

**BAYESIAN ANALYSIS
OF SPATIAL AND SURVIVAL MODELS
WITH APPLICATIONS
OF COMPUTATION TECHNIQUES**

A Dissertation presented to
the Faculty of the Graduate School
at the University of Missouri

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
YAJUN LIU
Dr. Dongchu Sun, Dissertation Supervisor
JULY 2012

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

BAYESIAN ANALYSIS OF SPATIAL AND SURVIVAL MODELS
WITH APPLICATIONS OF COMPUTATION TECHNIQUES

Presented by Yajun Liu,

A candidate for the degree of Doctor of Philosophy

And hereby certify that in their opinion it is worthy of acceptance.

Professor Dongchu Sun

Professor (Chong) Zhuoqiong He

Professor Paul Speckman

Professor Athanasios Micheas

Professor Jeannette Jackson-Thompson

ACKNOWLEDGEMENTS

I would like to acknowledge a sincere and deep debt of gratitude to my adviser and mentor, Dr. Dongchu Sun. This work would not have been made possible without his generous support, his earnest guidance, his encouragement and his patience throughout my graduate life. I am also very grateful to Dr. Zhuoqiong He, for her guidance and inspiring help in the survival analysis project.

I would like to express my appreciation to the other members of my advisory committee, Drs. Paul Speckman, Athanasios Micheas, and Jeannette Jackson-Thompson for their excellent comments and generous suggestions on my dissertation. I would also thank Drs. Christopher Wikle and Fei Liu, who served on my committee earlier, for their valuable inputs and supports.

I owe thanks to the Statistics Department faculty members, staffs and my graduate fellows for their support, assistance and friendships. This is a great and unforgettable experience in my life.

Foremost, I am forever indebted to my parents Xiaodi Zhang and Yongxin Liu. It is to them that I dedicate this work. Also, I wish to acknowledge all the love from my husband Yifan for his patience and support.

Finally, I wish to express my love to my daughter Juphie. She is the best reward in my graduate life.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	x
CHAPTER	
1 Introduction	1
1.1 Spatial and Survival Analysis Models	2
1.1.1 Spatial Statistics	2
1.1.2 Survival Models	2
1.2 Computation	3
1.2.1 A General Ancillarity Sufficiency Interweaving Scheme (ASIS)	3
1.2.2 BLAS/LAPACK	4
1.3 Overview	5
2 ASIS Application in Estimating Hunting Success Rates	7
2.1 Introduction	7
2.2 A Bayesian Hierarchical Model	8
2.3 ASIS Steps	10
2.3.1 ASIS Steps for δ_0 and δ_1	10
2.4 Computation and Result	13
2.5 Discussion	15
3 Hierarchical Bayes Models for QoL in Survivors of Breast Cancer	21

3.1	Introduction	21
3.2	Data Description	24
3.3	A Bayesian Hierarchical Model	26
3.3.1	Multinomial Distribution	26
3.3.2	Discrete Hazard Rate	26
3.3.3	Generalized Linear Mixed Model	27
3.4	Bayesian Computation	29
3.4.1	Conditional Posterior Densities	29
3.4.2	Computation	30
3.4.3	Estimation	31
3.4.4	Comparison	32
3.4.5	Robustness	35
3.5	Using an ASIS	35
3.5.1	An Algorithm via ASIS for Model (3.6)	35
3.5.2	Numerical Comparisons	37
3.6	Discussion	37
4	Hierarchical Model for Colorectal Cancer Survival Data	51
4.1	Introduction	51
4.2	Data Description	53
4.3	Bayesian Models Based on the Weibull Distribution	55
4.3.1	Survival Models	55
4.3.2	A Hierarchical Linear Mixed Model	57
4.3.3	A CAR Model	58
4.3.4	Other Priors	59

4.4	Computation	60
4.4.1	Model Selection	60
4.4.2	Full Conditional Posterior Distribution	61
4.4.3	Sampling by Ratio of Uniforms Method	68
4.5	Application	74
4.6	Discussion	75
5	A Latent Competing Risk Approach with A Hierarchical CAR Model for Colorectal Cancer Survival Data	85
5.1	Introduction	85
5.2	Data Description	86
5.3	Modeling	86
5.3.1	Latent Competing Risk Hazard Model	86
5.3.2	A Linear Mixed Model	89
5.3.3	A CAR Model	89
5.3.4	Priors	90
5.4	Computation	90
5.4.1	Identifiability	90
5.4.2	Model Selection	90
5.4.3	Sampling	91
5.5	Result	94
5.6	Discussion	94
6	Discussion on Spatial Confounding	100
6.1	Introduction	100
6.2	A Linear Mixed Model	103

6.2.1	Generalized Least Square Estimator	103
6.2.2	Bayes Estimator	104
6.2.3	\mathbf{X}_0 and \mathbf{u} are Independent	106
6.3	Comparison of Estimators of the Regression Coefficient $\boldsymbol{\beta}$	108
6.3.1	\mathbf{X}_0 and \mathbf{u} are Independent	108
6.3.2	\mathbf{X}_0 and \mathbf{u} are Correlated	111
6.4	Posterior Predictor of the Random Effect \mathbf{u}	117
6.5	Posterior Prediction	119
6.6	Bayes Factor	120
	BIBLIOGRAPHY	122
	VITA	132

LIST OF TABLES

Table	Page
2.1 Comparisons of Potential Scale Reduction Factor \hat{R}	17
2.2 Comparison of Bayesian Estimates and Running Times of δ_0 and δ_1 .	17
3.1 QoL Score Levels and Empty Cells for each Level	39
3.2 Bayesian Estimates of $(\delta_0, \delta_1, \rho, \theta_1, \theta_2)$ and their Standard Errors . . .	39
3.3 Hyper Parameters for Five Different Priors	39
3.4 Bayesian Estimates and their Standard Errors for Five Priors	40
3.5 First Order Correlations, Second Order Correlations, and Acceptance Rates based on the ASIS	40
4.1 Cause of Death Summary	77
4.2 Stage Grouping	77
4.3 Tumor Size Grouping	77
4.4 Positive Regional Lymph Nodes Grouping	77
4.5 Model Specifications	77
4.6 DIC Comparison and Common Parameter Estimates of Different Models	78
4.7 The Cure Rate Estimates of Cure Rate Models with Different Effects	78
4.8 The Estimated Effects and their Standard Errors	78
5.1 Parameter Estimates and their Standard Errors	96

LIST OF FIGURES

Figure	Page
2.1 Comparison of Sample Paths of δ_0	18
2.2 Comparison of Sample Paths of δ_1	19
2.3 Comparison of Posterior Densities of δ_0 and δ_1	20
3.1 Histograms of PCS Scores	41
3.2 Parameter Traces for Control Group - Regular Gibbs Sampling	42
3.3 Parameter Traces for Cancer Group - Regular Gibbs Sampling	43
3.4 Gelman-Rubin Convergence Diagnosis Plots	44
3.5 Comparison of Posterior Densities of the QoL Effects	45
3.6 Comparison of Posterior Densities of $\delta_i, i = 0, 1$	45
3.7 Comparison of Posterior Densities of ρ	46
3.8 Comparison of Spatial Effects Z_i	46
3.9 Comparison of Frequency and Bayesian Estimates of p_{ij}	47
3.10 Robust Stretching Plots for Different Priors	48
3.11 Comparison of Trace Plots of $\delta_k, k = 0, 1$ for the Control Group	49
3.12 Comparison of Trace Plots of $\delta_k, k = 0, 1$ for the Cancer Group	49
3.13 Density Comparison of δ_0 and δ_1 with the Reliable Densities	50
4.1 Kaplan-Meier Survival Function Estimates with Different Stages	79

4.2	Correlation Scatterplot Matrix of Important Effects	80
4.3	Behavior of $g(w)$	81
4.4	Acceptance Region with the Ratio-of-Uniforms Method When Sampling α for Cure Rate Model.	81
4.5	Posterior Densities of the Effects	82
4.6	Estimated Spatial Effect and Standard Errors	83
4.7	The Survival Estimates for Stages with Different Causes of Death	84
5.1	Estimated Spatial Effect and Standard Errors	97
5.2	Posterior Density Comparison of the Shape and Hazard Rate Parameters	98
5.3	Trace Plots for Hazard Rates	99

ABSTRACT

The discussion of spatial data and failure time data is developing very fast in recent years. Bayesian hierarchical linear mixed modeling is widely used in the analysis. This dissertation discusses the methodologies of applying Bayesian hierarchical models to different data with geographical characteristics or with right-censored failure time. A conditional autoregressive (CAR) prior is used for the model to capture spatial effects. Markov chain Monte Carlo (MCMC) methods are used in the sampling. The Ancillary-Sufficient Interweaving Strategy (ASIS) is applied to improve the performance for some parameters. The convergence of some of the parameters improved greatly, but the others do not have very significant improvement. However, the overall performance has improved greatly since it needs much fewer iterations than using regular Gibbs sampling to achieve convergence. For the survival analysis, we propose a generalized linear mixed model with different effects for the hazard rates, and adopt a cure rate model in Chen et al. (1999) for the hazards. A ratio-of-uniforms method is used to get the posterior density of some parameters that cannot be simply sampled by common methods. Both the Weibull model and cure rate models are compared. Moreover, for the same data set, competing risks model is considered by incorporating spatial effect to a latent competing risk model from Gelfand et al. (2000). The sampling method mentioned in Berger & Sun (1993) is adapted for efficiency. Finally, spatial confounding occurs when incorporating spatial effects in a regression model. Several estimators of the coefficients are compared for their Mean Squared Errors. The corresponding prediction errors are also discussed.

Chapter 1

Introduction

The discussion of spatial data and failure time data, or survival data, has developed very fast in recent years. Statistical analysis is commonly used for survival data, and most survival data include geographical components and demographic characteristics such as age, gender, and race, etc.

The development of Markov chain Monte Carlo (MCMC) methods such as Gibbs sampling has enabled Bayesian analysis to combine these two types of data and present more meaningful understanding. Because Bayesian statistics can effectively use researchers' prior information, it is able to provide more flexible analysis than obtained by frequentist methods.

This dissertation focuses on the methodology of applying the Bayesian hierarchical linear mixed models on different data with geographical characteristics or with right-censored failure time. MCMC methods are used for different applications with various models including the linear mixed model, Weibull and the cure rate models, and the competing risks model. Some advanced sampling methods are implemented to improve sampling efficiency.

In this chapter, Section 1.1 reviews the models used for spatial and survival analysis. Section 1.2 discusses some sampling techniques and the programming packages. Finally, Section 1.3 gives the overview of the chapters in this dissertation.

1.1 Spatial and Survival Analysis Models

1.1.1 Spatial Statistics

Spatial statistics arises when researchers in diverse areas need to do statistical inference and prediction from geographically referenced data. There are three basic types of spatial data sets: point-referenced data, areal data and point pattern data. Cressie (1993) provided a comprehensive guide for spatial statistical methods up to about 20 years ago. Since MCMC methods made full Bayesian analysis of complex models available, Banerjee, Carlin & Gelfand (2003) presented a practical treatment of hierarchical modeling and data analysis for complex spatial data sets. Cressie & Wikle (2011) presented spatio-temporal processes, which bridge classic ideas with modern hierarchical statistical modeling concepts and the latest computational methods.

This dissertation considers areal data structures. For the hierarchical models, a conditional autoregressive (CAR) prior originally developed by Besag (1974) is used to incorporate geographical information. It is computationally convenient for Gibbs sampling in Bayesian model fitting, according to Banerjee, Carlin & Gelfand (2003).

1.1.2 Survival Models

Survival data have been discussed by many researchers using different models. Since Kalbfleisch (1978) gave a Bayesian analysis of the semi-parametric regression and life

model of Cox (1972) and modeled the cumulative hazard function as a gamma process, Bayesian models have been widely developed. Arjas & Gasbarra (1994) considered simple right censored survival data with a common unknown hazard rate, which is modeled nonparametrically as a jump process. Aslanidou et al. (1998) extended the application by analyzing multivariate survival data from a Bayesian perspective using Markov-chain Monte Carlo methods.

Weibull model is one of the most common models used in survival analysis. It is a special case of the proportional hazard rate model (Cox (1972)). Many other models are derived from it. Chapter 4 compares the Weibull model and the cure rate model extended from Chen et al. (1999), incorporating a hierarchical mixed model.

The theory of competing risks models were introduced by David & Moeschberger (1978). Since then, many investigations have been carried out for competing risk models. The problem arises when there are several potential causes for failure, but there is insufficient information to tell the decisive cause. Chapter 5 discusses a latent competing risk approach adopted from Gelfand et al. (2000) to deal with survival data with geographical characteristics.

1.2 Computation

1.2.1 A General Ancillarity Sufficiency Interweaving Scheme (ASIS)

The Ancillarity Sufficiency Interweaving Scheme (ASIS) was introduced by Yu & Meng (2011). It is an interesting method using two types of data augmentation to sample from a posterior distribution $p(\boldsymbol{\theta} | Y_{obs}) \propto p(Y_{obs} | \boldsymbol{\theta})p(\boldsymbol{\theta})$, where $p(Y_{obs} | \boldsymbol{\theta})$ is the likelihood and $p(\boldsymbol{\theta})$ is the prior for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J)$. If there is a latent

variable/parameter or missing datum, Y_{mis} , such that $(Y_{obs} | \boldsymbol{\theta}, Y_{mis})$ is independent of $\boldsymbol{\theta}$, the sampling from $(\boldsymbol{\theta}, Y_{mis})$ is called a Sufficient Augmentation (SA); on the other hand, if there is another missing datum \tilde{Y}_{mis} so that $(\tilde{Y}_{mis} | \boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$, then the sampling from $(\boldsymbol{\theta}, Y_{mis})$ is called Ancillary Augmentation (AA). The ASIS combines both SA and AA in one Gibbs cycle with a one to one mapping, and does the Gibbs sampling on the basis of conditional densities of both augmentations.

The main steps for the component-wise interweaving strategy is illustrated below.

- Suppose we have the updated $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_J^{(t)})$.
- For $j = 1, \dots, J$, let Y_{Sj} and Y_{Aj} be conditional SA and conditional AA schemes for θ_j . Assume one-to-one transformations $Y_{Aj} = M_{Aj}(Y_{mis}; \theta_j)$ and $Y_{Sj} = M_{Sj}(Y_{mis}; \theta_j)$.
- Draw $Y_{mis} | \theta_j^{(t)}$ and iterate the pair of steps:
- Step $(j + 1)_A$. Compute Y_{Aj} with the current $\boldsymbol{\theta}$ and Y_{mis} ; draw

$$\theta_j^{(t+0.5)} \sim P(\theta_j | \theta_{<j}^{(t+1)}, \theta_{>j}^{(t)}, Y_{Aj}),$$

update Y_{mis} from the new $\boldsymbol{\theta}$ and Y_{Aj} by the inverse transform.

- Step $(j + 1)_S$. Compute Y_{Sj} with the updated Y_{mis} and $\boldsymbol{\theta}$ from Step $(j + 1)_A$, draw

$$\theta_j^{(t+1)} \sim P(\theta_j | \theta_{<j}^{(t+1)}, \theta_{>j}^{(t)}, Y_{Sj});$$

update Y_{mis} from the new $\boldsymbol{\theta}$ and Y_{Sj} by the inverse transform.

1.2.2 BLAS/LAPACK

Intel® Math Kernel Library threaded BLAS and LAPACK routines are used for matrix computation including matrix inversion and factorization in the projects. The

code is programmed with Intel® Fortran Compiler Professional Edition.

To improve the efficiency of matrix computation, there are two ways by using BLAS/LAPACK libraries. The first one is by using Sparse BLAS Routines to take advantage of the sparse matrices. Matrix inversion can be performed by the *PARDISO** (Parallel Direct Sparse Solver Interface, Schenk & Gärtner (2004)) package. However, *PARDISO** is not an efficient way from the performance and scalability points of view although the sparse storage format might be efficient. We found that using the packed storage format for symmetric matrices and utilize the routines in both BLAS and LAPACK libraries is preferable. The storage might be less efficient, but more threaded routines as inversion and factorization can be used to improve program efficiency.

According to Intel® (2009), the *ICOF* sampling method for Gaussian and multivariate Gaussian random generating works best with the generator *MCG59* (Intel® (2008)).

The program is run on an Intel® Xeon® X5560 2.80GHz 16-processor with 16G memory. The performance might be affected since the server is shared with other users.

1.3 Overview

The following chapters are organized as below. Chapter 2 applies the Ancillarity-Sufficiency Interweaving Strategy (ASIS in Yu & Meng (2011)) algorithm to improve the convergence of two variance parameters in He & Sun (2000). The performance of the ASIS algorithm is compared with regular Gibbs sampling by the potential scale reduction factor, trace plots and posterior estimates. Chapter 3 proposes a hierarchical logistic linear mixed model to estimate spatial effects and Quality of Life

levels for breast cancer survivors in Missouri. The Discrete Hazard Rate is adapted to simplify the sampling distribution. Chapter 4 compares a special case of the classic Cox model with Weibull hazard and the cure rate model originally developed by Chen et al. (1999). In both models, we proposed a generalized linear mixed model with different effects to the hazard rates. Chapter 5 discusses the spatial effect and latent competing risks of right-censored survival data. The spatial effect is included in the latent competing risk model, and the sampling method mentioned in Berger & Sun (1993) is adapted. In Chapter 6, the spatial confounding between coefficient covariates and the spatial covariates in a linear mixed model is discussed.

Chapter 2

ASIS Application in Estimating Hunting Success Rates

2.1 Introduction

Since Gelfand & Smith (1990), the Markov chain Monte Carlo (MCMC) method has been a powerful tool in Bayesian computation. However, the MCMC algorithm often converges slowly. Yu & Meng (2011) proposed an ancillarity-sufficiency interweaving strategy (ASIS) which significantly boosts MCMC efficiency. In linear mixed model or hierarchical generalized linear mixed model, some parameters often converge slowly. Here we apply this strategy to the MCMC steps of He & Sun (2000), and the convergence comparison between before and after using this strategy suggests that this is quite efficient for some parameters, but less obvious for other parameters.

In Section 2.2, we briefly describe the data, model and the MCMC steps used in He & Sun (2000). We apply the ASIS strategy in Section 2.3 on two parameters in the model and give the detailed steps. We evaluate the performance of these parameters from different perspectives and compare the results in Section 2.4. In Section 2.5, we

discuss the advantages of this strategy, and suggest possible improvement.

2.2 A Bayesian Hierarchical Model

He & Sun (2000) used a Bayesian hierarchical generalized linear model to estimate the Missouri spring turkey hunting success rates at the subarea level for postseason harvest surveys. Since each hunter can kill at most one turkey each week, they assumed that the number of killed turkeys y_{ij} in county i during week j has a binomial distribution with parameters n_{ij} and p_{ij} , where n_{ij} and p_{ij} are the total number of trips in the sample and the success rate per trip to county i during week j , respectively. They modeled the hunting rate by a logistic linear mixed model containing a residual effect,

$$\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \alpha_j + Z_i + e_{ij}, \quad i = 1, \dots, I = 114, \quad j = 1, 2, \quad (2.1)$$

in which α_j represents the effect of the j th week, Z_i the effect of the i th county, and e_{ij} the residual effects with $N(0, \delta_0)$.

Further stages of the model include priors on (α_j, Z_i, e_{ij}) and hyper priors on the parameters of the Z_i and e_{ij} .

Distribution of Z_i . Define $\mathbf{C} = (C_{kl})$ to be the $I \times I$ symmetric adjacency matrix, $C_{kl} = 1$ if counties k and l share a common boundary; $C_{kl} = 0$ otherwise, including $C_{kk} = 0$. Assume $\mathbf{Z} = (Z_1, \dots, Z_I)'$ follow the simultaneous CAR prior of Clayton & Kaldor (1987),

$$(\mathbf{Z} \mid \delta_1, \rho) \sim N_I(\mathbf{0}, \delta_1(\mathbf{I} - \rho\mathbf{C})^{-1}), \quad (2.2)$$

where δ_1 and ρ are the variation parameter and correlation parameter among neigh-

borhoods.

Let λ_1 and λ_I be the minimum and maximum eigenvalues of the matrix \mathbf{C} , respectively. If $\lambda_1^{-1} < \rho < \lambda_I^{-1}$, $\mathbf{B} = \mathbf{I} - \rho\mathbf{C}$ is positive definite. For the Missouri data, λ_1 and λ_I are -2.8931 and 5.6938 , respectively. So ρ is in the range $(-0.3457, 0.1756)$. Assume that

- (a) $\alpha_j \stackrel{indep.}{\sim} N(\mu_j, \tau_j)$, $j = 1, 2$.
- (b) $\delta_k \stackrel{indep.}{\sim} IG(a_k, b_k)$ with density $\pi(\delta_k) \propto \delta^{-(a_k+1)} \exp(-b_k/\delta_k)$, $k = 0, 1$.
- (c) $\rho \sim Uniform(\lambda_1^{-1}, \lambda_I^{-1})$.

He & Sun (2000) used the following hyper-parameters:

$$(\mu_1, \tau_1, \mu_2, \tau_2) = (-2.2, 1.5, -2.67, 1.5), (a_0, b_0) = (a_1, b_1) = (2.15, 0.09). \quad (2.3)$$

We define $v_{ij} = \text{logit}(p_{ij}) = \log\{p_{ij}/(1 - p_{ij})\}$, and write

$$\mathbf{p} = (p_{11}, \dots, p_{I1}, p_{12}, \dots, p_{I2})', \mathbf{Y} = (y_{11}, \dots, y_{I1}, y_{12}, \dots, y_{I2})'.$$

Fact 1. *The full conditional distributions are as follows.*

- (i) $[v_{ij} \mid \boldsymbol{\alpha}, \mathbf{Z}, \delta_0, \delta_1, \rho; \mathbf{Y}] \stackrel{indep.}{\propto} \exp\left\{v_{ij}y_{ij} - n_{ij} \log(1 + e^{v_{ij}}) - \frac{(v_{ij} - \alpha_j - Z_i)^2}{2\delta_0}\right\}$.
- (ii) $(\alpha_j \mid \mathbf{p}, \mathbf{Z}, \delta_1, \rho; \mathbf{Y}) \stackrel{indep.}{\sim} N\left(\frac{\tau_j \sum_i (v_{ij} - Z_i) + \delta_0 \mu_j}{I\tau_j + \delta_0}, \frac{\delta_0 \tau_j}{I\tau_j + \delta_0}\right)$.
- (iii) $[\mathbf{Z} \mid \mathbf{p}, \boldsymbol{\alpha}, \delta_0, \delta_1, \rho; \mathbf{Y}] \sim N\left(\frac{1}{\delta_0} \left(\frac{2}{\delta_0} \mathbf{I} + \frac{1}{\delta_1} \mathbf{B}\right)^{-1} \mathbf{c}, \left(\frac{2}{\delta_0} \mathbf{I} + \frac{1}{\delta_1} \mathbf{B}\right)^{-1}\right)$, where $\mathbf{c} = (c_1, \dots, c_I)^T$ and $c_i = \sum_{j=1}^2 (v_{ij} - \alpha_j)$.
- (iv) $(\delta_0 \mid \mathbf{p}, \boldsymbol{\alpha}, \mathbf{Z}, \delta_1, \rho; \mathbf{Y}) \sim IG\left(a_0 + \frac{IJ}{2}, b_0 + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J (v_{ij} - \alpha_j - Z_i)^2\right)$.
- (v) $(\delta_1 \mid \mathbf{p}, \boldsymbol{\alpha}, \mathbf{Z}, \delta_0, \rho; \mathbf{Y}) \sim IG\left(a_1 + \frac{I}{2}, b_1 + \frac{1}{2} \mathbf{Z}^T \mathbf{B} \mathbf{Z}\right)$.

$$(vi) [\rho \mid \mathbf{p}, \boldsymbol{\alpha}, \mathbf{Z}, \delta_0, \delta_1; \mathbf{Y}] \propto |\mathbf{I} - \rho \mathbf{C}|^{1/2} \exp \left\{ -\frac{1}{2\delta_1} \mathbf{Z}^T (\mathbf{I} - \rho \mathbf{C}) \mathbf{Z} \right\}, \text{ for } \rho \in (\lambda_1^{-1}, \lambda_I^{-1}).$$

From He & Sun (2000), the conditional density functions of v_{ij} and ρ in (i) and (vi) are log concave. So the adaptive rejection algorithm in Gilks & Wild (1992) can be used.

2.3 ASIS Steps

The convergence of the parameters δ_0 and δ_1 is relatively slow. Here we give the Gibbs sampling steps and the conditional posterior densities for the ASIS algorithm.

2.3.1 ASIS Steps for δ_0 and δ_1

In our model, we are interested in improving δ_0 and δ_1 , so we use the ASIS algorithm on the basis of the standard Gibbs sampling steps in Fact 1 and insert the Ancillarity Augmentation sampling steps before Steps (iv) and (v).

ASIS Algorithm. Before the cycle $(t + 1)$, we have $(v_{ij}^{(t)}, \boldsymbol{\alpha}^{(t)}, \mathbf{Z}^{(t)}, \delta_0^{(t)}, \delta_1^{(t)}, \rho^{(t)})$.

Step 1: Draw independent $v_{ij}^{(t+1)} \mid (\alpha_j^{(t)}, Z_i^{(t)}, \delta_0^{(t)}, \delta_1^{(t)}, \rho^{(t)}; \mathbf{Y})$.

Step 2: Draw $\boldsymbol{\alpha}^{(t+1)}$ given $(\mathbf{p}^{(t+1)}, \mathbf{Z}^{(t)}, \delta_0^{(t)}, \delta_1^{(t)}, \rho^{(t)}; \mathbf{Y})$ as Step (ii) in Fact 1.

Step 3: Draw $\mathbf{Z}^{(t+1)}$ given $(\mathbf{p}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \delta_0^{(t)}, \delta_1^{(t)}, \rho^{(t)}; \mathbf{Y})$ as Step (iii) in Fact 1.

Step 4A: Define $\beta_{ij} = \alpha_j^{(t+1)} + Z_i^{(t+1)}$, and transform $v_{ij}^{(t+1)}$ to $\tilde{\eta}_{ij} = (v_{ij}^{(t+1)} - \beta_{ij}) / \sqrt{\delta_0^{(t)}}$.

Let $\tilde{\boldsymbol{\eta}} = (\tilde{\eta}_{11}, \dots, \tilde{\eta}_{I1}, \tilde{\eta}_{12}, \dots, \tilde{\eta}_{I2})'$. Draw $\delta_0^{(t+0.5)}$ from the density

$$[\delta_0 \mid \tilde{\boldsymbol{\eta}}, \boldsymbol{\alpha}^{(t+1)}, \mathbf{Z}^{(t+1)}] \propto \frac{1}{\delta_0^{a_0+1}} e^{-b_0/\delta_0} \left\{ \prod_{i,j} \frac{e^{y_{ij}(\beta_{ij} + \sqrt{\delta_0} \tilde{\eta}_{ij})}}{(1 + e^{\beta_{ij} + \sqrt{\delta_0} \tilde{\eta}_{ij}})^{n_{ij}}} \right\}. \quad (2.4)$$

Step 4S: Transform $\tilde{\eta}_{ij}$ back to $\tilde{v}_{ij} = \beta_{ij} + \sqrt{\delta_0^{(t+0.5)}} \tilde{\eta}_{ij}$. Given $(\mathbf{p}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \mathbf{Z}^{(t+1)})$, draw $\delta_0^{(t+1)}$ from $IG(a_0 + \frac{1}{2}IJ, b_0 + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J (\tilde{v}_{ij} - \beta_{ij})^2)$.

Step 5A: Transform $Z_i^{(t+1)}$ to $\tilde{\zeta}_i = Z_i^{(t+1)} / \sqrt{\delta_1^{(t)}}$. Let $\tilde{\boldsymbol{\zeta}} = (\tilde{\zeta}_1, \dots, \tilde{\zeta}_I)'$. Draw $\delta_1^{(t+0.5)}$ from the conditional density

$$[\delta_1 \mid \boldsymbol{\alpha}^{(t+1)}, \tilde{\boldsymbol{\zeta}}, \delta_0^{(t+1)}] \propto \frac{1}{\delta_1^{a_1+1}} e^{-b_1/\delta_1} \exp \left\{ -\frac{1}{2\delta_0} \sum_{i=1}^I \sum_{j=1}^J (v_{ij} - \tilde{\zeta}_i \sqrt{\delta_1} - \alpha_j)^2 \right\}, \quad (2.5)$$

where the superscriptions are omitted for simplicity.

Step 5S: Transform $\tilde{\zeta}_i$ to $\tilde{Z}_i = \sqrt{\delta_1^{(t+0.5)}} \tilde{\zeta}_i$, draw $\delta_1^{(t+1)}$ from $IG(a_1 + \frac{I}{2}, b_1 + \frac{1}{2} \tilde{\mathbf{Z}}^T \mathbf{B} \tilde{\mathbf{Z}})$, where $\tilde{\mathbf{Z}} = (\tilde{Z}_1, \dots, \tilde{Z}_I)^T$.

Step 6: Draw $\rho^{(t+1)} \mid (\delta_1^{(t+1)}, \mathbf{Z}^{(t+1)})$ as Step (vi) in Fact 1.

If we treat $(a_0, b_0) = (-1, 0)$, and $(a_1, b_1) = (-1, 0)$ for the prior distributions of δ_0 and δ_1 , the conditional densities in (2.4) and (2.5) are reduced to

$$p_0(\delta_0 \mid \cdot) \propto \prod_{i,j} \frac{e^{y_{ij}(\beta_{ij} + \sqrt{\delta_0} \tilde{\eta}_{ij})}}{(1 + e^{\beta_{ij} + \sqrt{\delta_0} \tilde{\eta}_{ij}})^{n_{ij}}}, \quad (2.6)$$

$$p_1(\delta_1 \mid \cdot) \propto \exp \left\{ -\frac{1}{2\delta_0} \sum_{i=1}^I \sum_{j=1}^J (v_{ij} - \tilde{\zeta}_i \sqrt{\delta_1} - \alpha_j)^2 \right\}. \quad (2.7)$$

Proposition 1. For $k = 0, 1$, let $\gamma_k = \sqrt{\delta_k}$ and denote $\tilde{p}_k(\gamma_k \mid \cdot)$ the conditional density of γ_k derived from $p_k(\delta_k \mid \cdot)$. Then $\tilde{p}_k(\gamma_k \mid \cdot)$ is log concave.

Proof. When $k = 0$, let $h(\gamma_0)$ be the logarithm of $\tilde{p}_0(\gamma_0 | \cdot)$,

$$h(\gamma_0) = C_0 + \log(\gamma_0) + \sum_{i,j} \left[y_{ij}(\beta_{ij} + \gamma_0 \tilde{\eta}_{ij}) - n_{ij} \log(1 + e^{\beta_{ij} + \gamma_0 \tilde{\eta}_{ij}}) \right], \quad (2.8)$$

where C_0 is a constant. It is easy to verify

$$\frac{\partial^2}{\partial \gamma_0^2} h(\gamma_0) = -\frac{1}{\gamma_0^2} - \sum_{i,j} \frac{n_{ij} \tilde{\eta}_{ij}^2 e^{\beta_{ij} + \gamma_0 \tilde{\eta}_{ij}}}{(1 + e^{\beta_{ij} + \gamma_0 \tilde{\eta}_{ij}})^2}. \quad (2.9)$$

It is negative since $n_{ij} \geq 0$. The result holds.

If $k = 1$, let $h(\gamma_1)$ be the logarithm of $\tilde{p}_1(\gamma_1 | \cdot)$,

$$h(\gamma_1) = C_1 + \log(\gamma_1) - \frac{1}{2\delta_0} \sum_{i,j} (v_{ij} - \tilde{\zeta}_i \gamma_1 - \alpha_j)^2, \quad (2.10)$$

where C_1 is a constant. Clearly,

$$\frac{\partial^2}{\partial \gamma_1^2} h(\gamma_1) = -\frac{1}{\gamma_1^2} - \frac{J}{\delta_0} \sum_{i=1}^I \tilde{\zeta}_i^2 < 0. \quad (2.11)$$

The result holds. □

Proposition 1 shows that both the densities of transformed random variables $\gamma_0 = \sqrt{\delta_0}$ and $\gamma_1 = \sqrt{\delta_1}$ are log concave. Therefore, if we treat $(a_0, b_0) = (a_1, b_1) = (-1, 0)$ in (2.4) and (2.5), we can use the adaptive rejection algorithm in Gilks & Wild (1992) for log concave functions, and sample γ_0 and γ_1 first, then transform them back to δ_0 and δ_1 .

A Rejection Method for Sampling (γ_0, γ_1)

Firstly, we make a constraint of $(a_0, b_0) = (a_1, b_1) = (-1, 0)$ in (2.4) and (2.5). Then, we consider the general case of (a_0, b_0, a_1, b_1) and draw δ_0 and δ_1 in Steps 4A and 5A

in Section 2.3.1 by using the rejection method.

From Step 4A, the unnormalized conditional density function of δ_0 is

$$f(\delta_0 \mid \tilde{\boldsymbol{\xi}}, \boldsymbol{\theta}, \mathbf{Z}) = p_0(\delta_0 \mid \cdot) \cdot Ig(\delta_0 \mid a_0, b_0) \leq p_0(\delta_0 \mid \cdot) \cdot M_0, \quad (2.12)$$

where $p_0(\delta_0 \mid \cdot)$ is defined in (2.6), $Ig(\delta_0 \mid a_0, b_0)$ is the inverse-gamma density of δ_0 with shape parameter a_0 and rate parameter b_0 , and M_0 is its maximum. It is well known that the mode of $Ig(\delta_0 \mid a_0, b_0)$ is $b_0/(a_0 + 1)$, so

$$M_0 = \frac{b_0^{a_0}}{\Gamma(a_0)} \left(\frac{a_0 + 1}{b_0} \right)^{a_0+1} e^{-(a_0+1)} = \frac{1}{\Gamma(a_0)b_0} \left(\frac{a_0 + 1}{e} \right)^{a_0+1}. \quad (2.13)$$

We first sample δ_0 from the density $p_0(\delta_0 \mid \cdot)$, then use the rejection method to get the sample from $f(\delta_0 \mid \tilde{\boldsymbol{\xi}}, \boldsymbol{\theta}, \mathbf{Z})$. The sampling steps are as follows:

Step 4A1: Draw δ_0^* from $p_0(\delta_0 \mid \cdot)$ by Gilks & Wild (1992) Adaptive Rejection Algorithm.

Step 4A2: Draw u from $U(0, 1)$, if

$$u \leq \frac{f(\delta_0^* \mid \tilde{\boldsymbol{\xi}}, \boldsymbol{\theta}, \mathbf{Z})}{M_0 p_0(\delta_0^* \mid \cdot)} = \left(\frac{b_0}{(a_0 + 1)\delta_0^*} \right)^{a_0+1} e^{-b_0/\delta_0^* + a_0+1},$$

take $\delta_0^{(t+0.5)} = \delta_0^*$. otherwise, continue with Step 4A1.

Similarly, δ_1 can be sampled from Steps 5A.

2.4 Computation and Result

The potential scale reduction factor \hat{R} (Gelman et al. (2003)) is calculated for both algorithms to evaluate convergence.

For a single chain test, it takes about 2.5 CPU minutes to sample 150,000 samples with 10,000 burn-in cycles by regular Gibbs sampling.

We apply the ASIS algorithm on both parameters δ_0 and δ_1 . With 50,000 iterations and 1,000 burn-in cycles, the estimates are almost the same as the regular Gibbs sampling as listed in Table 2.2.

A pilot study ran three chains with different iterations and burn-in cycles for each chain. Three parameters δ_0 , δ_1 and ρ were evaluated. As \hat{R} tends to 1, convergence was achieved. The potential scale reduction factor is listed in Table 2.1.

Note that δ_0 converges very fast with ASIS algorithm. In fact, with as few as 150 iterations after 20 burn-in cycles, \hat{R} reaches 1.00. Nevertheless, δ_1 does not have such good performance. With shorter burn-in cycles, the regular algorithm performs better for δ_1 . However, with longer burn-in cycles, the overall performance of the ASIS algorithm gets better. The convergence of other parameters (e.g., ρ) is affected a little by δ_0 and δ_1 .

We compare the trace plots based on the two sampling schemes. Figures 2.1 (a) and (c) the 2000 iterations of δ_0 after 100 burn-in cycles with and without ASIS. It is obvious that the ASIS method is much better than regular Gibbs sampling. From Figures 2.1 (b) and (c), we see that in order to get the equivalent performance, regular Gibbs sampling would take about 5 times as many iterations. On the other hand, δ_1 does not show much improvement by using ASIS algorithm. The trace plots of 4000 iterations of δ_1 after 500 Gibbs cycles are shown in Figure 2.2. There is no clear winner for this parameter.

The posterior densities of δ_0 and δ_1 are plotted in Figure 2.3. Clearly, the posterior densities of δ_0 with and without ASIS are almost identical, while those of δ_1 are slightly different. The result from ASIS is more concentrated towards its mean and has a slightly thinner tail. The posterior densities from the two algorithms will be

eventually the same if we run many more iterations, since they theoretically come from the same posterior density.

We have run both algorithms with different iterations and burn-in cycles. To see the effects of the sampling schemes on Bayesian estimates of δ_0 and δ_1 , we compare Bayesian estimates and their standard deviations, together with the running times in Table 2.2. For δ_0 , the ASIS algorithm takes about 10,000 iterations to get a consistent estimate, while the regular algorithm takes many more cycles to reach the same accuracy. As for δ_1 , neither of the algorithms has better performance based on 10,000, 30,000 or 50,000 iterations after 1,000 burn-in cycles. The ASIS algorithm does takes more time to finish the same cycles as the regular algorithm, but since it converges faster, it can undoubtedly compensate for this deficiency.

2.5 Discussion

Apparently, the ASIS method has improved the convergence of the residual variance δ_0 greatly, but for the spatial variance δ_1 , we do not see significant improvement. One reason is that δ_0 is the variance for independent parameters v_{ij} , while δ_1 arises from the dependent parameters Z_i . During the sampling steps, δ_0 would have been affected less by the previous sampled v_{ij} 's than δ_1 by Z_i 's, $i = 1, \dots, 114$. However, the overall performance has been improved by using the ASIS algorithm.

This is a simple example of applying the ASIS algorithm to regular Gibbs sampling. A difficult step of sampling is using ancillarity augmentation. Usually, the working parameter to insert between the sufficient augmentation steps does not have a regular density. If the density is log-concave, or related with a log-concave density, as in our example, sampling is feasible. In this case, we improve the best convergence with the least extra effort. In other cases however, the researcher might spend much

time trying to find some effective sampling method to get the algorithm working. So whether it worth the tough work has to be considered before one uses this method.

As we discussed in Section 2.4, some parameters related with independent parameters might have better performance using the ASIS method than those with dependent parameters. The efficiency might also be related with the particular dataset. More research needs to be done to make a more accurate conclusion.

In this chapter, the ASIS method is only applied to the variation parameters. If more parameters are included, convergence might have some improvement, but the sampling time might be longer. Moreover, for other parameters, the interweaving steps become much harder because for Gibbs sampling, the parametrization will result in more complicated densities that might not be able to be implemented. In this case, we have to make some compromise to improve the best convergence with least extra effort.

Finally, the sampling time can also be improved by introducing MPI techniques. We do not have a proper facility to apply it, but there is no doubt it would spend less time sampling parallel chains with this implementation. If this is the case, the over-time that has been spent on sampling the non-regular conditional posterior densities of the interweaving steps for all chains can be greatly induced and hence ignorable.

Table 2.1: Comparisons of Potential Scale Reduction Factor \hat{R} for Two Algorithms in 3 Chains with Different Iterations and Burn-in Cycles

	δ_0	δ_1	ρ	Iterations (Burn-in Cycles)
Without ASIS	1.06	1.08	1.01	500 (20)
With Asis	1.00	1.20	1.05	
Without ASIS	1.06	1.03	1.00	500 (100)
With Asis	1.01	1.10	1.02	
Without ASIS	1.01	1.03	1.00	500 (200)
With Asis	1.00	1.03	1.01	
Without ASIS	1.03	1.04	1.01	500 (500)
With Asis	1.00	1.00	1.00	

Table 2.2: Comparison of Bayesian Estimates (standard deviations) and Running Times of δ_0 and δ_1 with and without ASIS. All Iterations are with 1,000 burn-in period, and time is recorded in minutes.

	δ_0	δ_1	Iterations	Time
Without ASIS	0.0220 (0.0079)	0.0559 (0.0139)	50,000	1.18
	0.0220 (0.0080)	0.0561 (0.0141)	30,000	0.68
	0.0229 (0.0085)	0.0564 (0.0138)	10,000	0.27
With ASIS	0.0221 (0.0079)	0.0560 (0.0137)	50,000	1.48
	0.0221 (0.0079)	0.0559 (0.0136)	30,000	0.91
	0.0221 (0.0079)	0.0562 (0.0138)	10,000	0.32

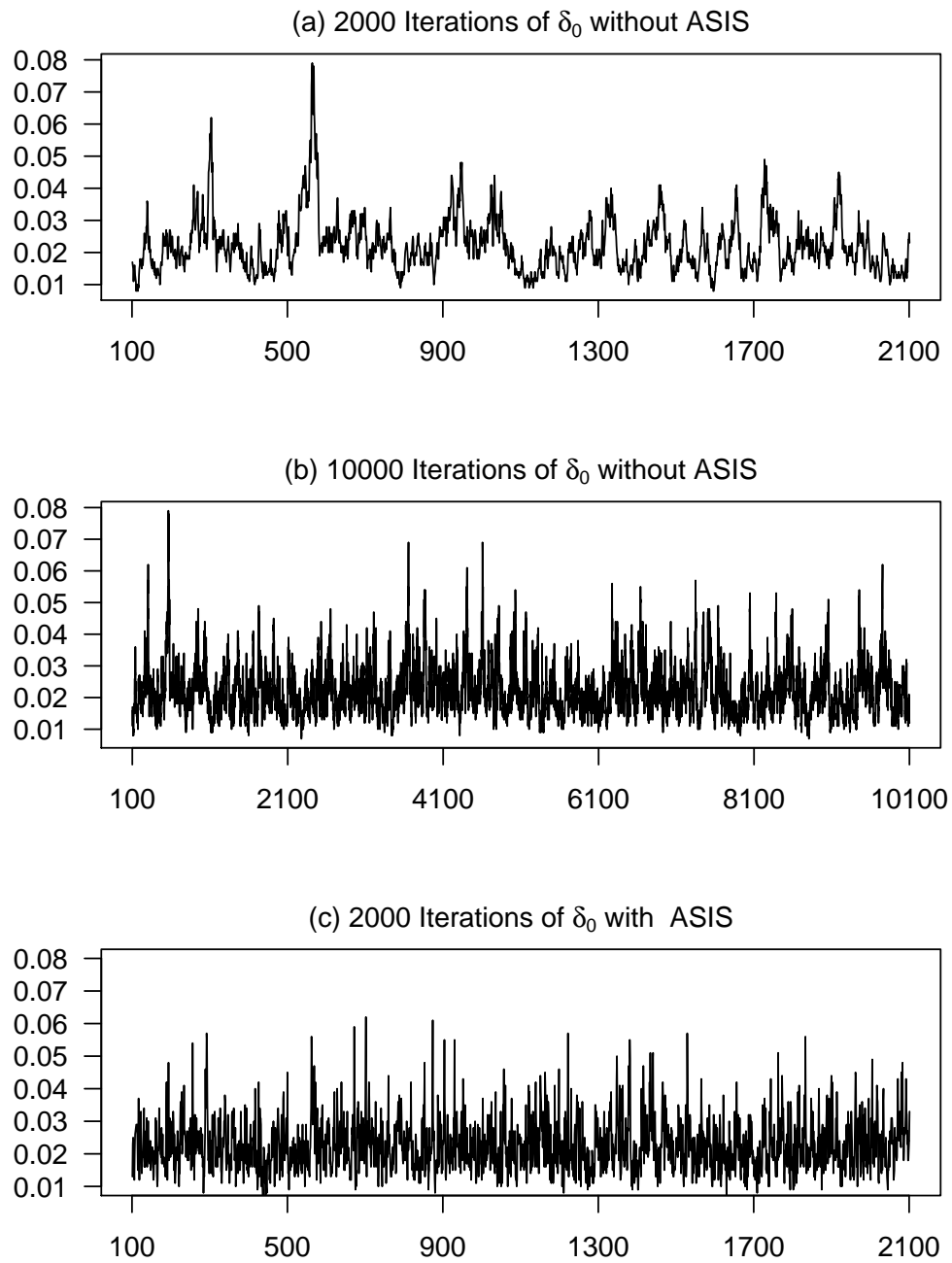


Figure 2.1: Comparison of Sample Paths of δ_0 from Gibbs Sampling with and without ASIS for Different Gibbs Cycles after 100 Burn-in Cycles

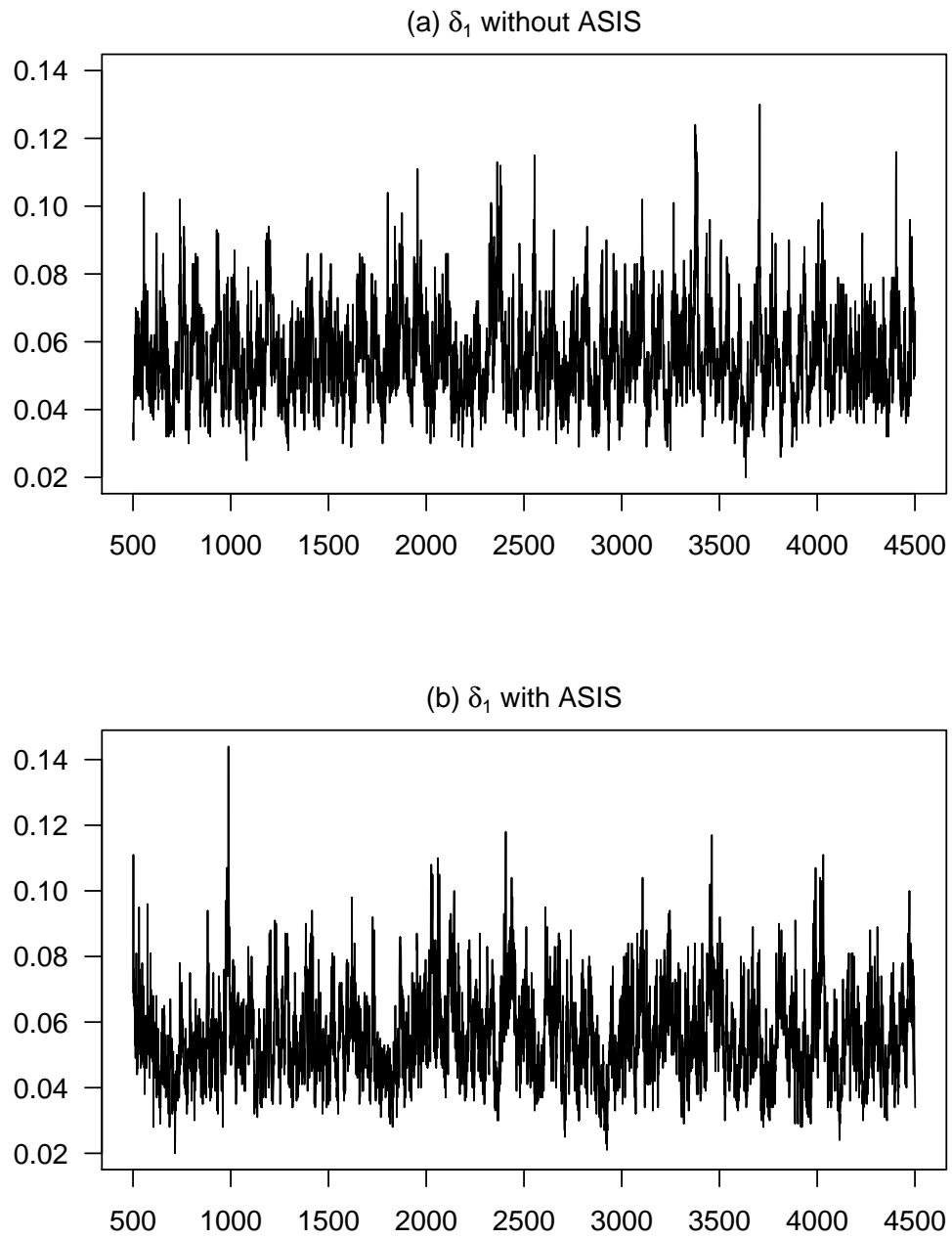


Figure 2.2: Comparison of Sample Paths of δ_1 from Gibbs Sampling with and without ASIS based on 4000 Iterations after 500 Burn-in Cycles

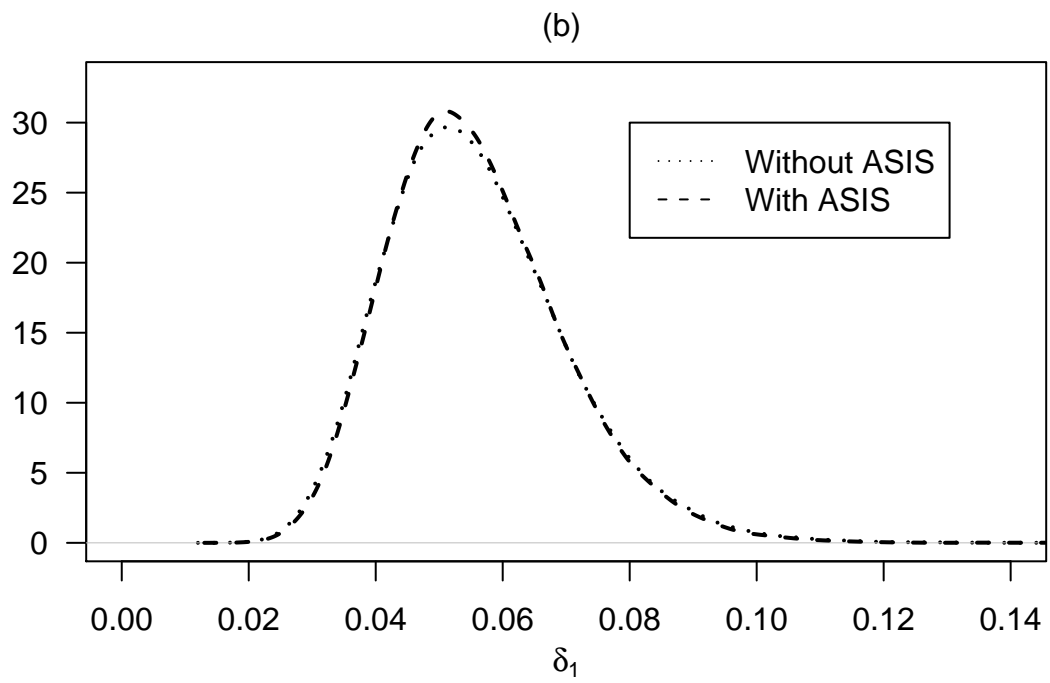
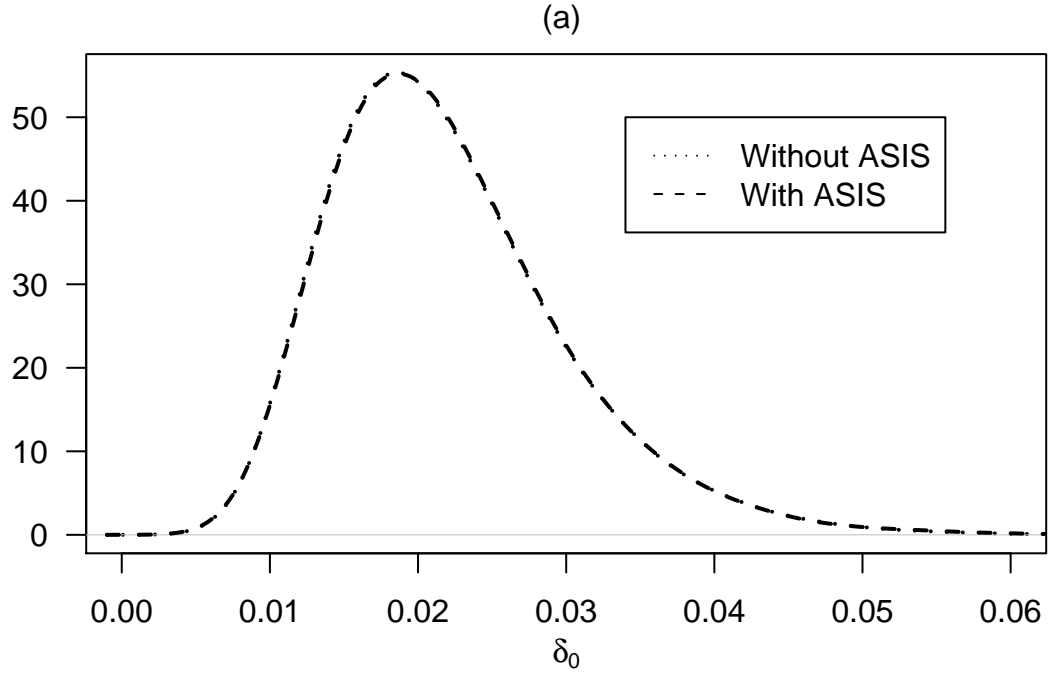


Figure 2.3: Comparison of Posterior Densities of δ_0 and δ_1 from Gibbs Sampling with and without ASIS.

Chapter 3

Hierarchical Bayes Models for QoL in Survivors of Breast Cancer

3.1 Introduction

Breast cancer is the most common cancer among women in the United States. It is also one of the leading causes of death among women. In the U.S. in 2008, 210,203 women were diagnosed with breast cancer, and 40,589 died from the disease (USCS Working Group (2012)). Yet deaths from breast cancer have decreased significantly by 2.0% per year among women (Eheman et al. (2012)). Moreover, women with breast cancer are the largest group of female survivors of cancer (Howlader et al. (2011)). With declining breast cancer mortality rates and improved survival following breast cancer, there is increasing focus on the health-related quality of life (QoL) of breast cancer survivors. Rustøen & Begnum (2000) discussed the importance of the research on the quality of life in breast cancer patients to the actual care of these women. In addition to breast cancer survivors, quality of life of survivors of other cancers has also been studied. There are numerous studies based on socioeconomic factors such

as Penson et al. (2001), Montazeri et al. (2003), Singh et al. (2002a) and Singh et al. (2002b). Specifically in breast cancer survivors, increasing emphasis has been placed on identifying determinants of reduced QoL among women with breast cancer. These determinants include treatment (Dorval et al. (1998)), social deprivation (Pollock & Vickers (1997)), age (Wenzel et al. (1999)), education and marital status (King et al. (2000)), ethnicity (Ashing-Giwa et al. (1999)), family history (Vacek et al. (2003)), race (Ashing-Giwa et al. (2007)), and psychological perspective (Kissane et al. (1998), Ganz et al. (2003), and Ganz et al. (2005)), etc.

Neighborhood disadvantage affects many areas such as adult health behaviors, breast cancer stage at diagnosis and survival, and QoL of the general population independent of individual-level factors. Many studies among the general population have shown that living in a deprived neighborhood adversely affects QoL (Schootman et al. (2004)). Neighborhood factors such as personal socioeconomic characteristics (Ross & Mirowsky (2001)), neighborhood residence and health (O'Campo (2003)), social characteristics (Sampson (2003)), ecological studies within groups (Roux (2003)), clinical demographic, and socioeconomic factors (Djibuti & Shakarishvili (2003)), and individual and neighborhood characteristics (Ross et al. (2004)) are all considered. However, few studies have considered a geographic clustering effect on the quality of life.

The aim of this chapter is to estimate the extent of geographic clustering effect of different QoL levels based on a prospective, population-based study at the neighborhood level among women with breast cancer as compared to community dwelling women without breast cancer (Schootman et al. (2004)). To estimate neighborhood effect, Ashing-Giwa et al. (2007) used a socio-ecologically and culturally contextual theoretical model by conducting descriptive, bivariate, and multivariate regression analysis. Singh et al. (2002a) and Singh et al. (2002b) used principal components

analysis and join-point regression analysis to model and identify statistically significant changes in the mortality trends. Schootman & Sun (2004) investigated geographic variability in the increase in breast cancer incidence observed among large areas and examined the variability among small areas in the incidence over time.

In this chapter, the discrete hazard rate is adapted to simplify the sampling distribution. A hierarchical logistic linear mixed model based on the discrete hazard rate is proposed to estimate both the neighborhood effect and the quality of life extent. Since the geographic effect is concerned, a conditional autoregressive (CAR) prior (Besag (1974), Cressie (1993)) is used to include geographical relationship among counties. The ASIS algorithm is used to improve convergence. The hierarchical model has the advantage of allowing information from all counties to be used for estimating success rates within each county; in our case, it is used for estimating all QoL level partitions.

In Section 3.2, we describe the data used in the study and explain the motivation of using the proposed model. In Section 3.3, we propose a logistic linear mixed model for the discrete hazard rate α_{ij} with a fixed QoL level effect and a random spatial effect. Prior distributions for the variance components and spatial correlations are also specified. In Section 3.4, Bayesian computation via Gibbs sampling is implemented with FORTRAN. In addition, the full conditional distributions of all the parameters are discussed. Bayesian estimates for the parameters in the model are reported, and their effects are interpreted. The comparison figures are plotted using CRAN-R. Robustness is also discussed based on several different priors. In Section 3.5, we consider an interweaving scheme of conditional sufficient augmentation and conditional ancillary augmentation for improving the convergence of some parameters in the Gibbs Sampling. Section 3.6 discusses the limitation of the model and gives a guideline for future work.

3.2 Data Description

The data used in the study were obtained from interviews conducted with 1,150 women with breast cancer (denoted as the “cancer group”) and 1,157 women without any type of reportable cancer (denoted as the “control group”). Women in the cancer group, identified from cases reported to the Missouri Cancer Registry (MCR)¹ (now the Missouri Cancer Registry and Research Center (MCR-ARC)), gave permission to MCR to release their confidential information to researchers and signed an informed consent form prior to being interviewed approximately one year following their diagnosis of breast cancer. A population-based control group of community-dwelling women without breast cancer or other cancers was identified through random-digit dialing and frequency matched by neighborhood (county of residence), age, and race (in selected area). Their data will be compared to that of women with breast cancer at the same time points. Women in both groups were 25 years of age or older and residents of Missouri at the time of diagnosis (cancer group) and interview (both groups).

The RAND 36-Item Health Survey 1.0 (a.k.a. SF-36) of Hayes et al. (1993), which consists of eight subscales, were completed by both groups of female respondents (Schoutman et al. (2004)). We include this survey to compare QoL between the women in the control and cancer groups, since this measure has been used in studies of both healthy and patient populations. Reliability and validity of the subscales have been established in studies of both general and patient populations in Stewart et al. (1988), Wells et al. (1989), and Ware & Sherbourne (1992). General population

¹The Missouri Cancer Registry (now the Missouri Cancer Registry and Research Center) is a Gold-certified member of, the North American Association of Central Cancer Registries (NAACCR). MCR-ARC is a collaborative partnership between the Missouri Department of Health and Senior Services (DHSS) and the University of Missouri. Since 1995, MCR-ARC has received financial support from the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC).

norms for this instrument are available in Ware (2000).

Two composite scores, measuring physical component summary (PCS) and mental component summary (MCS) were computed in Taft et al. (2001) from the eight subscales, corresponding respectively to physical and mental health measurements following a standard three-step procedure. In this chapter, we consider PCS only. First, all eight subscale scores ranging from 0 to 100 are standardized using a linear z-score transformation; second, z-scores are multiplied by the subscale factor score coefficients and summed over all eight subscales; finally, *t*-scores are calculated by multiplying the obtained PCS sums by 10 and adding 50 to the product, to yield a mean of 50 and a standard deviation of 10 for the US norm population. The PCS scores range from 0 to 100. We will analyze the scores as outcome measures potentially associated with neighborhood effects. The SF-36 subscales are reliable instruments for assessing QoL across race according to Ashing-Giwa et al. (2004).

Although the data had been complete at the time of the study, there were still some counties where no women were interviewed (5 out of 115 for the control group and 13 out of 115 for the cancer group.) Also, more than half of the counties have very small sample sizes. (Both control and cancer groups have 65 counties out of 115 that have less than 5 observations.) To investigate the quality of life of these women, we divide the scores into several QoL levels for both women in the control and cancer groups and consider the group frequencies. These levels can be treated as one factor for the PCS scores and are considered as the QoL level effect. In the chapter, we only treat the QoL level effect and the spatial effect as covariates.

The relative frequency of the response scores are given in Figure 3.1. We divide the scores into three levels in such a way that approximately the same number of subjects (including both the control and cancer groups) are included in each level. The frequencies of each level for each county and group are computed. People in

Level 1 have the worst quality of life, considered as “Poor”; people in Level 3 have better quality of life, and can be viewed as “Good”; and people in the middle level can be treated as having “Adequate” quality of life. For each group, there are 115 counties, so there are 115 cells for each level, or 345 (115×3) cells for each case group. However, there are still many empty cells using this strategy as shown in Table 3.1.

3.3 A Bayesian Hierarchical Model

3.3.1 Multinomial Distribution

For $i = 1, \dots, I = 115$, and $j = 1, \dots, J = 3$, let y_{ij} be the frequency for the j^{th} QoL level in the i^{th} county, n_i represents the total number of responses for the i^{th} county. Naturally, (y_{i1}, y_{i2}, y_{i3}) has a multinomial distribution,

$$(y_{i1}, y_{i2}, y_{i3}) \sim \text{Multinomial}(n_i; p_{i1}, p_{i2}, p_{i3}), \quad \sum_{j=1}^3 p_{ij} = 1, p_{ij} > 0. \quad (3.1)$$

The joint probability mass function of (y_{i1}, y_{i2}, y_{i3}) is

$$L(y_{i1}, y_{i2}, y_{i3} \mid p_{i1}, p_{i2}, p_{i3}) = \frac{n_i!}{y_{i1}! y_{i2}! y_{i3}!} p_{i1}^{y_{i1}} p_{i2}^{y_{i2}} p_{i3}^{y_{i3}} = n_i! \prod_{j=1}^3 \frac{p_{ij}^{y_{ij}}}{y_{ij}!}. \quad (3.2)$$

3.3.2 Discrete Hazard Rate

For the multinomial distribution, There is one constraint, $\sum_{j=1}^3 p_{ij} = 1$, so there are only 2 “true” parameters for each i . Without information loss, one can consider the discrete hazard rates α_{ij} , which offers a one-to-one mapping to p_{ij} . In particular, we

define

$$\alpha_{i1} = p_{i1}, \quad \alpha_{i2} = \frac{p_{i2}}{p_{i2} + p_{i3}} = \frac{p_{i2}}{1 - p_{i1}}. \quad (3.3)$$

Literally speaking, α_{i1} is the proportion of people in Level 1 in the whole population, and α_{i2} is the proportion of Level 2 people in those who not in Level 1 (or who are in Level 2 and 3). It is easy to show that $0 < \alpha_{ij} < 1$. Clearly, p_{ij} can be transformed back from α_{ij} ,

$$p_{i1} = \alpha_{i1}, \quad p_{i2} = \alpha_{i2}(1 - \alpha_{i1}), \quad p_{i3} = (1 - \alpha_{i1})(1 - \alpha_{i2}). \quad (3.4)$$

The joint probability mass function of $(y_{ij}, j = 1, \dots, J)$ then becomes

$$L(y_{i1}, y_{i2}, y_{i3} \mid \alpha_{i1}, \alpha_{i2}) = \frac{n_i!}{y_{i1}! y_{i2}! y_{i3}!} \alpha_{i1}^{y_{i1}} (1 - \alpha_{i1})^{y_{i2} + y_{i3}} \alpha_{i2}^{y_{i2}} (1 - \alpha_{i2})^{y_{i3}}. \quad (3.5)$$

From (3.3) to (3.5), it is clear that the likelihood function of p_{ij} based on the multinomial distribution is equivalent to the likelihood function of α_{ij} based on an “independent” binomial variable $y_{ij} \mid \alpha_{ij} \sim \text{Bin}(n_{ij}, \alpha_{ij})$, $j = 1, 2$, where $n_{i1} = n_i$ and $n_{i2} = n_i - y_{i1}$.

3.3.3 Generalized Linear Mixed Model

We use a hierarchical logistic linear mixed model for the problem. Since α_{ij} is between 0 and 1, we can use logit function to construct the model with the spatial effect Z_i , the QoL level effect θ_j , and a residual effect ϵ_{ij} . The residual effect includes some other sources of variability not explained by the linear part and can include higher order interaction as well as unknown sources of variability in α_{ij} . We write the model

as:

$$\text{logit}(\alpha_{ij}) = \log\left(\frac{\alpha_{ij}}{1 - \alpha_{ij}}\right) = Z_i + \theta_j + \epsilon_{ij}, \quad i = 1, \dots, I = 115, \quad j = 1, 2. \quad (3.6)$$

Distribution of Z_i . Assume $\mathbf{Z} = (Z_1, \dots, Z_I)'$ follow the simultaneous CAR prior of (2.2).

For the Breast Cancer dataset, λ_1 and λ_I are -2.8932 and 5.6941 , respectively. So ρ is in the range $(-0.3456, 0.1756)$.

QoL Level Effect. There are three QoL levels described in Section 3.2. From (3.3), it's sufficient to consider only the first two levels, Level 1 and Level 2, while Level 3 can be evaluated by the other two. We assume

$$\theta_j \stackrel{\text{indep}}{\sim} N(\mu_j, \tau_j), \quad j = 1, 2, \quad (3.7)$$

where (μ_j, τ_j) are fixed constants.

Residual Effect. We assume the residual effects have a normal distribution,

$$\epsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \delta_0), \quad i = 1, \dots, 115, \quad j = 1, 2. \quad (3.8)$$

We also make the following distribution assumptions:

- δ_k follows $IG(a_k, b_k)$, with density $Ig(\delta_k | a_k, b_k)$, for $k = 0, 1$,
- $\rho \sim U(\lambda_1^{-1}, \lambda_I^{-1})$.

Here (a_k, b_k) are fixed constants.

We write $\boldsymbol{\alpha} = (\alpha_{11}, \dots, \alpha_{I1}, \alpha_{12}, \dots, \alpha_{I2})'$, $\boldsymbol{\theta} = (\theta_1, \theta_2)'$, and assume the conditional independence. Given $(\boldsymbol{\theta}, \mathbf{Z}, \delta_0)$, $\boldsymbol{\alpha}$ is conditional independent of (δ_1, ρ) ; given (δ_1, ρ) , \mathbf{Z} is independent of $(\boldsymbol{\theta}, \delta_0)$; and $(\boldsymbol{\theta}, \delta_0, \delta_1, \rho)$ are mutually independent.

Note that the data of the control and cancer groups are fit with the same model separately. We will report the estimates for each group and compare them.

The hyper-parameters are chosen to be the same for both control and cancer groups

$$(\mu_1, \tau_1) = (\mu_2, \tau_2) = (0, 10), \quad (a_0, b_0) = (a_1, b_1) = (2.03, 0.30). \quad (3.9)$$

3.4 Bayesian Computation

3.4.1 Conditional Posterior Densities

Here we give the full conditional distributions in using Gibbs sampling (Geman & Geman (1994)). We define $v_{ij} = \text{logit}(\alpha_{ij}) = \log\{\alpha_{ij}/(1 - \alpha_{ij})\}$, $\mathbf{B} = \mathbf{I} - \rho\mathbf{C}$,

$$\mathbf{v} = (v_{11}, \dots, v_{I1}, v_{12}, \dots, v_{I2})', \text{ and } \mathbf{Y} = (y_{11}, \dots, y_{I1}, y_{12}, \dots, y_{I2})'.$$

The full conditional distributions are as follows.

Proposition 2.

(a) Given $(\boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1, \rho; \mathbf{Y})$, v_{ij} 's are independent and the conditional density of v_{ij} is

$$[v_{ij} \mid \cdot] \propto \exp \left\{ v_{ij} y_{ij} - n_{ij} \log(1 + e^{v_{ij}}) - \frac{(v_{ij} - \theta_j - Z_i)^2}{2\delta_0} \right\}. \quad (3.10)$$

(b) Given $(\mathbf{v}, \boldsymbol{\theta}, \mathbf{Z}, \delta_1, \rho; \mathbf{Y})$, the conditional posterior distribution of δ_0 is

$$IG\left(a_0 + \frac{I(J-1)}{2}, b_0 + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^{J-1} (v_{ij} - \theta_j - Z_i)^2\right). \quad (3.11)$$

(c) Given $(\mathbf{v}, \mathbf{Z}, \delta_0, \delta_1, \rho; \mathbf{Y})$, the conditional posterior distribution of θ_j is

$$N\left(\frac{\tau_j \sum_i (v_{ij} - Z_i) + \delta_0 \mu_j}{I\tau_j + \delta_0}, \frac{\delta_0 \tau_j}{I\tau_j + \delta_0}\right). \quad (3.12)$$

(d) Given $(\mathbf{v}, \boldsymbol{\theta}, \delta_0, \delta_1, \rho; \mathbf{Y})$, the conditional posterior distribution of \mathbf{Z} is

$$N\left(\frac{1}{\delta_0} \left(\frac{2}{\delta_0} \mathbf{I} + \frac{1}{\delta_1} \mathbf{B}\right)^{-1} \mathbf{c}, \left(\frac{2}{\delta_0} \mathbf{I} + \frac{1}{\delta_1} \mathbf{B}\right)^{-1}\right),$$

where

$$\mathbf{c} = (c_1, \dots, c_I)' \text{ and } c_i = \sum_{j=1}^{J-1} (v_{ij} - \theta_j). \quad (3.13)$$

(e) Given $(\mathbf{v}, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \rho; \mathbf{Y})$, the conditional posterior distribution of δ_1 is

$$IG\left(a_1 + \frac{I}{2}, b_1 + \frac{1}{2} \mathbf{Z}' \mathbf{B} \mathbf{Z}\right). \quad (3.14)$$

(f) Given $(\mathbf{v}, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1; \mathbf{Y})$, the conditional posterior density of ρ is

$$|\mathbf{I} - \rho \mathbf{C}|^{1/2} \exp\left\{-\frac{1}{2\delta_1} \mathbf{Z}' (\mathbf{I} - \rho \mathbf{C}) \mathbf{Z}\right\}, \text{ for } \rho \in (\lambda_1^{-1}, \lambda_I^{-1}). \quad (3.15)$$

3.4.2 Computation

From He & Sun (2000), we know that the conditional density functions of v_{ij} and ρ in (3.10) and (3.15) are log concave. So we can use the adaptive rejection algorithm in Gilks & Wild (1992) to sample these v_{ij} 's and ρ . We also adopted the more efficient algorithm mentioned in He & Sun (2000) to sample δ_1 and \mathbf{Z} by eigenvalues.

The number of parameters for $(\mathbf{v}, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1, \rho)$ is $(J-1) \times I + (J-1) + I + 3 = 350$,

$I = 115, J = 3$. MCMC algorithms are coded in FORTRAN.

We generate 100,000 samples with 10,000 burn-in for the control and cancer groups. It takes about 6 CPU minutes for each case to compute the 100,000 cycles.

3.4.3 Estimation

Convergence. By checking the traces of the first 10,000 samples after the burn-in period, the Markov chain of δ_0 and δ_1 does not converge very fast as shown in Figures 3.2 and 3.3.

Since δ_0 and δ_1 are two key parameters in the model, a Gelman-Rubin Convergence Diagnosis (Gelman & Rubin (1992)) in Figure 3.4 is done for the last 50,000 iterations of 100,000 iterations. It turns out that all the parameter medians converge to 1 and the 97.5% percentiles are all less than 1.05, which suggests that the parameters converge after 100,000 iterations.

The Bayesian estimates of the parameters are listed in Table 3.2 for both control and cancer groups.

Estimation of θ . As is suggested in Section 3.3.3, the two QoL level effects θ_1 and θ_2 correspond to Level 1 and Level 2 respectively, and Level 3 can be explained by these two. Figure 3.5 shows the QoL effect density plots. From the estimates of the parameters, estimates for proportions of three QoL levels without spatial effect can be obtained by using (3.4) and the inverse of the logit function, $\hat{\alpha}_{ij} = \text{logit}^{-1}(\hat{\theta}_j) = e^{\hat{\theta}_j} / (1 + e^{\hat{\theta}_j})$. For the control group, the proportion estimates are (36.6%, 32.9%, 30.5%); and for the cancer group, they are (41.6%, 36.0%, 22.4%). Clearly, for both groups, proportions decrease with the levels go up, indicating that the grouping strategy is taking effect. The plot also suggests that proportion gaps

between different levels are larger for the cancer group than for the control group. It can be interpreted that women in the cancer group tend to have lower QoL levels (Level 1 and Level 2).

Estimation of δ_0 and δ_1 . The density plots for δ_0 and δ_1 are shown in Figure 3.6. Recall that δ_0 is the error variance and δ_1 is the spatial variance, and they indicate the variability of the error term ϵ and the spatial effect \mathbf{Z} . For δ_0 , the estimates of both groups are similar, while the variance for the cancer group is smaller than that for the control group. The same pattern applies for δ_1 . Both δ_0 and δ_1 have similar estimates, which suggests that the model is satisfactory.

Estimation of ρ . By checking the 95% credible intervals, neither the ρ for the control group or for the cancer group is significant ($(-0.184, 0.168)$ for the control group, and $(-0.221, 0.171)$ for the cancer group). Although the ρ for both groups tends to be positive on the plot, it is not significantly strong to suggest spatial correlation at the county level.

3.4.4 Comparison

For parameters Z_i, α_{ij} and p_{ij} , we plot maps concerning their values for different counties to see the patterns among different counties and to compare the effect for the different groups.

Spatial Effect Comparison. For Z_i , we plot two maps in Figure 3.8 for women in both control and cancer groups. For the control group, Z_i in most northern counties and some southern counties are positive, which suggests that these counties tend to have positive neighborhood effects on the response, or equivalently, they have relatively larger proportions on QoL levels. More counties in central Missouri tend to have negative effects, which means smaller proportions. This is similar with the

cancer group, although about one fourth counties have Z_i 's around 0, and the pattern is more obvious that central counties tend to have negative spatial effects, while the northern and southern boundary tend to have positive ones. That is to say, a city on the northern boundary tends to have more people than a city in the middle for the same QoL level (Level 1 and Level 2).

Between the two groups, the most populous counties such as St. Louis County, Boone, Clay and Jackson, tend to have negative Z_i 's for both groups. For some other counties, such as Nodaway, Dallas, especially the counties in the southwestern Missouri, the effects are opposite. Most counties have similar effects for both groups.

This distinction might be caused by other covariates like social, economical, or populous factors, which are not included in the model. Nevertheless, we can still conclude that the QoL of both groups have some relationship with their neighborhood.

QoL Response Comparison. Besides the spatial effects, we can also analyze the QoL responses. As illustrated in Section 3.3.2, α_{ij} and p_{ij} can be transformed by a one-to-one mapping. Therefore, it suffices by comparing p_{ij} 's to understand the patterns of the quality of life of both control and cancer groups. We compare the Bayesian estimates of p_{ij} with the frequency estimates on the Missouri map as shown in Figure 3.9. In these plots, (a_j) and (c_j) , (b_j) and (d_j) in Figure 3.9 are frequency and Bayesian estimates of p_{ij} 's, $i = 1, \dots, I$, $j = 1, \dots, J$.

Obviously, through the combinations of the QoL levels and the counties, the empty cells for both groups are shown in Table 3.1, as we can also see in (a_j) and (c_j) of Figure 3.9. These empty cells cause a lot of shrinkage on the maps. On the other hand, the Bayesian estimates have covered all the counties including the empty cells. Moreover, by borrowing strength from neighboring counties, the Bayesian estimates have the similar proportion patterns with the Frequency estimates so that we may have comprehensive interpretations over the QoL level proportions over all the counties.

Note that p_{i1} indicates the proportion of QoL Level 1, p_{i2} represents QoL Level 2, and p_{i3} represent QoL Level 3, which is a complement to the first two proportions. The pattern is very clear and indicates different QoL levels.

For both control and cancer groups, there is an obvious pattern that the proportions decrease as the levels go up. This indicates that more women have worse quality of life in both groups. Level 2, in which people have “Adequate” quality of life, however, shares one third of the population for both groups, although there tend to be fewer people in the control group than in the cancer group. For Level 1, the proportions of women in the cancer group are larger than those in the control group in almost all the counties, while for Level 3, the opposite situation can be observed. Therefore, we conclude that women in cancer group tend to have worse quality of life.

Particularly, there is little variability across counties in all the QoL levels for women in the cancer group although there is variability among women in the control group. In the central counties for the control group, there tend to be fewer people on Level 1, and more people in Level 3. This pattern is similar to that estimated from the spatial effect, which is also consistent with the spatial coefficient ρ , whose variance for the control group is smaller than the cancer group. Nevertheless, no trend can be seen for the cancer group. There are two exceptions, St. Louis County and Boone County for the cancer group, in which the proportions are different from other counties. The reason might lie in the different other factors, such as medical facilities, social environment, or economical reasons, which are not discussed in this study.

3.4.5 Robustness

In (3.9), we gave a prior for the sampling. Now we show that this prior is robust for Bayesian estimators. We give a list of different priors, same for both control and cancer groups, and the scatter plots of the Bayesian estimates of p_{ij} based on various priors.

The plots are shown in Figure 3.10 for the control and cancer groups. A diagonal line is plotted for comparison. The x labels, indicated as $\hat{p}_{i1}, \hat{p}_{i2}, \hat{Z}_i$ are the estimates for the first prior, the y labels, indicated as $\tilde{p}_{i1}, \tilde{p}_{i2}, \tilde{Z}_i$, are the estimates of other priors. The estimates of other parameters for different priors are listed in Table 3.4.

3.5 Using an ASIS

As we discussed in Section 3.4.3, regular Gibbs Sampling method does not lead to fast convergence. As shown in Figures 3.2 and 3.3, the Markov chains for the parameters δ_0 and δ_1 show a serious auto-correlation. We adapt an Ancillarity Sufficiency Interweaving Scheme (ASIS) introduced in Yu & Meng (2011) to improve the convergence of the MCMC algorithm.

3.5.1 An Algorithm via ASIS for Model (3.6)

For our model, we are interested in improving δ_0 and δ_1 . It is easy to see that v_{ij} and $\xi_{ij} = \epsilon_{ij}/\sqrt{\delta_0}$ as SA and AA for δ_0 ; \mathbf{Z} and $\mathbf{Z}/\sqrt{\delta_1}$ are SA and AA for δ_1 . In particular, the ASIS algorithm was constructed by inserting the AA steps before Steps (b) and (e) in Section 3.4.1, respectively.

ASIS Algorithm. Before the cycle $(t + 1)$, we have $(\mathbf{v}^{(t)}, \boldsymbol{\theta}^{(t)}, \mathbf{Z}^{(t)}, \delta_0^{(t)}, \delta_1^{(t)}, \rho^{(t)})$.

Step 1: Draw $\mathbf{v}^{(t+0.5)} \mid (\boldsymbol{\theta}^{(t)}, \mathbf{Z}^{(t)}, \delta_0^{(t)}, \delta_1^{(t)}, \rho^{(t)}; \mathbf{Y})$ as Proposition 2 Part (a) .

Step 2*: Transform $v_{ij}^{(t+0.5)}$ to $\tilde{\xi}_{ij} = (v_{ij}^{(t+0.5)} - \beta_{ij})/\sqrt{\delta_0^{(t)}}$, where $\beta_{ij} = \theta_j^{(t)} + Z_i^{(t)}$. Let $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_{11}, \dots, \tilde{\xi}_{I1}, \tilde{\xi}_{12}, \dots, \tilde{\xi}_{I2})'$. Draw $\delta_0^{(t+0.5)}$ from the density

$$[\delta_0 \mid \tilde{\boldsymbol{\xi}}, \boldsymbol{\theta}, \mathbf{Z}] \propto \frac{1}{\delta_0^{a_0+1}} e^{-b_0/\delta_0} \prod_{i,j} \frac{e^{y_{ij}(\beta_{ij} + \sqrt{\delta_0} \tilde{\xi}_{ij})}}{[1 + e^{\beta_{ij} + \sqrt{\delta_0} \tilde{\xi}_{ij}}]^{n_{ij}}}. \quad (3.16)$$

Step 2: Transform $\tilde{\xi}_{ij}$ back to $v_{ij}^{(t+1)} = \beta_{ij} + \sqrt{\delta_0^{(t+0.5)}} \tilde{\xi}_{ij}$. As Proposition 2 Part (b), draw $\delta_0^{(t+1)} \mid (\mathbf{v}^{(t+1)}, \boldsymbol{\theta}^{(t)}, \mathbf{Z}^{(t)})$.

Step 3: Draw $\boldsymbol{\theta}^{(t+1)}$ given $(\mathbf{v}^{(t+1)}, \mathbf{Z}^{(t)}, \delta_0^{(t+1)}, \delta_1^{(t)}, \rho^{(t)}; \mathbf{Y})$ as Proposition 2 Part (c).

Step 4: Draw $\mathbf{Z}^{(t+0.5)}$ given $(\mathbf{v}^{(t+1)}, \boldsymbol{\theta}^{(t+1)}, \delta_0^{(t+1)}, \delta_1^{(t)}, \rho^{(t)}; \mathbf{Y})$ as Proposition 2 Part (d).

Step 5*: Transform $Z_i^{(t+0.5)}$ to $\tilde{\zeta}_i = Z_i^{(t+0.5)}/\sqrt{\delta_1^{(t)}}$. Let $\tilde{\boldsymbol{\zeta}} = (\tilde{\zeta}_1, \dots, \tilde{\zeta}_I)'$. Draw $\delta_1^{(t+0.5)}$ from the conditional density

$$[\delta_1 \mid \boldsymbol{\theta}^{(t+1)}, \tilde{\boldsymbol{\zeta}}, \delta_0^{(t+1)}] \propto \frac{1}{\delta_1^{a_1+1}} e^{-b_1/\delta_1} \exp \left\{ -\frac{1}{2\delta_0^{(t+1)}} \sum_{i=1}^I \sum_{j=1}^2 (v_{ij}^{(t+1)} - \tilde{\zeta}_i \sqrt{\delta_1} - \theta_j^{(t+1)})^2 \right\}. \quad (3.17)$$

Step 5: Transform $\tilde{\zeta}_i$ to $Z_i^{(t+1)} = \sqrt{\delta_1^{(t+0.5)}} \tilde{\zeta}_i$, and draw $\delta_1^{(t+1)}$ as Proposition 2 Part (e).

Step 6: Draw $\rho^{(t+1)} \mid (\delta_1^{(t+1)}, \mathbf{Z}^{(t+1)})$ as Proposition 2 Part (f).

A Rejection Method for Sampling (3.16) and (3.17) in Step 2* and 5*

To sample from (3.16) and (3.17), we can use exactly the same rejection method as described in Section 2.3.1 whose candidate density is a special case when $(a_0, b_0) = (a_1, b_1) = (-1, 0)$.

3.5.2 Numerical Comparisons

Estimates. The estimates of the parameters by using the ASIS agree with the estimates by regular Gibbs Sampling up to the third decimal point, and it takes almost the same time to compute the 100,000 cycles.

Convergence. The results are shown in Figures 3.11 and 3.12. By comparing 2,000 iterations from regular Gibbs sampling and the ASIS, we can see that the convergence improves greatly, especially for δ_0 . The acceptance rates for δ_0 and δ_1 are shown in Table 3.5.

Correlations. Besides the trace plots of parameters of interest, the correlation between iterations is another criterion for the performance of the convergence. For each case, the sampling computation is done for both regular Gibbs sampling and ASIS algorithm. Table 3.5 shows the first-order and second-order correlations of different groups. Obviously, the convergence improves.

Densities. By checking the density plots in Figure 3.6 for the first 3,000 iterations after burn-in for both regular Gibbs sampling and the ASIS algorithm, it is obvious that ASIS get close to the reliable density faster than the regular sampling does. The reliable density is the parameter density of the 100,000 iterations. After the MCMC, the two densities of the two algorithm match each other very well.

3.6 Discussion

We have applied a logistic hierarchical model with a CAR prior on Quality of Life of breast cancer survivors. The computation for the hierarchical models was improved significantly by using an Ancillarity-Sufficiency interweaving algorithm within the Gibbs Sampling. However, the conclusion that the spatial pattern can be detected is based on the visual check. There is no statistical statement to prove it. The

hypothesis testing or model selection can be performed for this purpose.

The project is completed by modeling the control and cancer groups separately. It is desirable to model the two groups together by adding a group indicator. The spatial effect shows the same pattern as that with two separate models. However, the sampling could not be satisfactorily convergent. One of the reasons is that there are too many missing cells for the model to have enough power fitting all the parameters. The sample size is too small compared to the large size of parameters. Grouping does reduce the parameter size, but it loses other individual characteristics, such as age, income, etc., that can contribute the model. Another reason is that the county population samples vary so much (from 0 to 523) that the model tends to borrow strength from the counties with larger populations and therefore larger samples, hence leading to inaccurate inference for counties with smaller populations and samples.

The research can potentially be enhanced if we reduce the size of the spatial effect by dividing the state map into meaningful larger regions, such as the seven regions used by DHSS and MCR-ARC to display behavioral risk factor and cancer incidence and mortality data (Wilson et al. (June 2010)). The seven regions are Kansas City Metro, St. Louis Metro, Central, Southwest, Southeast, Northwest and Northeast. A regional division will help to capture the spatial effect better than the county division, since fewer parameters will be included in the model, and the difference of the population sizes of the regional samples can be reduced and each region can be more representative.

Table 3.1: QoL Score Levels and Empty Cells for each Level. There are 115 cells for each level for both control and cancer groups, the empty percentages are given in parentheses

QoL Score Levels			Empty Cells	
PCS	Level	Condition	Control	Cancer
			Total: 96 (27.8%)	Total: 119 (34.5%)
[0, 41)	1	Poor	29 (25.2%)	34 (30.0%)
[41, 52)	2	Adequate	33 (28.7%)	37 (32.2%)
[52, 100]	3	Good	34 (30.0%)	48 (41.7%)

Table 3.2: Bayesian Estimates of $(\delta_0, \delta_1, \rho, \theta_1, \theta_2)$ and their Standard Errors (in Parentheses)

Groups	δ_0	δ_1	ρ	θ_1	θ_2
Control	0.092 (0.040)	0.104 (0.045)	0.051 (0.096)	-0.551 (0.100)	0.076 (0.119)
Cancer	0.079 (0.034)	0.095 (0.039)	0.043 (0.111)	-0.341 (0.100)	0.473 (0.120)

Table 3.3: Hyper Parameters for Five Different Priors

Priors	$(\mu_1, \tau_1) = (\mu_2, \tau_2)$	$(a_0, b_0) = (a_1, b_1)$
Prior 1	(0, 10)	(2.03, 0.30)
Prior 2	(1, 50)	(2.003, 0.25)
Prior 3	(5, 10^4)	(2.01, 0.28)
Prior 4	(0, 100)	(2.09, 0.33)
Prior 5	(-1, 10)	(2.18, 0.35)

Table 3.4: Bayesian Estimates and their Standard Errors (in Parentheses) for Five Priors

Control	δ_0	δ_1	ρ	θ_1	θ_2
Prior 1	0.092 (0.040)	0.104 (0.045)	0.051 (0.096)	-0.551 (0.100)	0.076 (0.119)
Prior 2	0.083 (0.040)	0.096 (0.044)	0.051 (0.098)	-0.558 (0.099)	0.073 (0.116)
Prior 3	0.088 (0.039)	0.101 (0.046)	0.052 (0.096)	-0.554 (0.099)	0.076 (0.117)
Prior 4	0.094 (0.040)	0.109 (0.049)	0.053 (0.094)	-0.551 (0.102)	0.079 (0.120)
Prior 5	0.097 (0.040)	0.109 (0.046)	0.051 (0.096)	-0.550 (0.101)	0.077 (0.118)
Cancer	δ_0	δ_1	ρ	θ_1	θ_2
Prior 1	0.079 (0.034)	0.095 (0.039)	0.043 (0.111)	-0.341 (0.100)	0.473 (0.120)
Prior 2	0.072 (0.033)	0.087 (0.038)	0.040 (0.114)	-0.344 (0.098)	0.464 (0.119)
Prior 3	0.077 (0.034)	0.092 (0.039)	0.043 (0.111)	-0.343 (0.100)	0.470 (0.121)
Prior 4	0.084 (0.035)	0.099 (0.040)	0.045 (0.110)	-0.340 (0.103)	0.479 (0.122)
Prior 5	0.086 (0.036)	0.100 (0.040)	0.044 (0.108)	-0.341 (0.101)	0.477 (0.122)

Table 3.5: First Order Correlations: the correlations between $\delta_0^{(t)}$ and $\delta_0^{(t+1)}$, $\delta_1^{(t)}$ and $\delta_1^{(t+1)}$; Second order Correlations: the correlations between $\delta_0^{(t)}$ and $\delta_0^{(t+2)}$, $\delta_1^{(t)}$ and $\delta_1^{(t+2)}$; and Acceptance Rates for these Parameters based on the ASIS

Group		Control		Cancer	
Correlations		δ_0	δ_1	δ_0	δ_1
Regular	First Order	0.95	0.89	0.94	0.88
	Second Order	0.89	0.85	0.89	0.83
ASIS	First Order	0.55	0.83	0.49	0.82
	Second Order	0.31	0.77	0.25	0.74
	Acceptance Rate	0.39	0.54	0.19	0.54

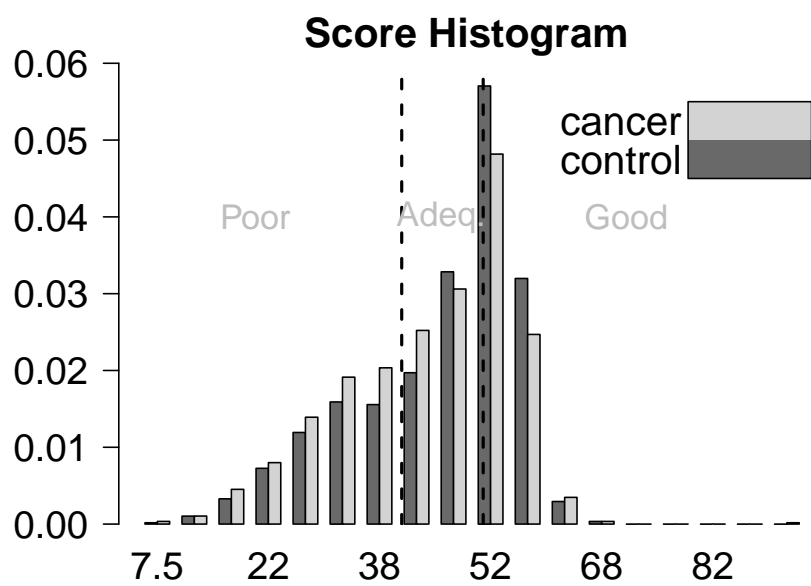


Figure 3.1: Histograms of PCS Scores, the dividing lines indicate the boundaries of different levels of QoL

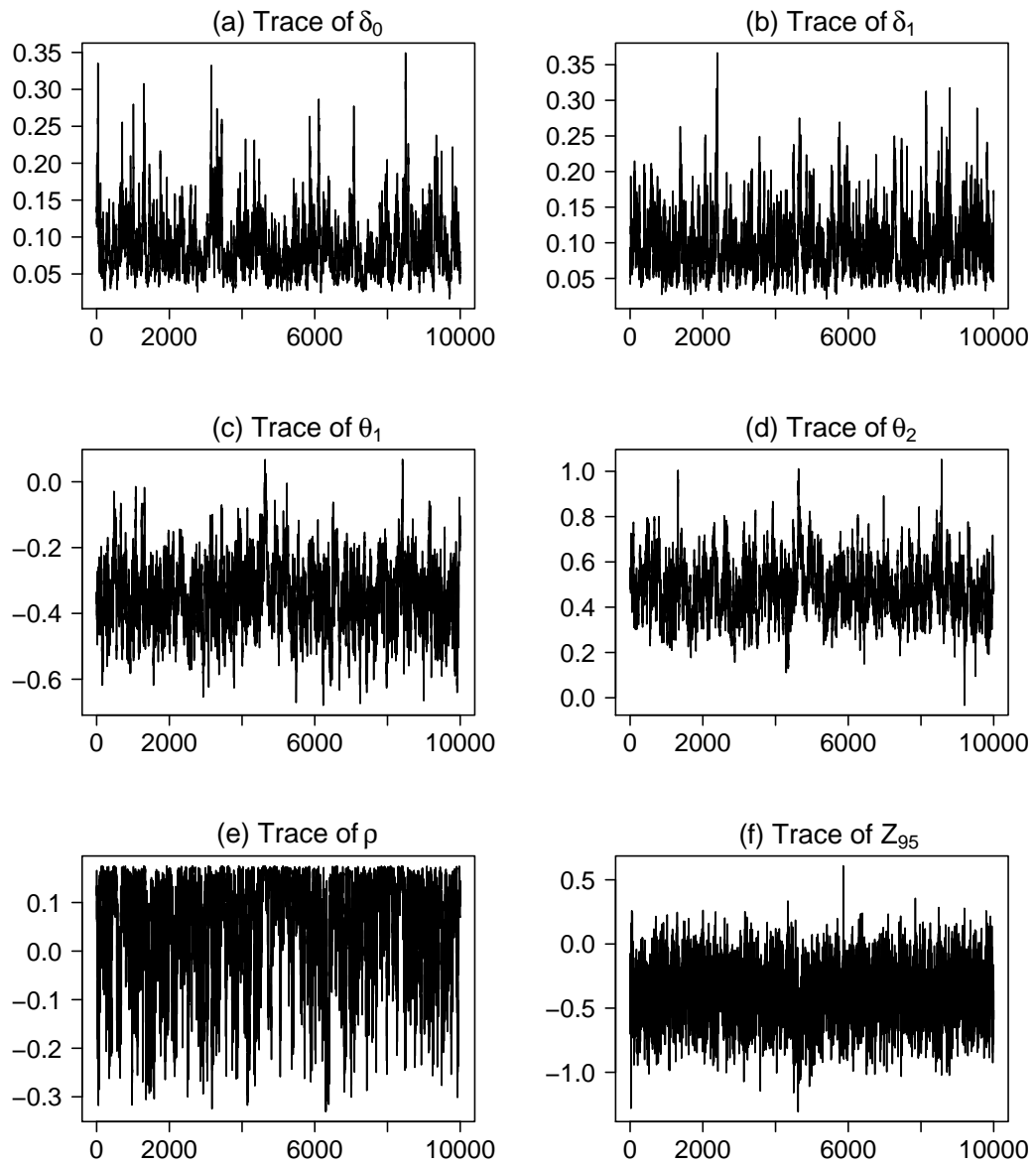


Figure 3.2: Parameter Traces for Control Group - Regular Gibbs Sampling

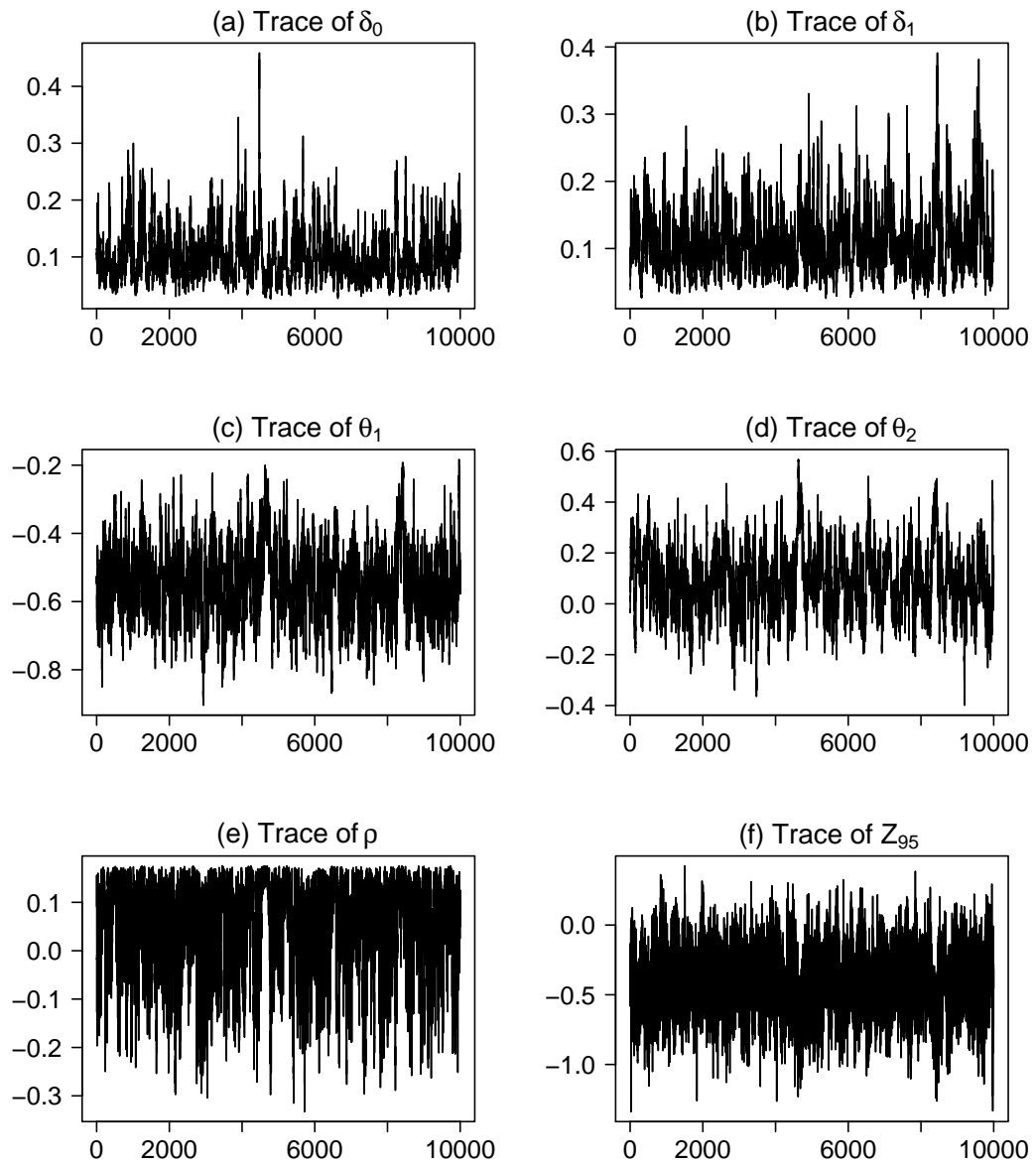


Figure 3.3: Parameter Traces for Cancer Group - Regular Gibbs Sampling

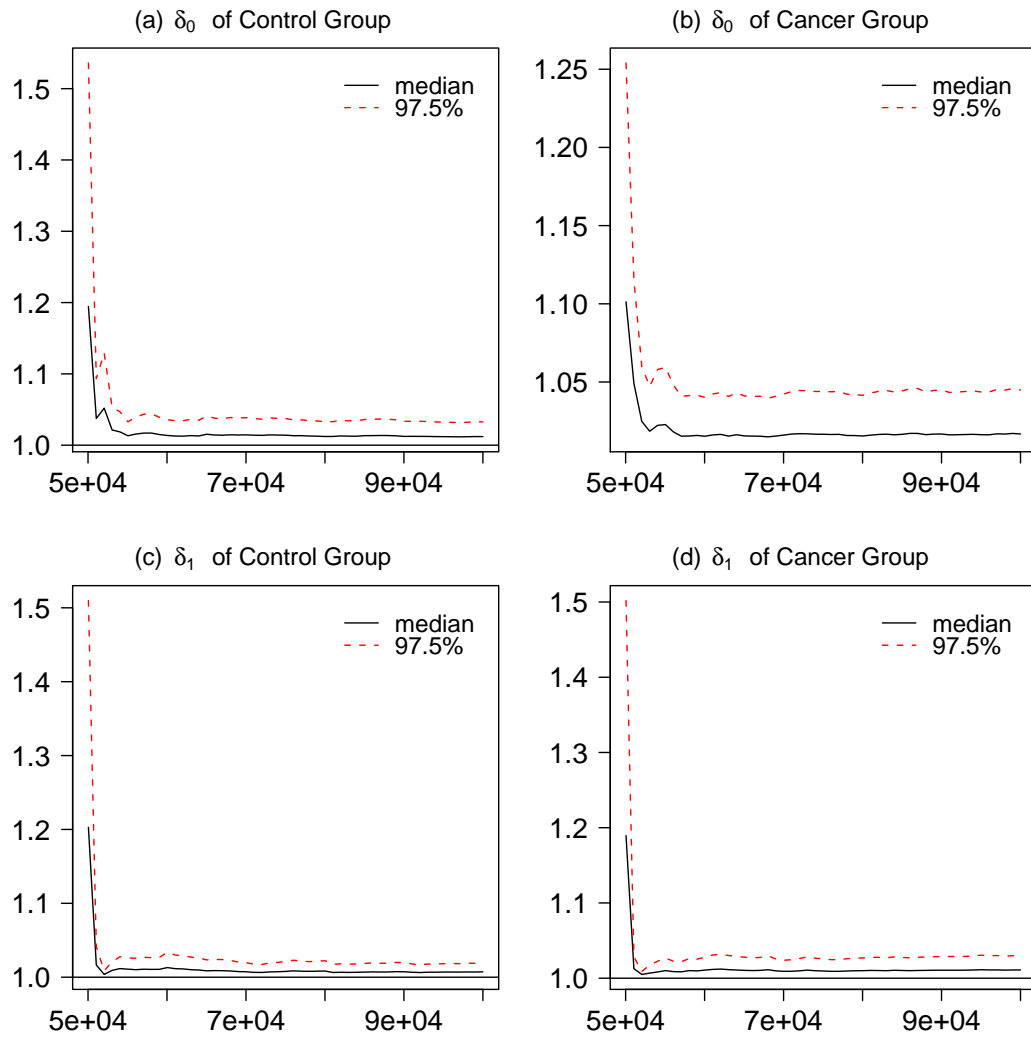


Figure 3.4: Gelman-Rubin Convergence Diagnosis Plots with Several Priors for last 50,000 Iterations

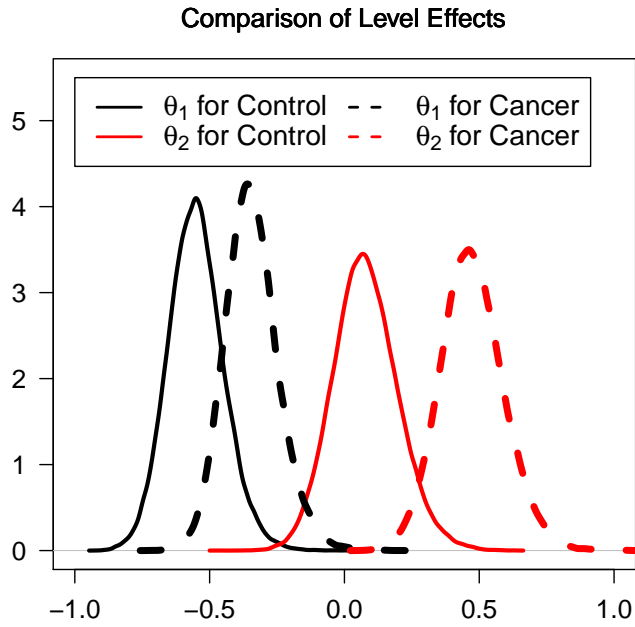


Figure 3.5: Comparison of Posterior Densities of θ_j , the QoL Effects for Both Control and Cancer Groups

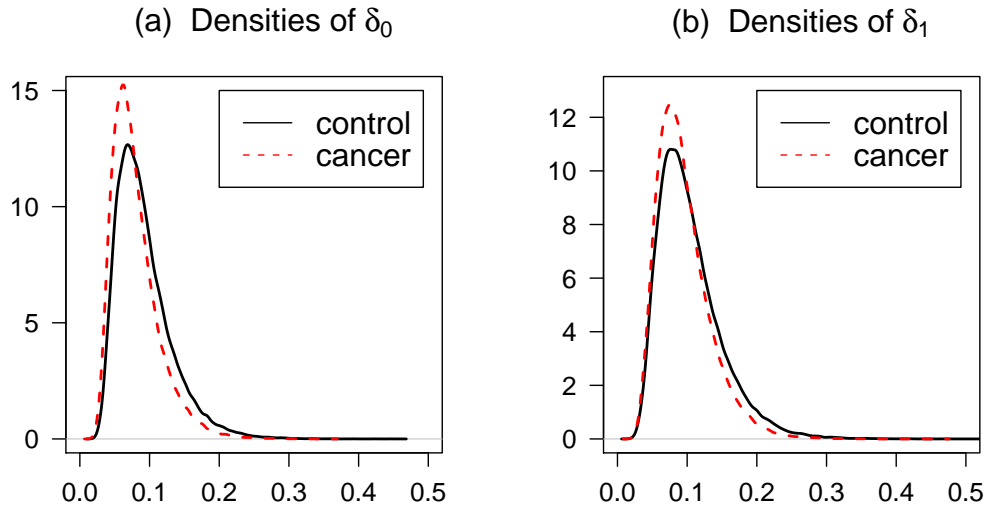


Figure 3.6: Comparison of Posterior Densities of δ_i , $i = 0, 1$, for Both Control and Cancer Groups

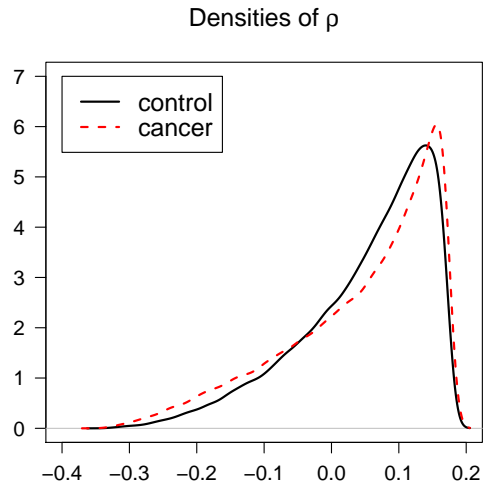


Figure 3.7: Comparison of Posterior Densities of ρ between Control and Cancer Groups

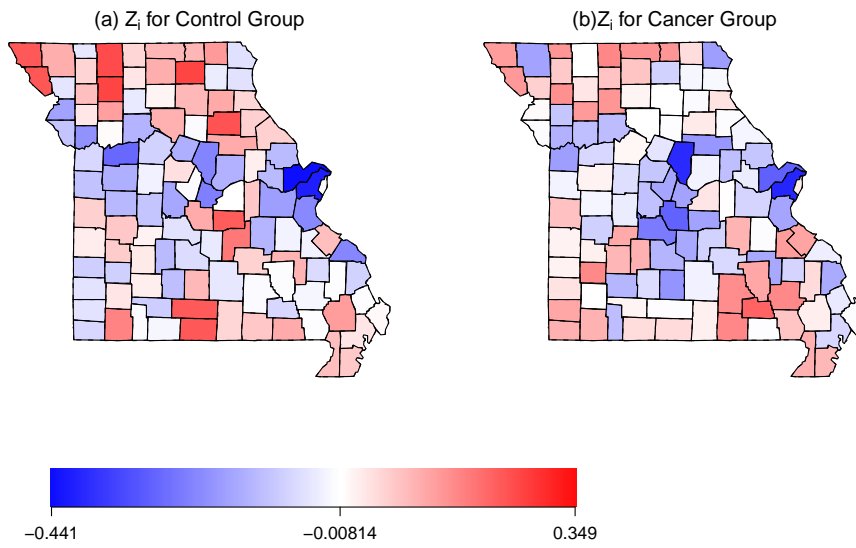


Figure 3.8: Comparison of Spatial Effects Z_i

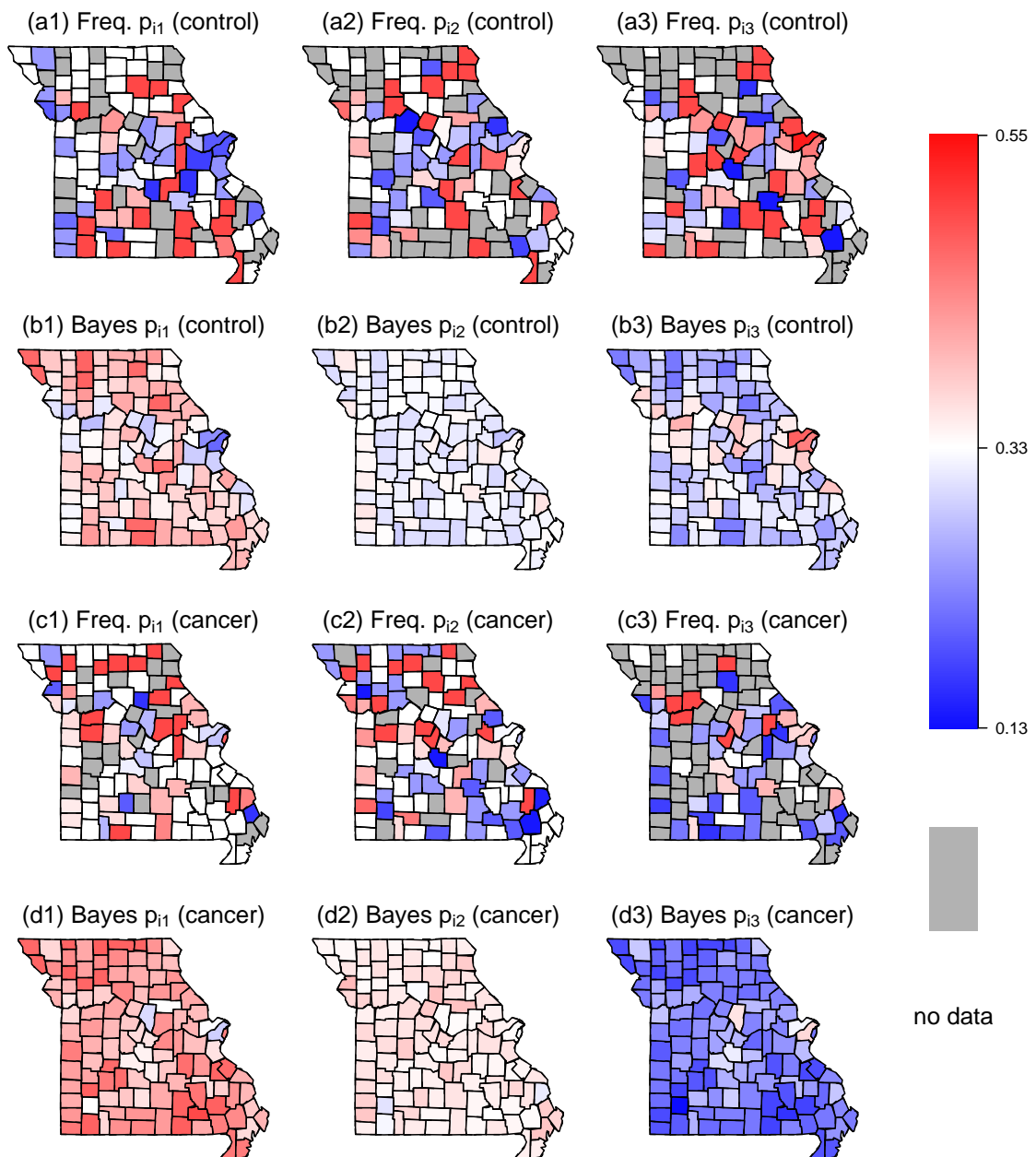


Figure 3.9: Comparison of Frequency and Bayesian Estimates of p_{ij}

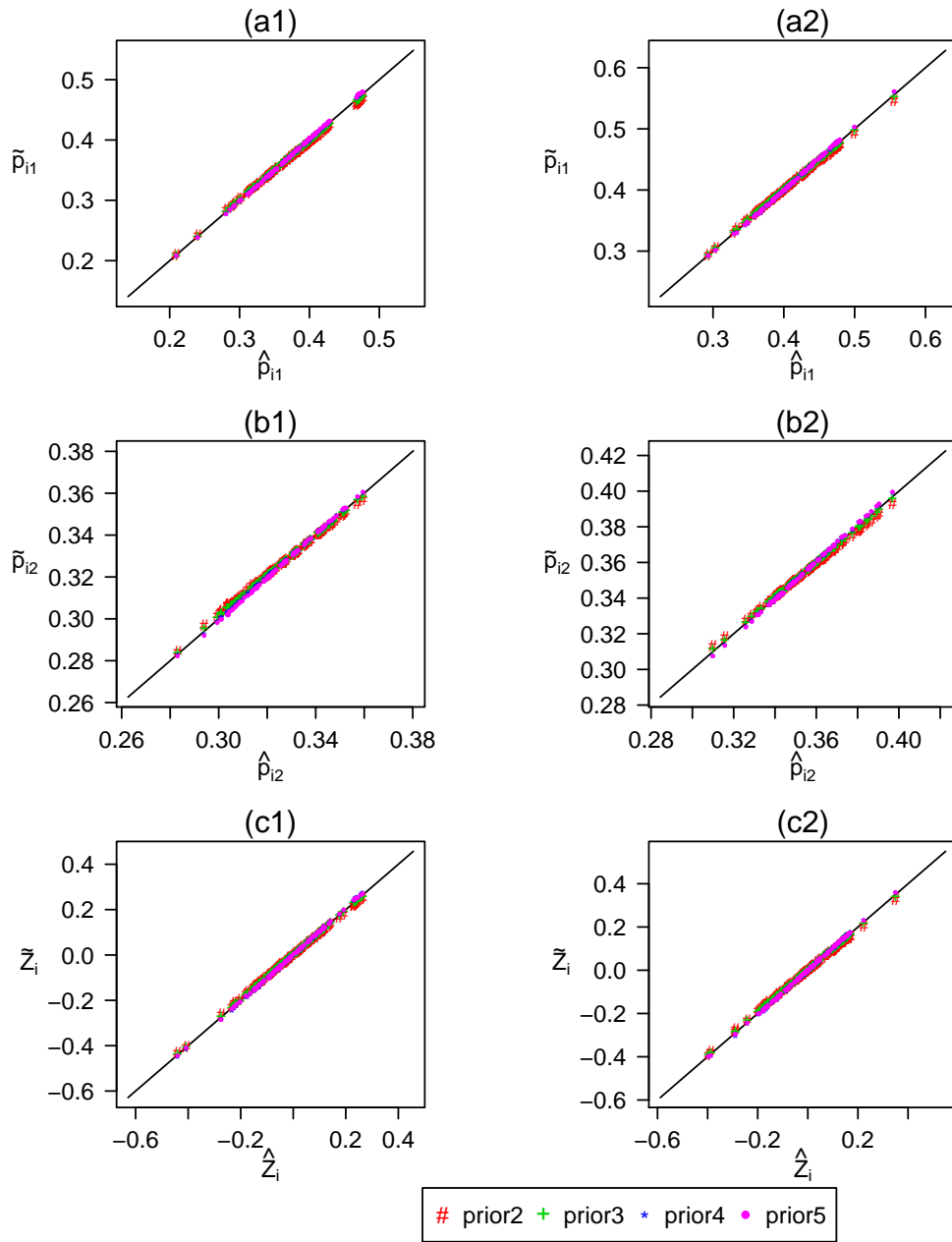


Figure 3.10: Robust Stretching Plots for Different Priors, (a1), (b1) and (c1) are for the Control Group; (a2), (b2) and (c2) are for the Cancer Group. The x labels, indicated as \hat{p}_{i1} , \hat{p}_{i2} , \hat{Z}_i are the estimates for the first prior, the y labels, indicated as \tilde{p}_{i1} , \tilde{p}_{i2} , \tilde{Z}_i , are the estimates of other priors.

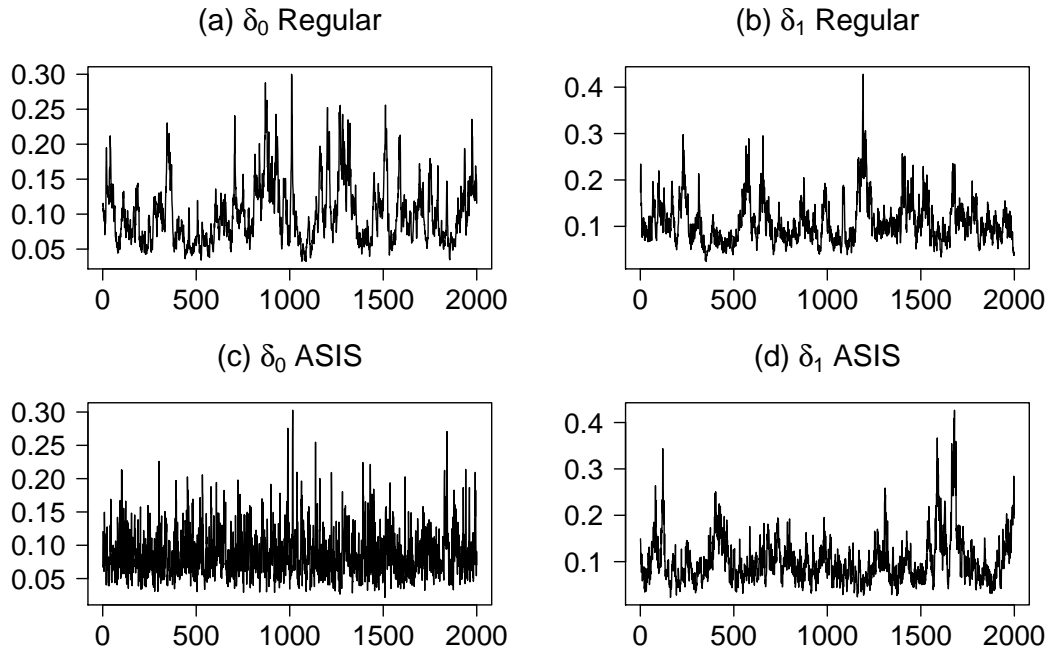


Figure 3.11: Comparison of Trace Plots of δ_k , $k = 0, 1$ based on Regular Gibbs Sampling and ASIS for the Control Group

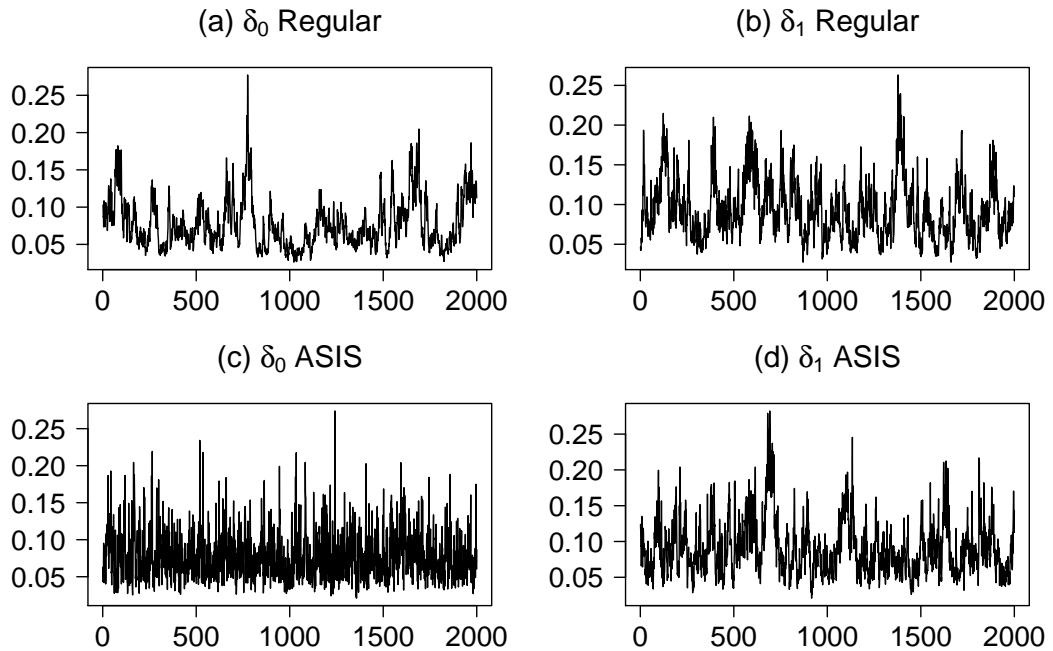


Figure 3.12: Comparison of Trace Plots of δ_k , $k = 0, 1$ based on Regular Gibbs Sampling and ASIS for the Cancer Group

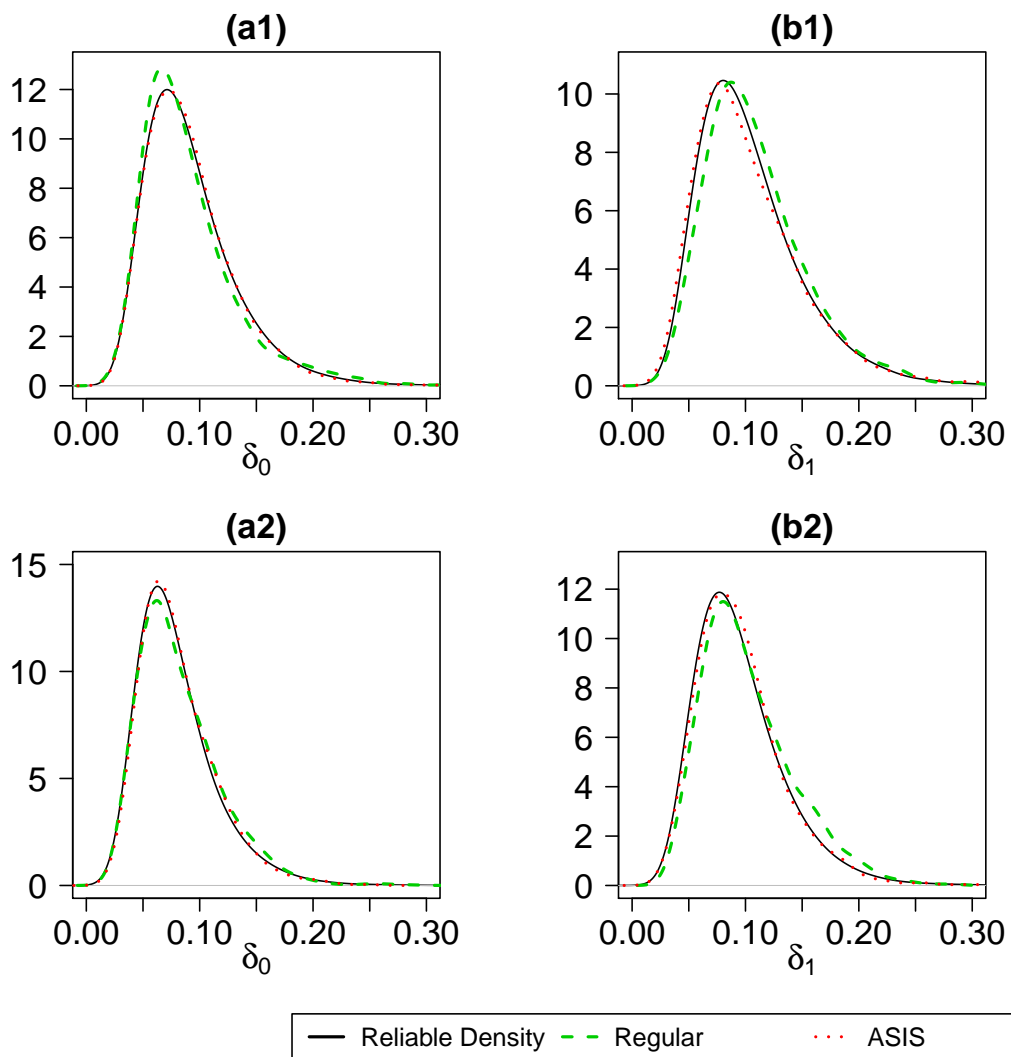


Figure 3.13: Density Comparison of first 3,000 iterations of δ_0 and δ_1 for Regular Gibbs Sampling and ASIS with the Reliable Density, (a1) and (b1) are for the Control Group, (a2) and (b2) are for the Cancer Group

Chapter 4

Hierarchical Model for Colorectal Cancer Survival Data

4.1 Introduction

Over that past twenty years, the statistical analysis of survival data has extensively developed to the application of Bayesian methodology. Breast cancer is a major cancer that interests many researchers. For instance, Omurlu et al. (2009) compared the performance of Cox regression analysis and Bayesian survival analysis by using simulations and breast cancer data. However, a very few statistical researches are based on the colorectal cancer. Nevertheless, of cancers affecting both men and women, colorectal cancer (cancer of the colon and rectum) is the second leading cancer killer in the United States. In the United States in 2008, 142,950 people were diagnosed with colorectal cancer, and 52,857 people died from it (USCS Working Group (2012)). Furthermore, colorectal cancer survivors make up the third largest group of cancer survivors, according to Ehemann et al. (2012). Therefore, it is necessary to pay more attention to it. This chapter will discuss the Bayesian survival analysis

on colorectal cancer data.

In modern survival analysis, more concerns are focused on geographical features of the data. Best et al. (2005) reviewed the main classes of spatial models that have been used for disease mapping within a Bayesian estimation paradigm. Banerjee, Wall & Carlin (2003) considered random effects corresponding to clusters that are spatially arranged, such as clinical sites or geographical regions.

For a large portion of censored data, Berkson & Gage (1952) pointed out that there exists a cure fraction to be considered in the survival analysis. Chen et al. (1999) derived the cure rate model by assuming that some latent risk with a Poisson distribution is generating the observations. Banerjee & Carlin (2004) extended the model with a spatial effect but treated the latent risk as having a binomial distribution in the smoking cessation setting. Cooner et al. (2008) propose a unifying class of cure rate models that facilitates flexible hierarchical model building while including both existing cure model classes as special cases.

In our model, we adopt the Poisson assumption of the latent risk and extend the model to the spatial context. The classic Weibull model can be applied as well to make the comparison.

The following sections are organized as follows. Section 4.2 describes the data specification and provides a summary for the random effects in the models. Section 4.3 illustrates both the Weibull model and the cure rate model in detail and proposes a hierarchical linear model with a CAR prior. Sampling and computation issues are discussed in Section 4.4. We use the ratio-of-uniforms method to sample a specific parameter. To validate the sampling, several theorems are proved. Model comparison and the analysis of the colorectal cancer survival are discussed in Section 4.5. Section 4.6 summarizes and discusses the future research.

4.2 Data Description

The data used in this chapter were obtained from the SEER 17 Registries Database of the Surveillance Epidemiology and End Results (SEER)¹, which contains full cancer records from 1973 to 2006 in the 17 population-based SEER registries. This chapter considers the Colon & Rectum Cancer (CRC) incidences in 99 counties of the state of Iowa (one of the 17 SEER registries) from 1992 to 2006, and censoring occurs at the end when the data collection was finished in 2006. Since the information of cancer patients is collected individually since 1973, it is type III censoring (Lee & Wang (2003)), or right censoring. According to Rowland et al. (2011), the largest numbers of survivors with colorectal cancer (625, 129) were aged 65 to 84 years in 2007, so we will take the threshold at 65 years or older. The 17, 669 subjects with first primary CRC, excluding death certificate only cases, are considered. The state and county neighboring data are from 2000 U.S. Census.

Table 4.1 to Table 4.3 list the frequency summaries of possible crucial effects of the survival time. Figure 4.1 shows the Kaplan-Meier survival function estimates across different stages. Figure 4.2 shows a comprehensive scatter plot matrix of these factors and the survival time.

We will consider the effect of age, county, positive regional lymph nodes, stage, tumor size and cause of death. To find the best fitting model, several reduced models are compared. By Table 4.1, the sample size for the “other” digestive system cancer cause is too small compared to the other causes, so we do not consider it separately but put it in the “other cancer” group instead.

For the censoring, we consider all the cases alive along with all the cases who died

¹Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Limited-Use Data (1973-2006), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2009, based on the November 2008 submission.

from causes other than colorectal cancer (including unknown causes, other cancer and non-cancer causes) as censored.

Suppose that for each county $i = 1, \dots, I$, there are n_i subjects, indicated by $j = 1, \dots, n_i$. The survival time t_{ij} and the right censoring time c_{ij} for the j^{th} subject in the i^{th} county (denoted by the $(i, j)^{\text{th}}$ subject) are observed. Define

$$\{y_{ij} = \min(t_{ij}, c_{ij}), d_{ij} = I(t_{ij} = y_{ij}); i = 1, \dots, I, j = 1, \dots, n_i\},$$

where $I(t_{ij} = y_{ij})$ is an indicator function with

$$\begin{cases} d_{ij} = 1 & \text{if } t_{ij} \leq c_{ij}, \\ d_{ij} = 0 & \text{if } t_{ij} > c_{ij}. \end{cases}$$

Therefore, the CRC data for the survival analysis can be represented by $K = \sum_{i=1}^I n_i$ pairs of random variables (y_{ij}, d_{ij}) .

Among all the CRC patients of 65 years or older in Iowa, as shown in Table 4.1, there are 7,165 censored ones, which accounts for 40.1% of the population. If all the non-CRC deaths are considered censored, the censoring rate would become 69.8%. The ages of the subjects ranges from 65 to 107. Based on the large sample size, we consider age as a continuous variable and the other factors as discrete. To simplify the models, it is also necessary to group some values of the effects. Kornprat et al. (2011) gave an optimal cut-off value of 45 millimeters within the whole group of colorectal cancer patients, but also suggested other cut-off values for different parts of the colon and rectum cancers. Based on this result and the histogram in Figure 4.2, we divide tumor size into five groups, as shown in Table 4.3. The positive regional lymph nodes effect is divided into five groups as well, as in Table 4.4, according to Edge et al. (2010).

4.3 Bayesian Models Based on the Weibull Distribution

Consider all the deaths of non CRC as censored, with continuous age, discrete county, stage, tumor size and positive regional lymph nodes. We consider two different survival models, one with the Cox model (Cox (1972)), and the other with the cure rate model (Chen et al. (1999)). Causes of death are only used to compare the results, since it makes no sense to include non-CRC deaths in the model when they are all considered alive.

4.3.1 Survival Models

A Weibull Survival Model

Assume that t_{ij} has a Weibull distribution $W(\alpha, \lambda_{ij})$, a special case of the Cox proportional hazard model (Cox (1972)). Then the density and hazard functions are given by $f(t_{ij} | \alpha, \lambda_{ij}) = \alpha \lambda_{ij} t_{ij}^{\alpha-1} \exp\{-\lambda_{ij} t_{ij}^\alpha\}$, and $h(t_{ij} | \alpha, \lambda_{ij}) = \alpha \lambda_{ij} t_{ij}^{\alpha-1}$, $t_{ij} > 0$, $\lambda_{ij} > 0$, $\alpha > 0$ for $i = 1, \dots, I$ and $j = 1, \dots, n_i$, respectively, where the λ_{ij} 's are the hazard rates.

We use a parametric approach to study the geographical variation and spatial effect. Let $\mathbf{y} = (y_{11}, \dots, y_{1n_1}, \dots, y_{I1}, \dots, y_{In_I})$, $\mathbf{d} = (d_{11}, \dots, d_{1n_1}, \dots, d_{I1}, \dots, d_{In_I})$, and $v_{ij} = \log(\lambda_{ij})$ for $i = 1, \dots, I$, $j = 1, \dots, n_i$. According to Dai et al. (2008b), the likelihood function of $\mathbf{v} = (v_{11}, \dots, v_{1n_1}, \dots, v_{I1}, \dots, v_{In_I})$ and α is

$$L(\alpha, \mathbf{v} | \mathbf{y}, \mathbf{d}) = \exp \left\{ \sum_{i=1}^I \sum_{j=1}^{n_i} \left(d_{ij} \left[\log(\alpha) + v_{ij} + (\alpha - 1) \log(y_{ij}) \right] - e^{v_{ij}} y_{ij}^\alpha \right) \right\}. \quad (4.1)$$

A Cure Rate Model

We adopt the Poisson assumption of the latent risk of the cure rate model in Chen et al. (1999), and suppose that N_{ij} is the number of fatal colorectal tumor cells in the $(i, j)^{th}$ individual, which is distributed as a Poisson random variable with mean κ_{ij} , i.e., $N_{ij} \mid \kappa_{ij} \sim \text{Poisson}(\kappa_{ij})$. An individual dies if $N_{ij} \geq 1$ and is considered censored otherwise. In this case, if one dies from cancers other than colorectal cancer, the event will still be considered as censored, which is consistent with our assumption. Let Z_{ijk} be the incubation time for the k^{th} fatal cell in the $(i, j)^{th}$ individual before activation, and assume that $Z_{ijk}, k = 1, 2, \dots$ are *i.i.d* distributed with *cdf* $F(t) = 1 - S(t)$ and are independent of N_{ij} . According to Chen et al. (1999), the survival function for the $(i, j)^{th}$ individual, hence for the populations is

$$S_c(t_{ij}) = \exp\{-\kappa_{ij}F(t_{ij})\} = \exp\{-\kappa_{ij}\} + (1 - \exp\{-\kappa_{ij}\})S^*(t_{ij}), \quad (4.2)$$

which is a standard cure rate model with $\exp\{-\kappa_{ij}\}$ as the cure rate and $S^*(\cdot)$ the survival function for the uncensored population. Recall from Section 4.2 that t_{ij} is the survival time of the $(i, j)^{th}$ individual. The corresponding density function is

$$f_c(t_{ij}) = \kappa_{ij}f(t_{ij}) \exp\{-\kappa_{ij}F(t_{ij})\}, \quad (4.3)$$

where $f(t) = (d/dt)F(t)$. In particular, the density and survival functions of t_{ij} given N_{ij} can be written as $f_c(t_{ij} \mid N_{ij}) = S(t_{ij})^{N_{ij}-1}(N_{ij}f(t_{ij}))$, and $S_c(t_{ij} \mid N_{ij}) = S(t_{ij})^{N_{ij}}$.

We assume a Weibull density for Z_{ijk} , with the survival and density functions as

$$S(t) = \exp\{-\lambda_{ij}t^\alpha\}, \text{ and } f(t \mid \alpha, \lambda_{ij}) = \alpha\lambda_{ij}t^{\alpha-1} \exp\{-\lambda_{ij}t^\alpha\}.$$

Letting $\mathbf{N} = (N_{11}, \dots, N_{1n_1}, \dots, N_{I1}, \dots, N_{In_I})$ and $\boldsymbol{\kappa} = (\kappa_{11}, \dots, \kappa_{1n_1}, \dots, \kappa_{I1})$, we continue using the notations in Section 4.3.1 about \mathbf{y} , \mathbf{d} and \mathbf{v} . Then the complete likelihood function can be written as

$$\begin{aligned} L(\alpha, \boldsymbol{\kappa}, \boldsymbol{\lambda} \mid \mathbf{y}, \mathbf{d}, \mathbf{N}) &= \prod_{i=1}^I \prod_{j=1}^{n_i} (N_{ij} f(y_{ij} \mid \alpha, \lambda_{ij}))^{d_{ij}} S(y_{ij} \mid \alpha, \lambda_{ij})^{N_{ij} - d_{ij}} \text{Poisson}(N_{ij} \mid \kappa_{ij}), \end{aligned}$$

where \mathbf{N} is an unobservable latent vector. By summing out \mathbf{N} , the likelihood function reduces to

$$\begin{aligned} L(\alpha, \boldsymbol{\kappa}, \boldsymbol{\lambda} \mid \mathbf{y}, \mathbf{d}) &= \prod_{i=1}^I \prod_{j=1}^{n_i} (f_c(y_{ij} \mid \alpha, \lambda_{ij}))^{d_{ij}} S_c(y_{ij} \mid \alpha, \lambda_{ij})^{1-d_{ij}} \\ &= \prod_{i=1}^I \prod_{j=1}^{n_i} (\kappa_{ij} f(y_{ij} \mid \alpha, \lambda_{ij}))^{d_{ij}} \exp \left\{ -\kappa_{ij} (1 - S(y_{ij} \mid \alpha, \lambda_{ij})) \right\} \\ &= \prod_{i=1}^I \prod_{j=1}^{n_i} (\kappa_{ij} \alpha \lambda_{ij} y_{ij}^{\alpha-1})^{d_{ij}} \\ &\quad \exp \left\{ -\left[d_{ij} \lambda_{ij} y_{ij}^\alpha + \kappa_{ij} (1 - \exp\{-\lambda_{ij} y_{ij}^\alpha\}) \right] \right\}, \quad (4.4) \end{aligned}$$

or equally

$$\begin{aligned} L(\alpha, \boldsymbol{\kappa}, \mathbf{v} \mid \mathbf{y}, \mathbf{d}) &= \exp \left\{ \sum_{i=1}^I \sum_{j=1}^{n_i} \left(d_{ij} \left[\log(\kappa_{ij}) + \log(\alpha) + v_{ij} + (\alpha - 1) \log(y_{ij}) \right] \right. \right. \\ &\quad \left. \left. - \left[d_{ij} e^{v_{ij}} y_{ij}^\alpha + \kappa_{ij} (1 - \exp\{-e^{v_{ij}} y_{ij}^\alpha\}) \right] \right) \right\}. \quad (4.5) \end{aligned}$$

4.3.2 A Hierarchical Linear Mixed Model

We consider the relationship between the transformed hazard rates v_{ij} 's and different effects including their interactions. The effects include:

- \mathbf{x}_{ij} , the vector of individual covariates including age, stage, tumor size, regional lymph nodes and interactions between these effects, etc.
- W_i , the spatial effect of the i^{th} county.

Therefore, the relationship can be formulated by a linear mixed model,

$$v_{ij} = \boldsymbol{\beta}' \mathbf{x}_{ij} + W_i + \epsilon_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, n_i, \quad (4.6)$$

where $\boldsymbol{\beta}$ is the parameter vector, and ϵ_{ij} is a residual effect including other sources of variability.

Define $\mathbf{X} = (\mathbf{x}'_{11}, \dots, \mathbf{x}'_{1n_1}, \dots, \mathbf{x}'_{I1}, \dots, \mathbf{x}'_{In_I})'$ as the covariates matrix, $\mathbf{W} = (W_1, \dots, W_I)'$, and $\boldsymbol{\epsilon} = (\epsilon_{11}, \dots, \epsilon_{In_I})'$. We can rewrite (4.6) into a matrix form,

$$\mathbf{v} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}\mathbf{W} + \boldsymbol{\epsilon}, \quad (4.7)$$

where $\boldsymbol{\eta} = \text{diag}(\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_I})$.

Suppose that ϵ_{ij} 's are i.i.d normally distributed with mean 0 and variance δ_0 . Then (4.7) follows a multivariate normal distribution,

$$(\mathbf{v} \mid \mathbf{X}; \boldsymbol{\beta}, \mathbf{W}, \delta_0) \sim N_K(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}\mathbf{W}, \delta_0 \mathbf{I}_K). \quad (4.8)$$

4.3.3 A CAR Model

Assume $\mathbf{W} = (W_1, \dots, W_I)'$ follow the simultaneous CAR prior of (2.2), in which \mathbf{Z} is replaced by \mathbf{W} . The spectral factorization of \mathbf{C} is

$$\mathbf{C} = \mathbf{G}\boldsymbol{\Lambda}\mathbf{G}', \quad (4.9)$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are the eigenvalues and \mathbf{G} is the orthogonal matrix whose columns are the corresponding eigenvectors. Let ζ_1 and ζ_I be minimum and maximum eigenvalues of the matrix \mathbf{C} , respectively. For the SEER dataset, ζ_1 and ζ_I are -3.282 and 6.545 (Sun et al. (1999)), respectively. So ρ is in the range $(-0.305, 0.153)$.

4.3.4 Other Priors

Since the latent risks \mathbf{N} are arising from the same cancer in the population, it is reasonable to make an assumption of common mean for all the N_{ij} , i.e., to use one common κ for all the κ_{ij} . Dai et al. (2008a) compared the linear mixed models on both the cure rate and the hazard rate and concluded that a linear mixed model on the hazard rate is much more preferable. Since we are interested in effects of geographical and other individual characteristics, which do not seem to be factors for the cure rate, the differences of the cancer risk in individuals can be ignored. In this case, $L(\alpha, \kappa, \boldsymbol{\lambda} \mid \mathbf{y}, \mathbf{d})$ in (4.4) becomes

$$\prod_{i=1}^I \prod_{j=1}^{n_i} (\kappa \alpha \lambda_{ij} y_{ij}^{\alpha-1})^{d_{ij}} \exp \left\{ - [d_{ij} \lambda_{ij} y_{ij}^{\alpha} + \kappa (1 - \exp\{-\lambda_{ij} y_{ij}^{\alpha}\})] \right\}. \quad (4.10)$$

Since $\exp(-\kappa)$ is the cure rate, whose range is between 0 and 1, we consider a gamma prior for κ with a distribution $gamma(a_c, b_c)$, which is conjugate.

Suppose $\alpha \sim gamma(a_s, b_s)$, where a_s is the shape parameter and b_s is the inverse scale parameter. We put a multivariate normal distribution on $\boldsymbol{\beta}$ with a diagonal covariance matrix $\boldsymbol{\tau}$ and mean $\boldsymbol{\mu}$ and independent inverse gamma distributions on covariance parameter: δ_0 and δ_1 , which are conjugate. Finally, we assume that $\rho \sim U(\zeta_1^{-1}, \zeta_I^{-1})$.

4.4 Computation

The computation is based on 3 chains of 20,000 samples with 5,000 burn-in period.

4.4.1 Model Selection

For the model fitting, we use the deviation information criterion (DIC) to compare the two different survival models, with the different combinations of the covariates and their interactions. According to (Gelman et al. (2004)), to calculate DIC, we need to find $\hat{D}_{avg}(y)$ and $D_{\hat{\theta}}(y)$. We know that

$$D(y, \boldsymbol{\theta}) = -2 \log(p(y | \boldsymbol{\theta})), \hat{D}_{avg}(y) = \frac{1}{L} \sum_{l=1}^L D(y, \boldsymbol{\theta}^l), \text{ and } D_{\hat{\theta}}(y) = D(y, \hat{\boldsymbol{\theta}}(y)),$$

where $\boldsymbol{\theta}$ is the general parameter, y is the data, and $l = 1, \dots, L$ is the simulation iterations. Therefore, for the Weibull model,

$$D(y, \boldsymbol{\theta}) = -2 \sum_{i=1}^I \sum_{j=1}^{n_i} \left(d_{ij} \left[\log(\alpha) + v_{ij} + \alpha \log(y_{ij}) - \log(y_{ij}) \right] - e^{v_{ij}} y_{ij}^{\alpha} \right),$$

and for the cure rate model,

$$D(y, \boldsymbol{\theta}) = 2K\kappa - 2 \sum_{i=1}^I \sum_{j=1}^{n_i} \left(d_{ij} \left[\log(\kappa) + \log(\alpha) + v_{ij} + (\alpha - 1) \log(y_{ij}) - e^{v_{ij}} y_{ij}^{\alpha} \right] + \kappa \exp\{-e^{v_{ij}} y_{ij}^{\alpha}\} \right).$$

Therefore, we have DIC and pD as

$$DIC = 2\hat{D}_{avg}(y) - D_{\hat{\theta}}(y), \text{ and } pD = \hat{D}_{avg}(y) - D_{\hat{\theta}}(y).$$

4.4.2 Full Conditional Posterior Distribution

The full conditional posterior distribution of all the parameters in both the Weibull and cure rate models are listed.

Weibull Model

Consider the observed data are $(\mathbf{y}, \mathbf{d}, \mathbf{X}, \mathbf{Z})$, the full conditional posterior distributions of all the parameters are as follows.

Proposition 3.

(a) Given $(\alpha, \boldsymbol{\beta}, \mathbf{W}, \delta_0; \text{data})$, v_{ij} 's are independent and the conditional posterior density is proportional to

$$\exp \left\{ d_{ij} v_{ij} - e^{v_{ij}} y_{ij}^\alpha - \frac{1}{2\delta_0} \left[v_{ij} - (\boldsymbol{\beta}' \mathbf{x}_{ij} + W_i) \right]^2 \right\}. \quad (4.11)$$

By Proposition 4, the densities are log-concave.

(b) Given $(\mathbf{v}; \mathbf{y}, \mathbf{d})$, the conditional posterior density of α is proportional to

$$\alpha^{a_s - 1 + \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij}} \cdot \exp \left\{ - \left(b_s - \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} \log y_{ij} \right) \alpha - \sum_{i=1}^I \sum_{j=1}^{n_i} e^{v_{ij}} y_{ij}^\alpha \right\}. \quad (4.12)$$

By Proposition 5, the density is log-concave.

(c) Given $(\mathbf{v}, \mathbf{W}, \delta_0; \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior distribution of $\boldsymbol{\beta}$ is

$$N \left(\left(\frac{1}{\delta_0} \mathbf{X}' \mathbf{X} + \boldsymbol{\tau}^{-1} \right)^{-1} \boldsymbol{\mu}_\beta, \left(\frac{1}{\delta_0} \mathbf{X}' \mathbf{X} + \boldsymbol{\tau}^{-1} \right)^{-1} \right), \quad (4.13)$$

where

$$\boldsymbol{\mu}_\beta = \frac{1}{\delta_0} \mathbf{X}'(\mathbf{v} - \boldsymbol{\eta}\mathbf{W}) + \boldsymbol{\tau}^{-1}\boldsymbol{\mu}.$$

(d) Define $\mathbf{B} = \mathbf{I}_I - \rho\mathbf{C}$. Given $(\mathbf{v}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{s}, \delta_0, \delta_1, \rho; \mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\eta})$, the conditional posterior distribution of \mathbf{W} is

$$N\left(\frac{1}{\delta_0} \left(\frac{1}{\delta_0} \boldsymbol{\eta}'\boldsymbol{\eta} + \frac{1}{\delta_1} \mathbf{B}\right)^{-1} \boldsymbol{\eta}'(\mathbf{v} - \mathbf{X}\boldsymbol{\beta}), \left(\frac{1}{\delta_0} \boldsymbol{\eta}'\boldsymbol{\eta} + \frac{1}{\delta_1} \mathbf{B}\right)^{-1}\right). \quad (4.14)$$

Obviously, $\boldsymbol{\eta}'\boldsymbol{\eta} = D$ where D is a diagonal matrix whose diagonal elements are (n_1, \dots, n_I) .

(e) Given $(\mathbf{v}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{s}, \mathbf{W}; \mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\eta})$, the conditional posterior distribution of δ_0 is

$$IG\left(a_0 + \frac{K}{2}, b_0 + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^{n_i} (v_{ij} - \boldsymbol{\beta}'\mathbf{x}_{ij} - W_i)^2\right), \quad (4.15)$$

or

$$IG\left(a_0 + \frac{K}{2}, b_0 + \frac{1}{2} (\mathbf{v} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\eta}\mathbf{W})'(\mathbf{v} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\eta}\mathbf{W})\right). \quad (4.16)$$

(f) Given (\mathbf{W}, ρ) , the conditional posterior distribution of δ_1 is

$$IG\left(a_1 + \frac{I}{2}, b_1 + \frac{1}{2} \mathbf{W}'(\mathbf{I}_I - \rho\mathbf{C})\mathbf{W}\right). \quad (4.17)$$

(g) Given (\mathbf{W}, δ_1) , the conditional posterior density of ρ is proportional to

$$|\mathbf{I}_I - \rho\mathbf{C}|^{1/2} \exp\left\{-\frac{1}{2\delta_1} \mathbf{W}'(\mathbf{I}_I - \rho\mathbf{C})\mathbf{W}\right\}. \quad (4.18)$$

Proposition 4. *The conditional density of v_{ij} given in (4.11) is log concave.*

Proof. Let $h(v_{ij})$ be the logarithm of the density in (4.11). Then we have

$$\frac{\partial^2}{\partial v_{ij}^2} h(v_{ij}) = -e^{v_{ij}} y_{ij}^\alpha - \frac{1}{\delta_0}, \quad (4.19)$$

which is negative. □

Proposition 5. *The conditional density of α given in (4.12) is log concave, and the α can be sampled by a rejection method if*

$$b_s - \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} \log y_{ij} > 0. \quad (4.20)$$

Proof. Let $h(\alpha)$ be the logarithm of the density in (4.12). Then we have

$$\frac{\partial^2}{\partial \alpha^2} h(\alpha) = -\frac{1}{\alpha^2} \left(a_s - 1 + \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} \right) - \sum_{i=1}^I \sum_{j=1}^{n_i} e^{v_{ij}} y_{ij}^\alpha (\log y_{ij})^2. \quad (4.21)$$

The second derivative is negative provided $a_s - 1 + \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} > 0$, which is always true if not all the subjects are censored.

However, we may use a rejection method to sample α . Note that

$$\exp \left\{ - \sum_{i=1}^I \sum_{j=1}^{n_i} e^{v_{ij}} y_{ij}^\alpha \right\} \leq 1,$$

so the unnormalized conditional density of α is $f(\alpha | \cdot) \leq M \cdot Ig(\alpha | \tilde{a}_s, \tilde{b}_s)$, where $\tilde{a}_s = a_s - 1 + \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij}$, $\tilde{b}_s = b_s - \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} \log y_{ij}$, and $M = \Gamma(\tilde{a}_s) / \tilde{b}_s^{\tilde{a}_s}$.

Consequently, the rejection method of sampling α is

Step 1: Draw α^* from $gamma(\tilde{a}_s, \tilde{b}_s)$,

Step 2: Draw u from $U(0, 1)$, if

$$u \leq \frac{f(\alpha^* | \cdot)}{M \cdot Ig(\alpha^* | \tilde{a}_s, \tilde{b}_s)} = \exp \left\{ - \sum_{i=1}^I \sum_{j=1}^{n_i} e^{v_{ij}} g_{ij}^{\alpha^*} \right\},$$

take α^* as the new α ; otherwise, go back to the first step.

This rejection method can be used as long as the condition (4.20) holds.

However, in practice, because of the large sample size, the above right-handed side term tends to be 0, the rejection method is not as efficient as the algorithm of Gilks & Wild (1992). Therefore, we do not use it in the sampling steps. \square

Proposition 6. *The conditional density of ρ given in (4.18) is log concave.*

Proof. By (4.9),

$$\mathbf{B} = \mathbf{I} - \rho \mathbf{C} = \mathbf{G}(\mathbf{I} - \rho \mathbf{\Lambda})\mathbf{G}', \text{ so } |\mathbf{B}| = \prod_{i=1}^I (1 - \rho \zeta_i).$$

Let $\mathbf{s} = (s_1, \dots, s_I)' = \mathbf{G}'\mathbf{W}$, we have

$$\mathbf{W}'\mathbf{B}\mathbf{W} = \mathbf{s}'(\mathbf{I} - \rho \mathbf{\Lambda})\mathbf{s} = \sum_{i=1}^I (1 - \rho \zeta_i) s_i^2. \quad (4.22)$$

Therefore,

$$(4.18) = \prod_{i=1}^I (1 - \rho \zeta_i)^{1/2} \exp \left\{ - \frac{1}{2\delta_0} \sum_{i=1}^I (1 - \rho \zeta_i) s_i^2 \right\}. \quad (4.23)$$

If $h(\rho)$ is the logarithm of the density in (4.18), we have

$$\frac{\partial^2 h(\rho)}{\partial \rho^2} = - \sum_{i=1}^I \frac{\zeta_i^2}{2(1 - \rho \zeta_i)^2} < 0,$$

so the result follows. \square

Cure Rate Model

The full conditional posterior distributions are as follows.

Proposition 7.

(a) Given $(\kappa, \alpha, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior density of v_{ij} 's are

$$[v_{ij} | \cdot] \propto \exp \left\{ d_{ij} v_{ij} - d_{ij} e^{v_{ij}} y_{ij}^\alpha + \kappa \exp \{ -e^{v_{ij}} y_{ij}^\alpha \} - \frac{1}{2\delta_0} (v_{ij} - \boldsymbol{\beta}' \mathbf{x}_{ij} - W_i)^2 \right\}. \quad (4.24)$$

Because of the extra term $\exp \{ \kappa \exp \{ -e^{v_{ij}} y_{ij}^\alpha \} \}$ compared to (4.11) (the extra d_{ij} for $e^{v_{ij}} y_{ij}^\alpha$ does not affect the log-concavity), the density function is not log-concave.

(b) Given $(\mathbf{v}, \kappa, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior density of α is

$$\begin{aligned} [\alpha | \mathbf{v}, \kappa; \mathbf{y}] &\propto \alpha^{a_s - 1 + \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij}} \exp \left\{ - \left(b_s - \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} \log y_{ij} \right) \alpha \right. \\ &\quad \left. - \sum_{i=1}^I \sum_{j=1}^{n_i} \left[d_{ij} e^{v_{ij}} y_{ij}^\alpha - \kappa \exp \{ -e^{v_{ij}} y_{ij}^\alpha \} \right] \right\}. \end{aligned} \quad (4.25)$$

For the same reason as in Proposition 7. (a), the density function is not log-concave either.

(c) Given $(\mathbf{v}, \alpha, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior distribution of κ is

$$\begin{aligned} (\kappa | \mathbf{v}, \alpha; \mathbf{y}, \mathbf{d}) &\sim \text{gamma} \left(a_c + \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij}, \right. \\ &\quad \left. b_c + K - \sum_{i=1}^I \sum_{j=1}^{n_i} \exp \{ -e^{v_{ij}} y_{ij}^\alpha \} \right). \end{aligned} \quad (4.26)$$

Since for every i and j , $\exp\{-e^{v_{ij}}y_{ij}^\alpha\} \leq 1$, the scale of the inverse gamma distribution is positive.

(d) The conditional posterior densities of other parameters are the same as listed in Proposition 3.

Sampling Strategies

The age variable is centered for computation, and the survival time is recorded by year.

Most of the conditional posterior densities are standard or log-concave, so they can be sampled directly from standard distributions or by using an adaptive rejection method. In this chapter, we use the Adaptive Rejection Algorithm (Gilks & Wild (1992)).

Proposition 8. *By the proof of Proposition 6, we give the sampling strategies for δ_1 and ρ . From (4.22), we get the steps of sampling the parameters.*

Step 1: Calculate $\mathbf{s} = \mathbf{G}'\mathbf{W}$ at the beginning of the sampling. \mathbf{G} does not change during the sampling.

Step 2: Find $\boldsymbol{\phi} = (1 - \rho\zeta_1, \dots, 1 - \rho\zeta_I)'$, sample δ_1 from $IG(a_1 + I/2, b_1 + \sum_{i=1}^I (1 - \rho\zeta_i)s_i^2)/2$.

Step 3: Sample ρ by the Adaptive Rejection Algorithm (Gilks & Wild (1992)) from (4.23).

Proposition 9. *The sampling of $\boldsymbol{\beta}$ based on Proposition 3. (c) can be simplified by separating $\boldsymbol{\beta}$ into different effects.*

Proof. Suppose there are L different types effects or interactions are included in the model, and $\boldsymbol{\beta}$ can be divided into $(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_L)$. $\boldsymbol{\beta}_l$ can either be a singular covariate for a continuous effect (or the intercept), or an indicator vector for a discrete effect. \mathbf{X} can be divided into different vectors or matrices $(\mathbf{X}_1, \dots, \mathbf{X}_L)$ correspondingly.

By this division, the $\boldsymbol{\beta}_l$ can be sampled separately as below. The conditional posterior distribution of $\boldsymbol{\beta}_l$ given $(\mathbf{v}, \boldsymbol{\beta}_{(-l)}, \mathbf{W}, \delta_0; \mathbf{X}, \boldsymbol{\eta})$ is

$$N\left(\left(\frac{1}{\delta_0}\mathbf{X}'_l\mathbf{X}_l + \boldsymbol{\tau}_l^{-1}\right)^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}_l}, \left(\frac{1}{\delta_0}\mathbf{X}'_l\mathbf{X}_l + \boldsymbol{\tau}_l^{-1}\right)^{-1}\right), \quad (4.27)$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_l} = \frac{1}{\delta_0}\mathbf{X}'_l(\mathbf{v} - \mathbf{X}_{(-l)}\boldsymbol{\beta}_{(-l)} - \boldsymbol{\eta}\mathbf{W}) + \boldsymbol{\tau}_l^{-1}\boldsymbol{\mu}. \quad (4.28)$$

Here $\boldsymbol{\beta}_{-l}$ indicates the vector removing $\boldsymbol{\beta}_l$, and $\mathbf{X}_{(-l)}$ the corresponding reduced covariate matrix.

What is interesting is that since $\boldsymbol{\beta}_l$ is an indicator vector or singular value, the corresponding \mathbf{X}_l would be simple enough so that $\mathbf{X}'_l\mathbf{X}_l$ becomes a singular value or a diagonal matrix, which would enhance the performance of the sampling greatly. \square

Proposition 10. *For the cure rate model samplings of v_{ij} , since the density function is not log-concave, we use a rejection method to sample them from the densities.*

Since $e^{v_{ij}y_{ij}^\alpha} \geq 0$, we have $\exp\{-e^{v_{ij}y_{ij}^\alpha}\} \leq 1$, so $\exp\{\kappa \exp(-e^{v_{ij}y_{ij}^\alpha})\} \leq e^\kappa$.

Let $f(v_{ij} | \cdot)$ be the unnormalized conditional density of v_{ij} in (4.24), and

$$g(v_{ij} | \cdot) = \exp\left\{d_{ij}v_{ij} - d_{ij}e^{v_{ij}y_{ij}^\alpha} - \frac{1}{2\delta_0}\left(v_{ij} - \beta_0 - \beta_1x_{ij} - \sum_{l=1}^L \xi_{ijl}s_l - \sum_{l=1}^L r_{ijl}\gamma_l - W_i\right)^2\right\},$$

then $f(v_{ij} | \cdot) \leq g(v_{ij} | \cdot)e^\kappa$. Since $g(v_{ij} | \cdot)$ is log-concave, which is obvious from Proposition 4, an exact sampling algorithm is as follows.

Step 1: Draw v_{ij}^* from $g(v_{ij} | \cdot)$.

Step 2: Draw u from $U(0, 1)$, if

$$u \leq \frac{f(v_{ij}^* | \cdot)}{e^\kappa \cdot g(v_{ij}^* | \cdot)} = \exp \left\{ -\kappa \left(1 - \exp \{ -e^{v_{ij}^*} y_{ij}^\alpha \} \right) \right\},$$

take v_{ij}^* as the new v_{ij} ; otherwise, go back to the first step.

When sampling α , a rejection method in the Gibbs Sampler is not very effective, so a ratio-of-uniforms method (Kinderman & Monahan (1977)) will be used to sample from (4.25).

4.4.3 Sampling by Ratio of Uniforms Method

For the cure rate model, the full conditional posterior density of α is not a standard form, nor is it log-concave. We alternatively use the ratio-of-uniforms method proposed by Kinderman & Monahan (1977) and later extended by Wakefield et al. (1991) to sample the shape parameter α from its conditional density.

The following corollary is derived directly from Wakefield et al. (1991).

Corollary 1. *Suppose f is a positive unimodal function on \mathcal{R} , and m is the mode. We assume $\sup_x [f(x)]^{1/(r+1)} < \infty$ and $\sup_x x [f(x)]^{1/(r+1)} < \infty$. If U, V are uniform on $\mathcal{A} = \{(u, v) : 0 < u < a(r), b^-(r) < v < b^+(r)\}$, where for $r \geq 0$,*

$$a(r) = \sup_x [f(x)]^{\frac{1}{r+1}}, \quad b^-(r) = \inf_{x \leq \mu} (x - \mu) [f(x)]^{\frac{1}{r+1}},$$

$$b^+(r) = \sup_{x \geq \mu} (x - \mu) [f(x)]^{\frac{1}{r+1}},$$

then $X = V/U + \mu$ has the density f , and the sampling acceptance rate is maximized at $\mu = m$.

Denoting the full conditional posterior density of α by $f(\alpha | \cdot)$, the following theorems will validate the steps of sampling α .

Theorem 1. Define $h_{1ij} = h(\alpha | d_{ij}, y_{ij}, v_{ij}) = d_{ij}e^{v_{ij}} - \kappa \exp\{-e^{v_{ij}}\}$, $K_{eij} = 1$ if $y_{ij} = 1$ and 0 otherwise, $K_{lij} = 1$ if $y_{ij} < 1$ and 0 otherwise, and $K_{gij} = 1$ if $y_{ij} > 1$ and 0 otherwise. Then $f(\alpha | \cdot)$ and $\alpha^{r+1}f(\alpha | \cdot)$ for $r \geq 0$ are bounded, given that \mathbf{v} , κ are fixed. An upper boundary of $f(\alpha | \cdot)$ is

$$C \left(\frac{A-1}{B} \right)^{A-1} \exp\{-(A-1)\} \exp \left\{ - \sum_{i=1}^I \sum_{j=1}^{n_i} [(K_{eij} + K_{gij})h_{1ij} + K_{lij}\kappa] \right\}. \quad (4.29)$$

An upper boundary of $\alpha^{r+1}f(\alpha | \cdot)$ is

$$C \left(\frac{A+1}{B} \right)^{A+1} \exp\{-(A+1)\} \exp \left\{ - \sum_{i=1}^I \sum_{j=1}^{n_i} [(K_{eij} + K_{gij})h_{1ij} + K_{lij}\kappa] \right\}. \quad (4.30)$$

Proof. Consider $f(\alpha | \cdot)$ first. Because of the complexity of this density function, different cases of y_{ij} s will be considered based on whether they are greater or less than 1.

Denote that

$$A = a_s + \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} > 0, \quad B = b_s - \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} \log y_{ij},$$

the normalized density can be written as

$$f(\alpha | \cdot) = C \alpha^{A-1} \exp\{-B\alpha\} \exp \left\{ - \sum_{i=1}^I \sum_{j=1}^{n_i} h(\alpha | d_{ij}, y_{ij}, v_{ij}) \right\}, \quad (4.31)$$

with

$$h(\alpha \mid d_{ij}, y_{ij}, v_{ij}) = d_{ij} e^{v_{ij}} y_{ij}^\alpha - \kappa \exp\{-e^{v_{ij}} y_{ij}^\alpha\},$$

C being the normalizing constant, and all $A, B, \mathbf{v}, \mathbf{y}, \kappa$ being fixed for each iteration. The difficulty of finding a boundary lies in the last term. A numerical method will be efficient and attractive, but we still give a rough boundary for theoretical base.

Note that $(A - 1)/B$ is the mode of $\alpha^{A-1} \exp\{-B\alpha\}$, we consider only the last term according to different y_{ij} s. We have

$$f(\alpha \mid \cdot) \leq C \left(\frac{A-1}{B} \right)^{A-1} \exp\{-(A-1)\} \exp \left\{ - \left(\sum_{y_{ij} < 1} + \sum_{y_{ij} > 1} + \sum_{y_{ij} = 1} \right) h(\alpha \mid d_{ij}, y_{ij}, v_{ij}) \right\}.$$

- If $y_{ij} = 1$, $h_{1ij} = h(\alpha \mid d_{ij}, y_{ij} = 1, v_{ij})$. We have

$$\exp \left\{ - \sum_{y_{ij} = 1} h(\alpha \mid d_{ij}, y_{ij}, v_{ij}) \right\} = \exp \left\{ - \sum_{i=1}^I \sum_{j=1}^{n_i} K_{eij} h_{1ij} \right\}.$$

- If $y_{ij} < 1$, then $0 \leq y_{ij}^\alpha < 1$. We have

$$\kappa \exp\{-e^{v_{ij}}\} < \kappa \exp\{-e^{v_{ij}} y_{ij}^\alpha\} \leq \kappa, \quad -d_{ij} e^{v_{ij}} < -d_{ij} e^{v_{ij}} y_{ij}^\alpha \leq 0,$$

so $-h_{1ij} < -h(\alpha \mid d_{ij}, y_{ij}, v_{ij}) \leq \kappa$, hence

$$\begin{aligned} \exp \left(- \sum_{i=1}^I \sum_{j=1}^{n_i} K_{lij} h_{1ij} \right) &< \exp \left\{ - \sum_{y_{ij} < 1} h(\alpha \mid d_{ij}, y_{ij}, v_{ij}) \right\} \\ &\leq \exp \left\{ \sum_{i=1}^I \sum_{j=1}^{n_i} K_{lij} \kappa \right\}. \end{aligned}$$

- If $y_{ij} > 1$, then $1 < y_{ij}^\alpha < \infty$. We have

$$0 < \kappa \exp\{-e^{v_{ij}} y_{ij}^\alpha\} < \kappa \exp\{-e^{v_{ij}}\}, \quad -d_{ij} \times \infty < -d_{ij} e^{v_{ij}} y_{ij}^\alpha < -d_{ij} e^{v_{ij}},$$

so $-d_{ij} \times \infty < -h(\alpha \mid d_{ij}, y_{ij}, v_{ij}) < -h_{1ij}$, hence

$$0 < \exp\left\{-\sum_{y_{ij}>1} h(\alpha \mid d_{ij}, y_{ij}, v_{ij})\right\} < \exp\left\{-\sum_{i=1}^I \sum_{j=1}^{n_i} K_{g_{ij}} h_{1ij}\right\}.$$

Therefore, an upper boundary of $f(\alpha \mid \cdot)$ is (4.29). Similarly, an upper boundary of $\alpha^{r+1} f(\alpha \mid \cdot)$ is (4.30). \square

Thus $a(r)$, $b^-(r)$ and $b^+(r)$ all exist according to this theorem. However, because of the complexity of $f(\alpha \mid \cdot)$, it would be very difficult to find their theoretical values.

The following theorem provides the unimodal condition when looking for the numerical value of $a(r)$ by using the Newton-Raphson method.

Theorem 2. *$f(\alpha \mid \cdot)$ has only one mode if the censoring rate is less than $1 - \kappa e^{-2}$, given that \mathbf{v} , κ are fixed.*

Proof. To prove that $f(\alpha \mid \cdot)$ has only one mode, we consider the derivative of its logarithm. We continue using the notations in (4.31), and we have

$$\begin{aligned} \log f(\alpha \mid \cdot) = & \log C + (A - 1) \log \alpha - B\alpha \\ & - \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} e^{v_{ij}} y_{ij}^\alpha + \kappa \sum_{i=1}^I \sum_{j=1}^{n_i} \exp\{-e^{v_{ij}} y_{ij}^\alpha\}. \end{aligned}$$

Set its derivative to 0, and we get

$$\begin{aligned} \frac{\partial}{\partial \alpha} \log f(\alpha | \cdot) &= \frac{1}{\alpha}(A-1) - B - \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} e^{v_{ij}} y_{ij}^{\alpha} \log y_{ij} \\ &- \kappa \sum_{i=1}^I \sum_{j=1}^{n_i} e^{v_{ij}} y_{ij}^{\alpha} \log y_{ij} \exp\{-e^{v_{ij}} y_{ij}^{\alpha}\} = 0. \end{aligned} \quad (4.32)$$

Note that the left term tends to be $+\infty$ when $\alpha \rightarrow 0$, and it tends to be $-\infty$ when $\alpha \rightarrow +\infty$, because of the continuity, there must exist at least one value of α such that the equality holds. Therefore, $f(\alpha | \cdot)$ must have at least one critical points (maxima or minima).

To prove the critical point is unique and it's maxima, we only need to show the second derivative is negative hence the first derivative is monotonously decreasing.

We have

$$\begin{aligned} \frac{\partial^2}{\partial \alpha^2} \log f(\alpha | \cdot) &= -\frac{1}{\alpha^2}(A-1) - \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} e^{v_{ij}} y_{ij}^{\alpha} (\log y_{ij})^2 \\ &- \kappa \sum_{i=1}^I \sum_{j=1}^{n_i} e^{v_{ij}} y_{ij}^{\alpha} [1 - e^{v_{ij}} y_{ij}^{\alpha}] (\log y_{ij})^2 \exp\{-e^{v_{ij}} y_{ij}^{\alpha}\} \\ &= -\frac{1}{\alpha^2}(A-1) - \sum_{i=1}^I \sum_{j=1}^{n_i} e^{v_{ij}} y_{ij}^{\alpha} (\log y_{ij})^2 \\ &\quad \left[d_{ij} + \kappa \exp\{-e^{v_{ij}} y_{ij}^{\alpha}\} (1 - e^{v_{ij}} y_{ij}^{\alpha}) \right]. \end{aligned} \quad (4.33)$$

One sufficient condition of (4.33) being negative is that

$$\sum_{i=1}^I \sum_{j=1}^{n_i} \left[d_{ij} + \kappa \exp\{-e^{v_{ij}} y_{ij}^{\alpha}\} (1 - e^{v_{ij}} y_{ij}^{\alpha}) \right] \geq 0. \quad (4.34)$$

Note that $e^{v_{ij}} y_{ij}^{\alpha} \geq 0$, and recall that for $w \geq 0$, $g(w) = e^{-w}(1-w)$ has a behavior as in Figure 4.3. The minimum value for $g(w)$ is $-e^{-2}$ when $w = 2$. So a sufficient

condition of (4.34) is

$$\sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} \geq \sum_{i=1}^I \sum_{j=1}^{n_i} \kappa e^{-2}, \quad (4.35)$$

or

$$D/K \geq \kappa e^{-2},$$

where D is the total of death in the population and K is the total of the population. Therefore, the condition that the censoring rate is less than $1 - \kappa e^{-2}$ is a sufficient condition for the unimodal conditional posterior density of α . \square

Since the colorectal data has the censoring rate of 69.8%, as long as $\kappa \leq 2.23$, the conditional posterior density is unimodal.

It is also easy to show that the following corollary holds, and the unimodal condition can be used to find $b^-(r)$ and $b^+(r)$.

Corollary 2. $(\mu - x)^{r+1}f(x)$ is unimodal if $x \leq \mu$; $(x - \mu)^{r+1}f(x)$ is unimodal if $x \geq \mu$.

To generate samples of X , the algorithm is as follows.

Step 1: Find the mode m of $f(x)$ by the Newton-Raphson method.

Step 2: Find $a(r)$, $b^-(r)$ and $b^+(r)$ as described in Corollary 1.

Step 3: Simulate $U \sim Uniform(0, a(r))$, $V \sim Uniform(b^-(r), b^+(r))$, and compute

$$x = V/U + m.$$

Step 4: If $U < [f(x)]^{1/(r+1)}$, accept x as a new X ; otherwise, return to Step 3.

We may choose an appropriate r for a better acceptance rate. From our experience, $r = 1$ has already provided a good acceptance rate of about 71.4%, so an optimal r is not necessary. Figure 4.4 shows the acceptance region with the ratio-of-uniforms method when sampling α for a typical iteration.

4.5 Application

We applied this method to the Iowa colorectal cancer cases from the SEER data. Table 4.5 lists the model specifications with different effects and their interactions. Both the Weibull and cure rate models were run. Each model took less than one hour. The DIC of the models and the estimates of the common parameters α , δ_0 , δ_1 , and ρ are listed in Table 4.6.

For the cure rate models, the estimates of the parameter κ are listed in Table 4.7. There is a trend that the more effects that are included in the models, the greater the estimated mean of the latent fatal colorectal tumor cells becomes, hence the smaller the cure rate is. This can partially be explained by the reason that as more effects are considered, the hazard rates of the subjects tend to have less variability (the error variance tends to decrease) between censored and uncensored cases. So it is harder to detect the real cure rate from the model.

By the DIC comparison, the best model is Model W4 for the Weibull models. For this particular data set, the Weibull models perform better than the cure rate models. For both survival models, if only a few effects, say, age and county effects only included, the models do not converge well, and both have very large error variances. However, as we increase the amount of the effects, and their interactions correspondingly, the performance does not always get better. In fact, if all the interactions between different effects are included, neither the Weibull models nor the cure rate

models converge. The figure compares the posterior means of different effects.

Obviously, the different stages have very different effects. The more distant the stage is, the more effect it has on the response. The differences of different positive regional lymph nodes are also very obvious: more nodes lead to greater hazard rates. However, the tumor size effect does not have such a clear distinction. If the tumor is small, say, less than 20, it has a relatively smaller hazard rate; otherwise, the trend is not so clear-cut.

The estimated spatial effects of different counties and their standard errors can be visualized on an Iowa map, as shown in Figure 4.6.

We could not include the causes of death into the model since we considered all non-CRC death as censoring. However, we can still compare them with different stages. Figure 4.7 gives the Bayesian survival estimates of all the subjects for different causes of death, including cases still alive.

4.6 Discussion

In summary, we analyzed the colorectal cancer survival data in Iowa with both the Weibull and the cure rate models. Several effects and the spatial effect are incorporated in the model. The Weibull model performs better for this particular data set with a high censoring rate. Moreover, stages and positive regional lymph nodes affect the survival response significantly.

Although the project shows some significant results, there are some limitations. For many cancers, including breast and colorectal cancers, survival depends on a variety of factors. Some, such as stage at diagnosis and tumor size, we were able to consider. We were not able to consider other important factors such as treatment (treatment type, length, intensity, etc.) in the model due to inadequate or incomplete

data. In addition, other factors such as income, family history, genetics, environmental exposures, comorbidities, etc. may also contribute to patients' survival. Without including some or all of these factors, we cannot expect full understanding of the data.

In further research, the tumor size effect can be grouped to see whether larger tumors may have a more significant effect than smaller tumors. The cure rate model can be considered again if treating all deaths (including non-CRC deaths), in which the censoring rate would be less, and the sampling procedure would be more convincing. Moreover, the causes of death can also be modeled with competing risks, so that we can have a deeper understanding of them.

Table 4.1: Cause of Death Summary, other digestive cancer are included in the “other Cancer” case, and the percentages are included in parentheses

COD	Alive	Colorectal	Other Cancer	Non-Cancer	Unknown
	7165(40.1%)	5334(30.2%)	793(4.5%[0.9%])	4287(24.3%)	90(0.1%)

Table 4.2: Stage Grouping, the percentages are included in parentheses

Stage	In Situ	Localized	Regional	Distant	Unknown
Frequency	663(3.8%)	6717(38.0%)	6505(36.8%)	2904(16.4%)	880(5.0%)

Table 4.3: Tumor Size (in mm) Grouping, the larger tumor sizes (≥ 60 mm) are grouped in the models because of their relatively small sample sizes

Group	1	2	3	4	5
Size (in mm)	0 – 19	20 – 39	40 – 59	over 60	Missing
Frequency	1076	3824	4261	3506	4406

Table 4.4: Positive Regional Lymph Nodes, the numbers are grouped in the models based on their sample sizes and their scatterplots in Figure 4.2

Group	1	2	3	4	5
Values	0	1 – 3	4 – 9	≥ 10	.
Freq.	8320	2838	1346	351	4218

Table 4.5: Model Specifications

Model	Effects included
Model 1	age+county
Model 2	age+stage+county
Model 3	age+stage+age \times stage+county
Model 4	age+stage+tumor+regional nodes+county

Table 4.6: DIC Comparison and Common Parameter Estimates of Different Models. W as Weibull model, C as the cure rate model. The standard errors are given in parentheses

Survival Model	DIC	α	δ_0	δ_1	ρ
W1	27866	0.971(0.018)	4.280(0.190)	0.050(0.013)	-0.008(0.056)
W2	25462	1.006(0.033)	1.680(0.215)	0.073(0.037)	0.004(0.056)
W3	24789	1.070(0.023)	1.588(0.146)	0.033(0.009)	-0.007(0.055)
W4	24655	1.061(0.022)	1.168(0.112)	0.032(0.008)	-0.006(0.055)
C1	28725	1.608(0.049)	4.974(0.356)	0.078(0.025)	-0.008(0.055)
C2	26163	1.135(0.020)	1.243(0.100)	0.040(0.010)	-0.007(0.055)
C3	26496	1.030(0.023)	0.621(0.127)	0.035(0.008)	-0.005(0.055)
C4	25784	1.045(0.019)	0.592(0.081)	0.031(0.008)	-0.005(0.055)

Table 4.7: The Cure Rate Estimates of the Cure Rate Models with Different Effects, the standard errors are given in parentheses

Model	C1	C2	C3	C4
κ	0.616(0.016)	2.551(0.110)	2.771(0.147)	3.410(0.234)
Cure Rate	54.01%	7.80%	6.26%	3.30%

Table 4.8: The Estimated Effects and their Standard Errors for Models W4 and C4, the standard errors are included in parentheses

Coefficients	Mean (Std) of Model W4		Mean (Std) of Model C4	
Intercept	-1.3195	(0.0732)	-2.3285	(0.0988)
Age	0.3944	(0.0206)	0.3867	(0.0197)
Stage (Unknown= 0)				
In Situ	-4.0842	(0.2190)	-3.9500	(0.2057)
Localized	-2.0580	(0.0989)	-2.0372	(0.0940)
Regional	-0.7252	(0.0908)	-0.7286	(0.0895)
Distant	1.6983	(0.0834)	1.6603	(0.0816)
Tumor Size (Missing= 0)				
0 - 19	-0.6200	(0.1211)	-0.5948	(0.1139)
20 - 39	-0.1299	(0.0627)	-0.1342	(0.0619)
40 - 59	-0.1430	(0.0610)	-0.1537	(0.0605)
≥ 60	-0.0311	(0.0611)	-0.0285	(0.0616)
Positive Regional Nodes (Missing= 0)				
0	-1.3265	(0.0604)	-1.2760	(0.0607)
1 - 3	-0.7953	(0.0684)	-0.7627	(0.0663)
4 - 9	-0.1335	(0.0725)	-0.1048	(0.0730)
≥ 10	0.4372	(0.1054)	0.3901	(0.1014)

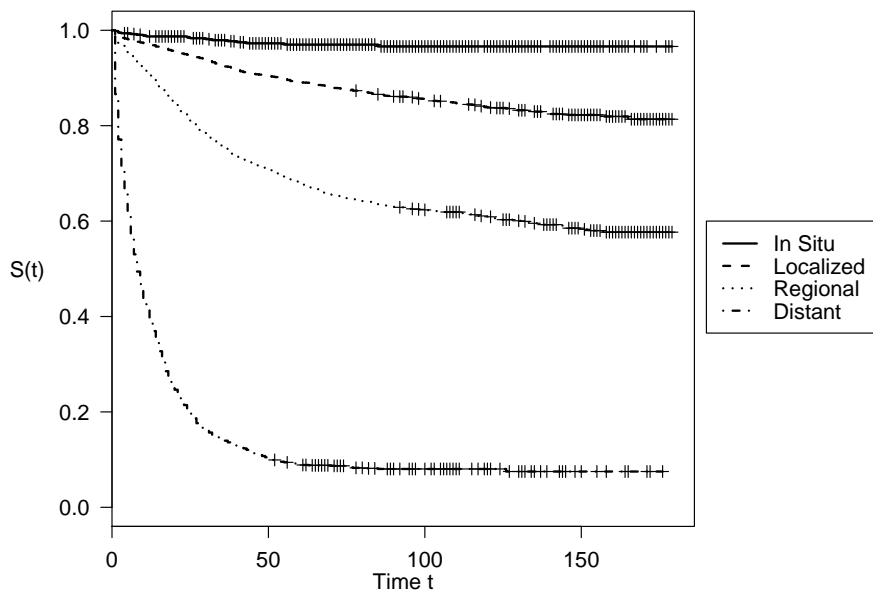


Figure 4.1: Kaplan-Meier Survival Function Estimates with Different Stages

Iowa State; population: 17669

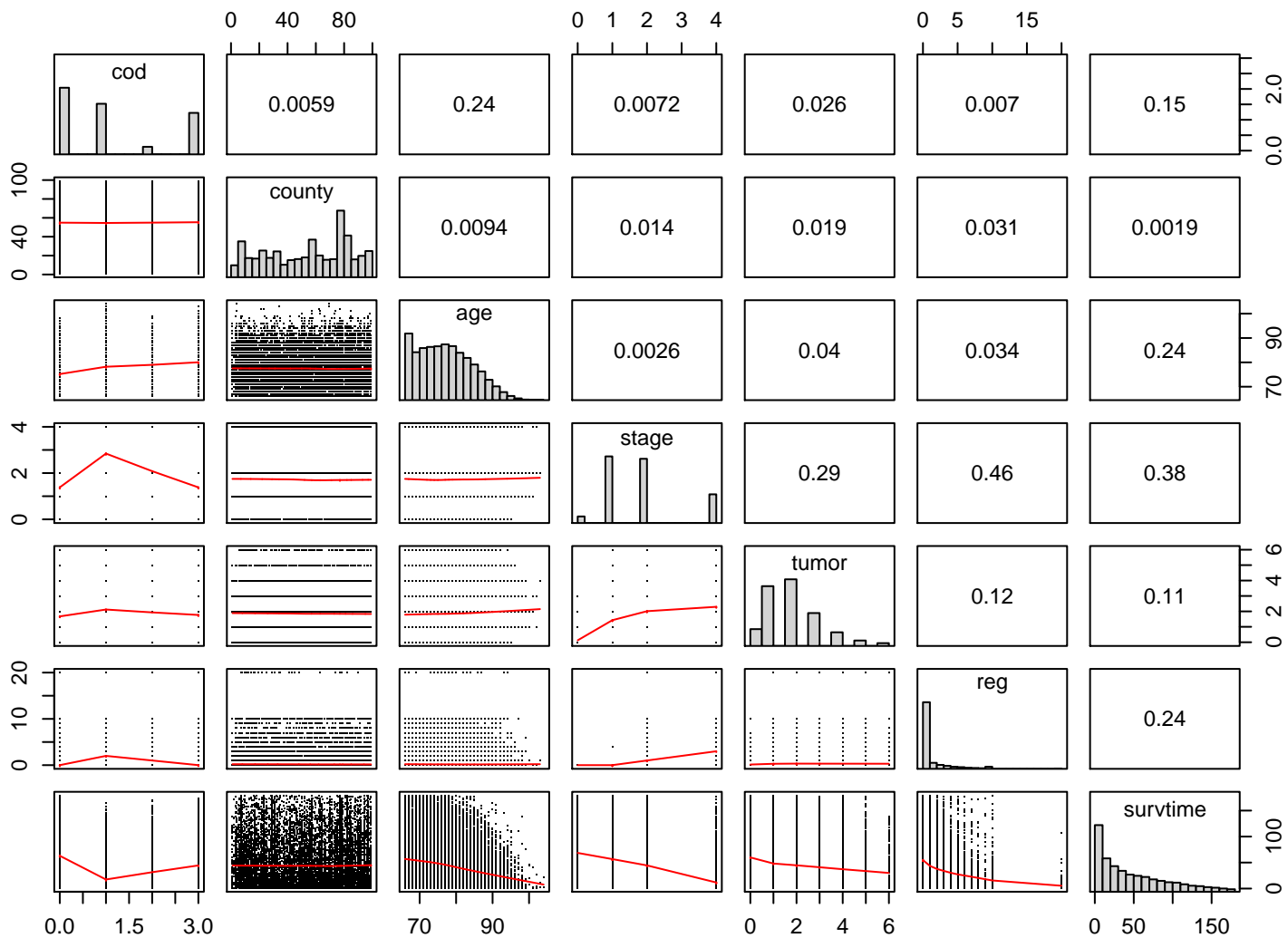


Figure 4.2: Correlation Scatterplot Matrix of Important Effects

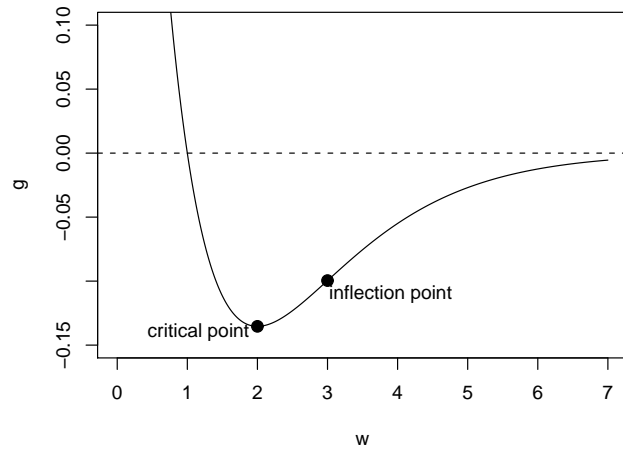


Figure 4.3: Behavior of $g(w)$, the critical point indicates the minimum value and at the inflection point, the function becomes concave from convex.

Acceptance Region for sampling of α

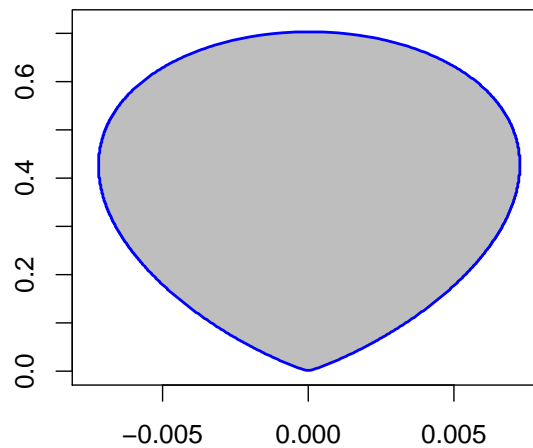
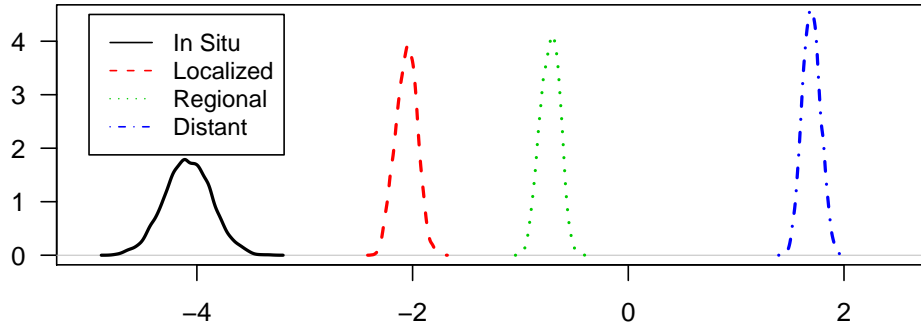
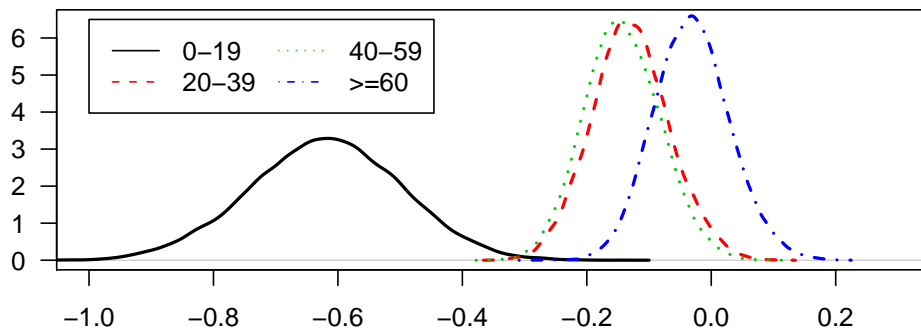


Figure 4.4: Acceptance Region with the Ratio-of-Uniforms Method When Sampling α for Cure Rate Model.

(a) Posterior Densities of Historic Stages Effect



(b) Posterior Densities of Tumor Size



(c) Posterior Densities of Positive Regional Nodes

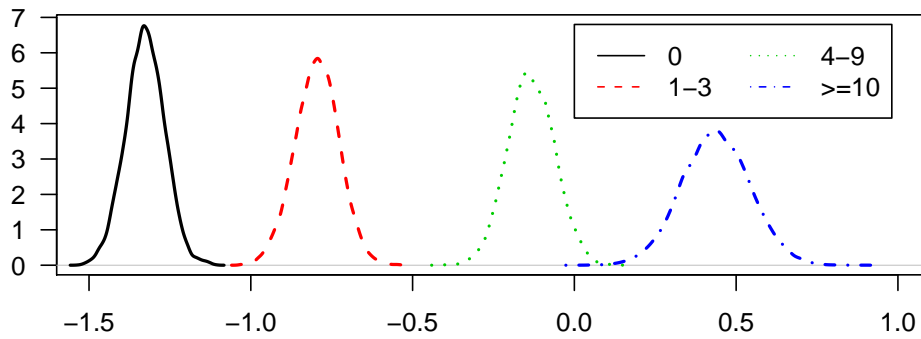


Figure 4.5: Posterior Densities of the Effects

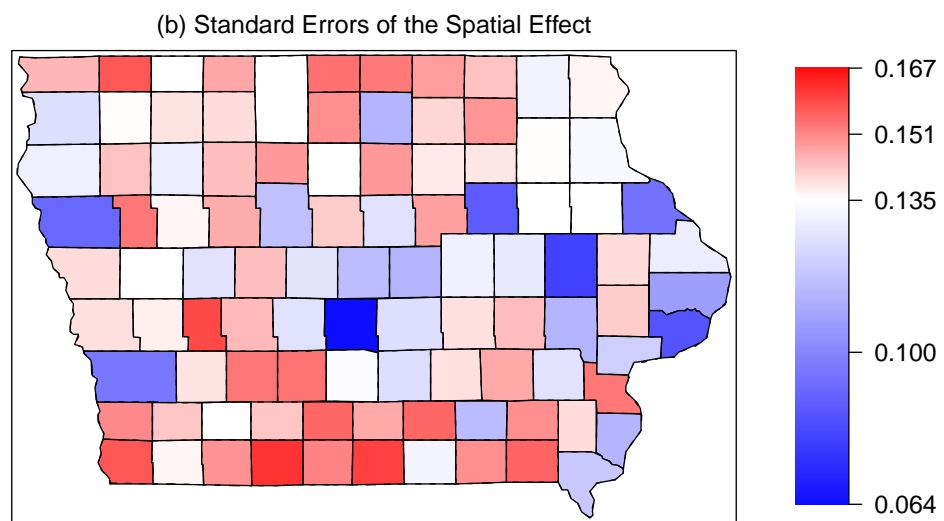
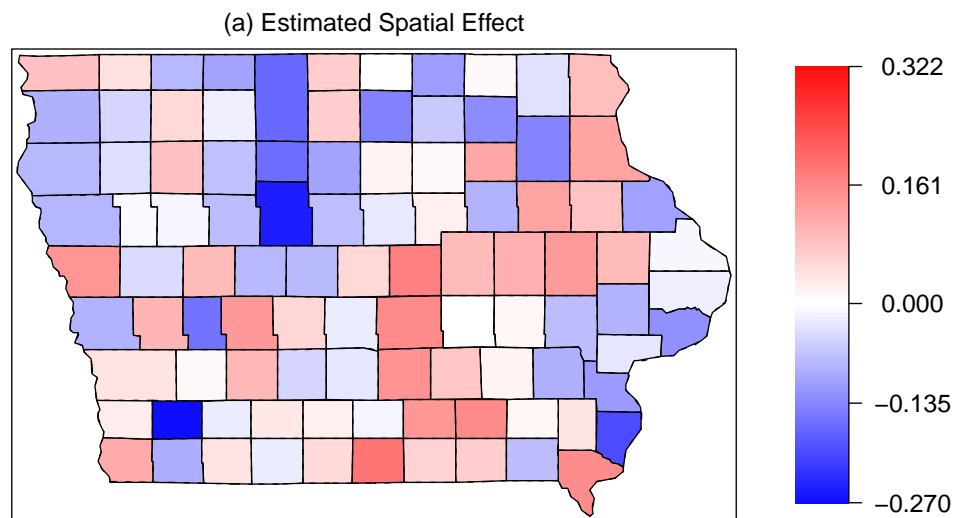


Figure 4.6: Estimated Spatial Effect and Standard Errors

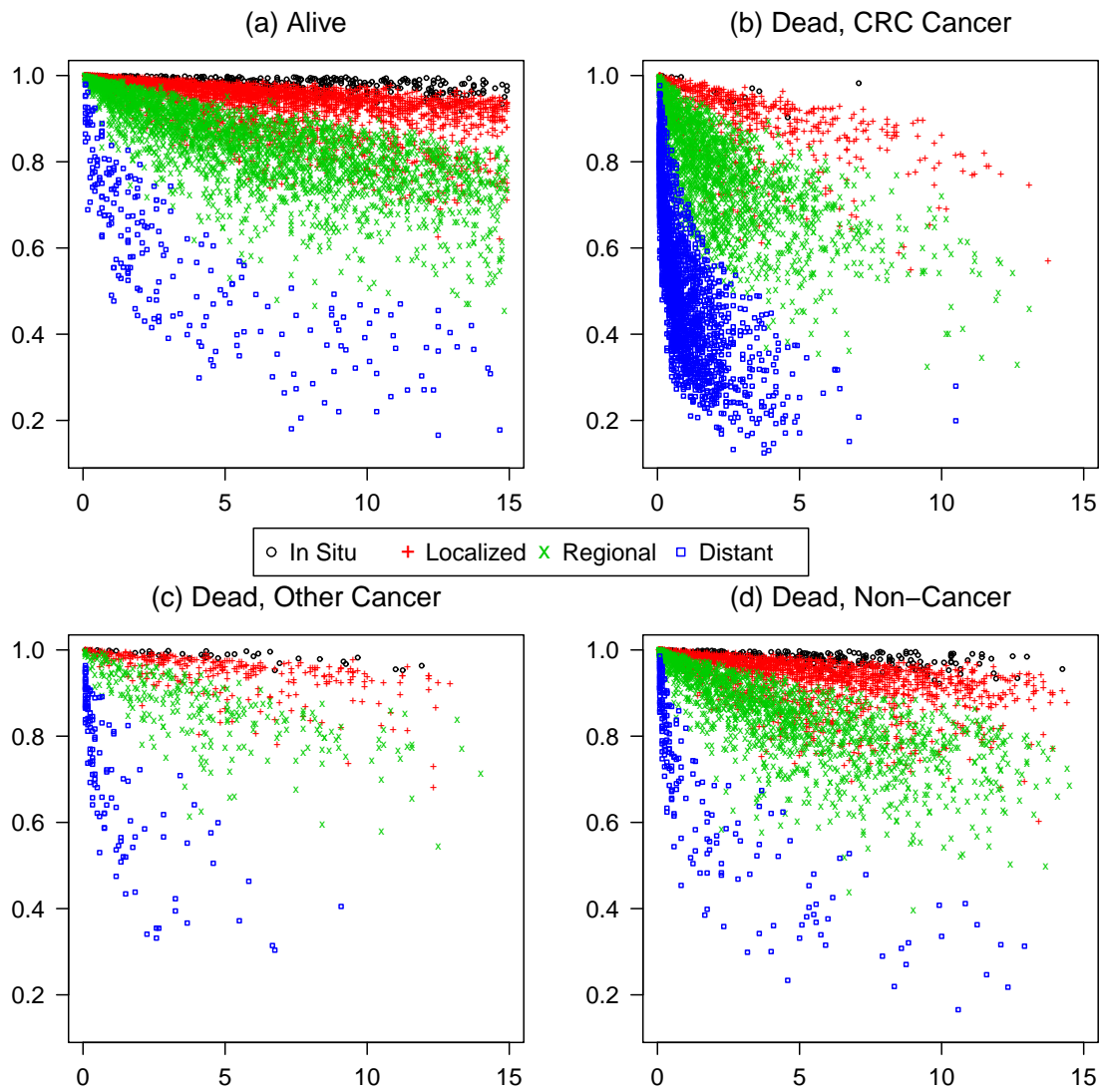


Figure 4.7: The Survival Estimates for Stages with Different Causes of Death, the x-axis is the survival years, and the y-axis is the survival estimates

Chapter 5

A Latent Competing Risk Approach with A Hierarchical CAR Model for Colorectal Cancer Survival Data

5.1 Introduction

We adopt a parametric proportional hazards specification in Gelfand et al. (2000) and suppose proportional hazard of the general form $h(t; v) = h_0(t)g(v)$, where $h_0 > 0$ is the baseline hazard, and $g(\cdot)$ is a strictly increasing function with v including covariates and random effects. Gelfand et al. (2000) suggested that $h_0(t)$ having the form $h_0(t) = \sum_{l=1}^r h_l(t)$, with r varying and $h_l(t)$ being different hazard for each l . We suppose $r = 3$ is fixed as we consider three competing risks. The first one is the specific cancer, e.g. colorectal cancer; the second competing risk is other cancers; and the third competing risk is causes other than cancer. Cases with unknown causes will not be considered. For simplicity, we suppose that all three hazards follow the family

of Weibull distribution. Berger & Sun (1993) introduced an efficient method to do Gibbs sampling for the latent variables arising from the competing risk setting. We will use their method.

For $g(v)$, we assume that v is a random variable of the linear combination of different covariates. The covariates include age (continuous), stage (4 levels), the county an individual belongs to (99 levels), etc. We chose $g(\cdot) = \exp(\cdot)$ since it is robust with respect to inference (Cox & Oakes (1984)).

5.2 Data Description

The same data set described in Chapter 4 Section 4.2 is used in this chapter. We continue using the notations of n_i , t_{ij} , and c_{ij} in Chapter 4.

We do not consider the unknown deaths from Table 4.1 since they are deaths with no reported case.

5.3 Modeling

5.3.1 Latent Competing Risk Hazard Model

Assume that $T_{ij} = \min\{U_{ij1}, \dots, U_{ijl}\}$ with some associated covariates v_{ij} . We can view U_{ijl} as the observed failure time for the l^{th} risk with the j^{th} patient in the i^{th} county. We suppose that U_{ijl} has a proportional hazard form $h_l(u)g(v)$. Then the density and survival functions are $f_l(u) = h_l(u)g(v) \exp\{-g(v)H_l(u)\}$ and $S_l(u) = \exp\{-H_l(u)g(v)\}$, respectively, with $H_l(u) = \int_0^u h_l(v)dv$. We also assume the U_{ijl} are independent from each other for different i , j and l .

We specify $h_l(u) = \alpha_l \lambda_l u^{\alpha_l - 1}$ as a Weibull hazard $W(\alpha_l, \lambda_l)$ for each l , and let

$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_l)$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_l)$. Then the likelihood without censoring for the failure time $\mathbf{T} = \{\mathbf{T}_1, \dots, \mathbf{T}_I\}$, where $\mathbf{T}_i = \{T_{i1}, \dots, T_{in_i}\}$, is given by

$$\begin{aligned} L(\boldsymbol{\alpha}, \boldsymbol{\lambda} \mid \mathbf{t}) &= \prod_{i=1}^I \prod_{j=1}^{n_i} g(v_{ij}) \sum_{l=1}^r h_l(t_{ij} \mid \alpha_l, \lambda_l) \prod_{l=1}^r S_l(t_{ij} \mid \alpha_l, \lambda_l) \\ &= \prod_{i=1}^I \prod_{j=1}^{n_i} g(v_{ij}) \sum_{l=1}^r \alpha_l \lambda_l t_{ij}^{\alpha_l - 1} \exp \left\{ - \sum_{l=1}^r H_l(t_{ij} \mid \alpha_l, \lambda_l) g(v_{ij}) \right\}. \end{aligned} \quad (5.1)$$

Define $\{y_{ij} = \min(t_{ij}, c_{ij}), d_{ij} = I(t_{ij} = y_{ij}); i = 1, \dots, I, j = 1, \dots, n_i\}$, where $I(t_{ij} = y_{ij})$ is an indicator function, i.e.,

$$d_{ij} = 1 \text{ if } t_{ij} \leq c_{ij}, \quad d_{ij} = 0 \text{ if } t_{ij} > c_{ij}.$$

Therefore, the CRC data for the survival analysis can be represented by $K = \sum_{i=1}^I n_i$ pairs of random variables (y_{ij}, d_{ij}) . We use a parametric approach to study the geographical variation and spatial effect. Let $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_I)$, $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})$, $\mathbf{d} = (\mathbf{d}_1, \dots, \mathbf{d}_I)$, and $\mathbf{d}_i = (d_{i1}, \dots, d_{in_i})$ for $i = 1, \dots, I, j = 1, \dots, n_i$. The likelihood of the observed data becomes

$$\begin{aligned} L(\boldsymbol{\alpha}, \boldsymbol{\lambda} \mid \mathbf{y}, \mathbf{d}) &= \prod_{i=1}^I \prod_{j=1}^{n_i} \left(g(v_{ij}) \sum_{l=1}^r h_l(y_{ij} \mid \alpha_l, \lambda_l) \prod_{l=1}^r S_l(y_{ij} \mid \alpha_l, \lambda_l) \right)^{d_{ij}} \\ &\quad \left(\prod_{l=1}^r S_l(y_{ij} \mid \alpha_l, \lambda_l) \right)^{1-d_{ij}}. \end{aligned} \quad (5.2)$$

For $i = 1, \dots, I, j = 1, \dots, n_i$, and $l = 1, \dots, r$, define $I_{ijl} = I(t_{ij} = u_{ijl})$, where $I(\cdot)$ is the indicator function, and denote $\mathbf{I} = (\mathbf{I}_1, \dots, \mathbf{I}_I)$, $\mathbf{I}_i = (\mathbf{I}_{i1}, \dots, \mathbf{I}_{in_i})$ and $\mathbf{I}_{ij} = (I_{ij1}, \dots, I_{ijl})$. These indicators can be used as ancillary random variables.

Let $\mathbf{U} = (\mathbf{U}_1, \dots, \mathbf{U}_I)$, $\mathbf{U}_i = (\mathbf{U}_{i1}, \dots, \mathbf{U}_{in_i})$, and $\mathbf{U}_{ij} = (U_{ij1}, \dots, U_{ijl})$. Gelfand et al. (2000) used this latent variable for sampling. If we consider \mathbf{U} as the complete

data, under the independence assumption, the likelihood becomes

$$\begin{aligned}
L(\boldsymbol{\alpha}, \boldsymbol{\lambda} \mid \mathbf{u}) &= \prod_{i=1}^I \prod_{j=1}^{n_i} \prod_{l=1}^r g(v_{ij}) h_l(u_{ijl} \mid \alpha_l, \lambda_l) S_l(u_{ijl} \mid \alpha_l, \lambda_l) \\
&= \prod_{i=1}^I \prod_{j=1}^{n_i} \left(g(v_{ij}) \right)^r \prod_{l=1}^r \alpha_l \lambda_l u_{ijl}^{\alpha_l - 1} \exp \left\{ - \sum_{l=1}^r H_l(u_{ijl} \mid \alpha_l, \lambda_l) g(v_{ij}) \right\} \\
&= \prod_{i=1}^I \prod_{j=1}^{n_i} \left(g(v_{ij}) \right)^r \prod_{l=1}^r \alpha_l \lambda_l u_{ijl}^{\alpha_l - 1} \exp \left\{ - \sum_{l=1}^r \lambda_l u_{ijl}^{\alpha_l} g(v_{ij}) \right\} \\
&= \prod_{i=1}^I \prod_{j=1}^{n_i} f(\mathbf{u}_{ij} \mid \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \prod_{i=1}^I \prod_{j=1}^{n_i} \prod_{l=1}^r f(u_{ijl} \mid \boldsymbol{\alpha}, \boldsymbol{\lambda}). \tag{5.3}
\end{aligned}$$

However, according to Berger & Sun (1993), using the indicator \mathbf{I} as a latent variable is much more efficient for sampling. We may write the likelihood for the complete data as follows.

$$\begin{aligned}
L(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I} \mid \mathbf{y}, \mathbf{d}) &= \prod_{i=1}^I \prod_{j=1}^{n_i} \prod_{l=1}^r \left(f_l(y_{ij} \mid \alpha_l, \lambda_l)^{I_{ijl}} S_l(y_{ij} \mid \alpha_l, \lambda_l)^{1-I_{ijl}} \right)^{d_{ij}} \\
&\quad S_l(y_{ij} \mid \alpha_l, \lambda_l)^{1-d_{ij}} \\
&= \prod_{i=1}^I \prod_{j=1}^{n_i} \prod_{l=1}^r f_l(y_{ij} \mid \alpha_l, \lambda_l)^{d_{ij} I_{ijl}} S_l(y_{ij} \mid \alpha_l, \lambda_l)^{1-d_{ij} I_{ijl}}. \tag{5.4}
\end{aligned}$$

If the Weibull baseline hazard is plugged in, the likelihood becomes

$$L(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I} \mid \mathbf{y}, \mathbf{d}) = \prod_{i=1}^I \prod_{j=1}^{n_i} \prod_{l=1}^r \left(\alpha_l \lambda_l y_{ij}^{\alpha_l - 1} g(v_{ij}) \right)^{d_{ij} I_{ijl}} \exp \left\{ - \lambda_l y_{ij}^{\alpha_l} g(v_{ij}) \right\}, \tag{5.5}$$

and the log likelihood $l(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I} \mid \mathbf{y}, \mathbf{d})$ is

$$\sum_{i=1}^I \sum_{j=1}^{n_i} \sum_{l=1}^r \left[d_{ij} I_{ijl} (\log \alpha_l + \log \lambda_l + (\alpha_l - 1) \log y_{ij} + \log g(v_{ij})) - \lambda_l y_{ij}^{\alpha_l} g(v_{ij}) \right]. \tag{5.6}$$

5.3.2 A Linear Mixed Model

For $g(v_{ij})$, we assume that $g(v_{ij}) = \exp(v_{ij})$, and that

$$v_{ij} = \boldsymbol{\beta}' \mathbf{x}_{ij} + W_i + \epsilon_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, n_i, \quad (5.7)$$

where $\boldsymbol{\beta}$ is the regression coefficient vector for the effects, W_i is the spatial effect for the i^{th} county, and ϵ_{ij} is a residual effect including other unknown sources of variability. There is no intercept in this linear model because it has been included in $\boldsymbol{\lambda}$. Define $\mathbf{v} = (v_{11}, \dots, v_{1n_1}, \dots, v_{I1}, \dots, v_{In_I})'$, $\mathbf{X} = (x_{11}, \dots, x_{1n_1}, \dots, x_{I1}, \dots, x_{In_I})'$, $\mathbf{W} = (W_1, \dots, W_I)'$ and $\boldsymbol{\epsilon} = (\epsilon_{11}, \dots, \epsilon_{1n_1}, \dots, \epsilon_{I1}, \dots, \epsilon_{In_I})'$. We can rewrite (5.7) in matrix form

$$\mathbf{v} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}\mathbf{W} + \boldsymbol{\epsilon}, \quad (5.8)$$

where $\boldsymbol{\eta} = \text{diag}(\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_I})$. It is obvious that $\mathbf{1}_K = \boldsymbol{\eta} \cdot \mathbf{1}_I$.

Assume that the ϵ_{ij} 's are i.i.d normally distributed with mean 0 and variance δ_0 . Then (5.8) follows a multivariate normal distribution,

$$(\mathbf{v} \mid \mathbf{X}; \boldsymbol{\beta}, \mathbf{W}, \delta_0) \sim N_K(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}\mathbf{W}, \delta_0 \mathbf{I}_K). \quad (5.9)$$

5.3.3 A CAR Model

Assume $\mathbf{W} = (W_1, \dots, W_I)'$ follows the simultaneous CAR prior of (2.2), in which \mathbf{Z} is replaced by \mathbf{W} . Let ζ_1 and ζ_I be the minimum and maximum eigenvalues of the matrix \mathbf{C} , respectively. Then they have the same values as in Section 4.3.3.

5.3.4 Priors

Suppose $\alpha_l \sim \text{gamma}(a_s, b_s)$, $l = 1, \dots, r$, where a_s is the shape parameter and b_s is the inverse scale parameter. To identify different hazards, we assume that the α_l 's are order statistics with $\alpha_r > \dots > \alpha_1$. Identifiability is discussed in Section 5.4.1. In the application, we take a flat prior with $a_s = 1$ and $b_s = 0$.

Let $\lambda_l \sim \text{gamma}(a_h, b_h)$, $l = 1, \dots, r$, where a_h is the shape parameter and b_h is the inverse scale parameter, and take a flat prior with $a_h = 1$ and $b_h = 0$. We also suppose that $\beta_m \sim N(\mu_m, \tau_m)$, $m = 1, 2$, $\delta_k \sim IG(a_k, b_k)$, $k = 0, 1$, and $\rho \sim U(\zeta_1^{-1}, \zeta_I^{-1})$.

5.4 Computation

5.4.1 Identifiability

To avoid the identifiability problem, the shapes for the Weibull hazards of different risks need to be ordered. If all α_l are greater than 1, the hazards are increasing. Since we are primarily interested in colorectal cancer survival, we assume that the colorectal cancer has the highest risk. From Table 4.1, the order of the risks goes by the colorectal cancer, other causes of deaths (including unknown cases), and other cancers. We assume $\alpha_1 < \dots < \alpha_r$.

5.4.2 Model Selection

For model selection, we use the deviation information criterion (DIC) to compare different models. According to (Gelman et al. (2004)), to calculate $DIC = 2\hat{D}_{avg}(y) -$

$D_{\hat{\theta}}(y)$, we need to find $\hat{D}_{avg}(y)$ and $D_{\hat{\theta}}(y)$. We know that

$$D(y, \boldsymbol{\theta}) = -2 \log(L(y | \boldsymbol{\theta})), \hat{D}_{avg}(y) = \frac{1}{L} \sum_{l=1}^L D(y, \boldsymbol{\theta}^l), \text{ and } D_{\hat{\theta}}(y) = D(y, \hat{\boldsymbol{\theta}}(y)),$$

where $\boldsymbol{\theta}$ is the general parameters, y is the data, and $l = 1, \dots, L$ is the simulation iterations, and that

$$\begin{aligned} D(y, \boldsymbol{\theta}) &= -2 \sum_{i=1}^I \sum_{j=1}^{n_i} \sum_{l=1}^r [d_{ij} I_{ijl} (\log \alpha_l + \log \lambda_l + (\alpha_l - 1) \log y_{ij} + v_{ij}) - \lambda_l y_{ij}^{\alpha_l} e^{v_{ij}}]. \end{aligned}$$

5.4.3 Sampling

We use Gibbs sampling to complete the sampling. Consider the observed data are $(\mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, and define $\boldsymbol{\lambda}_{(-l)} = \{\lambda_p, 1 \leq p \leq r, p \neq l\}$, $\boldsymbol{\alpha}_{(-l)} = \{\alpha_p, 1 \leq p \leq r, p \neq l\}$, $\mathbf{I}_{(-ij)} = \{\mathbf{I}_{mp}, 1 \leq m \leq I, 1 \leq p \leq n_m, m \neq i, p \neq j\}$, the full conditional posterior distributions of all the parameters are as follows.

Proposition 11.

(a) Given $(\boldsymbol{\alpha}_{(-l)}, \boldsymbol{\lambda}, \mathbf{I}, \mathbf{v}, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior density of α_l is

$$\begin{aligned} [\alpha_l | \lambda_l, \mathbf{v}; \text{data}] &\propto \alpha_l^{a_s - 1 + \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} I_{ijl}} \exp \left\{ - \left(b_s - \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} I_{ijl} \log y_{ij} \right) \alpha_l \right. \\ &\quad \left. - \sum_{i=1}^I \sum_{j=1}^{n_i} \lambda_l e^{v_{ij}} y_{ij}^{\alpha_l} \right\}, \end{aligned} \quad (5.10)$$

with $\alpha_{l-1} < \alpha_l < \alpha_{l+1}$, where $\alpha_0 = 0$ and $\alpha_{r+1} = \infty$.

(b) Given $(\boldsymbol{\alpha}, \boldsymbol{\lambda}_{(-l)}, \mathbf{I}, \mathbf{v}, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior density of λ_l is

$$(\lambda_l \mid \alpha_l, \mathbf{I}_l, \mathbf{v}; \text{data}) \sim \text{gamma}\left(a_h + \sum_{i=1}^I \sum_{j=1}^{n_i} d_{ij} I_{ijl}, b_h + \sum_{i=1}^I \sum_{j=1}^{n_i} e^{v_{ij}} y_{ij}^{\alpha_l}\right). \quad (5.11)$$

(c) Given $(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{v}, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, $\mathbf{I}_{ij}, i = 1, \dots, I, j = 1, \dots, n_i$ are independent. The dead and censored cases should be considered separately.

- If $d_{ij} = 0$, we have $\mathbf{I}_{ij} = \mathbf{0}$;
- otherwise, given $(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I}_{(-ij)}, \mathbf{v}, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \delta_1, \rho; \text{data})$, the conditional density of \mathbf{I}_{ij} follows a discrete distribution with

$$\pi_{ijl} = \pi(I_{ijl} = 1, \mathbf{I}_{ij(-l)} = \mathbf{0} \mid \boldsymbol{\alpha}, \boldsymbol{\lambda}; \mathbf{y}) = \frac{\alpha_l \lambda_l y_{ij}^{\alpha_l}}{\sum_{m=1}^r \alpha_m \lambda_m y_{ij}^{\alpha_m}}, \quad l = 1, \dots, r, \quad (5.12)$$

where $\mathbf{I}_{ij(-l)} = \{I_{ijm} : m \neq l\}$. Note that $\sum_{m=1}^r \pi_{ijm} = 1$, and it is always true that only one element in \mathbf{I}_{ij} is 1.

(d) Given $(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I}, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, v_{ij} 's are independent and the conditional posterior density is

$$[v_{ij} \mid \cdot] \propto \exp\left\{d_{ij} v_{ij} \sum_{l=1}^r I_{ijl} - e^{v_{ij}} \sum_{l=1}^r \lambda_l y_{ij}^{\alpha_l} - \frac{1}{2\delta_0} \left[v_{ij} - (\boldsymbol{\beta}' \mathbf{x}_{ij} + W_i)\right]^2\right\}. \quad (5.13)$$

(e) Given $(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I}, \mathbf{v}, \mathbf{W}, \delta_0, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior distribution of $\boldsymbol{\beta}$ is

$$N\left(\left(\frac{1}{\delta_0} \mathbf{X}' \mathbf{X} + \boldsymbol{\tau}^{-1}\right)^{-1} \boldsymbol{\mu}_\beta, \left(\frac{1}{\delta_0} \mathbf{X}' \mathbf{X} + \boldsymbol{\tau}^{-1}\right)^{-1}\right), \quad (5.14)$$

where

$$\boldsymbol{\mu}_\beta = \left[\frac{1}{\delta_0} \mathbf{X}'(\mathbf{v} - \boldsymbol{\eta} \mathbf{W}) + \boldsymbol{\tau}^{-1} \boldsymbol{\mu} \right]. \quad (5.15)$$

(f) Given $(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I}, \mathbf{v}, \boldsymbol{\beta}, \delta_0, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, and define $\mathbf{B} = \mathbf{I}_I - \rho \mathbf{C}$, the conditional posterior distribution of \mathbf{W} is

$$(\mathbf{W} \mid \mathbf{v}, \boldsymbol{\beta}, \delta_0; \boldsymbol{\eta}) \sim N(\tilde{\boldsymbol{\mu}}_w, \tilde{\boldsymbol{\tau}}_w), \quad (5.16)$$

where

$$\tilde{\boldsymbol{\mu}}_w = \frac{1}{\delta_0} \left(\frac{1}{\delta_0} \boldsymbol{\eta}' \boldsymbol{\eta} + \frac{1}{\delta_1} \mathbf{B} \right)^{-1} \boldsymbol{\eta}' (\mathbf{v} - \mathbf{X} \boldsymbol{\beta}) \quad \text{and} \quad \tilde{\boldsymbol{\tau}}_w = \left(\frac{1}{\delta_0} \boldsymbol{\eta}' \boldsymbol{\eta} + \frac{1}{\delta_1} \mathbf{B} \right)^{-1}.$$

Obviously, $\boldsymbol{\eta}' \boldsymbol{\eta} = D$ where D is a diagonal matrix whose diagonal elements are (n_1, \dots, n_I) .

(g) Given $((\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I}, \mathbf{v}, \boldsymbol{\beta}, \mathbf{W}, \delta_1, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior distribution of δ_0 is

$$(\delta_0 \mid \mathbf{v}, \boldsymbol{\beta}, \mathbf{W}; \mathbf{X}) \sim IG\left(a_0 + \frac{K}{2}, b_0 + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^{n_i} (v_{ij} - \boldsymbol{\beta}' \mathbf{x}_{ij} - W_i)^2\right). \quad (5.17)$$

(h) Given $((\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I}, \mathbf{v}, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \rho; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior distribution of δ_1 is

$$(\delta_1 \mid \mathbf{W}, \rho) \sim IG\left(a_1 + \frac{I}{2}, b_1 + \frac{1}{2} \mathbf{W}' (\mathbf{I}_I - \rho \mathbf{C}) \mathbf{W}\right). \quad (5.18)$$

(i) Given $((\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{I}, \mathbf{v}, \boldsymbol{\beta}, \mathbf{W}, \delta_0, \delta_1; \mathbf{y}, \mathbf{d}, \mathbf{X}, \boldsymbol{\eta})$, the conditional posterior density of ρ

is

$$[\rho \mid \mathbf{W}, \delta_1] \propto |\mathbf{I}_I - \rho \mathbf{C}|^{1/2} \exp \left\{ -\frac{1}{2\delta_1} \mathbf{W}'(\mathbf{I}_I - \rho \mathbf{C})\mathbf{W} \right\}. \quad (5.19)$$

It is easy to show that all the non-regular distributions are log concave and can be sampled by the Adaptive Rejection Method (Gilks & Wild (1992)).

5.5 Result

The computation is based on 3 chains of 20,000 samples with 5,000 burn-in period.

The DIC for 3 risks is 57794. Parameter estimates and their standard errors are listed in Table 5.1. The spatial effect can be observed from Figure 5.1. No particular trend is found from the map, which indicates that the response does not have a significant geographical factor.

From Figure 5.2, it can be seen that the shape of the third risk is much different from the other two, leading to a smaller hazard rate. The first two risks, however, does not have a clear shape difference, although they have different hazard rates. This might be a result of a identifiability problem or a convergence problem, as shown in Figure 5.3. It is hard to conclude the hazard rate parameters are convergent according to the trace plots. Moreover, plot (c) suggests that there might be identifiability problems that have not been detected yet.

5.6 Discussion

This chapter discusses the competing risks model with application of colorectal cancer in SEER data. The Weibull hazard has been used. For further exploration, other

different hazard functions can be added to the model for comparison.

If we consider more competing risks, the model will be more powerful for detecting behaviors of different risks. However, the identifiability problem will become worse. To reduce its effect, the data have to show great differences between different risks.

Table 5.1: Parameter Estimates and their Standard Errors, the standard errors are in parentheses

Shapes			Stage (Unknown= 0)		
α_1	1.0965	(0.1071)	In Situ	-2.9196	(0.1469)
α_2	1.1809	(0.0339)	Localized	-2.6633	(0.1119)
α_3	3.4650	(0.2633)	Regional	-1.9625	(0.0948)
			Distant	0.5969	(0.0731)
Hazard Rates			Other Parameters		
λ_1	0.0606	(0.0602)	Age	0.1884	(0.2053)
λ_2	0.4782	(0.1236)	δ_0	1.8667	(0.2094)
λ_3	0.0012	(0.0008)	δ_1	0.0180	(0.0067)
			ρ	-0.0962	(0.1140)

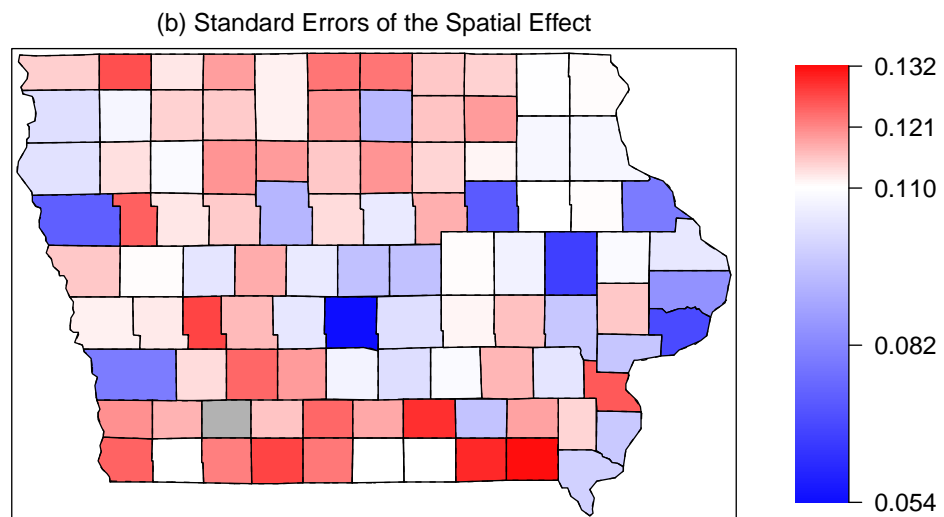
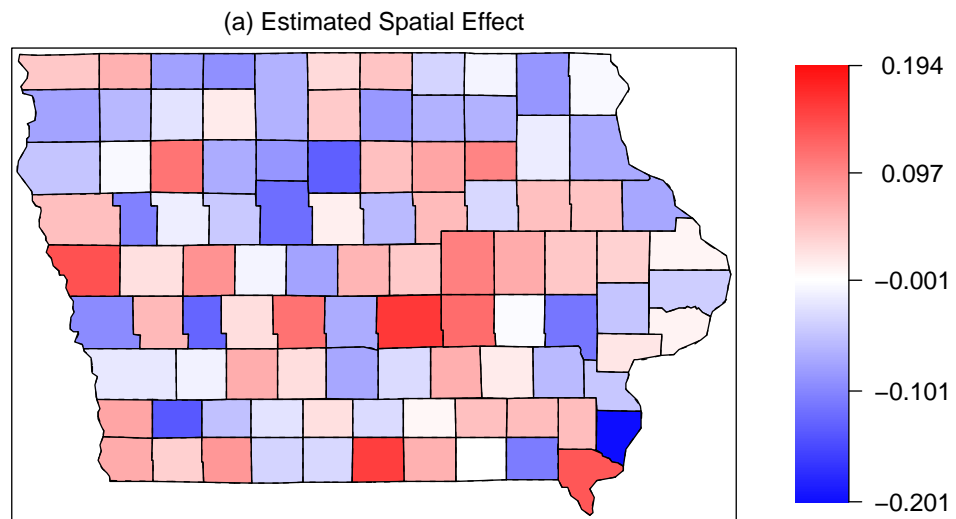


Figure 5.1: Estimated Spatial Effect and Standard Errors

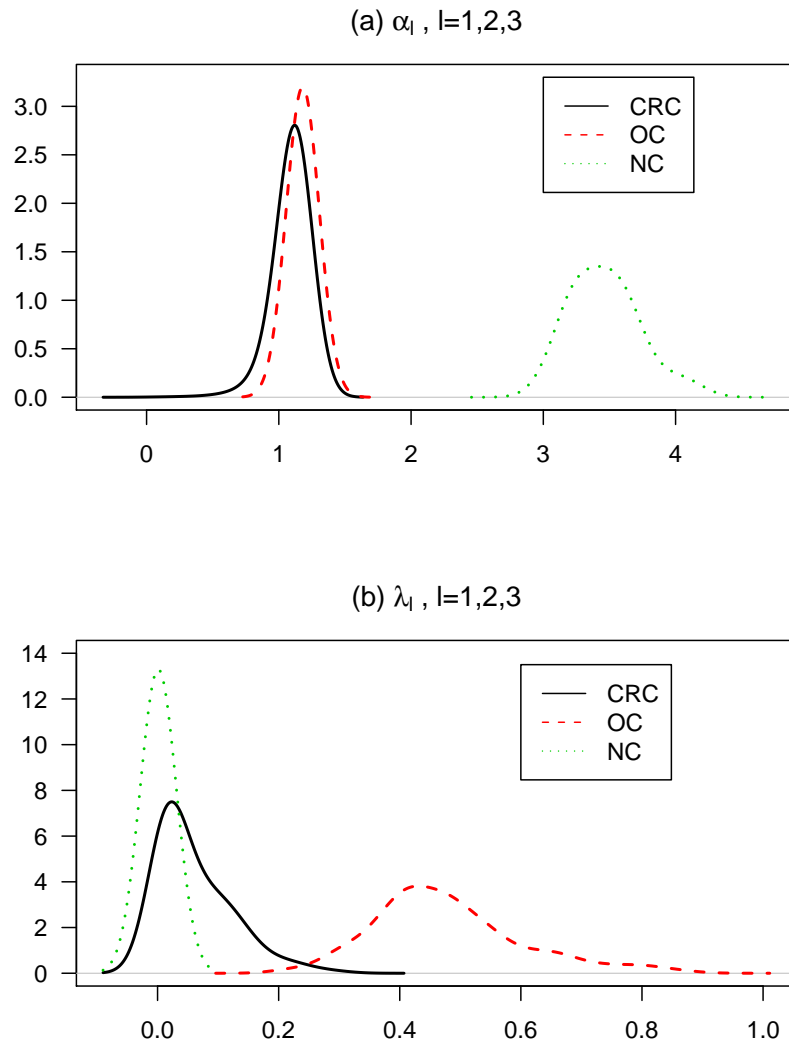


Figure 5.2: Posterior Density Comparison of the Shape and Hazard Rate Parameters

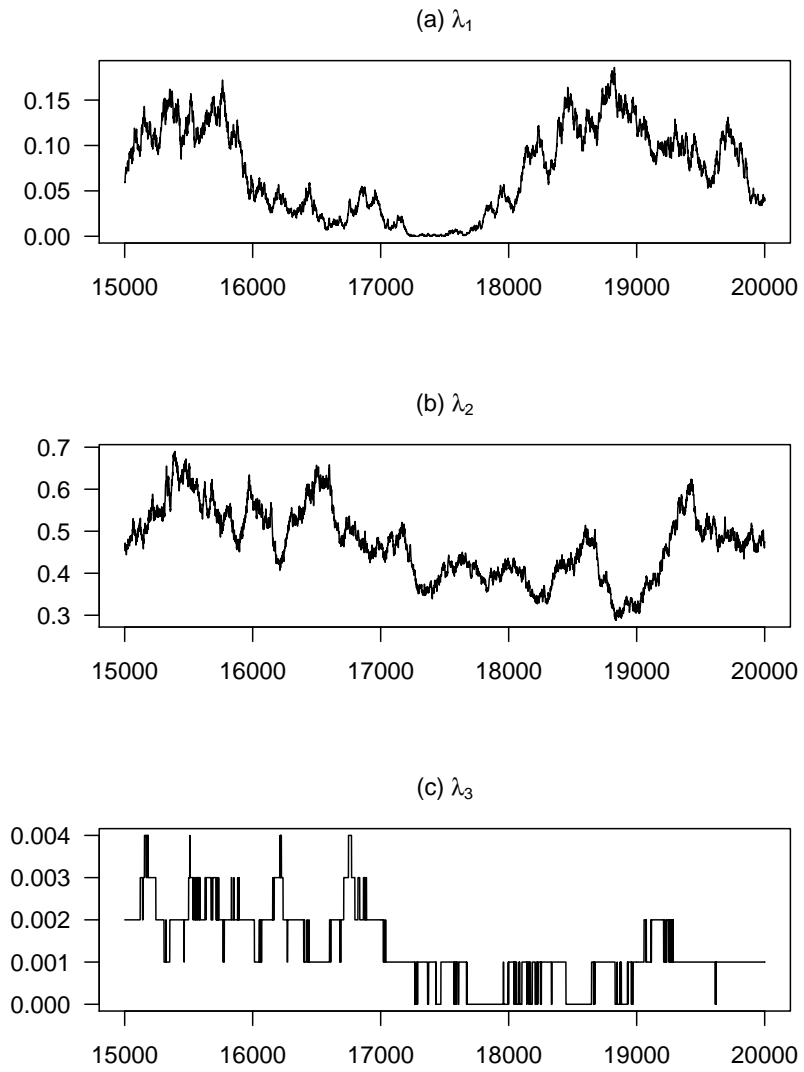


Figure 5.3: Trace Plots for Hazard Rates

Chapter 6

Discussion on Spatial Confounding

6.1 Introduction

Spatial analysis is motivated when the error term in a regression model will tend to be spatially correlated. Anselin (2002) provided an overview of the motivation for including spatial effects in regression models, both from a theory-driven as well as from a data-driven perspective. Researchers have modeled the spatial structure by adding a spatial random effect with zero mean in the model (e.g., Breslow & Clayton (1993), Ver Hoef et al. (1993), etc.). The spatial analysis has been applied to various areas. For example, Carlin & Banerjee (2002) discussed the spatial process models for multivariate survival data sets. Cressie & Wikle (2011) gave a comprehensive demonstration of spatio-temporal processes on environmental processes, climate trends and public-health data, etc.

However, most researches are based on the assumption that the spatial random effect is independent with the other regressors. In reality, nevertheless, this spatial effect is unmeasured, and might be correlated with other random variables in the

model. When this occurs it is not easy to separate the effect of a regression factor from a spatially varying error. This inability to separate factor effects from spatial error is referred to as *spatial confounding*. When the confounder is unmeasured, bias in estimation of regression coefficients can not be avoided. Paciorek (2010) noted this issue and showed the bias dependence on the spatial scales of the covariate and the residual.

This chapter discusses the confounding problem arising from the linear mixed model with an unobserved random effect, i.e., the spatial random effect. We compare the variances and mean square errors of different estimators of the regression coefficient for the linear mixed model in the effort of reducing the bias caused by the confounding issue.

We suppose that \mathbf{y} is a random variable with n elements, $\boldsymbol{\beta}$ is a $p \times 1$ unknown regression coefficient vector, \mathbf{u} is a q -vector corresponding to an unobserved random effect. Assume that \mathbf{X}_0 is a $n \times p$ matrix of covariates, and \mathbf{X}_1 is a fixed design matrix with $n \times q$ elements. Consider the linear mixed model

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta} + \mathbf{X}_1\mathbf{u} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \delta_0\boldsymbol{\Psi}_0), \quad (6.1)$$

$$\mathbf{u} \sim N_q(\mathbf{0}, \delta_1\boldsymbol{\Psi}_1), \quad (6.2)$$

where $\boldsymbol{\epsilon}$ is the error term, and $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$ are known positive definite matrices. We often assume that \mathbf{X}_i , $i = 0, 1$ are fixed or prespecified with full column ranks. Without loss of generality, we assume that $\boldsymbol{\Psi}_0 = \mathbf{I}_n$. Here $\boldsymbol{\Psi}_1$ is fixed but might not always be identity. Note that \mathbf{u} and $\boldsymbol{\epsilon}$ are both unobservable. Thus, \mathbf{y} has the marginal distribution

$$\mathbf{y} \mid \mathbf{X}_0 \sim N(\mathbf{X}_0\boldsymbol{\beta}, \delta_0\boldsymbol{\Sigma}), \quad (6.3)$$

where

$$\boldsymbol{\Sigma} = \mathbf{I}_n + \frac{1}{\eta} \mathbf{X}_1 \boldsymbol{\Psi}_1 \mathbf{X}_1' \text{ and } \eta = \frac{\delta_0}{\delta_1}. \quad (6.4)$$

Both δ_0 and δ_1 are often unknown, but η may be known.

Speckman (1988) considered the partial linear model

$$y_i = \boldsymbol{\xi}_i' \boldsymbol{\beta} + f(t_i) + \epsilon_i \quad (1 \leq i \leq n), \quad (6.5)$$

where the $\boldsymbol{\xi}_i$ are fixed known p vectors, $\boldsymbol{\beta}$ is an unknown vector of parameters, and the t_i are in some bounded domain $D \subset \mathbb{R}^k$. He proposed two methods of estimation based on methods for scatterplot smoothing in the simplified model of (6.5) with $\boldsymbol{\beta} = \mathbf{0}$. The first estimator is obtained from the method of penalized least squares introduced by Engle et al. (1986) among others. The second method is to estimate $\boldsymbol{\beta}$ from partial residuals after adjustment of y_i and $\boldsymbol{\xi}_i$ for dependence on t_i . Both methods use a smoother matrix for ordinary scatterplot smoothing. Under suitable assumptions, the asymptotic bias and variance are obtained for both methods, and it is shown that estimating $\boldsymbol{\beta}$ by partial residuals results in improved bias with no asymptotic loss in variance.

Note that the term $(f(t_1), \dots, f(t_n))'$ in (6.5) can be reparameterized as $\mathbf{X}_1 \mathbf{u}$. Two types of adjustment for partial residuals are applied to \mathbf{y} and \mathbf{X}_0 in Model (6.1) under the assumption that \mathbf{X}_0 and \mathbf{u} are independent. The first adjustment is obtained by the conditional posterior distribution of \mathbf{u} given \mathbf{y} and $\boldsymbol{\beta}$, and the second by the marginal posterior distribution of \mathbf{u} given \mathbf{y} .

In Section 6.2, we discuss the linear mixed model (6.1) and its estimators of the covariate coefficient $\boldsymbol{\beta}$. Besides the generalized least square estimator and the Bayes estimator, two other estimators motivated by Speckman (1988) are discussed.

The corresponding predictors of the random effect \mathbf{u} and the variable \mathbf{y} are also computed. Section 6.3 compares the variances and MSEs of the estimators under different conditions. We give two examples for illustration. Sections 6.4 discusses the orders of the variances and MSEs of \mathbf{u} . The variances of the prediction error are compared in Section 6.5. Finally, the Bayes Factor of testing the regression coefficient is discussed in Section 6.6.

6.2 A Linear Mixed Model

6.2.1 Generalized Least Square Estimator

For model (6.1), the commonly used GLS estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_G = (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{y}. \quad (6.6)$$

Henderson (1975) gave the Best Linear Unbiased Predictor (BLUP) of \mathbf{u} ,

$$\hat{\mathbf{u}}_G = \boldsymbol{\Lambda}_1^{-1} \mathbf{X}'_1 (\mathbf{y} - \mathbf{X}_0 \hat{\boldsymbol{\beta}}_G),$$

where

$$\boldsymbol{\Lambda}_1 = \mathbf{X}'_1 \mathbf{X}_1 + \eta \boldsymbol{\Psi}_1^{-1}. \quad (6.7)$$

It is easy to show that $\boldsymbol{\Sigma}^{-1} = \mathbf{I}_n - \mathbf{X}_1 \boldsymbol{\Lambda}_1^{-1} \mathbf{X}'_1$, hence $\boldsymbol{\Lambda}_1^{-1} \mathbf{X}'_1 = \frac{1}{\eta} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \boldsymbol{\Sigma}^{-1}$. We define

$$\mathbf{C}_G = \mathbf{I}_n - \boldsymbol{\Sigma}^{-1/2} \mathbf{X}_0 (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-1/2}. \quad (6.8)$$

Then

$$\hat{\mathbf{u}}_G = \frac{1}{\eta} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \boldsymbol{\Sigma}^{-1/2} \mathbf{C}_G \boldsymbol{\Sigma}^{-1/2} \mathbf{y}. \quad (6.9)$$

Therefore the predictor of \mathbf{y} can be written as

$$\hat{\mathbf{y}}_G = \mathbf{X}_0 \hat{\boldsymbol{\beta}}_G + \mathbf{X}_1 \hat{\mathbf{u}}_G = (\mathbf{I}_n - \boldsymbol{\Sigma}^{-1/2} \mathbf{C}_G \boldsymbol{\Sigma}^{-1/2}) \mathbf{y}. \quad (6.10)$$

6.2.2 Bayes Estimator

Consider the Bayes estimator for $\boldsymbol{\beta}$ given η using a constant prior for $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta}) \propto 1$, the prior (6.2) for \mathbf{u} , and a prior of δ_0 . The joint posterior distribution of $(\boldsymbol{\beta}', \mathbf{u}')'$ given $(\mathbf{y}; \eta)$ is

$$N_{p+q} \left(\mathbf{V}^{-1} (\mathbf{X}'_0, \mathbf{X}'_1)' \mathbf{y}, \delta_0 \mathbf{V}^{-1} \right), \quad (6.11)$$

where

$$\mathbf{V} = \begin{pmatrix} \mathbf{X}'_0 \mathbf{X}_0 & \mathbf{X}'_0 \mathbf{X}_1 \\ \mathbf{X}'_1 \mathbf{X}_0 & \mathbf{X}'_1 \mathbf{X}_1 + \eta \boldsymbol{\Psi}_1^{-1} \end{pmatrix}. \quad (6.12)$$

We define

$$\mathbf{P}_0 = \mathbf{X}_0 (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0, \quad (6.13)$$

and

$$\boldsymbol{\Phi}_1 = \mathbf{X}'_1 (\mathbf{I}_n - \mathbf{P}_0) \mathbf{X}_1 + \eta \boldsymbol{\Psi}_1^{-1}. \quad (6.14)$$

Both \mathbf{X}_0 and \mathbf{X}_1 have full column rank, \mathbf{V} is invertible and

$$\mathbf{V}^{-1} = \begin{bmatrix} \mathbf{V}_0 & \mathbf{V}_1 \\ \mathbf{V}_1' & \Phi_1^{-1} \end{bmatrix},$$

where

$$\mathbf{V}_0 = (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 (\mathbf{I}_n + \mathbf{X}_1 \Phi_1^{-1} \mathbf{X}'_1) \mathbf{X}_0 (\mathbf{X}'_0 \mathbf{X}_0)^{-1}, \quad \mathbf{V}_1 = -(\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \mathbf{X}_1.$$

The Bayes estimator of $\boldsymbol{\beta}$ is the first p -vector of $\mathbf{V}^{-1}(\mathbf{X}'_0, \mathbf{X}'_1)' \mathbf{y}$, i.e.,

$$\hat{\boldsymbol{\beta}}_B = E(\boldsymbol{\beta} \mid \mathbf{y}; \eta) = (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 (\mathbf{I}_n - \mathbf{X}_1 \mathbf{B}) \mathbf{y}. \quad (6.15)$$

Alternatively, the Bayes estimator can be given by

$$E(\boldsymbol{\beta} \mid \mathbf{y}; \eta) = E[E(\boldsymbol{\beta} \mid \mathbf{y}; \eta, \mathbf{u}) \mid \mathbf{y}; \eta] = E[(\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 (\mathbf{y} - \mathbf{X}_1 \mathbf{u}) \mid \mathbf{y}; \eta].$$

Defining

$$\mathbf{B} = \Phi_1^{-1} \mathbf{X}'_1 (\mathbf{I}_n - \mathbf{P}_0), \quad (6.16)$$

it is easy to verify that the marginal posterior distribution of \mathbf{u} given \mathbf{y} is

$$(\mathbf{u} \mid \mathbf{y}) \sim N(\mathbf{B} \mathbf{y}, \delta_0 \Phi_1^{-1}). \quad (6.17)$$

So the Bayes estimator has the expression (6.15). The predictor of the random effect \mathbf{u} would be

$$\hat{\mathbf{u}}_B = \mathbf{B} \mathbf{y}. \quad (6.18)$$

Therefore, the predictor of \mathbf{y} becomes

$$\begin{aligned}
\hat{\mathbf{y}}_B &= [\mathbf{X}_0(\mathbf{X}'_0\mathbf{X}_0)^{-1}\mathbf{X}'_0(\mathbf{I}_n - \mathbf{X}_1\mathbf{B}) + \mathbf{X}_1\mathbf{B}]\mathbf{y} \\
&= [(\mathbf{I}_n - \mathbf{P}_0)\mathbf{X}_1\Phi_1^{-1}\mathbf{X}'_1(\mathbf{I}_n - \mathbf{P}_0) + \mathbf{P}_0]\mathbf{y} \\
&= \left(\left[\mathbf{I}_n + \frac{1}{\eta}(\mathbf{I}_n - \mathbf{P}_0)\mathbf{X}_1\Psi_1^{-1}\mathbf{X}'_1(\mathbf{I}_n - \mathbf{P}_0) \right]^{-1} - (\mathbf{I}_n - \mathbf{P}_0) \right) \mathbf{y}. \quad (6.19)
\end{aligned}$$

6.2.3 \mathbf{X}_0 and \mathbf{u} are Independent

The basic assumption of the usual linear mixed model (6.1) and (6.2) is that \mathbf{X}_0 and \mathbf{u} are independent. In this case, $\hat{\boldsymbol{\beta}}_G$ is unbiased. Motivated by Speckman (1988), we consider several estimators of $\boldsymbol{\beta}$.

Estimate of $\boldsymbol{\beta}$ from the Conditional Posterior Distribution

For model (6.1) and (6.2), we can treat (6.2) as a prior. Then the posterior distribution of \mathbf{u} given $(\mathbf{y}; \boldsymbol{\beta}, \delta_0, \delta_1)$ is

$$(\mathbf{u} \mid \mathbf{y}, \boldsymbol{\beta}) \sim N(\mathbf{c}_1, \delta_0\boldsymbol{\Lambda}_1^{-1}), \quad (6.20)$$

where

$$\mathbf{c}_1 = \boldsymbol{\Lambda}_1^{-1}\mathbf{X}'_1(\mathbf{y} - \mathbf{X}_0\boldsymbol{\beta}). \quad (6.21)$$

If we replace \mathbf{u} in (6.1) by $\mathbf{c}_1 = E(\mathbf{u} \mid \mathbf{y}, \boldsymbol{\beta})$, we get $\mathbf{y} - \mathbf{X}_1\mathbf{c}_1 = \mathbf{X}_0\boldsymbol{\beta} + \boldsymbol{\epsilon}$. This is equivalent to

$$\boldsymbol{\Sigma}^{-1}\mathbf{y} = \boldsymbol{\Sigma}^{-1}\mathbf{X}_0\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (6.22)$$

Noting that $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \delta_0 \mathbf{I}_n)$, the corresponding ‘‘OLS’’ and ‘‘GLS’’ of $\boldsymbol{\beta}$ for model (6.22) are identical and given by

$$\tilde{\boldsymbol{\beta}}_O = \tilde{\boldsymbol{\beta}}_G = (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{y}. \quad (6.23)$$

Consequently, \mathbf{u} can be estimated by replacing $\boldsymbol{\beta}$ in (6.21) by $\tilde{\boldsymbol{\beta}}_O$:

$$\tilde{\mathbf{u}}_O = \frac{1}{\eta} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}_0 \tilde{\boldsymbol{\beta}}_O). \quad (6.24)$$

If we define

$$\mathbf{C}_O = \mathbf{I}_n - \boldsymbol{\Sigma}^{-1} \mathbf{X}_0 (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-1}, \quad (6.25)$$

we get

$$\tilde{\mathbf{u}}_O = \frac{1}{\eta} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \mathbf{C}_O \boldsymbol{\Sigma}^{-1} \mathbf{y}. \quad (6.26)$$

Then the predictor of \mathbf{y} would be

$$\tilde{\mathbf{y}}_O = (\mathbf{I}_n - \mathbf{C}_O \boldsymbol{\Sigma}^{-1}) \mathbf{y}. \quad (6.27)$$

Estimate of $\boldsymbol{\beta}$ from the Marginal Posterior Distribution

In model (6.1) and (6.2), we choose a constant prior for $\boldsymbol{\beta}$ and treat (6.2) as the prior for \mathbf{u} . The marginal distribution of \mathbf{u} given \mathbf{y} is given by (6.17). We can replace \mathbf{u} in (6.1) with \mathbf{c}_1^* and get another approximating model,

$$\mathbf{B}^* \mathbf{y} = \mathbf{X}_0 \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (6.28)$$

where

$$\mathbf{B}^* = \mathbf{I}_n - \mathbf{X}_1 \mathbf{B}. \quad (6.29)$$

It is easy to show that $\mathbf{B}^* = (\mathbf{I}_n + \frac{1}{\eta} \mathbf{X}_1 \boldsymbol{\Psi}_1 \mathbf{X}_1' (\mathbf{I}_n - \mathbf{P}_0))^{-1}$. The corresponding ‘‘OLS’’ and ‘‘GLS’’ estimates of $\boldsymbol{\beta}$ for model (6.17) are identical and given by (6.15).

6.3 Comparison of Estimators of the Regression Coefficient $\boldsymbol{\beta}$

So far, we have listed three methods of estimating $\boldsymbol{\beta}$ and their corresponding predictors, and it would be interesting to compare their MSEs. The estimators of $\boldsymbol{\beta}$ are given by (6.6), (6.15), and (6.23).

6.3.1 \mathbf{X}_0 and \mathbf{u} are Independent

Since $E(\mathbf{u} \mid \mathbf{X}_0) = E(\mathbf{u}) = \mathbf{0}$, (6.6) and (6.23) are unbiased estimators. Their MSEs are

$$MSE(\hat{\boldsymbol{\beta}}_G \mid \mathbf{X}_0, \boldsymbol{\beta}) = Var(\hat{\boldsymbol{\beta}}_G \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0 (\mathbf{X}_0' \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1}, \quad (6.30)$$

$$\begin{aligned} MSE(\tilde{\boldsymbol{\beta}}_O \mid \mathbf{X}_0, \boldsymbol{\beta}) &= Var(\tilde{\boldsymbol{\beta}}_O \mid \mathbf{X}_0, \boldsymbol{\beta}) \\ &= \delta_0 (\mathbf{X}_0' \boldsymbol{\Sigma}^{-2} \mathbf{X}_0)^{-1} \mathbf{X}_0' \boldsymbol{\Sigma}^{-3} \mathbf{X}_0 (\mathbf{X}_0' \boldsymbol{\Sigma}^{-2} \mathbf{X}_0)^{-1}. \end{aligned} \quad (6.31)$$

The MSE of the Bayes estimator can be computed by

$$\begin{aligned} MSE(\hat{\boldsymbol{\beta}}_B \mid \mathbf{X}_0, \boldsymbol{\beta}) &= Var(\hat{\boldsymbol{\beta}}_B \mid \mathbf{X}_0, \boldsymbol{\beta}) \\ &\quad + [Bias(\hat{\boldsymbol{\beta}}_B \mid \mathbf{X}_0, \boldsymbol{\beta})] [Bias(\hat{\boldsymbol{\beta}}_B \mid \mathbf{X}_0, \boldsymbol{\beta})]', \end{aligned} \quad (6.32)$$

where

$$\text{Var}(\widehat{\boldsymbol{\beta}}_B | \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0 (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 (\mathbf{I}_n - \mathbf{X}_1 \mathbf{B}) \boldsymbol{\Sigma} (\mathbf{I}_n - \mathbf{X}_1 \mathbf{B})' \mathbf{X}_0 (\mathbf{X}'_0 \mathbf{X}_0)^{-1},$$

$$\text{and } \text{Bias}(\widehat{\boldsymbol{\beta}}_B | \mathbf{X}_0, \boldsymbol{\beta}) = (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \mathbf{X}_1 \mathbf{B} \mathbf{X}_0 \boldsymbol{\beta}.$$

Lemma 1. *Suppose that \mathbf{S} and \mathbf{T} are two $m \times n$ matrices satisfying $\mathbf{S}'\mathbf{S} \leq \mathbf{T}'\mathbf{T}$.*

For any positive definite (p.d.) matrix \mathbf{A} , $\mathbf{S}'\mathbf{A}\mathbf{S} \leq \mathbf{T}'\mathbf{A}\mathbf{T}$.

Proof. Since \mathbf{A} is positive definite, there exist orthogonal projection matrices \mathbf{P}_i , $i = 1, \dots, n$, such that

$$\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{P}_i, \quad (6.33)$$

where the λ_i are the eigenvalues of \mathbf{A} . Since $\sum_{i=1}^n \mathbf{P}_i = \mathbf{I}_n$, we have

$$\mathbf{S}'\mathbf{S} \leq \mathbf{T}'\mathbf{T} \Leftrightarrow \sum_{i=1}^n \mathbf{S}'\mathbf{P}_i\mathbf{S} \leq \sum_{i=1}^n \mathbf{T}'\mathbf{P}_i\mathbf{T}.$$

This implies that

$$\sum_{i=1}^n \lambda_i \mathbf{S}'\mathbf{P}_i\mathbf{S} \leq \sum_{i=1}^n \lambda_i \mathbf{T}'\mathbf{P}_i\mathbf{T} \Leftrightarrow \mathbf{S}'\mathbf{A}\mathbf{S} \leq \mathbf{T}'\mathbf{A}\mathbf{T},$$

and the lemma is proved. □

Theorem 3. *Assume that in model (6.1) \mathbf{X}_0 and \mathbf{u} are independent. (6.6), (6.15), and (6.23) are three estimators for $\boldsymbol{\beta}$ indicated by $\widehat{\boldsymbol{\beta}}_G$, $\widetilde{\boldsymbol{\beta}}_O$, and $\widehat{\boldsymbol{\beta}}_B$. Then there are following results:*

(a) $\widehat{\boldsymbol{\beta}}_G$ and $\widetilde{\boldsymbol{\beta}}_O$ are unbiased estimators.

(b) $\widehat{\boldsymbol{\beta}}_G$ has the smallest MSE among the unbiased estimators, i.e.,

$$\text{Var}(\widehat{\boldsymbol{\beta}}_G | \mathbf{X}_0, \boldsymbol{\beta}) \leq \text{Var}(\widetilde{\boldsymbol{\beta}}_O | \mathbf{X}_0, \boldsymbol{\beta}). \quad (6.34)$$

(c) Furthermore, $\widehat{\boldsymbol{\beta}}_B$ is biased, but it has the smallest variance, i.e.,

$$\text{Var}(\widehat{\boldsymbol{\beta}}_B | \mathbf{X}_0, \boldsymbol{\beta}) \leq \text{Var}(\widehat{\boldsymbol{\beta}}_G | \mathbf{X}_0, \boldsymbol{\beta}). \quad (6.35)$$

Proof. Part (a) is obvious.

For Part (b), because of (a), the MSEs are equal to the variances. We know that $\boldsymbol{\Sigma} > \mathbf{I}_n$, so there exists a series of projection matrices \mathbf{P}_i , $i = 1, \dots, n$, such that $\boldsymbol{\Sigma} = \sum_{i=1}^n \lambda_i \mathbf{P}_i$, where the λ_i s are eigenvalues of $\boldsymbol{\Sigma}$. It is easy to show that $\lambda_i > 1$ for any $i = 1, \dots, n$, and that $\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^n \frac{1}{\lambda_i} \mathbf{P}_i$. Therefore,

$$\begin{pmatrix} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-3} \mathbf{X}_0 & \mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0 \\ \mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0 & \mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0 \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} \frac{1}{\lambda_i^3} & \frac{1}{\lambda_i^2} \\ \frac{1}{\lambda_i^2} & \frac{1}{\lambda_i} \end{pmatrix} \otimes \mathbf{X}'_0 \mathbf{P}_i \mathbf{X}_0 \geq 0.$$

According to Bhatia (2006), we have

$$\mathbf{X}'_0 \boldsymbol{\Sigma}^{-3} \mathbf{X}_0 \geq (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0)(\mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0,$$

or

$$(\mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1} \leq (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-3} \mathbf{X}_0 (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0)^{-1}.$$

Therefore, $\text{Var}(\widehat{\boldsymbol{\beta}}_G | \mathbf{X}_0, \boldsymbol{\beta}) \leq \text{Var}(\widetilde{\boldsymbol{\beta}}_O | \mathbf{X}_0, \boldsymbol{\beta})$.

For Part (c), the MSE and the variance of the biased estimator $\widehat{\boldsymbol{\beta}}_B$ are not equal. Since

$$\begin{pmatrix} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0 & \mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0 \\ \mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0 & \mathbf{X}'_0 \mathbf{X}_0 \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} \frac{1}{\lambda_i^2} & \frac{1}{\lambda_i} \\ \frac{1}{\lambda_i} & 1 \end{pmatrix} \otimes \mathbf{X}'_0 \mathbf{P}_i \mathbf{X}_0 \geq 0,$$

we have

$$(\mathbf{X}'_0 \mathbf{X}_0)^{-1} \leq (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-2} \mathbf{X}_0 (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1}.$$

From Lemma 1, we have

$$(\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma} \mathbf{X}_0 (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \leq (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0 (\mathbf{X}'_0 \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1}.$$

Since $\mathbf{B}^* \leq \mathbf{I}_n$, it is easy to show that

$$\text{Var}(\widehat{\boldsymbol{\beta}}_B | \mathbf{X}_0, \boldsymbol{\beta}) \leq (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \boldsymbol{\Sigma} \mathbf{X}_0 (\mathbf{X}'_0 \mathbf{X}_0)^{-1},$$

hence the inequality is proved. □

6.3.2 \mathbf{X}_0 and \mathbf{u} are Correlated

We assume that $\mathbf{X}_0 = (\mathbf{x}_1, \dots, \mathbf{x}_p)$, and suppose that at least one of \mathbf{x}_i , $i = 1, \dots, p$, and \mathbf{u} are not independent. Define

$$\mathbf{x} = \text{vec}(\mathbf{X}_0) = (\mathbf{x}'_1, \dots, \mathbf{x}'_p)', \quad (6.36)$$

the vectorization of \mathbf{X}_0 . Then the relationship between \mathbf{X}_0 and \mathbf{u} can be transformed to that of \mathbf{x} and \mathbf{u} . In fact, there exists a one-to-one transformation between \mathbf{X}_0 and \mathbf{x} , i.e.,

$$\mathbf{X}_0 = \sum_{i=1}^p (\mathbf{U}'_i \otimes \mathbf{I}_n) \mathbf{x} \mathbf{U}'_i, \quad \text{and} \quad \mathbf{x} = \sum_{i=1}^p (\mathbf{U}_i \otimes \mathbf{I}_n) \mathbf{X}_0 \mathbf{U}_i,$$

where \mathbf{U}_i is a $p \times 1$ vector whose i^{th} element is 1 and 0 otherwise.

Define

$$\boldsymbol{\gamma} = E(\mathbf{u} \mid \mathbf{x}, \boldsymbol{\beta}), \quad \tilde{\boldsymbol{\Sigma}} = Var(\mathbf{u} \mid \mathbf{x}, \boldsymbol{\beta}). \quad (6.37)$$

Then $\boldsymbol{\gamma} \neq \mathbf{0}$, and $\tilde{\boldsymbol{\Sigma}} > 0$ if \mathbf{x} and \mathbf{u} are correlated. Therefore,

$$E(\mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}) = \mathbf{X}_0\boldsymbol{\beta} + \mathbf{X}_1\boldsymbol{\gamma}, \quad Var(\mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0\mathbf{I}_n + \mathbf{X}_1\tilde{\boldsymbol{\Sigma}}\mathbf{X}_1'.$$

Note that if \mathbf{x} and \mathbf{u} are independent, we have

$$\boldsymbol{\gamma} = E(\mathbf{u}) = \mathbf{0}, \quad \tilde{\boldsymbol{\Sigma}} = Var(\mathbf{u}) = \delta_1\boldsymbol{\Psi}_1, \quad (6.38)$$

hence

$$Var(\mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0\boldsymbol{\Sigma}. \quad (6.39)$$

Define

$$\boldsymbol{\Sigma}_V^* = \mathbf{I}_n + \frac{1}{\delta_0}\mathbf{X}_1\tilde{\boldsymbol{\Sigma}}\mathbf{X}_1', \quad \boldsymbol{\Sigma}_M^* = \boldsymbol{\Sigma}_V^* + \frac{1}{\delta_0}\mathbf{X}_1\boldsymbol{\gamma}\boldsymbol{\gamma}'\mathbf{X}_1'. \quad (6.40)$$

Then the expected values, variances and MSEs of the estimators $(\hat{\boldsymbol{\beta}}_G, \tilde{\boldsymbol{\beta}}_O, \hat{\boldsymbol{\beta}}_B)$ are

$$E(\hat{\boldsymbol{\beta}}_G \mid \mathbf{X}_0, \boldsymbol{\beta}) = \boldsymbol{\beta} + (\mathbf{X}_0'\boldsymbol{\Sigma}^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1\boldsymbol{\gamma}, \quad (6.41)$$

$$Var(\hat{\boldsymbol{\beta}}_G \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0(\mathbf{X}_0'\boldsymbol{\Sigma}^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_V^*\boldsymbol{\Sigma}^{-1}\mathbf{X}_0(\mathbf{X}_0'\boldsymbol{\Sigma}^{-1}\mathbf{X}_0)^{-1}, \quad (6.42)$$

$$MSE(\hat{\boldsymbol{\beta}}_G \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0(\mathbf{X}_0'\boldsymbol{\Sigma}^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_M^*\boldsymbol{\Sigma}^{-1}\mathbf{X}_0(\mathbf{X}_0'\boldsymbol{\Sigma}^{-1}\mathbf{X}_0)^{-1}; \quad (6.43)$$

$$E(\tilde{\boldsymbol{\beta}}_O \mid \mathbf{X}_0, \boldsymbol{\beta}) = \boldsymbol{\beta} + (\mathbf{X}_0'\boldsymbol{\Sigma}^{-2}\mathbf{X}_0)^{-1}\mathbf{X}_0'\boldsymbol{\Sigma}^{-2}\mathbf{X}_1\boldsymbol{\gamma}, \quad (6.44)$$

$$Var(\tilde{\boldsymbol{\beta}}_O \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0(\mathbf{X}_0'\boldsymbol{\Sigma}^{-2}\mathbf{X}_0)^{-1}\mathbf{X}_0'\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_V^*\boldsymbol{\Sigma}^{-2}\mathbf{X}_0(\mathbf{X}_0'\boldsymbol{\Sigma}^{-2}\mathbf{X}_0)^{-1}, \quad (6.45)$$

$$MSE(\tilde{\boldsymbol{\beta}}_O \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0(\mathbf{X}_0'\boldsymbol{\Sigma}^{-2}\mathbf{X}_0)^{-1}\mathbf{X}_0'\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_M^*\boldsymbol{\Sigma}^{-2}\mathbf{X}_0(\mathbf{X}_0'\boldsymbol{\Sigma}^{-2}\mathbf{X}_0)^{-1}; \quad (6.46)$$

$$E(\hat{\boldsymbol{\beta}}_B \mid \mathbf{X}_0, \boldsymbol{\beta}) = (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{B}^*(\mathbf{X}_0\boldsymbol{\beta} + \mathbf{X}_1\boldsymbol{\gamma}), \quad (6.47)$$

$$Var(\widehat{\beta}_B | \mathbf{X}_0, \beta) = \delta_0(\mathbf{X}'_0\mathbf{X}_0)^{-1}\mathbf{X}'_0\mathbf{B}^*\Sigma_V^*\mathbf{B}'\mathbf{X}_0(\mathbf{X}'_0\mathbf{X}_0)^{-1}, \quad (6.48)$$

$$\begin{aligned} MSE(\widehat{\beta}_B | \mathbf{X}_0, \beta) &= Var(\widehat{\beta}_B | \mathbf{X}_0, \beta) \\ &+ [Bias(\widehat{\beta}_B | \mathbf{X}_0, \beta)][Bias(\widehat{\beta}_B | \mathbf{X}_0, \beta)]'. \end{aligned} \quad (6.49)$$

We can now get a theorem about the order of the MSEs and variances of all the estimators $(\widehat{\beta}_G, \widetilde{\beta}_O, \widehat{\beta}_B)$.

Theorem 4. *For model (6.1), consider three estimators for β : (6.6), (6.15), and (6.23) indicated by $\widehat{\beta}_G$, $\widetilde{\beta}_O$, and $\widehat{\beta}_B$. If \mathbf{X}_0 and \mathbf{u} are correlated, all the estimators are biased, and we have the following facts:*

(a) *The variances satisfy*

$$Var(\widehat{\beta}_B | \mathbf{X}_0, \beta) \leq Var(\widehat{\beta}_G | \mathbf{X}_0, \beta) \leq Var(\widetilde{\beta}_O | \mathbf{X}_0, \beta). \quad (6.50)$$

(b) *The MSEs satisfy*

$$MSE(\widehat{\beta}_B | \mathbf{X}_0, \beta) \leq MSE(\widehat{\beta}_G | \mathbf{X}_0, \beta) \leq MSE(\widetilde{\beta}_O | \mathbf{X}_0, \beta). \quad (6.51)$$

Proof. It is obvious that all the estimators are biased. It is easy to show from Lemma 1 that the orders of the variances and the MSEs of $\widehat{\beta}_G$, and $\widetilde{\beta}_O$ do not change from Theorem 3 (b).

Moreover, $\mathbf{B}^* \leq \mathbf{I}_n$, hence

$$\begin{aligned} &[Bias(\widehat{\beta}_B | \mathbf{X}_0, \beta)][Bias(\widehat{\beta}_B | \mathbf{X}_0, \beta)]' \\ &\leq (\mathbf{X}'_0\mathbf{X}_0)^{-1}\mathbf{X}'_0\mathbf{B}^*\mathbf{X}_1\gamma\gamma'\mathbf{X}'_1\mathbf{B}'\mathbf{X}_0(\mathbf{X}'_0\mathbf{X}_0)^{-1}. \end{aligned}$$

From the proof of Theorem 3, the inequality (6.51) holds. □

We give two examples of Theorem 4.

Example 1. (Bivariate Normal Distribution) Suppose that in Model (6.1), $\mathbf{X}_1 = \mathbf{I}_n$, $n = q$, and

$$\mathbf{x}_i \stackrel{\text{indep.}}{\sim} N_n(\mathbf{0}, \sigma_{\mathbf{x}_i}^2 \mathbf{R}_{\mathbf{x}_i}), \mathbf{i} = 1, \dots, \mathbf{p}. \quad (6.52)$$

Define

$$\mathbf{R}_x^{1/2} = \text{diag}(\mathbf{R}_{x_1}^{1/2}, \dots, \mathbf{R}_{x_p}^{1/2}), \Sigma_{xx} = \text{diag}(\sigma_{x_1}^2, \dots, \sigma_{x_p}^2),$$

and

$$\boldsymbol{\sigma}_x = (\rho_1 \sigma_{x_1}, \dots, \rho_p \sigma_{x_p})', \quad (6.53)$$

and consider the bivariate normal relationship between \mathbf{x} and \mathbf{u}

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{u} \end{pmatrix} \sim N_{(p+1)n} \left(\mathbf{0}, \begin{pmatrix} (\Sigma_{xx} \otimes \mathbf{I}_n) \mathbf{R}_x & \sqrt{\delta_1} \mathbf{R}_x^{1/2} (\boldsymbol{\sigma}_x \otimes \mathbf{I}_n) \boldsymbol{\Psi}_1^{1/2} \\ \sqrt{\delta_1} \boldsymbol{\Psi}_1^{1/2} (\boldsymbol{\sigma}_x' \otimes \mathbf{I}_n) \mathbf{R}_x^{1/2} & \delta_1 \boldsymbol{\Psi}_1 \end{pmatrix} \right). \quad (6.54)$$

In this setting,

$$\boldsymbol{\sigma}_x' \Sigma_{xx}^{-1} = \left(\frac{\rho_1}{\sigma_{x_1}}, \dots, \frac{\rho_p}{\sigma_{x_p}} \right),$$

hence

$$\mathbf{u} \mid \mathbf{x} \sim N_n(\boldsymbol{\Psi}_1^{1/2} (\sqrt{\delta_1} \boldsymbol{\sigma}_x' \Sigma_{xx}^{-1} \otimes \mathbf{I}_n) \mathbf{R}_x^{-1/2} \mathbf{x}, \delta_1 (1 - \sum_{i=1}^p \rho_i^2) \boldsymbol{\Psi}_1). \quad (6.55)$$

Therefore, $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\Sigma}}$ defined in (6.37) have the expression

$$\boldsymbol{\gamma} = \sqrt{\delta_1} \sum_{i=1}^p \frac{\rho_i}{\sigma_{x_i}} \boldsymbol{\Psi}_1^{1/2} \mathbf{R}_x^{-1/2} \mathbf{x}, \quad \tilde{\boldsymbol{\Sigma}} = \delta_1 \left(\mathbf{I} - \sum_{i=1}^p \rho_i^2 \boldsymbol{\Psi}_1 \right).$$

Hence,

$$\text{Var}(\mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0 \left(\boldsymbol{\Sigma} - \frac{1}{\eta} \sum_{i=1}^p \rho_i^2 \boldsymbol{\Psi}_1 \right).$$

Corollary 3. Assume that in model (6.1), \mathbf{x} defined in (6.36) and \mathbf{u} satisfies (6.54). $\hat{\boldsymbol{\beta}}_B$ has the smallest variance and MSE. The order of the variances follows (6.50), and the order of the MSEs follows (6.51).

Furthermore, if

$$\boldsymbol{\Psi}_1 = \mathbf{R}_{x_1} = \cdots = \mathbf{R}_{x_p} = \mathbf{R}, \quad (6.56)$$

we have

$$\boldsymbol{\gamma} = \sqrt{\delta_1} \sum_{i=1}^p \frac{\rho_i}{\sigma_{x_i}} \mathbf{x}.$$

Note that Page et al. (2011) discussed a special case of this example by assuming $p = 2$, and the first column of \mathbf{X}_0 is the vector of ones. In this case

$$\text{Bias}(\hat{\boldsymbol{\beta}}_G \mid \mathbf{X}_0, \boldsymbol{\beta}) = \sqrt{\delta_1} \frac{\rho_1}{\sigma_{x_1}} (\mathbf{X}_0' \boldsymbol{\Sigma}^{-1} \mathbf{X}_0)^{-1} \mathbf{X}_0' \boldsymbol{\Sigma}^{-1} \mathbf{x},$$

which is the same as Paciorek (2010). Moreover, it is easy to show that

$$\boldsymbol{\Sigma}_M^* - \mathbf{I}_n = \rho_1^2 \frac{\delta_0}{\eta} \left(\frac{\mathbf{x} \mathbf{x}'}{\sigma_{x_1}^2} - \boldsymbol{\Psi}_1 \right).$$

Example 2. (Linear Relationship) A common relationship between \mathbf{x} and \mathbf{u} would be a linear relationship. Suppose that

$$\mathbf{x} = \mathbf{C}\mathbf{u} + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \delta_2 \mathbf{\Psi}_2), \quad (6.57)$$

where \mathbf{C} is an $np \times q$ matrix. Then $\mathbf{x} \mid \mathbf{u} \sim N(\mathbf{C}\mathbf{u}, \delta_2 \mathbf{\Psi}_2)$. It is easy to show that

$$\mathbf{u} \mid \mathbf{x} \sim N\left(\mathbf{G}_2^{-1} \mathbf{C}' \mathbf{\Psi}_2^{-1} \mathbf{x}, \delta_2 \mathbf{G}_2^{-1}\right), \quad (6.58)$$

where

$$\mathbf{G}_2 = \mathbf{C}' \mathbf{\Psi}_2^{-1} \mathbf{C} + \frac{1}{\eta_{12}} \mathbf{\Psi}_1^{-1}, \quad \text{and } \eta_{12} = \frac{\delta_1}{\delta_2}. \quad (6.59)$$

Therefore, $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\Sigma}}$ defined in (6.37) have the expressions

$$\boldsymbol{\gamma} = \mathbf{G}_2^{-1} \mathbf{C}' \mathbf{\Psi}_2^{-1} \mathbf{x}, \quad \tilde{\boldsymbol{\Sigma}} = \delta_2 \mathbf{G}_2^{-1}.$$

We have

$$\text{Var}(\mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0 \left(\mathbf{I}_n + \frac{1}{\eta_2} \mathbf{X}_1 \mathbf{G}_2^{-1} \mathbf{X}_1' \right),$$

where

$$\eta_2 = \frac{\delta_0}{\delta_2}. \quad (6.60)$$

Corollary 4. Assume that in model (6.1), \mathbf{x} is defined in (6.36) and \mathbf{u} is linearly related as in (6.57).

(a) $\hat{\boldsymbol{\beta}}_B$ has the smallest variance, and the order of variances follows (6.50).

(b) All estimators are biased, and $\widehat{\boldsymbol{\beta}}_B$ has the smallest MSE. Moreover, the order of MSEs follows (6.51).

Note that if δ_2 tends to infinity, which leads to independence of \mathbf{x} and \mathbf{u} , $\boldsymbol{\gamma}$ would tend to be $\mathbf{0}$, and $\widetilde{\boldsymbol{\Sigma}}$ would tend to be $\delta_1 \boldsymbol{\Psi}_1$. Hence, $\boldsymbol{\Sigma}_V^*$ defined in (6.40) would tend to be $\boldsymbol{\Sigma}$, and Corollary 4 would become Theorem 3 exactly.

6.4 Posterior Predictor of the Random Effect \mathbf{u}

The estimators of \mathbf{u} corresponding to the estimators of $\boldsymbol{\beta}$ are given in (6.9), (6.18), and (6.26). We suppose \mathbf{X} and \mathbf{u} are correlated without loss of generality. We define

$$\mathbf{E}_y = \mathbb{E}(\mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}), \quad \mathbf{V}_y = \text{Var}(\mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}), \quad (6.61)$$

and use $\boldsymbol{\gamma}$ as defined in (6.37). Then the expected values, variances and MSEs of the estimators of $\widehat{\mathbf{u}}_G$, $\widetilde{\mathbf{u}}_O$ and $\widehat{\mathbf{u}}_B$ are

$$\mathbb{E}(\widehat{\mathbf{u}}_G \mid \mathbf{X}_0, \boldsymbol{\beta}) = \text{Bias}(\widehat{\mathbf{u}}_G \mid \mathbf{X}_0, \boldsymbol{\beta}) = \frac{1}{\eta} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \boldsymbol{\Sigma}^{-1/2} \mathbf{C}_G \boldsymbol{\Sigma}^{-1/2} \mathbf{X}_1 \boldsymbol{\gamma}, \quad (6.62)$$

$$\text{Var}(\widehat{\mathbf{u}}_G \mid \mathbf{X}_0, \boldsymbol{\beta}) = \frac{1}{\eta^2} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \boldsymbol{\Sigma}^{-1/2} \mathbf{C}_G \boldsymbol{\Sigma}^{-1/2} \mathbf{V}_y \boldsymbol{\Sigma}^{-1/2} \mathbf{C}_G \boldsymbol{\Sigma}^{-1/2} \mathbf{X}_1 \boldsymbol{\Psi}_1, \quad (6.63)$$

$$\mathbb{E}(\widetilde{\mathbf{u}}_O \mid \mathbf{X}_0, \boldsymbol{\beta}) = \text{Bias}(\widetilde{\mathbf{u}}_O \mid \mathbf{X}_0, \boldsymbol{\beta}) = \frac{1}{\eta} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \mathbf{C}_O \boldsymbol{\Sigma}^{-1} \mathbf{X}_1 \boldsymbol{\gamma}, \quad (6.64)$$

$$\text{Var}(\widetilde{\mathbf{u}}_O \mid \mathbf{X}_0, \boldsymbol{\beta}) = \frac{1}{\eta^2} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \mathbf{C}_O \boldsymbol{\Sigma}^{-1} \mathbf{V}_y \boldsymbol{\Sigma}^{-1} \mathbf{C}_O \mathbf{X}_1 \boldsymbol{\Psi}_1, \quad (6.65)$$

$$\mathbb{E}(\widehat{\mathbf{u}}_B \mid \mathbf{X}_0, \boldsymbol{\beta}) = \boldsymbol{\Phi}_1^{-1} \mathbf{X}'_1 (\mathbf{I}_n - \mathbf{P}_0) \mathbf{E}_y, \quad (6.66)$$

$$\text{Var}(\widehat{\mathbf{u}}_B \mid \mathbf{X}_0, \boldsymbol{\beta}) = \boldsymbol{\Phi}_1^{-1} \mathbf{X}'_1 (\mathbf{I}_n - \mathbf{P}_0) \mathbf{V}_y (\mathbf{I}_n - \mathbf{P}_0) \mathbf{X}_1 \boldsymbol{\Phi}_1^{-1}. \quad (6.67)$$

The MSEs can be computed by

$$\text{MSE} = \text{Variance} + \text{Bias} \cdot \text{Bias}'.$$

Lemma 2. *Suppose \mathbf{S} and \mathbf{T} are two different $n \times n$ idempotent matrices with the same rank r . Then \mathbf{S} and \mathbf{T} cannot be ordered.*

Proof. Since \mathbf{S} is idempotent, there exist an orthogonal matrix \mathbf{P} such that $\mathbf{PSP}' = \text{diag}(\mathbf{I}_r, \mathbf{0})$. It is easy to show that \mathbf{PTP}' is idempotent also, and $\text{rank}(\mathbf{PTP}') = r$. Therefore, the diagonal elements of \mathbf{PTP}' are between 0 and 1 inclusively, and if they are equal to 0 and 1, the corresponding row and column would be zero.

If $\mathbf{S} \leq \mathbf{T}$, then

$$\text{diag}(\mathbf{I}_r, \mathbf{0}) - \mathbf{PTP}' \leq \mathbf{0}.$$

Hence the first r elements of \mathbf{PTP}' are 1, and $\mathbf{PTP}' = \text{diag}(\mathbf{I}_r, \mathbf{0})$, so

$$\mathbf{S} - \mathbf{T} = \mathbf{0}. \tag{6.68}$$

On the other hand, if $\mathbf{T} \leq \mathbf{S}$, then the same result as (6.68) can be derived. Therefore, the two matrices can not be ordered. \square

Noting that \mathbf{C}_G and \mathbf{C}_O are idempotent matrices with rank p , we consider

$$\Sigma^{-1/2}\mathbf{C}_G\Sigma^{-1}\mathbf{C}_G\Sigma^{-1/2} \quad \text{and} \quad \mathbf{C}_O\Sigma^{-2}\mathbf{C}_O.$$

According to Lemma 2, they cannot be ordered. Then Lemma 1 cannot be applied here, and the variances of the estimators of the random effect \mathbf{u} cannot be ordered.

Moreover, both the expected values and the variances are decided by the known parameter η . If δ_1 , the variance coefficient of the random effect, has larger scale than the covariance coefficient δ_0 , they tend to be larger; whereas if δ_1 has smaller scale than δ_0 , they would remain in an acceptable range.

We consider the special case that \mathbf{X}_0 and \mathbf{u} are independent. Section 6.3.2 noted that (6.38) and (6.39) hold for the independence case. Therefore, (6.62), (6.64) and (6.66) become

$$E(\hat{\mathbf{u}}_G | \mathbf{X}_0, \boldsymbol{\beta}) = \mathbf{0}, \quad (6.69)$$

$$MSE(\hat{\mathbf{u}}_G | \mathbf{X}_0, \boldsymbol{\beta}) = Var(\hat{\mathbf{u}}_G | \mathbf{X}_0, \boldsymbol{\beta}) = \frac{\delta_0}{\eta^2} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \boldsymbol{\Sigma}^{-1/2} \mathbf{C}_G \boldsymbol{\Sigma}^{-1/2} \mathbf{X}_1 \boldsymbol{\Psi}_1; \quad (6.70)$$

$$E(\hat{\mathbf{u}}_O | \mathbf{X}_0, \boldsymbol{\beta}) = \mathbf{0}, \quad (6.71)$$

$$MSE(\hat{\mathbf{u}}_O | \mathbf{X}_0, \boldsymbol{\beta}) = Var(\hat{\mathbf{u}}_O | \mathbf{X}_0, \boldsymbol{\beta}) = \frac{\delta_0}{\eta^2} \boldsymbol{\Psi}_1 \mathbf{X}'_1 \mathbf{C}_O \boldsymbol{\Sigma}^{-1} \mathbf{C}_O \mathbf{X}_1 \boldsymbol{\Psi}_1; \quad (6.72)$$

$$E(\hat{\mathbf{u}}_B | \mathbf{X}_0, \boldsymbol{\beta}) = \boldsymbol{\Phi}_1^{-1} \mathbf{X}'_1 (\mathbf{I}_n - \mathbf{P}_0) \mathbf{X}_0 \boldsymbol{\beta}, \quad (6.73)$$

$$Var(\hat{\mathbf{u}}_B | \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0 \boldsymbol{\Phi}_1^{-1} \mathbf{X}'_1 (\mathbf{I}_n - \mathbf{P}_0) \boldsymbol{\Sigma} (\mathbf{I}_n - \mathbf{P}_0) \mathbf{X}_1 \boldsymbol{\Phi}_1^{-1}, \quad (6.74)$$

$$\begin{aligned} MSE(\hat{\mathbf{u}}_B | \mathbf{X}_0, \boldsymbol{\beta}) &= Var(\hat{\mathbf{u}}_B | \mathbf{X}_0, \boldsymbol{\beta}) \\ &+ [Bias(\hat{\mathbf{u}}_B | \mathbf{X}_0, \boldsymbol{\beta})][Bias(\hat{\mathbf{u}}_B | \mathbf{X}_0, \boldsymbol{\beta})]'. \end{aligned} \quad (6.75)$$

6.5 Posterior Prediction

Consider the predictors of \mathbf{y} given by (6.10), (6.19), and (6.27) from the three different methods of estimation. Define

$$\mathbf{Q} = (\mathbf{I}_n - \mathbf{P}_0)(\mathbf{I}_n - \mathbf{X}_1 \boldsymbol{\Phi}_1^{-1} \mathbf{X}'_1)(\mathbf{I}_n - \mathbf{P}_0). \quad (6.76)$$

Then the variances of the prediction error are

$$Var(\hat{\mathbf{y}}_G - \mathbf{y} | \mathbf{X}_0, \boldsymbol{\beta}) = \boldsymbol{\Sigma}^{-1/2} \mathbf{C}_G \boldsymbol{\Sigma}^{-1/2} \mathbf{V}_y \boldsymbol{\Sigma}^{-1/2} \mathbf{C}_G \boldsymbol{\Sigma}^{-1/2}, \quad (6.77)$$

$$Var(\hat{\mathbf{y}}_O - \mathbf{y} | \mathbf{X}_0, \boldsymbol{\beta}) = \mathbf{C}_O \boldsymbol{\Sigma}^{-1} \mathbf{V}_y \boldsymbol{\Sigma}^{-1} \mathbf{C}_O, \quad (6.78)$$

$$Var(\hat{\mathbf{y}}_B - \mathbf{y} | \mathbf{X}_0, \boldsymbol{\beta}) = \mathbf{Q} \mathbf{V}_y \mathbf{Q}. \quad (6.79)$$

The scale of the variances of the prediction error is decided by \mathbf{V}_y , while \mathbf{V}_y depends

on the relationship between \mathbf{X}_0 and \mathbf{u} , so to determine the variances, the relationship should be examined first.

We consider the special case that \mathbf{X}_0 and \mathbf{u} are independent. Then (6.77), (6.78) and (6.79) become

$$\text{Var}(\widehat{\mathbf{y}}_G - \mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0 \boldsymbol{\Sigma}^{-1/2} \mathbf{C}_G \boldsymbol{\Sigma}^{-1/2}, \quad (6.80)$$

$$\text{Var}(\widehat{\mathbf{y}}_O - \mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}) = \delta_0 \mathbf{C}_O \boldsymbol{\Sigma}^{-1} \mathbf{C}_O, \quad (6.81)$$

$$\text{Var}(\widehat{\mathbf{y}}_B - \mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta}) = \mathbf{Q} \boldsymbol{\Sigma} \mathbf{Q}. \quad (6.82)$$

6.6 Bayes Factor

In Model (6.1), we would like to consider the Bayes factor in testing

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \text{ vs } H_1 : \boldsymbol{\beta} \neq \mathbf{0}.$$

Let the hypotheses be M_0 and M_1 . The Bayes factor of M_1 to M_0 for (6.1) is defined by

$$B_{10} = \frac{m_1(\mathbf{y})}{m_0(\mathbf{y})} = \frac{\int f_1(\mathbf{y} \mid \boldsymbol{\theta}_1) \pi_1(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1}{\int f_0(\mathbf{y} \mid \boldsymbol{\theta}_0) \pi_0(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0}, \quad (6.83)$$

where $m_j(\mathbf{y})$ is the marginal density of \mathbf{y} under M_j , and $\boldsymbol{\theta}_j$ is other parameters, $j = 0, 1$. In addition, we assume $\boldsymbol{\theta}_j$ is fixed, and the prior of (δ_0, β_0) is

$$(\delta_0, \beta_0) \propto \frac{1}{\delta_0}. \quad (6.84)$$

We continue using the notations defined in (6.61). Since $E(\mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\beta} = \mathbf{0}) = \mathbf{X}_1 \boldsymbol{\gamma}$, for the Bayes Factor in testing H_0 , we have

$$m_0(\mathbf{y} \mid \mathbf{X}_0, \boldsymbol{\theta}_j) = (2\pi)^{-n/2} |\mathbf{V}_y|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\gamma})' \mathbf{V}_y^{-1} (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\gamma}) \right\}.$$

Let $\omega_{yy} = (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\gamma})' \mathbf{V}_y^{-1} (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\gamma})$. Then

$$m_0(\mathbf{y}) = (2\pi)^{-n/2} |\mathbf{V}_y|^{-1/2} e^{-\omega_{yy}/2}.$$

On the other hand, under M_1 , suppose that $\boldsymbol{\beta}$ follows the g-prior in Zellner (1986),

$$[\boldsymbol{\beta} \mid g, \delta_0] \sim N_p(0, g\delta_0(\mathbf{X}'_0\mathbf{X}_0/n)^{-1}),$$

for some fixed $g > 0$. A hierarchical prior for $\boldsymbol{\beta}$ is often obtained from assuming a prior density $\pi(g)$ for g . For example, Zellner & Siow (1980) assumed a Cauchy prior for $\boldsymbol{\beta}$, which is a special case here by assuming $g \sim \chi_1^2$.

We suppose that all the parameters other than $\boldsymbol{\beta}$ are fixed. Define

$$\boldsymbol{\omega}_{xy} = \mathbf{X}'_0\mathbf{V}_y^{-1}(\mathbf{y} - \mathbf{X}_1\boldsymbol{\gamma}), \quad \boldsymbol{\omega}_{xx} = \mathbf{X}'_0\mathbf{V}_y^{-1}\mathbf{X}_0,$$

and

$$\boldsymbol{\Omega}(g) = (\boldsymbol{\omega}_{xx} + \mathbf{X}'_0\mathbf{X}_0/(ng\delta_0))^{-1}. \quad (6.85)$$

Then for H_1 , we have

$$\begin{aligned} m_1(\mathbf{y}) &= \int_g \int_{\boldsymbol{\beta}} f_1(\mathbf{y} \mid \boldsymbol{\beta} \neq \mathbf{0}) \pi_1(\boldsymbol{\beta} \mid g) \pi(g) d\boldsymbol{\beta} dg \\ &= (2\pi)^{-(n+p)/2} |\mathbf{V}_y|^{-1/2} \left| \frac{\mathbf{X}'\mathbf{X}}{n\delta_0} \right|^{1/2} e^{-\omega_{yy}/2} \int_0^\infty \pi(g) \exp \left\{ \frac{1}{2} \boldsymbol{\omega}'_{xy} \boldsymbol{\Omega}(g) \boldsymbol{\omega}_{xy} \right\} \\ &\quad \int_{\boldsymbol{\beta}} \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta} - \boldsymbol{\omega}'_{xy} \boldsymbol{\Omega}(g) \right]' \boldsymbol{\Omega}(g)^{-1} \left[\boldsymbol{\beta} - \boldsymbol{\omega}'_{xy} \boldsymbol{\Omega}(g) \right] \right\} d\boldsymbol{\beta} dg \\ &= (2\pi)^{-n/2} |\mathbf{V}_y|^{-1/2} \left| \frac{\mathbf{X}'\mathbf{X}}{n\delta_0} \right|^{1/2} e^{-\omega_{yy}/2} \\ &\quad \int_0^\infty \pi(g) \left| \boldsymbol{\Omega}(g) \right|^{1/2} \exp \left\{ \frac{1}{2} \boldsymbol{\omega}'_{xy} \boldsymbol{\Omega}(g) \boldsymbol{\omega}_{xy} \right\} dg. \end{aligned} \quad (6.86)$$

Therefore,

$$B_{10} = \left| \frac{\mathbf{X}'\mathbf{X}}{n\delta_0} \right|^{1/2} \int_0^\infty \pi(g) \left| \boldsymbol{\Omega}(g) \right|^{1/2} \exp \left\{ \frac{1}{2} \boldsymbol{\omega}'_{xy} \boldsymbol{\Omega}(g) \boldsymbol{\omega}_{xy} \right\} dg. \quad (6.87)$$

A Monte Carlo simulation can be done based on the conclusion to calculate B_{10} .

Bibliography

- Anselin, L. (2002), ‘Under the hood issues in the specification and interpretation of spatial regression models’, *Agricultural Economics* **27**(3), 247–267.
- Arjas, E. & Gasbarra, D. (1994), ‘Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler’, *Statistica Sinica* **4**, 505–524.
- Ashing-Giwa, K., Ganz, P. A. & Petersen, L. (1999), ‘Quality of life of African-American and white long term breast carcinoma survivors’, *Cancer* **85**, 418–426.
- Ashing-Giwa, K., Padilla, G., Tejero, J. & Kim, J. (2004), ‘Breast cancer survivorship in a multiethnic sample: challenges in recruitment and measurement’, *Cancer* **101**, 450–465.
- Ashing-Giwa, K. T., Tejero, J. S., Kim, J., Padilla, G. V. & Hellemann, G. (2007), ‘Examining predictive models of HRQOL in a population-based, multiethnic sample of women with breast carcinoma’, *Quality of Life Research* **16**, 413–428.
- Aslanidou, H., Dey, D. K. & Sinha, D. (1998), ‘Bayesian analysis of multivariate survival data using Monte Carlo methods’, *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **26**, 33–48.
- Banerjee, S. & Carlin, B. P. (2004), ‘Parametric spatial cure rate models for interval-censored time-to-relapse data’, *Biometrics* **60**, 268–275.

- Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2003), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC.
- Banerjee, S., Wall, M. M. & Carlin, B. P. (2003), ‘Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota’, *Biostatistics* **4**, **1**, 123–142.
- Berger, J. O. & Sun, D. (1993), ‘Bayesian analysis for the poly-Weibull distribution’, *Journal of the American Statistical Association* **88**, 1412–1418.
- Berkson, J. & Gage, R. P. (1952), ‘Survival curve for cancer patients following treatment’, *Journal of American Statistical Association* **47**, 501–515.
- Besag, J. (1974), ‘Spatial interaction and the statistical analysis of lattice systems’, *Journal of the Royal Statistical Society, Series B (Methodological)* **36**, 192–236.
- Best, N., Richardson, S. & Thomson, A. (2005), ‘A comparison of Bayesian spatial models for disease mapping’, *Statistical Methods in Medical Research* **14**, 35–59.
- Bhatia, R. (2006), *Positive Definite Matrices*, Princeton University Press.
- Breslow, N. E. & Clayton, D. G. (1993), ‘Approximate inference in generalized linear mixed models’, *Journal of the American Statistical Association* **88**, 9–24.
- Carlin, B. P. & Banerjee, S. (2002), ‘Hierarchical multivariate CAR models for spatio-temporally correlated survival’, *Bayesian Statistics* **45**, 45–63.
- Chen, M.-H., Ibrahim, J. G. & Sinha, D. (1999), ‘A new Bayesian model for survival data with a surviving fraction’, *Journal of the American Statistical Association* **94**, 909–919.

- Clayton, D. & Kaldor, J. (1987), ‘Empirical Bayes estimates of age-standardized relative risks for use in disease mapping’, *Biometrics* **43**, 671–681.
- Cooner, F., Banerjee, S., Carlin, B. P. & Sinha, D. (2008), ‘Flexible cure rate modeling under latent activation schemes’, *Journal of the American Statistical Association* **102(478)**, 560–572.
- Cox, D. (1972), ‘Regression models and life-tables’, *Journal of the Royal Statistical Society B* **34**, 187–220.
- Cox, D. R. & Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman and Hall.
- Cressie, N. (1993), *Statistics for Spatial Data*, Wiley-Interscience; Rev Sub edition (January 15, 1993).
- Cressie, N. & Wikle, C. K. (2011), *Statistics for Spatio-Temporal Data*, Wiley Series in Probability and Statistics.
- Dai, L., He, Z., Sun, D. & Schootman, M. (2008*a*), ‘Bayesian nonparametric spatial modeling of breast cancer survival with cure fractions’, *Ph.D Thesis* pp. 129–180.
- Dai, L., He, Z., Sun, D. & Schootman, M. (2008*b*), ‘Hierarchical models for detecting multi-level geographical effects in cancer survival’, *Ph.D Thesis* pp. 76–86.
- David, H. A. & Moeschberger, M. L. (1978), *The theory of competing risks*, New York: Macmillan Pub. Co., Inc.
- Djibuti, M. & Shakarishvili, R. (2003), ‘Influence of clinical, demographic, and socioeconomic variables on quality of life in patients with epilepsy: findings from georgian study’, *J Neurol Neurosurg Psychiatry* **74**, 570–573.

- Dorval, M., Maunsell, E., Deschenes, L., Brisson, J. & Masse, B. (1998), 'Long-term quality of life after breast cancer: comparison of 8-year survivors with population controls', *Journal of Clinical Oncology* **16**, 487–494.
- Edge, S., Byrd, D., Compton, C., Fritz, A., Greene, F. & Trotti, A. (2010), *AJCC: Colon and rectum*, New York, NY: Springer.
- Eheman, C., Henley, S., Ballard-Barbash, R., Jacobs, E., Schymura, M., Noone, A., Pan, L., Anderson, R., Fulton, J., Kohler, B., Jemal, A., Ward, E., Plescia, M., Ries, L. & Edwards, B. (2012), 'Annual report to the nation on the status of cancer, 1975-2008, featuring cancers associated with excess weight and lack of sufficient physical activity', *Cancer* **118(9)**, 2338–2366.
- Engle, R. F., Granger, C. W. J., Rice, J. & Weiss, A. (1986), 'Semiparametric estimates of the relation between weather and electricity sales', *Journal of the American Statistical Association* **81, No. 394**, 310–320.
- Ganz, P. A., Coscarelli, A., Fred, C., Kahn, B., Polinsky, M. L. & Petersen, L. (2005), 'Breast cancer survivors: Psychosocial concerns and quality of life', *Breast Cancer Research and Treatment* **38**, 183–199.
- Ganz, P., Guadagnoli, E., Landrum, M., Lash, T., Rakowski, W. & Silliman, R. (2003), 'Breast cancer in older women: quality of life and psychosocial adjustment in the 15 months after diagnosis', *Journal of Clinical Oncology* **21**, 4027–33.
- Gelfand, A. E., Ghost, S. K., Christiansen, C., Soumarai, S. B. & McLaughlin, T. J. (2000), 'Proportional hazards models: A latent competing risk approach', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **49, Issue 3**, 385–397.
- Gelfand, A. E. & Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**, 398–409.

- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2003), *Bayesian Data Analysis, Second Edition*, Chapman and Hall/CRC; 2 edition.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2004), *Bayesian Data Analysis (2nd ed.)*, Boca Raton, Florida: Chapman & Hall/CRC. Jeffreys, H. Oxford University Press, London.
- Gelman, A. & Rubin, D. B. (1992), ‘Inference from iterative simulation using multiple sequences’, *Statistical Science* **7**, 457–472.
- Geman, S. & Geman, D. (1994), ‘Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images’, *I.E.E.E. Transactions: Pattern Analysis and Machine Intelligence* **12**, 609–628.
- Gilks, W. R. & Wild, P. (1992), ‘Adaptive rejection sampling for Gibbs sampling’, *Applied Statistics* **41**, 337–348.
- Hayes, R., Sherbourne, C. & Mazel, R. (1993), ‘The RAND 36-item health survey 1.0’, *Health Economics* **2**, 217–227.
- He, Z. & Sun, D. (2000), ‘Hierarchical Bayes estimation of hunting success rates with spatial correlations’, *Biometrics* **56**, 360–367.
- Henderson, C. (1975), ‘Best linear unbiased estimation and prediction under a selection model’, *Biometrics* **31**, 423–447.
- Howlader, N., Noone, A., Krapcho, M., Neyman, N., Aminou, R., Waldron, W., Altekruse, S., Kosary, C., Ruhl, J., Tatalovich, Z., Cho, H., Mariotto, A., Eisner, M., Lewis, D., Chen, H., Feuer, E., Cronin, K. & Edwards, B. (2011), *Seer cancer statistics review, 1975–2008*, Technical report, National Cancer Institute. Bethesda, MD.

- Intel® (2008), ‘Vector statistical library notes’, *Intel® Math Kernel Library – Documentation*, Intel Corporation pp. web–page.
- Intel® (2009), ‘Vector statistical library (VSL) performance data’, *Intel® Math Kernel Library 10.2 – Documentation*, Intel Corporation pp. web–page.
- Kalbfleisch, J. D. (1978), ‘Non-parametric Bayesian analysis of survival time data’, *Journal of the Royal Statistical Society* **B 40 No. 2**, 214–221.
- Kinderman, A. J. & Monahan, J. F. (1977), ‘Computer generation of random variables using the ratio of uniforms deviates’, *ACM Transactions on Mathematical Software* **3**, 257–260.
- King, M., Kenny, P., Shiell, A., Hall, J. & Boyages, J. (2000), ‘Quality of life three months and one year after first treatment for early stage breast cancer: Influence of treatment and patient characteristics’, *Quality of Life Research* **9**, 789–800.
- Kissane, D., Clarke, D., Ikin, J., Bloch, S., Smith, G., Vitetta, L. & McKenzie, D. (1998), ‘Psychological morbidity and quality of life in australian women with early-stage breast cancer: a cross-sectional survey’, *The Medical Journal of Australia* **169**, 192–196.
- Kornprat, P., Pollheimer, M., Lindtner, R., Schlemmer, A., Rehak, P. & Langner, C. (2011), ‘Value of tumor size as a prognostic variable in colorectal cancer: a critical reappraisal’, *Am J Clin Oncol.* **34(1)**, 43–49.
- Lee, E. T. & Wang, J. W. (2003), *Statistical methods for survival data analysis*, Wiley-Interscience; 3 edition.
- Montazeri, A., Hole, D., Milroy, R., McEwen, J. & Gillis, C. (2003), ‘Quality of

- life in lung cancer patients: does socioeconomic status matter?', *Health Qual Life Outcomes* **Jun 9**, 1–19.
- O'Campo, P. (2003), 'Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health', *American Journal of Epidemiology* **157**, 9–13.
- Omurlu, I. K., Ozdamar, K. & Ture, M. (2009), 'Comparison of Bayesian survival analysis and Cox regression analysis in simulated and breast cancer data sets', *Expert Systems with Applications* **36**, 11341–11346.
- Paciorek, C. J. (2010), 'The importance of scale for spatial-confounding bias and precision of spatial regression estimators', *Statistical Science* **25**, 107–125.
- Page, G. L., Liu, Y. & Sun, D. (2011), 'A solution to spatial confounding', *In Preparation* .
- Penson, D., Stoddard, M., Pasta, D., Lubeck, D., Flanders, S. & Litwin, M. (2001), 'The association between socioeconomic status, health insurance coverage, and quality of life in men with prostate cancer', *Journal of Clinical Epidemiology* **54**, 350–358.
- Pollock, A. M. & Vickers, N. (1997), 'Breast, lung and colorectal cancer incidence and survival in south thames region, 1987-1992: the effect of social deprivation', *Journal of Public Health* **19**, 288–294.
- Ross, C. E. & Mirowsky, J. (2001), 'Neighborhood disadvantage, disorder, and health', *Journal of Health and Social Behavior* **42**, 258–276.
- Ross, N., Tremblay, S. & Graham, K. (2004), 'Neighbourhood influences on health in Montral, Canada', *Social Science & Medicine* **59**, Issue 7, 1485–1494.

- Roux, D. A. (2003), *The examination of neighborhood effects on health: Conceptual and methodological issues related to the presence of multiple levels of organization, Neighborhoods and health*, Oxford: Oxford University Press.
- Rowland, J., Mariotto, A., Alfano, C., Pollack, L., Weir, H. & White, A. (2011), ‘Cancer survivors — united states, 2007’, *Morbidity and Mortality Weekly Report (MMWR)* **60(09)**, 269–272.
- Rustøen, T. R. & Begnum, S. (2000), ‘Quality of life in women with breast cancer: a review of the literature and implications for nursing practice’, *Cancer Nursing* **23**, 416–421.
- Sampson, R. (2003), ‘The neighborhood context of well-being’, *Perspectives in Biology and Medicine* **46 (3 Suppl)**, S53–64.
- Schenk, O. & Gärtner, K. (2004), ‘Solving unsymmetric sparse systems of linear equations with PARDISO’, *J. of Future Generation Computer Systems* **20(3)**, 475–487.
- Schootman, M. & Sun, D. (2004), ‘Small-area incidence trends in breast cancer’, *Epidemiology* **15**, 300–307.
- Schootman, M., Sun, D. & Jackson-Thompson, J. (2004), Neighborhood effects on quality of life in breast cancer, proposal, submitted to the National Cancer Institute.
- Singh, G. K., Miller, B. A. & Hankey, B. F. (2002b), ‘Changing area socioeconomic patterns in U.S. cancer mortality, 1950-1998: Part II-lung and colorectal cancers’, *JNCI Journal of the National Cancer Institute* **94**, 916–925.
- Singh, G. K., Miller, B. A., Hankey, B. F., Feuer, E. J. & Pickle, L. W. (2002a), ‘Changing area socioeconomic patterns in U.S. cancer mortality, 1950-1998: Part

- I-all cancers among men', *JNCI Journal of the National Cancer Institute* **94**, 916–925.
- Speckman, P. (1988), 'Kernel smoothing in partial linear model', *Journal of the Royal Statistical Society. Series B (Methodological)* **50**, 413–436.
- Stewart, A., Hays, R. & Ware, J. (1988), 'The MOS short-form general health survey: Reliability and validity in a patient population', *Medical Care* **26**, 724–735.
- Sun, D., Tsutakawa, R. & Speckman, P. (1999), 'Bayesian inference for CAR (1) models with noninformative priors', *Biometrika* **86**, 341–350.
- Taft, C., Karlsson, J. & Sullivan, M. (2001), 'Do SF-36 summary component scores accurately summarize subscale scores?', *Quality of Life Research* **10**, 395–404.
- USCS Working Group (2012), United states cancer statistics: 1999-2008 incidence and mortality web-based report, Technical report, Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute.
- Vacek, P. M., Winstead-Fry, P., Secker-Walker, R. H., Hooper, G. J. & Plante, D. A. (2003), 'Factors influencing quality of life in breast cancer survivors', *Quality of Life Research* **12**, 527–537.
- Ver Hoef, J. M., Cressie, N. A. C. & Glenn-Lewin, D. C. (1993), 'Spatial models for spatial statistics: some unification', *Journal of Vegetation Science* **4(4)**, 441–452.
- Wakefield, J. C., Gelfand, A. E. & Smith, A. F. M. (1991), 'Efficient generation of random variates via the ratio-of-uniforms method', *Statistics and Computing* **1, Number 2**, 129–133.
- Ware, J. J. (2000), 'SF-36 health survey update', *Spine* **25**, 3130–3139.

- Ware, J. & Sherbourne, C. (1992), ‘The MOS 36-item Short-Form Health Survey (SF-36)’, *Medical Care* **30**, 473–483.
- Wells, K., Stewart, A. & Hayes, R. (1989), ‘The functioning and well-being of depressed patients: Results from the medical outcomes study’, *Journal of the American Medical Association* **262**, 914–919.
- Wenzel, L. B., Fairclough, D. L., Brady, M. J., Cella, D., Garrett, K. M., Kluhsman, B. C., Crane, L. A. & Marcus, A. C. (1999), ‘Age-related differences in the quality of life of breast carcinoma patients after treatment’, *Cancer* **86**, 1768–1774.
- Wilson, J., Wells, W. T., Browne, A., Yun, S., Pashi, A. & Hogrefe, W. (June 2010), 2008 missouri behavioral risk factor surveillance system data report, Technical report, Jefferson City, MO: Missouri Department of Health and Senior Services. Office of Epidemiology.
- Yu, Y. & Meng, X.-L. (2011), ‘To center or not to center: That is not the question . an ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency’, *Journal of Computational and Graphical Statistics* **20(3)**, 531–570.
- Zellner, A. (1986), *On assessing prior distributions and Bayesian regression analysis with g-prior distributions*, North-Holland/Elsevier.
- Zellner, A. & Siow, A. (1980), Posterior odds ratios for selected regression hypothesis, *in* J. Bernardo, M. Degroot, D. Lindley & A. Smith, eds, ‘Bayesian Statistics: Proceedings of the First International meeting held in Valencia’, University of Valencia Press.

VITA

Yajun Liu was born in February 19, 1983 in Hohhot, Inner Mongolia, China. She received a B.A. in Mathematics from Inner Mongolia University of China in 2002. Two years later in 2004, she received a B.S. in Software Engineering from Tsinghua University of China. Before she joined the graduate program in the Department of Statistics at the University of Missouri in August of 2007, she had worked as a software engineer for three years in Beijing Risesoft Co., an IT company in Beijing, China. Yajun received a M.A. in Statistics from University of Missouri in 2009. She is expecting to receive her Ph.D. in Statistics from the University of Missouri in July of 2012. She will join University of Cincinnati as a visiting Assistant Professor as of August of 2012.