, \boldsymbol{\zeta}_3^{(g_5)}, \boldsymbol{\zeta}_4^{(g_5)}, \boldsymbol{\zeta}_5^{(g_5)}, \boldsymbol{y}^{*(g_5)})}.
\end{aligned}
$$

Using a the similar calculation, we have that

$$
\begin{aligned}
&\hat{p}(\hat{\boldsymbol{\zeta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2) \\
&= S^{-1}\sum_{s=1}^{S}\{\alpha(\boldsymbol{\zeta}_3^{(s)}, \hat{\boldsymbol{\zeta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_4^{(s)}, \boldsymbol{\zeta}_5^{(s)}, \boldsymbol{y}^{*(s)}) \\
&\qquad \frac{\times q(\boldsymbol{\zeta}_3^{(s)}, \hat{\boldsymbol{\zeta}}_3|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_4^{(s)}, \boldsymbol{\zeta}_5^{(s)}, \boldsymbol{y}^{*(s)})\}}{G_6^{-1}\sum_{g_6=1}^{G_6}\alpha(\hat{\boldsymbol{\zeta}}_3, \boldsymbol{\zeta}_3^{(g_6)}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\zeta}}_2, \boldsymbol{\zeta}_4^{(g_6)}, \boldsymbol{\zeta}_5^{(g_6)}, \boldsymbol{y}^{*(g_6)})},
\end{aligned}
$$

$$
\begin{aligned}
&\hat{p}(\hat{\boldsymbol{\zeta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3) \\
&= S^{-1}\sum_{s=1}^{S}\{\alpha(\boldsymbol{\zeta}_4^{(s)}, \hat{\boldsymbol{\zeta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \boldsymbol{\zeta}_5^{(s)}, \boldsymbol{y}^{*(s)}) \\
&\qquad \frac{\times q(\boldsymbol{\zeta}_4^{(s)}, \hat{\boldsymbol{\zeta}}_4|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \boldsymbol{\zeta}_5^{(s)}, \boldsymbol{y}^{*(s)})\}}{G_7^{-1}\sum_{g_7=1}^{G_7}\alpha(\hat{\boldsymbol{\zeta}}_4, \boldsymbol{\zeta}_4^{(g_7)}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \boldsymbol{\zeta}_5^{(g_7)}, \boldsymbol{y}^{*(g_7)})},
\end{aligned}
$$

$$
\begin{aligned}
&\hat{p}(\hat{\boldsymbol{\zeta}}_5|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \hat{\boldsymbol{\zeta}}_4) \\
&= S^{-1}\sum_{s=1}^{S}\{\alpha(\boldsymbol{\zeta}_5^{(s)}, \hat{\boldsymbol{\zeta}}_5|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \hat{\boldsymbol{\zeta}}_4, \boldsymbol{y}^{*(s)}) \\
&\qquad \frac{\times q(\boldsymbol{\zeta}_5^{(s)}, \hat{\boldsymbol{\zeta}}_5|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\Sigma}^{-1}}, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \hat{\boldsymbol{\zeta}}_4, \boldsymbol{y}^{*(s)})\}}{G_8^{-1}\sum_{g_8=1}^{G_8}\alpha(\hat{\boldsymbol{\zeta}}_5, \boldsymbol{\zeta}_5^{(g_8)}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3, \hat{\boldsymbol{\zeta}}_4, \boldsymbol{y}^{*(g_8)})}.
\end{aligned}
$$

Appendix C

Derivation in Chapter 4

C.1 Derivation of Algorithm 1

C.1.1 Full Conditional Distribution of ϕ

The full conditional distribution of ϕ is

$$p(\phi|\cdots) \propto \exp\left[-\frac{1}{2}(\phi-\tilde{\phi})'\tilde{\omega}^{-1}(\phi-\tilde{\phi})\right] 1_{\{\phi\in(-1,1)\}}$$
$$\times \prod_{i=1}^{n}\prod_{t=1}^{T}\exp\left\{-\frac{1}{2}\sigma^{-2}\left[y^*_{it}-(\phi y^*_{i(t-1)}+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\boldsymbol{w}'_i\boldsymbol{\delta})\right]'\left[y^*_{it}-(\phi y^*_{i(t-1)}+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\boldsymbol{w}'_i\boldsymbol{\delta})\right]\right\}$$
$$\propto \exp\left\{-\frac{1}{2}\phi^2\left(\tilde{\omega}^{-1}+\sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T}y^{*2}_{i(t-1)}\right)\right\}$$
$$\times \exp\left\{\phi\left[\tilde{\omega}^{-1}\tilde{\phi}+\sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T}y^*_{i(t-1)}(y^*_{it}-\boldsymbol{x}'_{it}\boldsymbol{\beta}-\boldsymbol{w}'_i\boldsymbol{\delta})\right]\right\}.$$

Therefore, we have

$$\phi|\cdots \sim \mathrm{N}(\hat{\phi},\hat{\omega})1_{\{\phi\in(-1,1)\}},$$

where

$$\hat{\omega} = \left(\tilde{\omega}^{-1} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{i(t-1)}^{*2} \right)^{-1},$$

$$\hat{\phi} = \hat{\omega} \left[\tilde{\omega}^{-1} \tilde{\phi} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{i(t-1)}^{*} (y_{it}^{*} - \boldsymbol{x}_{it}' \boldsymbol{\beta} - \boldsymbol{w}_i' \boldsymbol{\delta}) \right].$$

C.1.2 Full Conditional Distribution of $\boldsymbol{\beta}$

The full conditional distribution of $\boldsymbol{\beta}$ is

$$p(\boldsymbol{\beta}|\cdots) \propto \exp\left[-\frac{1}{2} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' \tilde{\boldsymbol{B}}^{-1} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) \right]$$
$$\times \prod_{i=1}^{n} \prod_{t=1}^{T} \exp\left\{ -\frac{1}{2} \sigma^{-2} \left[y_{it}^{*} - (\phi y_{i(t-1)}^{*} + \boldsymbol{x}_{it}' \boldsymbol{\beta} + \boldsymbol{w}_i' \boldsymbol{\delta}) \right]' \left[y_{it}^{*} - (\phi y_{i(t-1)}^{*} + \boldsymbol{x}_{it}' \boldsymbol{\beta} + \boldsymbol{w}_i' \boldsymbol{\delta}) \right] \right\}$$
$$\propto \exp\left\{ -\frac{1}{2} \boldsymbol{\beta}' \left(\tilde{\boldsymbol{B}}^{-1} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \boldsymbol{x}_{it} \boldsymbol{x}_{it}' \right) \boldsymbol{\beta} \right\}$$
$$\times \exp\left\{ \boldsymbol{\beta}' \left[\tilde{\boldsymbol{B}}^{-1} \tilde{\boldsymbol{\beta}} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \boldsymbol{x}_{it} (y_{it}^{*} - \phi y_{i(t-1)}^{*} - \boldsymbol{w}_i' \boldsymbol{\delta}) \right] \right\}.$$

Therefore, we have

$$\boldsymbol{\beta}|\cdots \sim \mathrm{N}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{B}}),$$

where

$$\hat{\boldsymbol{B}} = \left(\tilde{\boldsymbol{B}}^{-1} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \boldsymbol{x}_{it} \boldsymbol{x}_{it}' \right)^{-1},$$

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{B}} \left[\tilde{\boldsymbol{B}}^{-1} \tilde{\boldsymbol{\beta}} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \boldsymbol{x}_{it} (y_{it}^{*} - \phi y_{i(t-1)}^{*} - \boldsymbol{w}_i' \boldsymbol{\delta}) \right].$$

C.1.3 Full Conditional Distribution of $\boldsymbol{\delta}$

The full conditional distribution of $\boldsymbol{\delta}$ is

$$
\begin{aligned}
p(\boldsymbol{\delta}|\cdots) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})'\tilde{\boldsymbol{D}}^{-1}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}})\right] \\
\times \prod_{i=1}^{n}\prod_{t=1}^{T} &\exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{it}^{*}-(\phi y_{i(t-1)}^{*}+\boldsymbol{x}_{it}'\boldsymbol{\beta}+\boldsymbol{w}_i'\boldsymbol{\delta})\right]'\left[y_{it}^{*}-(\phi y_{i(t-1)}^{*}+\boldsymbol{x}_{it}'\boldsymbol{\beta}+\boldsymbol{w}_i'\boldsymbol{\delta})\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\delta}'\left(\tilde{\boldsymbol{D}}^{-1}+\sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T}\boldsymbol{w}_i\boldsymbol{w}_i'\right)\boldsymbol{\delta}\right\} \\
&\propto \exp\left\{\boldsymbol{\delta}'\left[\tilde{\boldsymbol{D}}^{-1}\tilde{\boldsymbol{\delta}}+\sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T}\boldsymbol{w}_i(y_{it}^{*}-\phi y_{i(t-1)}^{*}-\boldsymbol{x}_{it}'\boldsymbol{\beta})\right]\right\}.
\end{aligned}
$$

Therefore, we have

$$\boldsymbol{\delta}|\cdots \sim \mathrm{N}(\hat{\boldsymbol{\delta}},\hat{\boldsymbol{D}}),$$

where

$$
\begin{aligned}
\hat{\boldsymbol{D}} &= \left(\tilde{\boldsymbol{D}}^{-1}+\sigma^{-2}T\sum_{i=1}^{n}\boldsymbol{w}_i\boldsymbol{w}_i'\right)^{-1}, \\
\hat{\boldsymbol{\delta}} &= \hat{\boldsymbol{D}}\left[\tilde{\boldsymbol{D}}^{-1}\tilde{\boldsymbol{\delta}}+\sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T}\boldsymbol{w}_i(y_{it}^{*}-\phi y_{i(t-1)}^{*}-\boldsymbol{x}_{it}'\boldsymbol{\beta})\right].
\end{aligned}
$$

C.1.4 Full Conditional Distribution of $\boldsymbol{\beta}_0$

The full conditional distribution of $\boldsymbol{\beta}_0$ is

$$\begin{aligned}
p(\boldsymbol{\beta}_0|\cdots) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_0-\tilde{\boldsymbol{\beta}}_0)'\tilde{\boldsymbol{B}}_0^{-1}(\boldsymbol{\beta}_0-\tilde{\boldsymbol{\beta}}_0)\right] \\
&\times \prod_{i=1}^{n}\exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{i0}^*-(\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0+\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]'\left[y_{i0}^*-(\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0+\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}_0'\left(\tilde{\boldsymbol{B}}_0^{-1}+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{x}_{i0}\boldsymbol{x}_{i0}'\right)\boldsymbol{\beta}_0+\boldsymbol{\beta}_0'\left[\tilde{\boldsymbol{B}}_0^{-1}\tilde{\boldsymbol{\beta}}_0+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{x}_{i0}(y_{i0}^*-\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\}.
\end{aligned}$$

Therefore, we have

$$\boldsymbol{\beta}_0|\cdots\sim \mathrm{N}(\hat{\boldsymbol{\beta}}_0,\hat{\boldsymbol{B}}_0),$$

where

$$\begin{aligned}
\hat{\boldsymbol{B}}_0 &= \left(\tilde{\boldsymbol{B}}_0^{-1}+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{x}_{i0}\boldsymbol{x}_{i0}'\right)^{-1}, \\
\hat{\boldsymbol{\beta}}_0 &= \hat{\boldsymbol{B}}_0\left[\tilde{\boldsymbol{B}}_0^{-1}\tilde{\boldsymbol{\beta}}_0+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{x}_{i0}(y_{i0}^*-\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right].
\end{aligned}$$

C.1.5 Full Conditional Distribution of $\boldsymbol{\delta}_0$

The full conditional distribution of $\boldsymbol{\delta}_0$ is

$$\begin{aligned}
p(\boldsymbol{\delta}_0|\cdots) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{\delta}_0-\tilde{\boldsymbol{\delta}}_0)'\tilde{\boldsymbol{D}}_0^{-1}(\boldsymbol{\delta}_0-\tilde{\boldsymbol{\delta}}_0)\right] \\
&\times \prod_{i=1}^{n}\exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{i0}^*-(\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0+\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]'\left[y_{i0}^*-(\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0+\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\delta}_0'\left(\tilde{\boldsymbol{D}}_0^{-1}+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{w}_i\boldsymbol{w}_i'\right)\boldsymbol{\delta}_0+\boldsymbol{\delta}_0'\left[\tilde{\boldsymbol{D}}_0^{-1}\tilde{\boldsymbol{\delta}}_0+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{w}_i(y_{i0}^*-\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0)\right]\right\}.
\end{aligned}$$

Therefore, we have

$$\boldsymbol{\delta}_0|\cdots \sim \mathrm{N}(\hat{\boldsymbol{\delta}}_0, \hat{\boldsymbol{D}}_0),$$

where

$$\hat{\boldsymbol{D}}_0 = \left(\tilde{\boldsymbol{D}}_0^{-1} + \sigma^{-2}\sum_{i=1}^{n} \boldsymbol{w}_i \boldsymbol{w}_i'\right)^{-1},$$
$$\hat{\boldsymbol{\delta}}_0 = \hat{\boldsymbol{D}}_0 \left[\tilde{\boldsymbol{D}}_0^{-1}\tilde{\boldsymbol{\delta}}_0 + \sigma^{-2}\sum_{i=1}^{n} \boldsymbol{w}_i (y_{i0}^* - \boldsymbol{x}_{i0}'\boldsymbol{\beta}_0)\right].$$

C.1.6 Full Conditional Distribution of σ^2

The full conditional distribution of σ^2 is

$$p(\sigma^{-2}|\cdots) \propto (\sigma^2)^{-\tilde{c}-1}\exp(-\frac{\tilde{d}}{\sigma^2})$$
$$\times \prod_{i=1}^{n}\prod_{t=1}^{T} \exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{it}^* - (\phi y_{i(t-1)}^* + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]'\left[y_{it}^* - (\phi y_{i(t-1)}^* + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]\right\}$$
$$\times \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{i0}^* - (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]'\left[y_{i0}^* - (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\}$$
$$\propto (\sigma^2)^{-\{\tilde{c}+\frac{n}{2}(T+1)\}-1}$$
$$\times \exp\left[-\sigma^{-2}\{\tilde{d} + \frac{1}{2}[\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}^* - \phi y_{i(t-1)}^* - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{\delta})^2 + \sum_{i=1}^{n}(y_{i0}^* - \boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 - \boldsymbol{w}_i'\boldsymbol{\delta}_0)^2]\}\right].$$

Therefore, we have

$$\sigma^{-2}|\cdots \sim \mathrm{gamma}(\hat{c}, \hat{d}),$$

where

$$\hat{c} = \tilde{c} + \frac{n}{2}(T+1),$$

$$\hat{d} = \tilde{d} + \frac{1}{2}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}^{*} - \phi y_{i(t-1)}^{*} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{\delta})^2 + \sum_{i=1}^{n}(y_{i0}^{*} - \boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 - \boldsymbol{w}_i'\boldsymbol{\delta}_0)^2\right].$$

C.1.7 Full Conditional Distribution of $\boldsymbol{y}_0^*$

The full conditional distribution of y_{i0}^* is

$$\begin{aligned} p(y_{i0}^*|\cdots) &\propto \exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{i1}^* - (\phi y_{i0}^* + \boldsymbol{x}_{i1}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]'\left[y_{i1}^* - (\phi y_{i0}^* + \boldsymbol{x}_{i1}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]\right\} \\ &\times \exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{i0}^* - (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]'\left[y_{i0}^* - (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\} 1_{\{y_{i0}^*\in(\gamma_{i(j-1)},\gamma_{ij}]\}} \\ &\propto \exp\left\{-\frac{1}{2}(\frac{1+\phi^2}{\sigma^2})y_{i0}^{*2}\right\} \\ &\times \exp\left\{y_{i0}^*\sigma^{-2}\left[\phi(y_{i1}^* - \boldsymbol{x}_{i1}'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{\delta}) + (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\} 1_{\{y_{i0}^*\in(\gamma_{i(j-1)},\gamma_{ij}]\}}. \end{aligned}$$

Therefore, we have

$$y_{i0}^*|\cdots \sim \mathrm{N}(\hat{y}_{i0}^*, \frac{\sigma^2}{1+\phi^2})1_{\{y_{i0}^*\in(\gamma_{i(j-1)},\gamma_{ij}]\}},$$

where

$$\hat{y}_{i0}^* = \frac{\phi(y_{i1}^* - \boldsymbol{x}_{i1}'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{\delta}) + (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)}{1+\phi^2}.$$

C.1.8 Full Conditional Distribution of $\boldsymbol{y}^*$

The full conditional distribution of y^*_{it} is

$$p(y^*_{it}|\cdots) \propto \exp\left\{-\frac{1}{2}\sigma^2[y^*_{i(t+1)} - (\phi y^*_{it} + \boldsymbol{x}'_{i(t+1)}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})]'[y^*_{i(t+1)} - (\phi y^*_{it} + \boldsymbol{x}'_{i(t+1)}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})]\right\}$$
$$\times \exp\left\{-\frac{1}{2}\sigma^2[y^*_{it} - (\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})]'[y^*_{it} - (\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})]\right\}$$
$$\times 1_{\{y^*_{it}\in(\gamma_{i(j-1)},\gamma_{ij}]\}}$$
$$\propto \exp\left\{-\frac{1}{2}(\frac{1+\phi^2}{\sigma^2})y^{*2}_{it}\right\} \times 1_{\{y^*_{it}\in(\gamma_{i(j-1)},\gamma_{ij}]\}}$$
$$\times \exp\left\{y^*_{it}\sigma^{-2}\left[(\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta}) + \phi(y^*_{i(t+1)} - \boldsymbol{x}'_{i(t+1)}\boldsymbol{\beta} - \boldsymbol{w}'_i\boldsymbol{\delta})\right]\right\}.$$

Therefore, we have

$$y^*_{it}|\cdots \sim \mathrm{N}(\hat{y}^*_{it}, \frac{\sigma^2}{1+\phi^2})1_{\{y^*_{it}\in(\gamma_{i(j-1)},\gamma_{ij}]\}},\ t=1,\ldots,T-1,\ i=1,\ldots,n,$$

where

$$\hat{y}^*_{it} = \frac{(\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta}) + \phi(y^*_{i(t+1)} - \boldsymbol{x}'_{i(t+1)}\boldsymbol{\beta} - \boldsymbol{w}'_i\boldsymbol{\delta})}{1+\phi^2},$$

and

$$y^*_{iT}|\cdots \sim \mathrm{N}(\phi y^*_{i(T-1)} + \boldsymbol{x}'_{iT}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta}, \sigma^2)1_{\{y^*_{iT}\in(\gamma_{i(j-1)},\gamma_{ij}]\}}.$$

C.1.9 Full Conditional Distribution of $\boldsymbol{\zeta}$

The full conditional distribution of $\boldsymbol{\nu}$ is

$$
\begin{aligned}
p(\boldsymbol{\nu}|\cdots) &= p(\boldsymbol{\nu})p(\boldsymbol{\gamma}|\cdots)\prod_{j=2}^{J-2}\frac{(1-\gamma_{j-1})\exp(\nu_j)}{(1+\exp(\nu_j))^2} \\
&\propto \exp\left[-\frac{1}{2}(\boldsymbol{\nu}-\boldsymbol{\nu}_0)'\boldsymbol{V}_0^{-1}(\boldsymbol{\nu}-\boldsymbol{\nu}_0)\right] \\
&\times \prod_{i=1}^{n}\prod_{t=0}^{T}\left[\Phi(\frac{\gamma_2-\tilde{y^*}_{it}}{\sigma})-\Phi(-\frac{\tilde{y^*}_{it}}{\sigma})\right]^{1(y_{it}=2)} \\
&\times \prod_{i=1}^{n}\prod_{t=0}^{T}\left[\Phi(\frac{\gamma_3-\tilde{y^*}_{it}}{\sigma})-\Phi(\frac{\gamma_2-\tilde{y^*}_{it}}{\sigma})\right]^{1(y_{it}=3)} \\
&\times\cdots\times \prod_{i=1}^{n}\prod_{t=0}^{T}\left[\Phi(\frac{1-\tilde{y^*}_{it}}{\sigma})-\Phi(\frac{\gamma_{J-2}-\tilde{y^*}_{it}}{\sigma})\right]^{1(y_{it}=J-1)} \\
&\times \prod_{j=2}^{J-2}\frac{(1-\gamma_{j-1})\exp(\nu_j)}{(1+\exp(\nu_j))^2},
\end{aligned}
$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution and

$$
\begin{cases}
\tilde{y^*}_{i0} = \boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0,\ t=0, \\
\tilde{y^*}_{it} = \phi y^*_{i(t-1)} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta},\ t=1,\ldots,T.
\end{cases}
$$

Let ν_j^s denote the value of ν_j at the sth iteration. Using the Metropolis-Hastings algorithm, we can sample ν_j by drawing $\nu_j^{s+1} \sim q(\nu_j|\cdots)$ from a Student-t proposal density $q(\nu_j|\cdots) = f_T(\nu_j|\hat{\nu}_j, \hat{\boldsymbol{G}}_j, \boldsymbol{n})$, where $\boldsymbol{n}$ is the degrees of freedom, and

$$\hat{\nu}_j = \arg\max_{\nu_j} \log p(\nu_j|\cdots)$$
$$\hat{\boldsymbol{G}}_j = \left\{ \left[-\frac{\partial^2 \log p(\nu_j|\cdots)}{\partial \nu_j \partial \nu_j'} \right]_{\nu_j=\hat{\nu}_j} \right\}^{-1}.$$

Return ν_j^{s+1}, with probability of $\min\left\{1, \dfrac{p(\nu_j^{s+1}|\cdots)}{p(\nu_j^{s}|\cdots)} \dfrac{f_T(\nu_j^s|\hat{\nu}_j, \hat{\boldsymbol{G}}_j, \boldsymbol{n})}{f_T(\tilde{\nu}_j^{s+1}|\hat{\nu}_j, \hat{\boldsymbol{G}}_j, \boldsymbol{n})}\right\}$; otherwise, repeat using the previous value of ν_j^s.

Then, γ_j can be calculated using the following equation:

$$\gamma_j = \frac{\gamma_{j-1} + \exp(\nu_j)}{1 + \exp(\nu_j)}.$$

C.2 Derivation of Algorithm 2

C.2.1 Full Conditional Distribution of ϕ

The full conditional distribution of ϕ is

$$\begin{aligned}
p(\phi|\cdots) &\propto \exp\left[-\frac{1}{2}(\phi - \tilde{\phi})'\tilde{\omega}^{-1}(\phi - \tilde{\phi})\right] 1_{\{\phi\in(-1,1)\}} \\
&\times \prod_{i=1}^{n}\prod_{t=1}^{T} \exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{it}^* - (\phi y_{i(t-1)}^* + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]'\left[y_{it}^* - (\phi y_{i(t-1)}^* + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\phi^2\left(\tilde{\omega}^{-1} + \sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{i(t-1)}^{*2}\right)\right\} \\
&\times \exp\left\{\phi\left[\tilde{\omega}^{-1}\tilde{\phi} + \sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{i(t-1)}^*(y_{it}^* - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{\delta})\right]\right\}.
\end{aligned}$$

Therefore, we have

$$\phi|\cdots \sim \mathrm{N}(\hat{\phi}, \hat{\omega}) 1_{\{\phi\in(-1,1)\}},$$

where

$$\hat{\omega} = \left(\tilde{\omega}^{-1} + \sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{i(t-1)}^{*2}\right)^{-1},$$
$$\hat{\phi} = \hat{\omega}\left[\tilde{\omega}^{-1}\tilde{\phi} + \sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{i(t-1)}^{*}(y_{it}^{*} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{\delta})\right].$$

C.2.2 Full Conditional Distribution of $\boldsymbol{\beta}$

The full conditional distribution of $\boldsymbol{\beta}$ is

$$\begin{aligned}
p(\boldsymbol{\beta}|\cdots) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})'\tilde{\boldsymbol{B}}^{-1}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})\right] \\
&\times \prod_{i=1}^{n}\prod_{t=1}^{T}\exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{it}^{*} - (\phi y_{i(t-1)}^{*} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]'\left[y_{it}^{*} - (\phi y_{i(t-1)}^{*} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}'\left(\tilde{\boldsymbol{B}}^{-1} + \sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)\boldsymbol{\beta}\right\} \\
&\times \exp\left\{\boldsymbol{\beta}'\left[\tilde{\boldsymbol{B}}^{-1}\tilde{\boldsymbol{\beta}} + \sigma^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T}\boldsymbol{x}_{it}(y_{it}^{*} - \phi y_{i(t-1)}^{*} - \boldsymbol{w}_i'\boldsymbol{\delta})\right]\right\}.
\end{aligned}$$

Therefore, we have

$$\boldsymbol{\beta}|\cdots \sim \mathrm{N}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{B}}),$$

where

$$\hat{\boldsymbol{B}} = \left(\tilde{\boldsymbol{B}}^{-1} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \boldsymbol{x}_{it} \boldsymbol{x}_{it}' \right)^{-1},$$

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{B}} \left[\tilde{\boldsymbol{B}}^{-1} \tilde{\boldsymbol{\beta}} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \boldsymbol{x}_{it} (y_{it}^* - \phi y_{i(t-1)}^* - \boldsymbol{w}_i' \boldsymbol{\delta}) \right].$$

C.2.3 Full Conditional Distribution of $\boldsymbol{\delta}$

The full conditional distribution of $\boldsymbol{\delta}$ is

$$\begin{aligned} p(\boldsymbol{\delta}|\cdots) &\propto \exp\left[-\frac{1}{2} (\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}})' \tilde{\boldsymbol{D}}^{-1} (\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}) \right] \\ &\times \prod_{i=1}^{n} \prod_{t=1}^{T} \exp\left\{ -\frac{1}{2} \sigma^{-2} \left[y_{it}^* - (\phi y_{i(t-1)}^* + \boldsymbol{x}_{it}' \boldsymbol{\beta} + \boldsymbol{w}_i' \boldsymbol{\delta}) \right]' \left[y_{it}^* - (\phi y_{i(t-1)}^* + \boldsymbol{x}_{it}' \boldsymbol{\beta} + \boldsymbol{w}_i' \boldsymbol{\delta}) \right] \right\} \\ &\propto \exp\left\{ -\frac{1}{2} \boldsymbol{\delta}' \left(\tilde{\boldsymbol{D}}^{-1} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \boldsymbol{w}_i \boldsymbol{w}_i' \right) \boldsymbol{\delta} \right\} \\ &\propto \exp\left\{ \boldsymbol{\delta}' \left[\tilde{\boldsymbol{D}}^{-1} \tilde{\boldsymbol{\delta}} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \boldsymbol{w}_i (y_{it}^* - \phi y_{i(t-1)}^* - \boldsymbol{x}_{it}' \boldsymbol{\beta}) \right] \right\}. \end{aligned}$$

Therefore, we have

$$\boldsymbol{\delta}|\cdots \sim \mathrm{N}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{D}}),$$

where

$$\hat{\boldsymbol{D}} = \left(\tilde{\boldsymbol{D}}^{-1} + \sigma^{-2} T \sum_{i=1}^{n} \boldsymbol{w}_i \boldsymbol{w}_i' \right)^{-1},$$

$$\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{D}} \left[\tilde{\boldsymbol{D}}^{-1} \tilde{\boldsymbol{\delta}} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \boldsymbol{w}_i (y_{it}^* - \phi y_{i(t-1)}^* - \boldsymbol{x}_{it}' \boldsymbol{\beta}) \right].$$

C.2.4 Full Conditional Distribution of $\boldsymbol{\beta}_0$

The full conditional distribution of $\boldsymbol{\beta}_0$ is

$$\begin{aligned}
p(\boldsymbol{\beta}_0|\cdots) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_0-\tilde{\boldsymbol{\beta}}_0)'\tilde{\boldsymbol{B}}_0^{-1}(\boldsymbol{\beta}_0-\tilde{\boldsymbol{\beta}}_0)\right] \\
&\times\prod_{i=1}^{n}\exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{i0}^{*}-(\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0+\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]'\left[y_{i0}^{*}-(\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0+\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}_0'\left(\tilde{\boldsymbol{B}}_0^{-1}+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{x}_{i0}\boldsymbol{x}_{i0}'\right)\boldsymbol{\beta}_0+\boldsymbol{\beta}_0'\left[\tilde{\boldsymbol{B}}_0^{-1}\tilde{\boldsymbol{\beta}}_0+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{x}_{i0}(y_{i0}^{*}-\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\}.
\end{aligned}$$

Therefore, we have

$$\boldsymbol{\beta}_0|\cdots\sim \mathrm{N}(\hat{\boldsymbol{\beta}}_0,\hat{\boldsymbol{B}}_0),$$

where

$$\begin{aligned}
\hat{\boldsymbol{B}}_0 &= \left(\tilde{\boldsymbol{B}}_0^{-1}+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{x}_{i0}\boldsymbol{x}_{i0}'\right)^{-1}, \\
\hat{\boldsymbol{\beta}}_0 &= \hat{\boldsymbol{B}}_0\left[\tilde{\boldsymbol{B}}_0^{-1}\tilde{\boldsymbol{\beta}}_0+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{x}_{i0}(y_{i0}^{*}-\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right].
\end{aligned}$$

C.2.5 Full Conditional Distribution of $\boldsymbol{\delta}_0$

The full conditional distribution of $\boldsymbol{\delta}_0$ is

$$\begin{aligned}
p(\boldsymbol{\delta}_0|\cdots) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{\delta}_0-\tilde{\boldsymbol{\delta}}_0)'\tilde{\boldsymbol{D}}_0^{-1}(\boldsymbol{\delta}_0-\tilde{\boldsymbol{\delta}}_0)\right] \\
&\times\prod_{i=1}^{n}\exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{i0}^{*}-(\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0+\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]'\left[y_{i0}^{*}-(\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0+\boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\delta}_0'\left(\tilde{\boldsymbol{D}}_0^{-1}+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{w}_i\boldsymbol{w}_i'\right)\boldsymbol{\delta}_0+\boldsymbol{\delta}_0'\left[\tilde{\boldsymbol{D}}_0^{-1}\tilde{\boldsymbol{\delta}}_0+\sigma^{-2}\sum_{i=1}^{n}\boldsymbol{w}_i(y_{i0}^{*}-\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0)\right]\right\}.
\end{aligned}$$

Therefore, we have

$$\boldsymbol{\delta}_0 | \cdots \sim \mathrm{N}(\hat{\boldsymbol{\delta}}_0, \hat{\boldsymbol{D}}_0),$$

where

$$\hat{\boldsymbol{D}}_0 = \left(\tilde{\boldsymbol{D}}_0^{-1} + \sigma^{-2} \sum_{i=1}^{n} \boldsymbol{w}_i \boldsymbol{w}_i' \right)^{-1},$$

$$\hat{\boldsymbol{\delta}}_0 = \hat{\boldsymbol{D}}_0 \left[\tilde{\boldsymbol{D}}_0^{-1} \tilde{\boldsymbol{\delta}}_0 + \sigma^{-2} \sum_{i=1}^{n} \boldsymbol{w}_i (y_{i0}^* - \boldsymbol{x}_{i0}' \boldsymbol{\beta}_0) \right].$$

C.2.6 Full Conditional Distribution of σ^2

The full conditional distribution of σ^2 is

$$p(\sigma^{-2} | \cdots) \propto (\sigma^2)^{-\tilde{c}-1} \exp(-\frac{\tilde{d}}{\sigma^2})$$
$$\times \prod_{i=1}^{n} \prod_{t=1}^{T} \exp\left\{ -\frac{1}{2} \sigma^{-2} \left[y_{it}^* - (\phi y_{i(t-1)}^* + \boldsymbol{x}_{it}' \boldsymbol{\beta} + \boldsymbol{w}_i' \boldsymbol{\delta}) \right]' \left[y_{it}^* - (\phi y_{i(t-1)}^* + \boldsymbol{x}_{it}' \boldsymbol{\beta} + \boldsymbol{w}_i' \boldsymbol{\delta}) \right] \right\}$$
$$\times \prod_{i=1}^{n} \exp\left\{ -\frac{1}{2} \sigma^{-2} \left[y_{i0}^* - (\boldsymbol{x}_{i0}' \boldsymbol{\beta}_0 + \boldsymbol{w}_i' \boldsymbol{\delta}_0) \right]' \left[y_{i0}^* - (\boldsymbol{x}_{i0}' \boldsymbol{\beta}_0 + \boldsymbol{w}_i' \boldsymbol{\delta}_0) \right] \right\}$$
$$\propto (\sigma^2)^{-\{\tilde{c} + \frac{n}{2}(T+1)\}-1}$$
$$\times \exp\left[-\sigma^{-2} \{ \tilde{d} + \frac{1}{2} [\sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it}^* - \phi y_{i(t-1)}^* - \boldsymbol{x}_{it}' \boldsymbol{\beta} - \boldsymbol{w}_i' \boldsymbol{\delta})^2 + \sum_{i=1}^{n} (y_{i0}^* - \boldsymbol{x}_{i0}' \boldsymbol{\beta}_0 - \boldsymbol{w}_i' \boldsymbol{\delta}_0)^2] \} \right].$$

Therefore, we have

$$\sigma^{-2} | \cdots \sim \mathrm{gamma}(\hat{c}, \hat{d}),$$

where

$$\hat{c} = \tilde{c} + \frac{n}{2}(T+1),$$

$$\hat{d} = \tilde{d} + \frac{1}{2}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}^{*} - \phi y_{i(t-1)}^{*} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{\delta})^2 + \sum_{i=1}^{n}(y_{i0}^{*} - \boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 - \boldsymbol{w}_i'\boldsymbol{\delta}_0)^2\right].$$

C.2.7 Full Conditional Distribution of $\boldsymbol{y}_0^*$

The full conditional distribution of y_{i0}^* is

$$\begin{aligned}
p(y_{i0}^*|\cdots) \propto \exp&\left\{-\frac{1}{2}\sigma^{-2}\left[y_{i1}^* - (\phi y_{i0}^* + \boldsymbol{x}_{i1}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]'\left[y_{i1}^* - (\phi y_{i0}^* + \boldsymbol{x}_{i1}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta})\right]\right\} \\
&\times \exp\left\{-\frac{1}{2}\sigma^{-2}\left[y_{i0}^* - (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]'\left[y_{i0}^* - (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\} 1_{\{y_{i0}^*\in(\gamma_{i(j-1)},\gamma_{ij}]\}} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1+\phi^2}{\sigma^2}\right)y_{i0}^{*2}\right\} \\
&\times \exp\left\{y_{i0}^*\sigma^{-2}\left[\phi(y_{i1}^* - \boldsymbol{x}_{i1}'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{\delta}) + (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)\right]\right\} 1_{\{y_{i0}^*\in(\gamma_{i(j-1)},\gamma_{ij}]\}}.
\end{aligned}$$

Therefore, we have

$$y_{i0}^*|\cdots \sim \mathrm{N}\left(\hat{y}_{i0}^*, \frac{\sigma^2}{1+\phi^2}\right)1_{\{y_{i0}^*\in(\gamma_{i(j-1)},\gamma_{ij}]\}},$$

where

$$\hat{y}_{i0}^* = \frac{\phi(y_{i1}^* - \boldsymbol{x}_{i1}'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{\delta}) + (\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0)}{1+\phi^2}.$$

C.2.8 Full Conditional Distribution of $\boldsymbol{y}^*$

The full conditional distribution of y^*_{it} is

$$p(y^*_{it}|\cdots) \propto \exp\left\{-\frac{1}{2}\sigma^2[y^*_{i(t+1)} - (\phi y^*_{it} + \boldsymbol{x}'_{i(t+1)}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})]'[y^*_{i(t+1)} - (\phi y^*_{it} + \boldsymbol{x}'_{i(t+1)}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})]\right\}$$
$$\times \exp\left\{-\frac{1}{2}\sigma^2[y^*_{it} - (\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})]'[y^*_{it} - (\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})]\right\}$$
$$\times 1_{\{y^*_{it}\in(\gamma_{i(j-1)},\gamma_{ij}]\}}$$
$$\propto \exp\left\{-\frac{1}{2}(\frac{1+\phi^2}{\sigma^2})y^{*2}_{it}\right\} \times 1_{\{y^*_{it}\in(\gamma_{i(j-1)},\gamma_{ij}]\}}$$
$$\times \exp\left\{y^*_{it}\sigma^{-2}\left[(\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta}) + \phi(y^*_{i(t+1)} - \boldsymbol{x}'_{i(t+1)}\boldsymbol{\beta} - \boldsymbol{w}'_i\boldsymbol{\delta})\right]\right\}.$$

Therefore, we have

$$y^*_{it}|\cdots \sim \mathrm{N}(\hat{y}^*_{it}, \frac{\sigma^2}{1+\phi^2})1_{\{y^*_{it}\in(\gamma_{i(j-1)},\gamma_{ij}]\}},\ t=1,\ldots,T-1,\ i=1,\ldots,n,$$

where

$$\hat{y}^*_{it} = \frac{(\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta}) + \phi(y^*_{i(t+1)} - \boldsymbol{x}'_{i(t+1)}\boldsymbol{\beta} - \boldsymbol{w}'_i\boldsymbol{\delta})}{1+\phi^2},$$

and

$$y^*_{iT}|\cdots \sim \mathrm{N}(\phi y^*_{i(T-1)} + \boldsymbol{x}'_{iT}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta}, \sigma^2)1_{\{y^*_{iT}\in(\gamma_{i(j-1)},\gamma_{ij}]\}}.$$

C.2.9 Full Conditional Distribution of $\boldsymbol{\zeta}$

In algorithm 2, the cutpoints are specified as follows:

$$\gamma_{ij} = \frac{\gamma_{i(j-1)} + \exp(\boldsymbol{\zeta}_j' \boldsymbol{z}_i)}{1 + \exp(\boldsymbol{\zeta}_j' \boldsymbol{z}_i)}.$$

Since $\boldsymbol{\zeta}_j \sim \mathrm{N}(\tilde{\boldsymbol{\zeta}}, \tilde{\boldsymbol{G}})$, the full conditional distribution of $\boldsymbol{\zeta}_j$ is

$$\begin{aligned} p(\boldsymbol{\zeta}_j|\cdots) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{\zeta}_j - \tilde{\boldsymbol{\zeta}})'\tilde{\boldsymbol{G}}^{-1}(\boldsymbol{\zeta}_j - \tilde{\boldsymbol{\zeta}})\right] \\ &\times \prod_{i=1}^{n}\prod_{t=0}^{T}\left[\Phi(\frac{\gamma_{i2} - \tilde{y^*}_{it}}{\sigma}) - \Phi(-\frac{\tilde{y^*}_{it}}{\sigma})\right]^{1(y_{it}=2)} \\ &\times \prod_{i=1}^{n}\prod_{t=0}^{T}\left[\Phi(\frac{\gamma_{i3} - \tilde{y^*}_{it}}{\sigma}) - \Phi(\frac{\gamma_{i2} - \tilde{y^*}_{it}}{\sigma})\right]^{1(y_{it}=3)} \\ &\times \cdots \times \prod_{i=1}^{n}\prod_{t=0}^{T}\left[\Phi(\frac{1 - \tilde{y^*}_{it}}{\sigma}) - \Phi(\frac{\gamma_{i(J-2)} - \tilde{y^*}_{it}}{\sigma})\right]^{1(y_{it}=J-1)}, \end{aligned}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution and

$$\begin{cases} \tilde{y^*}_{i0} = \boldsymbol{x}_{i0}'\boldsymbol{\beta}_0 + \boldsymbol{w}_i'\boldsymbol{\delta}_0, \ t = 0, \\ \tilde{y^*}_{it} = \phi y^*_{i(t-1)} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{w}_i'\boldsymbol{\delta}, \ t = 1, \ldots, T. \end{cases}$$

Let $\boldsymbol{\zeta}_j^s$ denote the value of $\boldsymbol{\zeta}_j$ at the sth iteration. Using the Metropolis-Hastings algorithm, we can sample $\boldsymbol{\zeta}_j$ by drawing $\boldsymbol{\zeta}_j^{s+1} \sim q(\boldsymbol{\zeta}_j|\cdots)$ from a Student-t proposal density $q(\boldsymbol{\zeta}_j|\cdots) = f_T(\boldsymbol{\zeta}_j|\hat{\boldsymbol{\zeta}}_j, \hat{\boldsymbol{G}}_j, \boldsymbol{n})$, where $\boldsymbol{n}$ is the degrees of freedom, and

$$\hat{\boldsymbol{\zeta}}_j = \arg\max_{\boldsymbol{\zeta}_j} \log p(\boldsymbol{\zeta}_j|\cdots)$$
$$\hat{\boldsymbol{G}}_j = \left\{ \left[-\frac{\partial^2 \log p(\boldsymbol{\zeta}_j|\cdots)}{\partial\boldsymbol{\zeta}_j \partial\boldsymbol{\zeta}_j'} \right]_{\zeta_j=\hat{\zeta}_j} \right\}^{-1}.$$

Return $\boldsymbol{\zeta}_j^{s+1}$, with probability of $\min\left\{1, \frac{p(\boldsymbol{\zeta}_j^{s+1}|\cdots)}{p(\boldsymbol{\zeta}_j^{s}|\cdots)} \frac{f_T(\boldsymbol{\zeta}_j^s|\hat{\boldsymbol{\zeta}}_j, \hat{\boldsymbol{G}}_j, \boldsymbol{n})}{f_T(\tilde{\boldsymbol{\zeta}}_j^{s+1}|\hat{\boldsymbol{\zeta}}_j, \hat{\boldsymbol{G}}_j, \boldsymbol{n})}\right\}$; otherwise, repeat using the previous value of $\boldsymbol{\zeta}_j^s$.

Then, γ_{ij} can be calculated using the following equation:

$$\gamma_{ij} = \frac{\gamma_{i(j-1)} + \exp(\boldsymbol{\zeta}_j' \boldsymbol{z}_i)}{1 + \exp(\boldsymbol{\zeta}_j' \boldsymbol{z}_i)}.$$

C.3 The Partial Effects of the Explanatory Variables on the Response Probabilities

C.3.1 The Partial Effects of Algorithm 2

The probability of observing $y_{it} = j$ can be wrriten as

$$p(y_{it} = j|\ldots) = p(\gamma_{i(j-1)} < y^*_{it} \leq \gamma_{ij})$$
$$\begin{cases} = \Phi(\frac{-\tilde{y^*}_{it}}{\sigma}),\ j = 1 \\ = \Phi(\frac{\gamma_{ij}-\tilde{y^*}_{it}}{\sigma}) - \Phi(\frac{\gamma_{i(j-1)}-\tilde{y^*}_{it}}{\sigma}),\ j = 2, \ldots, J-1 \\ = 1 - \Phi(\frac{1-\tilde{y^*}_{it}}{\sigma}),\ j = J \end{cases} \tag{C.1}$$

where $t = 0, 1, \ldots, T$, $i = 1, \ldots, n$ and

$$\begin{cases} \tilde{y^*}_{i0} = \boldsymbol{x}'_{i0}\boldsymbol{\beta}_0 + \boldsymbol{w}'_i\boldsymbol{\delta}_0, \ t = 0, \\ \tilde{y^*}_{it} = \phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta}, \ t = 1, \ldots, T. \end{cases}$$

When computing the partial effects, it is necessary to take the increments of cut points into account. If an explanatory variable x_k is continuous, we can calculate the partial effects by differentiating (C.1):

$$\frac{\partial p(y_{it} = j | \ldots)}{\partial x_k} = \begin{cases} -\frac{\beta_k}{\sigma} f(-\frac{1}{\sigma}(\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})), \ j = 1 \\ \frac{1}{\sigma} f(\frac{1}{\sigma}\{\gamma_{ij} - (\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})\})(\frac{\partial \gamma_{ij}}{\partial x_k} - \beta_k) \\ -\frac{1}{\sigma} f(\frac{1}{\sigma}\{\gamma_{i(j-1)} - (\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})\})(\frac{\partial \gamma_{i(j-1)}}{\partial x_k} - \beta_k), \ j = 2, \ldots, J-1 \\ \frac{\beta_k}{\sigma} f(\frac{1}{\sigma}\{1 - (\phi y^*_{i(t-1)} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta})\}), \ j = J \end{cases}$$

$$t = 1, \ldots, T, \ i = 1, \ldots, n,$$

and

$$\frac{\partial p(y_{i0} = j | \ldots)}{\partial x_k} = \begin{cases} -\frac{\beta_{0k}}{\sigma} f(-\frac{1}{\sigma}(\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0 + \boldsymbol{w}'_i\boldsymbol{\delta}_0)), \ j = 1 \\ \frac{1}{\sigma} f(\frac{1}{\sigma}\{\gamma_{i(j-1)} - (\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0 + \boldsymbol{w}'_i\boldsymbol{\delta}_0)\})(\frac{\partial \gamma_{ij}}{\partial x_k} - \beta_{0k}) \\ -\frac{1}{\sigma} f(\frac{1}{\sigma}\{\gamma_{ij} - (\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0 + \boldsymbol{w}'_i\boldsymbol{\delta}_0)\})(\frac{\partial \gamma_{i(j-1)}}{\partial x_k} - \beta_{0k}), \ j = 2, \ldots, J-1 \\ \frac{\beta_{0k}}{\sigma} f(\frac{1}{\sigma}\{1 - (\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0 + \boldsymbol{w}'_i\boldsymbol{\delta}_0)\}), \ j = J \end{cases}$$

$$i = 1, \ldots, n.$$

Further, for a time-variant discrete variable x_l, the partial effect of x_l arising from c_l to c_l+1 can be calculated by

$$\Delta p(y_{it}=j|\cdots)=p(y_{it}=j|x_l=c_l+1)-p(y_{it}=j|x_l=c_l).$$

$$\begin{cases} =\Phi(\frac{-\phi y^*_{i(t-1)}-(x_{it1}\beta_1+\cdots+(c_l+1)\beta_l+\cdots+x_{itk}\beta_k)-\boldsymbol{w}'_i\boldsymbol{\delta}}{\sigma}) \\ \quad -\Phi(\frac{-\phi y^*_{i(t-1)}-(x_{it1}\beta_1+\cdots+c_l\beta_l+\cdots+x_{itk}\beta_k)-\boldsymbol{w}'_i\boldsymbol{\delta}}{\sigma}),\ j=1 \\ =\Phi(\frac{\gamma_{ij}-(\phi y^*_{i(t-1)}+(x_{it1}\beta_1+\cdots+(c_l+1)\beta_l+\cdots+x_{itk}\beta_k)+\boldsymbol{w}'_i\boldsymbol{\delta})}{\sigma}) \\ \quad -\Phi(\frac{\gamma_{ij}-(\phi y^*_{i(t-1)}+(x_{it1}\beta_1+\cdots+c_l\beta_l+\cdots+x_{itk}\beta_k)+\boldsymbol{w}'_i\boldsymbol{\delta})}{\sigma}) \\ \quad +\Phi(\frac{\gamma_{i(j-1)}-(\phi y^*_{i(t-1)}+(x_{it1}\beta_1+\cdots+c_l\beta_l+\cdots+x_{itk}\beta_k)+\boldsymbol{w}'_i\boldsymbol{\delta})}{\sigma}) \\ \quad -\Phi(\frac{\gamma_{i(j-1)}-(\phi y^*_{i(t-1)}+(x_{it1}\beta_1+\cdots+(c_l+1)\beta_l+\cdots+x_{itk}\beta_k)+\boldsymbol{w}'_i\boldsymbol{\delta})}{\sigma}),\ j=2,\ldots,J-1 \\ =\Phi(\frac{1-(\phi y^*_{i(t-1)}+(x_{it1}\beta_1+\cdots+c_l\beta_l+\cdots+x_{itk}\beta_k)+\boldsymbol{w}'_i\boldsymbol{\delta})}{\sigma}) \\ \quad -\Phi(\frac{1-(\phi y^*_{i(t-1)}+(x_{it1}\beta_1+\cdots+(c_l+1)\beta_l+\cdots+x_{itk}\beta_k)+\boldsymbol{w}'_i\boldsymbol{\delta})}{\sigma}),\ j=J, \end{cases}$$

$$t=1,\ldots,T,\ i=1,\ldots,n,$$

and

$$\Delta p(y_{i0}=j|\cdots)=p(y_{i0}=j|x_l=c_l+1)-p(y_{i0}=j|x_l=c_l).$$

$$\begin{cases} =\Phi(\frac{-(x_{i01}\beta_{01}+\cdots+(c_l+1)\beta_{0l}+\cdots+x_{i0k}\beta_{0k}+\boldsymbol{w}'_i\boldsymbol{\delta}_0)}{\sigma}) \\ \quad -\Phi(\frac{-(x_{i01}\beta_{01}+\cdots+c_l\beta_{0l}+\cdots+x_{i0k}\beta_{0k}+\boldsymbol{w}'_i\boldsymbol{\delta}_0)}{\sigma}),\ j=1 \\ =\Phi(\frac{\gamma_{ij}-(x_{i01}\beta_{01}+\cdots+(c_l+1)\beta_{0l}+\cdots+x_{i0k}\beta_{0k}+\boldsymbol{w}'_i\boldsymbol{\delta}_0)}{\sigma}) \\ \quad -\Phi(\frac{\gamma_{ij}-(x_{i01}\beta_{01}+\cdots+c_l\beta_{0l}+\cdots+x_{i0k}\beta_{0k}+\boldsymbol{w}'_i\boldsymbol{\delta}_0)}{\sigma}) \\ \quad +\Phi(\frac{\gamma_{i(j-1)}-(x_{i01}\beta_{01}+\cdots+c_l\beta_{0l}+\cdots+x_{i0k}\beta_{0k}+\boldsymbol{w}'_i\boldsymbol{\delta}_0)}{\sigma}) \\ \quad -\Phi(\frac{\gamma_{i(j-1)}-(x_{i01}\beta_{01}+\cdots+(c_l+1)\beta_{0l}+\cdots+x_{i0k}\beta_{0k}+\boldsymbol{w}'_i\boldsymbol{\delta}_0)}{\sigma}),\ j=2,\ldots,J-1 \\ =\Phi(\frac{1-(x_{i01}\beta_{01}+\cdots+c_l\beta_{0l}+\cdots+x_{i0k}\beta_{0k}+\boldsymbol{w}'_i\boldsymbol{\delta}_0)}{\sigma}) \\ \quad -\Phi(\frac{1-(x_{i01}\beta_{01}+\cdots+(c_l+1)\beta_{0l}+\cdots+x_{i0k}\beta_{0k}+\boldsymbol{w}'_i\boldsymbol{\delta}_0)}{\sigma}),\ j=J. \end{cases}$$

Finally, for a time-invariant discrete variable w_l, the partial effect of w_l arising from c_l to c_l+1 can be calculated as (the case of $\boldsymbol{z}=\boldsymbol{w}$)

$$\Delta p(y_{it}=j|\cdots)=p(y_{it}=j|w_l=c_l+1)-p(y_{it}=j|w_l=c_l).$$

$$\begin{cases} =\Phi(\frac{-\phi y^*_{i(t-1)}-\boldsymbol{x}'_{it}\boldsymbol{\beta}-(w_{i1}\delta_1+\cdots+(c_l+1)\delta_l+\cdots+w_{ip}\delta_p)}{\sigma}) \\ \quad -\Phi(\frac{-\phi y^*_{i(t-1)}-\boldsymbol{x}'_{it}\boldsymbol{\beta}-(w_{i1}\delta_1+\cdots+c_l\delta_l+\cdots+w_{ip}\delta_p)}{\sigma}),\ j=1 \\ =\Phi(\frac{\frac{\gamma_{i(j-1)}+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+(c_l+1)\zeta_l+\cdots+z_{ip}\zeta_p)}{1+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+(c_l+1)\zeta_l+\cdots+z_{ip}\zeta_p)}-(\phi y^*_{i(t-1)}+\boldsymbol{x}'_{it}\boldsymbol{\beta}+(w_{i1}\delta_1+\cdots+(c_l+1)\delta_l+\cdots+w_{ip}\delta_p))}{\sigma}) \\ \quad -\Phi(\frac{\frac{\gamma_{i(j-1)}+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+c_l\zeta_l+\cdots+z_{ip}\zeta_p)}{1+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+c_l\zeta_l+\cdots+z_{ip}\zeta_p)}-(\phi y^*_{i(t-1)}+\boldsymbol{x}'_{it}\boldsymbol{\beta}+(w_{i1}\delta_1+\cdots+c_l\delta_l+\cdots+w_{ip}\delta_p))}{\sigma}) \\ \quad +\Phi(\frac{\frac{\gamma_{i(j-2)}+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+c_l\zeta_l+\cdots+z_{ip}\zeta_p)}{1+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+c_l\zeta_l+\cdots+z_{ip}\zeta_p)}-(\phi y^*_{i(t-1)}+\boldsymbol{x}'_{it}\boldsymbol{\beta}+(w_{i1}\delta_1+\cdots+c_l\delta_l+\cdots+w_{ip}\delta_p))}{\sigma}) \\ \quad -\Phi(\frac{\frac{\gamma_{i(j-2)}+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+(c_l+1)\zeta_l+\cdots+z_{ip}\zeta_p)}{1+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+(c_l+1)\zeta_l+\cdots+z_{ip}\zeta_p)}-(\phi y^*_{i(t-1)}+\boldsymbol{x}'_{it}\boldsymbol{\beta}+(w_{i1}\delta_1+\cdots+(c_l+1)\delta_l+\cdots+w_{ip}\delta_p))}{\sigma}), \\ \quad j=2,\ldots,J-1 \\ =\Phi(\frac{1-(\phi y^*_{i(t-1)}+\boldsymbol{x}'_{it}\boldsymbol{\beta}+(w_{i1}\delta_1+\cdots+c_l\delta_l+\cdots+w_{ip}\delta_p))}{\sigma}) \\ \quad -\Phi(\frac{1-(\phi y^*_{i(t-1)}+\boldsymbol{x}'_{it}\boldsymbol{\beta}+(w_{i1}\delta_1+\cdots+(c_l+1)\delta_l+\cdots+w_{ip}\delta_p))}{\sigma}),\ j=J, \end{cases}$$

$$t=1,\ldots,T,\ i=1,\ldots,n,$$

and

$$\Delta p(y_{i0}=j|\cdots)=p(y_{i0}=j|w_l=c_l+1)-p(y_{i0}=j|w_l=c_l).$$

$$\begin{cases} =\Phi(\frac{-\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0-(w_{i1}\delta_{01}+\cdots+(c_l+1)\delta_{0l}+\cdots+w_{ip}\delta_{0p})}{\sigma}) \\ \quad -\Phi(\frac{-\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0-(w_{i1}\delta_{01}+\cdots+c_l\delta_{0l}+\cdots+w_{ip}\delta_{0p})}{\sigma}),\ j=1 \\ =\Phi(\frac{\frac{\gamma_{i(j-1)}+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+(c_l+1)\zeta_l+\cdots+z_{ip}\zeta_p)}{1+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+(c_l+1)\zeta_l+\cdots+z_{ip}\zeta_p)}-(\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0+(w_{i1}\delta_{01}+\cdots+(c_l+1)\delta_{0l}+\cdots+w_{ip}\delta_{0p})}{\sigma}) \\ \quad -\Phi(\frac{\frac{\gamma_{i(j-1)}+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+c_l\zeta_l+\cdots+z_{ip}\zeta_p)}{1+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+c_l\zeta_l+\cdots+z_{ip}\zeta_p)}-(\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0+(w_{i1}\delta_{01}+\cdots+c_l\delta_{0l}+\cdots+w_{ip}\delta_{0p})}{\sigma}) \\ \quad +\Phi(\frac{\frac{\gamma_{i(j-2)}+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+c_l\zeta_l+\cdots+z_{ip}\zeta_p)}{1+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+c_l\zeta_l+\cdots+z_{ip}\zeta_p)}-(\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0+(w_{i1}\delta_{01}+\cdots+c_l\delta_{0l}+\cdots+w_{ip}\delta_{0p})}{\sigma}) \\ \quad -\Phi(\frac{\frac{\gamma_{i(j-2)}+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+(c_l+1)\zeta_l+\cdots+z_{ip}\zeta_p)}{1+\exp(\zeta_0+z_{i1}\zeta_1+\cdots+(c_l+1)\zeta_l+\cdots+z_{ip}\zeta_p)}-(\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0+(w_{i1}\delta_{01}+\cdots+(c_l+1)\delta_{0l}+\cdots+w_{ip}\delta_{0p})}{\sigma}), \\ \quad j=2,\ldots,J-1 \\ =\Phi(\frac{1-(\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0+(w_{i1}\delta_{01}+\cdots+c_l\delta_{0l}+\cdots+w_{ip}\delta_{0p})}{\sigma}) \\ \quad -\Phi(\frac{1-(\boldsymbol{x}'_{i0}\boldsymbol{\beta}_0+(w_{i1}\delta_{01}+\cdots+(c_l+1)\delta_{0l}+\cdots+w_{ip}\delta_{0p}))}{\sigma}),\ j=J. \end{cases}$$