Parametric and Semi-parametric Cure Rate Models with Spatial Frailties for Interval-Censored Data

Bao Yiqi

Parametric and Semi-parametric Cure Rate Models with Spatial Frailties for Interval-Censored Data

Bao Yiqi

***Orientador*: Prof. Dr. Vicente Garibay Cancho**

Monografia apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística.

São Carlos

Março de 2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

Bao, Yiqi
B221p Parametric and semi-parametric cure rate models with spatial frailties for interval-censored data / Bao Yiqi. -- São Carlos : UFSCar, 2016.
175 p.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2016.

1. Inferência Bayesiana. 2. Fração de cura. 3. Diagnósticos de influência. 4. Fragilidade espacial. 5. Modelos de sobrevivência. I. Título.

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de tese de doutorado da candidata Bao Yiqi, realizada em 31/05/2016:

Prof. Dr. Vicente Garibay Cancho
USP

Prof. Dr. Edwin Moises Marcos Ortega
ESALQ-USP

Prof. Dr. Francisco Louzada Neto
USP

Prof. Dr. Heleno Bolfarine
USP

Prof. Dr. Josemar Rodrigues
USP

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Prof. Dr. Heleno Bolfarine e, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do aluno Eder Angelo Milani.

Prof. Dr. Vicente Garibay Cancho
Presidente da Comissão Examinadora
USP

Aos meus pais, ***Bao Ke Wei*** *e* ***Wang Cuiqin****, pelo incentivo e esforço para minha formação.*

Aos todos meus amigos, pela amizade e apoio.

Agradecimentos

Primeiramente agradeço a Deus por tudo, pela minha saúde, minha família, e por ter me dado a oportunidade desta conquista.

Ao meu orientador, Vicente Garibay Cancho, pela orientação, sugestões, paciêncuia e amizade que contribuíram para o meu crescimento e minha formação acadêmica.

Aos meus pais que sempre me ajudam, apoiam e incentivam, e a quem sempre estarei em dívida.

Ao meu orientador exterior, Dipak Dey, que me deu uma excelente oportunidade para desenvolver um parte do meu trabalho no University of Connecticut, EUA.

Aos professores Francisco Louzada Neto, Josemar Rodrigues, Adriano K. Suzuki, Heleno Bolfarine and Edwin Ortega, pelas sugestões.

Aos professores e funcionários da pós-graduação do departamento de Estatística da Universidade Federal de São Carlos, pelo excelente convívio.

A todos os meus amigos da USP, UFSCAR e UCONN, Aline, Daiane, Guto, Willian, Dainela, Larissa, Marcos e Lourdes pelos momentos compartilhados durante a minha formação.

A todas as pessoas que não foram nominalmente mencionadas, mas que de alguma forma contribuíram para viabilizar este trabalho.

Finalmente, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro concedido para a realização deste trabalho no Brasil e no exterior.

"Feliz aquele que transfere o que sabe e aprende o que ensina".
"O saber se aprende com os mestres. A sabedoria, só com o corriqueiro da vida".

Cora Coralina

Resumo

Nesta tese, estendemos os modelos flexíveis de sobrevivência com fração de cura, tais como os modelos de sobrevivência com fração de cura geométricos, binomial negativa e séries de potências, para permitir correlações espaciais incluindo fragilidades espaciais para os dados de censura intervalar. Modelos de cura paramétricos e semi-paramétricos com as fragilidades espaciais independentes e dependentes são propostos e comparados. Os modelos propostos abrangem vários modelos de cura bem conhecidos como seus casos particulares. Uma vez que estes modelos de cura são obtidos considerando que a ocorrência de um evento de interesse é causada pela presença de quaisquer riscos não observados, estudamos também os modelos de cura complementares, nesse caso, os modelos são obtidos assumindo que a ocorrência de um evento de interesse é causada quando todos os riscos, não observados, são ativados. Uma nova medida de seleção de modelo, baseada no paradigma da perda do preditivo, para dados de censura intervalar é proposta. Métodos MCMC são utilizados em uma abordagem de inferência Bayesiana sendo que os critérios de seleção de modelos Bayesiano são utilizados para comparação de modelos. Além disso, realizamos um diagnóstico de influência para detectar as possíveis observações influentes ou extremas que podem causar distorções sobre os resultados da análise. Finalmente, os modelos propostos são aplicados para analisar um conjunto de dados real de abstenção tabágica.

Palavras-chave: Inferência Bayesiana; Fração de cura; Diagnósticos de influência; Fragilidade espacial; Modelos de sobrevivência.

Abstract

In this thesis, we extend some flexible cure rate models, such as the geometric, negative binomial and power series cure rate models, to allow for spatial correlations by including spatial frailties for the interval censored data setting. Parametric and semi-parametric cure rate models with independent and dependent spatial frailties are proposed and compared. The proposed models encompass several well-known cure rate models as its particular cases. Since these cure rate models are obtained by considering that the occurrence of an event of interest is caused by the presence of any non-observed risks, we also study the complementary cure model, which arises when the cure rate models are obtained by assuming the occurrence of an event of interest is caused when all of non-observed risks are activated. A new measure of model selection, based on the notion of predictive loss paradigm, for the interval-censoring data is also proposed. The MCMC method is used in a Bayesian inference approach and some Bayesian model selection criteria are used for model comparison. Moreover, we conduct an influence diagnostics to detect possible influential or extreme observations that can cause distortions on the results of analysis. Finally, the proposed models are applied to analyze a real dataset from a stop smoking study.

Keywords: Bayesian inference; Cure fraction; Influence diagnostics; Spatial frailty; Survival models.

List of Figures

1.1 Estimated survival functions considering the intervals and its midpoints. 4

1.2 Cities in which are collected the smoking cessation data. 5

2.1 (a) Survival function and (b) hazard function of Weibull distribution with $\lambda = 0$ and different selected values of α. 9

2.2 (a) Survival function and (b) hazard function of piecewise exponential distribution with different selected values of λ. 11

2.3 (a) Survival function and (b) hazard function of logGEV distribution with $\mu = 0$ and $\sigma = 1$. 12

2.4 (a) Survival function and (b) hazard function of logGEV distribution with $\mu = 0$ and $\sigma = 1.5$. 13

2.5 (a) Probability density function and (b) cumulative distribution function of gamma distribution with different selected values of α and β. 14

2.6 (a) Probability density function and (b) cumulative distribution function of gamma distribution with different values of α and β. 15

3.1 Index plots of Kullback-Leibler divergence measure from the fitted of the WGCR model considering prior 1. 60

3.2 Index plots of Kullback-Leibler divergence measure from the fitted WGCR model considering prior 2. 60

3.3 Index plots of J-distance from the fitted WGCR model considering prior 1. 61

3.4 Index plots of J-distance from the fitted WGCR model considering prior 2. 61

3.5 Index plots of L_1 norm distance from the fitted WGCR model considering prior 1. . . 62

3.6 Index plots of L_1 norm distance from the fitted WGCR model considering prior 2. . . 62

3.7 Index plots of χ^2-square divergence from the fitted WGCR model considering prior 1. 63

3.8 Index plots of χ^2-square divergence from the fitted WGCR model considering prior 2. 63

3.9 Index plots of Kullback-Leibler divergence measure from the fitted CWGCR model considering prior 1. 64

3.10 Index plots of Kullback-Leibler divergence measure from the fitted CWGCR model considering prior 2. 64

3.11 Index plots of J-distance from the fitted CWGCR model considering prior 1. 65

3.12 Index plots of J-distance from the fitted CWGCR model considering prior 2. 65

3.13 Index plots of L_1 norm distance from the fitted CWGCR model considering prior 1. . 66

3.14 Index plots of L_1 norm distance from the fitted CWGCR model considering prior 2. . 66

3.15 Index plots of χ^2-square divergence from the fitted CWGCR model considering prior 1. 67

3.16 Index plots of χ^2-square divergence from the fitted CWGCR model considering prior 2. 67

3.17 Index plots of Kullback-Leibler divergence measure from the fitted PHGCR model considering prior 1. 68

3.18 Index plots of Kullback-Leibler divergence measure from the fitted PHGCR model considering prior 2. 68

3.19 Index plots of J-distance from the fitted PHGCR model considering prior 1. 69

3.20 Index plots of J-distance from the fitted PHGCR model considering prior 2. 69

3.21 Index plots of L_1 norm distance from the fitted PHGCR model considering prior 1. . 70

3.22 Index plots of L_1 norm distance from the fitted PHGCR model considering prior 2. . 70

3.23 Index plots of χ^2-square divergence from the fitted PHGCR model considering prior 1. 71

3.24 Index plots of χ^2-square divergence from the fitted PHGCR model considering prior 2. 71

3.25 Index plots of Kullback-Leibler divergence measure from the fitted CPHGCR model considering prior 1. 72

3.26 Index plots of Kullback-Leibler divergence measure from the fitted CPHGCR model considering prior 2. 72

3.27 Index plots of J-distance from the fitted CPHGCR model considering prior 1. 73

3.28 Index plots of J-distance from the fitted CPHGCR mode considering prior 2. 73

3.29 Index plots of L_1 norm distance from the fitted CPHGCR model considering prior 1. . 74

3.30 Index plots of L_1 norm distance from the fitted CPHGCR model considering prior 2. . 74

3.31 Index plots of χ^2-square divergence from the fitted CPHGCR model considering prior 1. 75

3.32 Index plots of χ^2-square divergence from the fitted CPHGCR model considering prior 2. 75

3.33 Maps of posterior means for frailties $\boldsymbol{U}$ (left panel) and $\boldsymbol{V}$ (right panel) in model 15. . 78

3.34 Maps of posterior standard derivations for frailties $\boldsymbol{U}$ (left panel) and $\boldsymbol{V}$ (right panel) in model 15. 78

3.35 Estimates of ψ-divergence measures for Model 15 79

3.36 Posterior means and credibility intervals of α_i's 82

3.37 Maps of posterior means for frailties $\boldsymbol{U}$ (left panel) and $\boldsymbol{V}$ (right panel). 83

3.38 Maps of posterior standard derivations for frailties $\boldsymbol{U}$ (left panel) and $\boldsymbol{V}$ (right panel). 83

3.39 Estimates of ψ-divergence measures for Model 9 84

3.40 Posterior means and credibility intervals of α_i's for the data set without individuals 14 and 86. 85

3.41 Index plots of Kullback-Leibler divergence measure from the fitted WNBCR model considering prior 1. 104

3.42 Index plots of Kullback-Leibler divergence measure from the fitted WNBCR model considering prior 2. 104

3.43 Index plots of J-distance from the fitted WNBCR model considering prior 1. 105

3.44 Index plots of J-distance from the fitted WNBCR model considering prior 2. 105

3.45 Index plots of L_1 norm distance from the fitted WNBCR model considering prior 1. . 106

3.46 Index plots of L_1 norm distance from the fitted WNBCR model considering prior 2. . 106

3.47 Index plots of χ^2-square divergence from the fitted WNBCR model considering prior 1.107

3.48 Index plots of χ^2-square divergence from the fitted WNBCR model considering prior 2.107

3.49 Index plots of Kullback-Leibler divergence measure from the fitted CWNBCR model considering prior 1. 108

3.50 Index plots of Kullback-Leibler divergence measure from the fitted CWNBCR model considering prior 2. 108

3.51 Index plots of J-distance from the fitted CWNBCR model considering prior 1. 109

3.52 Index plots of J-distance from the fitted CWNBCR model considering prior 2. 109

3.53 Index plots of L_1 norm distance from the fitted CWNBCR model considering prior 1. 110

3.54 Index plots of L_1 norm distance from the fitted CWNBCR model considering prior 2. 110

3.55 Index plots of χ^2-square divergence from the fitted CWNBCR model considering prior 1. 111

3.56 Index plots of χ^2-square divergence from the fitted CWNBCR model considering prior 2. 111

3.57 Index plots of Kullback-Leibler divergence measure from the fitted PHNBCR model considering prior 1. 112

3.58 Index plots of Kullback-Leibler divergence measure from the fitted PHNBCR model considering prior 2. 112

3.59 Index plots of J-distance from the fitted PHNBCR model considering prior 1. 113

3.60 Index plots of J-distance from the fitted PHNBCR model considering prior 2. 113

3.61 Index plots of L_1 norm distance from the fitted PHNBCR model considering prior 1. . 114

3.62 Index plots of L_1 norm distance from the fitted PHNBCR model considering prior 2. . 114

3.63 Index plots of χ^2-square divergence from the fitted PHNBCR model considering prior 1.115

3.64 Index plots of χ^2-square divergence from the fitted PHNBCR model considering prior 2.115

3.65 Index plots of Kullback-Leibler divergence measure from the fitted CPHNBCR model considering prior 1. 116

3.66 Index plots of Kullback-Leibler divergence measure from the fitted PHNBCR model considering prior 2. 116

3.67 Index plots of J-distance from the fitted CPHNBCR model considering prior 1. 117

3.68 Index plots of J-distance from the fitted CPHNBCR model considering prior 2. 117

3.69 Index plots of L_1 norm distance from the fitted CPHNBCR model considering prior 1. 118

3.70 Index plots of L_1 norm distance from the fitted CPHNBCR model considering prior 2. 118

3.71 Index plots of χ^2-square divergence from the fitted CPHNBCR model considering prior 1. 119

3.72 Index plots of χ^2-square divergence from the fitted CPHNBCR model considering prior 2. 119

3.73 Posterior means and credible intervals of α_i's 123

3.74 Maps of posterior means for frailties $\boldsymbol{U}$ (upper-left panel) and $\boldsymbol{V}$ (upper-right panel) and posterior standard derivations for frailties $\boldsymbol{U}$ (lower-left panel) and $\boldsymbol{V}$ (lower-right panel). 125

3.75 Estimates of ψ-divergence measures for CPHNBCR model 126

4.1 Index plots of K-L and L_1 divergence measures from the fitted cure models. 140

4.2 Maps of posterior means for frailties U (upper-left panel) and V (upper-right panel) and posterior standard derivations for frailties U (lower-left panel) and V (lower-right panel). 142

4.3 Surviving function under the complementary promotion time cure model stratified by treatments and sex for patients who residence in Rochester city with three levels of duration of smoking habit and cigarette consumption. 142

4.4 Surviving function under the complementary promotion time cure model stratified by cigarette consumption and sex for patients who residence in Rochester city with duration of smoking habit equal 12, 30 and 46 years. 143

1 Posterior means of α_i's with $\Lambda_0 = 0.001$, 0.01, 0.1 and 1. 175

List of Tables

3.1 Simulation results for WGCR and CWGCR models with depended spatial fragilities . 51

3.2 Simulation results for PHGCR model and CPHGCR model with depended spatial fragilities 52

3.3 Simulation results of the perturbed cases for WGCR model 55

3.4 Simulation results of the perturbed cases for CWGCR model 56

3.5 Simulation results of the perturbed cases for PHGCR model 57

3.6 Simulation results of the perturbed cases for CPHGCR model 58

3.7 Divergence measures of the perturbed cases and DIC values for the simulated data sets. 59

3.8 Bayesian criteria for the fitted models. 76

3.9 Bayesian criteria for the fitted models. 77

3.10 Posterior summaries of the parameter of the model 15 for the smoking cessation data. 77

3.11 Possible influential observations are detected by four divergence measures 80

3.12 Posterior summaries of the parameter of the model 15 and relative variations adjusted for the smoking cessation data without detected individuals 72, 138, 151 and 199. . . 80

3.13 Bayesian criteria for the PHGCR and CPHGCR models. 81

3.14 Posterior summaries of the parameter of Model 9 for the smoking cessation data. . . . 81

3.15 Possible influential observations are detected the divergence measures 84

3.16 Posterior summaries of the parameter of Model 9 and RV adjusted for the smoking cessation data without detected individuals 14 and 86. 85

3.17 Simulation results for WNBCR and CWNBCR models with depended spatial fragilities 95

3.18 Simulation results for PHNBCR and CPHNBCR models with depended spatial fragilities 96

3.19 Simulation results of the perturbed cases for WNBCR model 99

3.20 Simulation results of the perturbed cases for CWNBCR model 100

3.21 Simulation results of the perturbed cases for PHNBCR model 101

3.22 Simulation results of the perturbed cases for CPHNBCR model 102

3.23 Divergence measures of the perturbed cases and DIC values for the simulated data sets.103

3.24 Posterior estimate of the elements of matrix $\mathbf{\Lambda}$ and $\mathbf{\Sigma}$ 121

3.25 DIC values for the fitted proposed cure rate models considering different priors for the parameters. 122

3.26 DIC values for the fitted cure rate models considering prior (iv) for the parameters. . 122

3.27 Posterior summaries of the parameter of the CPHNB model for the smoking cessation data. 123

3.28 Posterior summaries of the parameter of CPHNBCR model and RV adjusted for the smoking cessation data without detected individual 138. 127

4.1 DIC values for the fitted cure rate models 138

4.2 Posterior summaries of the parameter of the selected models for the smoking cessation data. 139

4.3 Posterior summaries of the parameter of the complementary promotion time cure model without detected individual 14. 141

5.1 Percentage of samples in which the adjusted survival model was indicated as the best model according to the criteria for the right censoring data sets 152

5.2 Percentage of samples in which the adjusted cure rate model was indicated as the best model according to the criteria for the right censoring data sets 154

5.3 Percentage of samples in which the adjusted cure rate model was indicated as the best model according to the criteria for the interval censoring data sets 156

5.4 Bayesian criteria for the fitted models for the melanoma data. 157

5.5 Bayesian criteria for the fitted models for the smoking cessation data. 158

1 Posterior summaries of the parameters of the WNBCR model considering the prior distribution $a \sim Uniform(0, 1)$ and different values for $\mathbf{\Lambda}_0$. 170

2 Posterior summaries of the parameters of the WNBCR model considering the prior distribution $a \sim \text{Beta}(18, 2)$ and different values for $\mathbf{\Lambda}_0$. 171

3 Posterior summaries of the parameters of the CWNBCR model considering the prior distribution $a \sim \text{Uniform}(0, 1)$ and different values for $\mathbf{\Lambda}_0$. 172

4 Posterior summaries of the parameters of the CWNBCR model considering the prior distribution $a \sim \text{Beta}(18, 2)$ and different values for $\mathbf{\Lambda}_0$. 173

5 Posterior summaries of the parameters of the CPHNBCR model considering the prior distribution $a \sim \text{Beta}(18, 2)$ and different values for $\mathbf{\Lambda}_0$. 174

Contents

1 Introduction **1**

1.1 Smoking cessation data 3

2 Basic concepts **7**

2.1 Some interesting distributions 8

2.1.1 Weibull distribution 8

2.1.2 Piecewise exponential distribution 9

2.1.3 Generalized extreme value distribution and Log generalized extreme value distribution 11

2.1.4 Gamma distribution and inverse gamma distribution 13

2.1.5 Wishart distribution 15

2.2 Interval Censoring 16

2.3 The likelihood functions 17

2.3.1 "Case 2" interval censoring 20

2.4 Cure rate model 25

2.5 Frailty model 29

2.5.1 Introduction 29

2.5.2 Univariate frailty model 30

2.5.3 Multivariate frailty models 34

2.6 Model comparison criteria 38

2.7 Bayesian case influence diagnostics 39

3 Spatial frailty in Cure rate models **41**

3.1 Geometric cure rate models with spatial frailties 41

3.1.1 Bayesian Inference 45

3.1.2 Simulation study 49

3.1.3 Application 76

3.1.4 Conclusions 84

3.2 Negative-Binomial cure rate models with spatial frailties 86

3.2.1 Bayesian Inference 89

3.2.2 Simulation study 93

3.2.3 Application 120

3.2.4 Conclusions 128

4 The Power Series Cure Rate Model for Spatially Correlated Interval-Censored Data based on Generalized Extreme Value Distribution **129**

4.1 Introduction 129

4.2 Power Series cure rate Model 130

4.3 Special cases of the PS cure rate model under first/last activation 132

4.3.1 **Mixture cure model** 132

4.3.2 **Promotion time cure model and complementary promotion time cure model** 133

4.3.3 **Geometric cure rate model and complementary geometric cure rate model** 133

4.3.4 **Logarithmic cure rate model and complementary logarithmic cure rate model** 134

4.4 Bayesian Inference 135

4.5 Application 137

4.6 Conclusions 143

5 D-Measure: A Bayesian model selection criterion for survival data **145**

5.1 Introduction . . . 145
5.2 D-measure . . . 146
5.2.1 Uncensored data . . . 147
5.2.2 Right censored data . . . 147
5.2.3 Interval censored data . . . 148
5.3 Simulation . . . 151
5.3.1 Survival model with right censored data . . . 151
5.3.2 Cure rate model with right censored data . . . 152
5.3.3 Cure rate model with interval censored data . . . 154
5.4 Application . . . 156
5.4.1 Melanoma data . . . 156
5.4.2 Smoking cessation data . . . 157
5.5 Conclusion . . . 158

6 Concluding Remarks **159**

Chapter 1

Introduction

With the development of medical and health sciences, the datasets collected from clinical studies pose some new challenges to statisticians. New statistical models which can incorporate these changes should be investigated. The most prevalent change noted in many clinical studies is that, more patients respond favorably to a treatment or, were not susceptible to the event of interest in the study, so they are considered cured or have prolonged disease-free survival. This proportion of patients is called the cure fraction. Incorporating the cure fraction in survival models leads to cure rate models or long-term survival models. These models have been widely developed in the biostatistics literature. One of the most famous cure rate models is the mixture cure model introduced by Berkson & Gage (1952). This model has been extensively discussed by several authors, including Farewell (1982), Maller & Zhou (1996), Ewell & Ibrahim (1997) and Stangl & Greenhouse (1998). Later, Yakovlev & Tsodikov (1996) and Chen *et al.* (1999) proposed the promotion time cure model or bounded cumulative hazard model in cancer relapse settings, assuming that a latent biological process of propagation of latent carcinogenic tumor cells is generating the observed failure (relapse). Recently, Cooner *et al.* (2007) generalized this framework to a flexible class of cure models under latent activation schemes, Rodrigues *et al.* (2009b) extend the promotion time cure model proposed by Chen *et al.* (1999) through the generating function of a real sequence introduced by Feller (1968) and Cancho *et al.* (2011) proposed a flexible cure rate model, that encompasses as special cases and the mixture model (Berkson & Gage, 1952), the promotion time cure model (Chen *et al.*, 1999) and the cure rate proportional odds model proposed by Gu *et al.* (2011).

The second challenge is the existence of incomplete (censoring) datasets. In many clinical trials, the patients are examined periodically for disease occurrence or progression. In this situation, the exact failure time of each patient cannot be observed. Rather, it can only be determined to lie in an interval obtained from a sequence of examination times. This time to event is known as

the interval censoring (Peto, 1973). The estimation methods available for right censored data, such as the Kaplan-Meier estimator, are not adequate for application to interval-censored data, because they can lead to biased estimation and invalid inferences. The interval censorship information should be taken into account in modeling (Rücker & Messerer, 1988; Lindsey & Ryan, 1998; Sun & Chen, 2010).

Another challenge is appearing with the development of geographic information systems (GIS) and computing technology. Datasets increasingly incorporate geographical information about the subjects under study. Adopting a traditional cure rate model by including random effects for each region fails to consider the correlations of the regions. Therefore, several researchers have developed survival models that account for spatial clustering and variation. Banerjee *et al.* (2003) investigated spatially correlated frailties in traditional parametric survival models. Later, Banerjee & Carlin (2004) introduced spatially correlated frailties in the parametric cure model. They developed a Bayesian approach to the mixture cure model (Berkson & Gage, 1952) with spatial random effects in the survival function for subjects at risk and spatial frailties using a multivariate conditionally autoregressive (MCAR) prior. Recently, Pan *et al.* (2014) proposed a Bayesian approach under a proportional hazards frailty model to analyze interval-censored survival data with spatial correlation. Li Dan & Dey (2015) proposed flexible cure rate models in analyzing univariate right-censored data based on the assumption that the logarithm of survival time follows a generalized extreme value distribution with spatial and nonlinear covariate effects.

Considering these three challenges, there are two main goals in this work. First more flexible cure rate models that account for spatial clustering and variation should be devolved and investigated for the censored datasets. Here, we assume two most natural activations schemes, the first and last activations schemes. The first activation scheme presents the situation where the presence of any of latent risk will ultimately lead to the occurrence of the event, while the last activation scheme presents a situation where the occurrence of the event will happen when all latent risks are activated. Thus, the proposed cure rate models are much more general and encompass several well-known cure models as special cases, such as some cure models introduced by Banerjee & Carlin (2004) and others were suggested as future investigations by the authors. To investigate the correlation between the hazard function and cure fraction, the covariates and frailties are incorporated into both of them, assuming the spatial frailties can be independent or dependent. The inference procedures are developed through a Bayesian perspective.

The second goal is propose a new measure for model selection for the interval-censored data, which measures the performance of a model by how close its predictions are to the observed data. Compared with the deviance information criterion (DIC) proposed by Spiegelhalter *et al.* (2002),

the proposed measure is based on the notion of predictive loss paradigm (Gelfand & Ghosh, 1998; Ibrahim *et al.*, 2001b) and only a very weak assumption about censoring is made for the computation of the new measure. Both the DIC and the new proposed criterion are used to compare the models. Furthermore, we also conduct influence diagnostics in order to check the model assumptions and conduct sensitivity analysis to detect possible influential or extreme observations that can cause distortions in the results. Here case deletion influence diagnostics are developed for the joint posterior distribution based on the ψ-divergence (Peng & Dey, 1995; Weiss, 1996). In this work, the proposed cure rate models are fitted to a real dataset (smoking cessation data) to illustrate their flexibility. Thus, we present the dataset in follow the section.

1.1 Smoking cessation data

In smoking cessation study, all of the patients (smokers) were randomized into either a smoking intervention (SI) group, or a usual care (UC) group which received no special anti-smoking intervention. The smoking intervention treatment program was conducted in Rochester, Minnesota, located in the center of the maps. The details of the program can be found in Murray *et al.* (1998). Here, each patient was observed once a year over the five year follow-up. Our event of interest is whether they relapse (resume smoking) or not. If a smoker resumed smoking after an initial attempt to quit, then only an approximate one-year time interval was observed from the previous observation to the current observation. Thus, the relapse times are interval-censored. In this analysis, we limit our attention to those patients who are known to have quit smoking at least once during the study period and who have an identifiable Minnesota Zip code of residence. Thus, the data consist of 223 patients who reside in 51 Zip codes in the southeastern corner of Minnesota, among them 65 patients having relapsed, which implies the empirical cure rate is approximately 71%. The map of cities which correspond the Zip codes is showed in Figure 4.2 and the covariate information for each patient considered in the study are

- intervention type SI/UC (1=special intervention [SI], 0=usual care [UC]);
- sex (0=male, 1=female);
- the average number of cigarettes smoked per day (5 to 60);
- duration of smoking habit in years (12 to 46 year);

To estimate the covariate effects on the success rate of smoking cessation as well as that on the smoking relapse time. Therefore, all recorded covariates are considered in both the cure rate

and the PH components of the studied cure models. The Figure 1.1 shows the survival functions estimated considering the intervals by Turnbull algorithm and using the midpoints of intervals by Kaplan-Meier method (Kaplan & Meier (1958)). We can note that the estimated curves are similar in some moments, but they are very different in many other moments. Moveover, it is also can observe that the curves were stabilized before 0.6, which confirmed the existence of a significant fraction of cured individuals. Recently, Ma & Xiang (2013) also confirmed the existence of a noneligible cure fraction in the population.

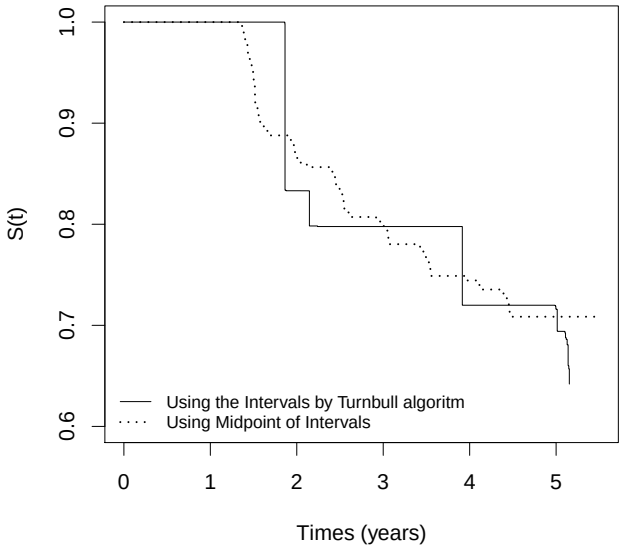

Figure 1.1: Estimated survival functions considering the intervals and its midpoints.

The dataset was also analyzed by Carlin & Banerjee (2003), they developed a Bayesian approach to the mixture cure model, assuming the failure times due to the latent risks (competing times) have Weibull and gamma distributions, with spatial random effects in the survival function for at-risk subjects. They showed that the models through assuming competing times having Weibull distribution have better fitting than gamma. In this work, we will compare our models with their models thought the Bayeisan DIC.

The remainder of our text is organized as follows. In Chapter 2, we will present some basic concepts in the survival analysis. The statistic models which are used in our work, some important definitions of statistic terms such as censoring and likelihood are described in detail. Moreover, some well known Bayesian comparison criteria and diagnostic measures based on the ψ-divergence are also showed. In Chapter 3, we propose two flexible cure rate models for spatial correlations by including spatial frailties for the interval censored data setting. For the proposed cure rate models, the Bayesian inferences are developed and the simulation studies are also conducted. To illustrate the flexibility of proposed models, they are fitted to a real data set (smoking cessation study). In

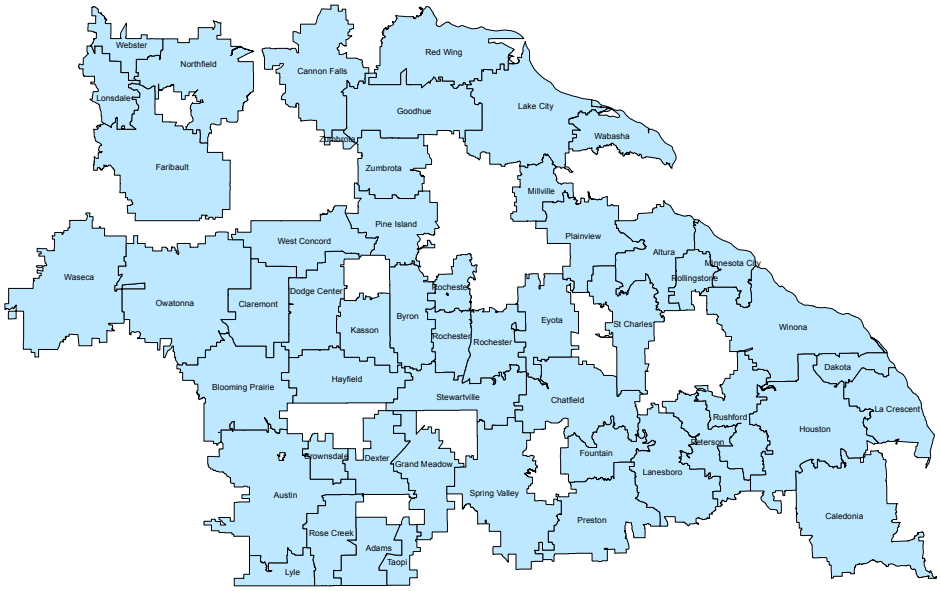

Figure 1.2: Cities in which are collected the smoking cessation data.

Chapter 4, we propose the Power Series cure rate survival model for spatially correlated interval-censored data based on the generalized extreme value distribution. This cure rate model is much more general than the cure models which are proposed in Chapter 3. A new measure based on survival function is proposed in Chapter 5. Finally, some general remarks and some perspectives for future work are listed in Chapter 6. The algorithm used in this work and the prior sensitive analysis studies are presented in Appendix.

Chapter 2

Basic concepts

This section describes some important results and defines the notations which build the basis for specific points in the later chapters. Moreover, some parametric models which are used in the work and its respective characteristics and properties are presented.

We consider a single non-negative random variable T, representing the lifetime or time to failure of an individual, usually, it is assumed to be continuous. The probability density function (p.d.f.) is denoted by f. The cumulative distribution function (c.d.f.) of T can be determined by its probability density function and it is denoted by F. The survival function of T is defined by:

$$S(t) = P[T \geq t] = 1 - F(t) = \int_t^\infty f(s)ds,$$

which is the probability of an individual to survive until time t. It is a continuous monotonically decreasing function with $S(0) = 1$ and $\lim_{t\to\infty} S(t) = 0$.

Another important is failure rate function (or hazard function), which specifies the instantaneous rate of failure or death of an individual at time t, given that it survives until time t. The function is useful to describe the lifetime distribution of the observations under study, and it is defined by:

$$h(t) = \lim_{\delta t\to 0} \frac{P[t \leq T \leq t + \delta t | T \geq t]}{\delta t} = \frac{f(t)}{S(t)}.$$

Sometimes, it is useful to deal with the cumulative hazard function

$$H(t) = \int_0^t h(s)ds.$$

The shape of a hazard function can take different forms: it can be increasing, decreasing, constant, unimodal or U-shaped. In applications, it is often have qualitative information about the form of

the the hazard function, which can be of help in selecting a appropriate model.

The relationship of the functions f, F, S, h and H is

$$H(t) = \int_0^t h(s)ds = \int_0^t \frac{f(s)}{S(s)}ds = \int_0^t \exp[-\ln(S(s))]f(s)ds = \ln(S(t))$$

and

$$S(t) = 1 - F(t) = \exp(-H(t)).$$

These relationships are very useful in the survival analysis.

2.1 Some interesting distributions

Some distributions which are used in the work will be presented as follow section.

2.1.1 Weibull distribution

The Weibull distribution was firstly introduced by Weibull (1939) and then was used in survival analyze by Weibull (1951). This function is an important generalization of the exponential model with two positive parameters, there are shape parameter and scale parameter. One of the main characteristics of this distribution is its flexibility in accommodating different forms in failure rate. Therefore, it is one of most widely used model in practice.

The random variable T has Weibull distribution with shape parameter $\alpha > 0$ and scale parameter λ, $\lambda \in \mathbb{R}$, denoted by $T \sim Weibull(\alpha, \lambda)$, and its probability density function (p.d.f.) is given

$$f(t|\alpha, \lambda) = \alpha t^{\alpha-1} \exp(\lambda - t^\alpha e^\lambda), \tag{2.1}$$

and the corresponding survival and hazard function are given by

$$S(t|\alpha, \lambda) = \exp(-t^\alpha e^\lambda) \qquad \text{and} \qquad h(t|\alpha, \lambda) = \alpha e^\lambda t^{\alpha-1}, \tag{2.2}$$

respectively.

The survival and hazard functions are presented in Figure 2.1, which illustrates that the hazard function of the Weibull distribution $h(t)$ is strictly increasing for $\alpha > 1$, strictly decreasing for $\alpha < 1$ e constant for $\alpha = 1$. In this case, T follows an Exponential distribution with parameter λ, which reveals a certain flexibility in the behavior of the hazard function.

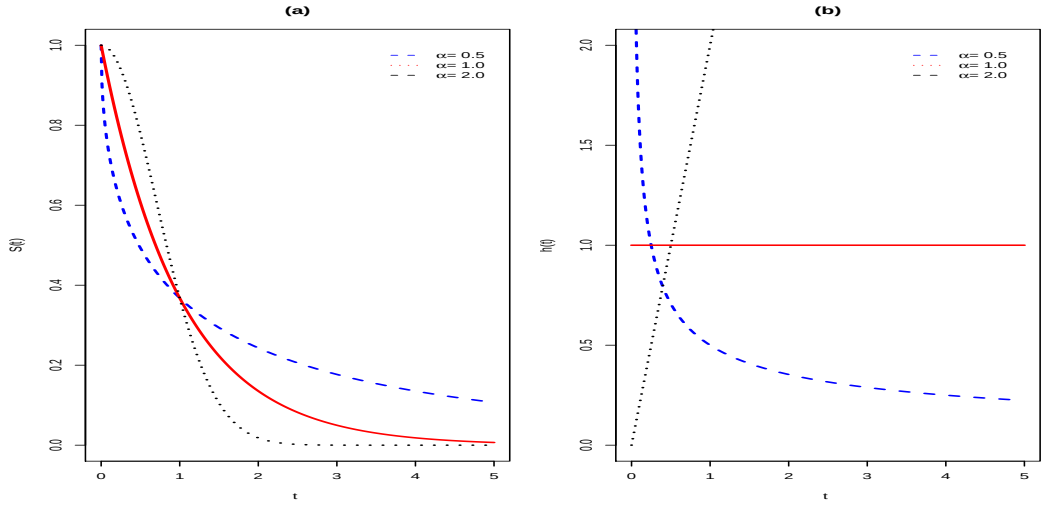

Figure 2.1: (a) Survival function and (b) hazard function of Weibull distribution with $\lambda = 0$ and different selected values of α.

The kth moment of the distribution is $E[T^k] = \exp\left(-\frac{\lambda k}{\alpha}\right)\Gamma\left(\frac{k}{\alpha}+1\right)$. So the mean and variance are thus

$$\begin{aligned} E[T] &= \exp\left(-\frac{\lambda}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}+1\right) \\ Var[T] &= \exp\left(-\frac{2\lambda}{\alpha}\right)\left\{\Gamma\left(\frac{2}{\alpha}+1\right) - \left[\Gamma\left(\frac{1}{\alpha}+1\right)\right]^2\right\} \end{aligned}$$

where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\mathrm{dt}$ is a gamma function.

The qth quantile of the Weibull distribution, obtained by inverting the cumulative distribution function of T, is given by

$$t_q = (-\log(1-q)/e^\lambda)^{1/\alpha},$$

and particularly, the median is $t_{1/2} = (\log(2)/e^\lambda)^{1/\alpha}$.

2.1.2 Piecewise exponential distribution

The Piecewise exponential distribution was firstly introduced by Feigl & Zelen (1965) and then it was used to analyze survival data with multiple covariates by Friedman (1982). The risk rate of the distribution is constant within each considered time interval. The Piecewise exponential distribution can be used to be an approximated distribution while the true distribution is unknown and the approximation becomes better when the length of each interval becomes smaller.

Let $0 = a_0 < a_1 < \ldots < a_Q = \infty$ be a partition of the time axis, assuming the risk rate is constant in each of these intervals. Let the vector $\boldsymbol{a} = (a_1, \ldots, a_{Q-1})$ with $0 < a_1 < \ldots < a_{Q-1} < \infty$

and define $a_0 = 0$ and $a_Q = \infty$. The random variable T has piecewise exponential distribution with parameter $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_Q)$ and partition vector $\boldsymbol{a}$, denoted by $T \sim \text{PExp}_{\boldsymbol{a}}(\boldsymbol{\lambda})$, and the corresponding probability density function (p.d.f.) is given by

$$f(t|\boldsymbol{\lambda}) = \kappa_q \lambda_q \exp\{-\lambda_q(t - a_{q-1})\}, \quad t \in (a_{q-1}, a_q], \quad q = 1, \ldots, Q, \tag{2.3}$$

where

$$\kappa_l = \begin{cases} 1, & \text{if } q = 1; \\ \exp\left\{-\sum_{i=1}^{q-1} \lambda_i (a_i - a_{i-1})\right\}, & \text{if } q = 2, \ldots, Q. \end{cases}$$

The corresponding survival and hazard function are given respectively by

$$S(t|\boldsymbol{\lambda}) = \exp\left\{-\sum_{q=1}^{Q} \lambda_q \Delta_q(t)\right\}, \quad t > 0, \tag{2.4}$$

$$h(t|\boldsymbol{\lambda}) = \lambda_q, \quad t \in (a_{q-1}, a_q], q = 1, \ldots, Q, \tag{2.5}$$

where

$$\Delta_q(t) = \begin{cases} 0, & \text{if } t < a_{q-1}; \\ t - a_{q-1} & \text{if } a_{q-1} \le t < a_q, \quad q = 1 \ldots, Q. \\ a_q - a_{q-1} & \text{if } t \ge a_q. \end{cases}$$

The kth moment of the distribution is

$$E[T^k] = \sum_{q=1}^{Q} \kappa_q \left(a_q + \frac{1}{\lambda_q}\right)^k \left(1 - \exp\{-\lambda_q(a_q - a_{q-1})\}\right).$$

Thus, the mean and variance are

$$E[T] = \sum_{q=1}^{Q} \kappa_q \left(a_q + \frac{1}{\lambda_q}\right) \left(1 - \exp\{-\lambda_q(a_q - a_{q-1})\}\right),$$

and

$$Var[T] = \left[\sum_{q=1}^{Q} \kappa_q \left(a_q + \frac{1}{\lambda_q}\right)^2 \left(1 - \exp\{-\lambda_q(a_q - a_{q-1})\}\right)\right] - E[T]^2,$$

respectively.

Note that the exponential distribution with a parameter λ is the particular case of the piecewise exponential distribution when $\lambda_q = \lambda$ for $q = 1, \ldots, Q$. The survival and hazard functions are presented in Figure 2.2. We selected $\boldsymbol{a} = (0.5, 1, 2, 3)$ and considered four different parameter vec-

tors, which lead to the hazard function of piecewise exponential distribution has constant, increasing, decreasing and U shapes.

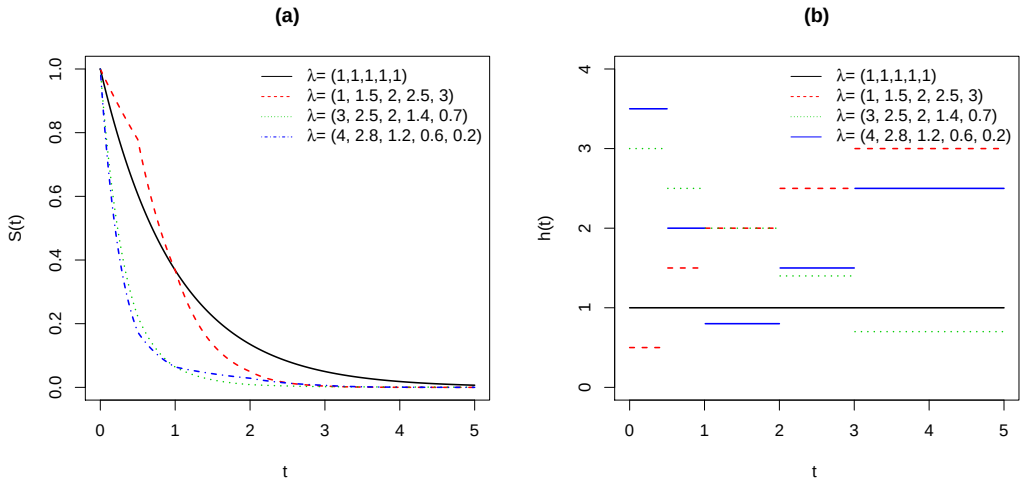

Figure 2.2: (a) Survival function and (b) hazard function of piecewise exponential distribution with different selected values of λ.

2.1.3 Generalized extreme value distribution and Log generalized extreme value distribution

The generalized extreme value (GEV) distribution is a family of continuous probability distributions under the extreme value theory which combine the Gumbel, Fréchet and Weibull families. It was introduced by Jenkinson (1955, 1969) and recommended by Natural Environment Research Council (1975) of Great Britain. The GEV distribution has gained popularity in many disciplines, but its use in survival modeling is relatively new (Li Dan & Dey, 2015). Its flexible hazard function is the main reason that it has gained attention in survival analysis. Recently, Roy & Dey (2014) showed that different shapes for the hazard function can be obtained by varying the shape parameter in the GEV distribution.

The random variable X has GEV distribution with incorporation of location and scale parameters are given by

$$F(x|\mu,\sigma,\varsigma) = \begin{cases} \exp\left\{-\left(1+\varsigma\frac{x-\mu}{\sigma}\right)_+^{-\frac{1}{\varsigma}}\right\}, & \text{if } \varsigma \neq 0, \\ \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, & \text{if } \varsigma = 0, \end{cases}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$ and $\varsigma \in \mathbb{R}$ are the location, scale and shape parameters respectively, and $x_+ = \max(0, x)$. In the survival analysis, we assume that $\log T \sim GEV(\mu, \sigma, \varsigma)$, where T denotes time to event of interest, i.e., let $T \sim logGEV(\mu, \sigma, \varsigma)$, the cumulative distribution function (c.d.f.)

of T is given by

$$F(t|\mu,\sigma,\varsigma) = \begin{cases} \exp\left\{-\left(1+\varsigma\frac{\log t-\mu}{\sigma}\right)_+^{-\frac{1}{\varsigma}}\right\}, & \text{if } \varsigma \neq 0, \\ \exp\left\{-\exp\left(-\frac{\log t-\mu}{\sigma}\right)\right\}, & \text{if } \varsigma = 0, \end{cases} \tag{2.6}$$

and the probability density function (p.d.f.) is given

$$f(t|\mu,\sigma,\varsigma) = \begin{cases} \frac{1}{\sigma t}\left(1+\varsigma\frac{\log t-\mu}{\sigma}\right)^{-\frac{1}{\varsigma}-1}\exp\left\{-\left(1+\varsigma\frac{\log t-\mu}{\sigma}\right)^{-\frac{1}{\varsigma}}\right\}, & t > \exp\left(\mu-\frac{\sigma}{\varsigma}\right) \text{ if } \varsigma > 0, \text{ or} \\ & t < \exp\left(\mu-\frac{\sigma}{\varsigma}\right) \text{ if } \varsigma < 0, \\ \frac{1}{\sigma t}\exp\left(-\frac{\log t-\mu}{\sigma}\right)\exp\left\{-\exp\left(-\frac{\log t-\mu}{\sigma}\right)\right\}, & 0 < t < \infty \text{ if } \varsigma = 0. \end{cases} \tag{2.7}$$

The corresponding survival function and hazard function are given respectively by

$$S(t|\mu,\sigma,\varsigma) = \begin{cases} 1-\exp\left\{-\left(1+\varsigma\frac{\log t-\mu}{\sigma}\right)_+^{-\frac{1}{\varsigma}}\right\}, & \text{if } \varsigma \neq 0, \\ 1-\exp\left\{-\exp\left(\frac{\log t-\mu}{\sigma}\right)\right\}, & \text{if } \varsigma = 0, \end{cases}$$

and

$$h(t|\mu,\sigma,\varsigma) = \begin{cases} \frac{1}{\sigma t}\left(1+\varsigma\frac{\log t-\mu}{\sigma}\right)_+^{-\frac{1}{\varsigma}-1}\left[\exp\left\{\left(1+\varsigma\frac{\log t-\mu}{\sigma}\right)_+^{-\frac{1}{\varsigma}}\right\}-1\right]^{-1}, & \text{if } \varsigma \neq 0, \\ \frac{1}{\sigma t}\left(1+\varsigma\frac{\log t-\mu}{\sigma}\right)\left[\exp\left\{\exp\left(-\frac{\log t-\mu}{\sigma}\right)\right\}-1\right]^{-1}, & \text{if } \varsigma = 0. \end{cases}$$

The survival and hazard functions are presented in Figure 2.3 and 2.4. We fixed location parameter $\mu = 0$, scale parameter $\sigma = 1.0$ and $\sigma = 1.5$, the shape parameter ς are selected 1.5, 0.5, 0.0 and -0.5 four different values. We note that the hazard function of logGEV distribution has increasing, decreasing, bell and U shapes. Inasmuch as in many practical situations, especially in cancer related studies, the hazard function is not monotone, the logGEV distribution could be more adequate than the usual parametric distributions, such as Weibull and Gamma distributions.

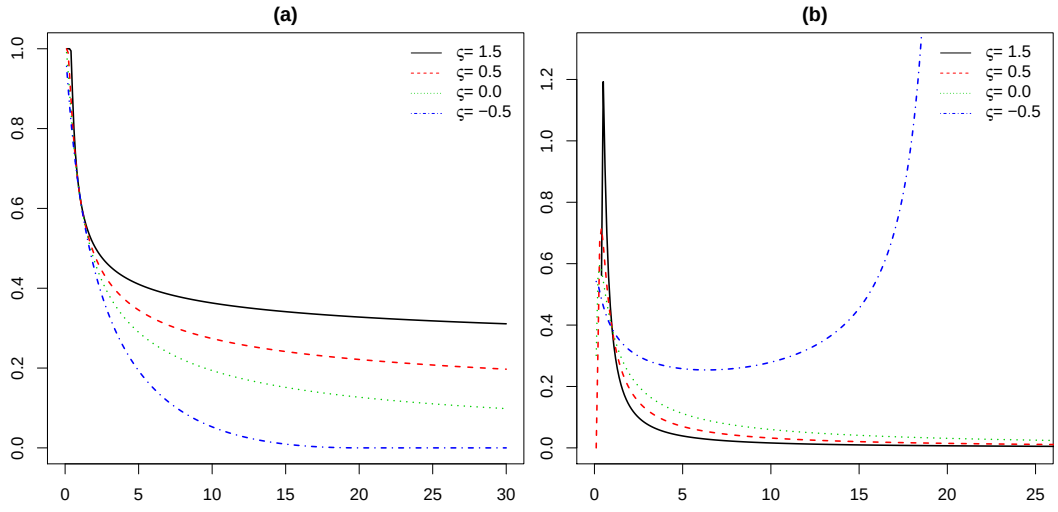

Figure 2.3: (a) Survival function and (b) hazard function of logGEV distribution with $\mu = 0$ and $\sigma = 1$.

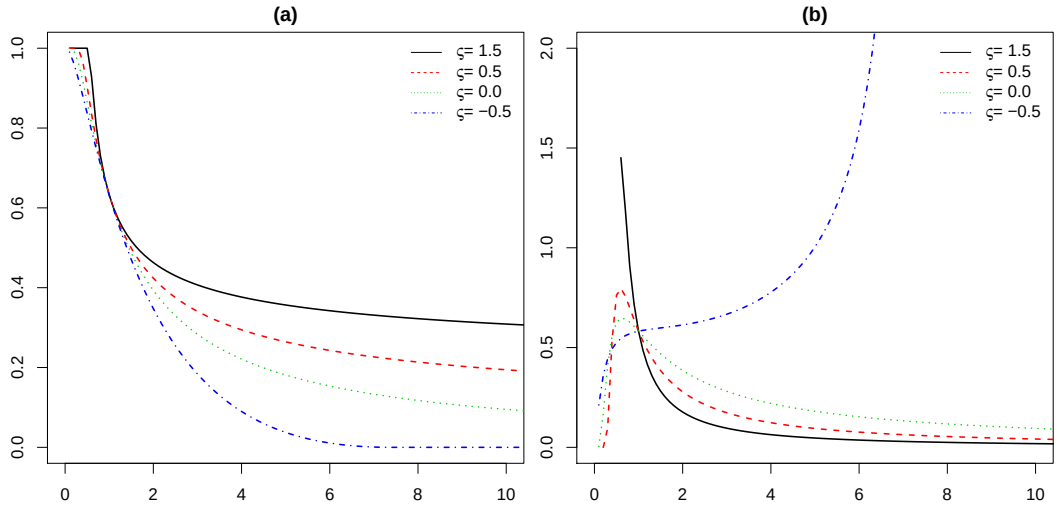

Figure 2.4: (a) Survival function and (b) hazard function of logGEV distribution with $\mu = 0$ and $\sigma = 1.5$.

2.1.4 Gamma distribution and inverse gamma distribution

The gamma distribution was firstly introduced by Pearson (1895). It includes the exponential distribution and chi-squared distribution are special cases. The gamma distribution is of limited use in survival analysis because the gamma models do not have closed form expressions for survival and hazard functions. Both include the incomplete gamma integral. Consequently, traditional maximum likelihood estimation is difficult and requires the calculation of such incomplete gamma integrals, which imposes additional numerical problems in parameter estimation.

The random variable T has gamma distribution with shape parameter $\alpha > 0$ and rate parameter (inverse scale parameter) $\beta > 0$, denoted by $T \sim Gamma(\alpha, \beta)$, and its probability density function (p.d.f.) is given

$$f(t|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\beta t), \quad t > 0, \tag{2.8}$$

and the corresponding c.d.f. and hazard function are given respectively by

$$F(t|\alpha, \lambda) = \frac{\gamma(\alpha, \beta t)}{\Gamma(\alpha)} \qquad \text{and} \qquad h(t|\alpha, \lambda) = \frac{\beta^\alpha t^{\alpha-1} \exp(-\beta t)}{\Gamma(\alpha) - \gamma(\alpha, \beta t)}, \tag{2.9}$$

where $\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$ is the lower completed function and $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is a gamma function.

The mean and variance of gamma distribution are

$$E[T] = \frac{\alpha}{\beta} \qquad \text{and} \qquad Var[T] = \frac{\alpha}{\beta^2},$$

The probability density function and cumulative distribution function of gamma distribution are presented in Figure 5.3.1.

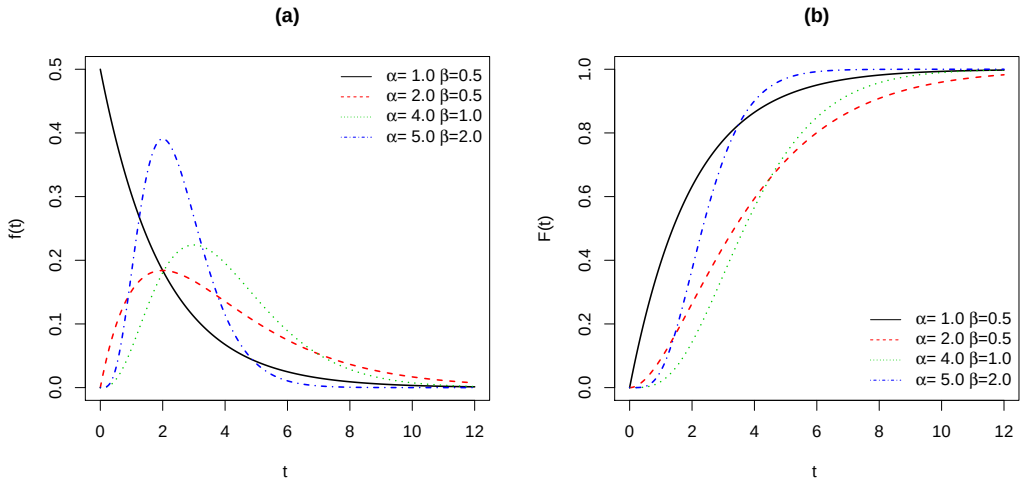

Figure 2.5: (a) Probability density function and (b) cumulative distribution function of gamma distribution with different selected values of α and β.

A random variable T takes the inverse gamma distribution with parameter shape α and scale β if $1/T$ has the gamma distribution with shape parameter α and scale parameter $1/\beta$. The p.d.f. of the inverse gamma distribution is given by

$$f(t|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{-\alpha-1}\exp(-\beta t^{-1}), \quad t>0, \tag{2.10}$$

and the corresponding c.d.f. and hazard function are given respectively by

$$F(t|\alpha,\lambda) = \frac{\Gamma(\alpha,\beta t^{-1})}{\Gamma(\alpha)} \qquad \text{and} \qquad h(t|\alpha,\lambda) = \frac{\beta^\alpha t^{-\alpha-1}\exp(-\beta t^{-1})}{\left(1-\frac{\Gamma(\alpha,\beta t^{-1})}{\Gamma(\alpha)}\right)\Gamma(\alpha)}, \tag{2.11}$$

where $\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$ is the upper completed function and $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ is a gamma function.

The mean and variance of the distribution are

$$E[T] = \frac{\beta}{\alpha-1}, \quad \alpha>1 \qquad \text{and} \qquad Var[T] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \quad \alpha>2.$$

The probability density function and cumulative distribution function of inverse gamma distribution are presented in Figure 2.6.

In Bayesian statistics, the inverse gamma distribution is the conjugate prior of the unknown variance of a normal distribution. It is usual to set a low value for its parameters such as 1 or 0.01

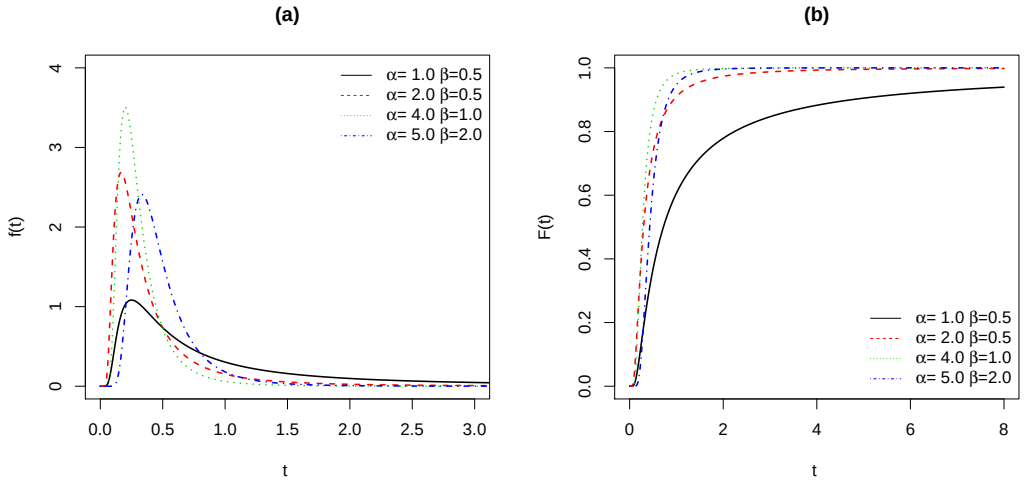

Figure 2.6: (a) Probability density function and (b) cumulative distribution function of gamma distribution with different values of α and β.

or 0.001 in order to let it be an weakly informative prior distribution.

2.1.5 Wishart distribution

A Wishart distribution was introduced by Wishart (1928). It is a generalization to multiple dimensions of the chi-squared distribution, or, in the case of non-integer degrees of freedom, of the gamma distribution. This distribution is very important in the estimation of covariance matrices in multivariate analysis.

Let $\boldsymbol{X}_i$, $i = 1, \ldots, n$ be a p-dimensional random vector, which is independently drawn from a p-variate normal distribution with zero mean and symmetric positive definite covariance matrix $\boldsymbol{\Lambda}_0$ ($\boldsymbol{X}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Lambda}_0)$). Then the Wishart distribution is the probability distribution of the $p \times p$ symmetric positive definite random matrix $\boldsymbol{S} = \sum_{i=1}^{n_0} \boldsymbol{X}_i^\top \boldsymbol{X}_i$, denoted by $\boldsymbol{S} \sim W_p(n_0, \Lambda_0)$, with scala matrix Λ_0 and degrees of freedom $n \geq p$. The p.d.f. of Wishart distribution is given by

$$f(\boldsymbol{S}) = \frac{|\boldsymbol{S}|^{\frac{n_0-p-1}{2}}}{2^{\frac{n_0 p}{2}} |\boldsymbol{\Lambda}_0|^{\frac{n_0}{2}} \Gamma_p(\frac{n_0}{2})} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{S})\right\}, \tag{2.12}$$

where $tr(\ )$ is the trace function and Γ_p is the multivariate gamma function defined as

$$\Gamma_p\left(\frac{n_0}{2}\right) = \pi^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(\frac{n}{2} + \frac{1-j}{2}\right).$$

If $p = S = 1$ then this distribution is a chi-squared distribution with n_0 degrees of freedom.

The mean and variance of the distribution are

$$E[\boldsymbol{S}] = n_0\boldsymbol{\Lambda}_0 \qquad \text{and} \qquad Var[\boldsymbol{S}_{ij}] = n_0(\lambda_{0ij}^2 + \lambda_{0ii}\lambda_{0jj}),$$

respectively, where λ_{0ij} denotes the element of Λ_0 matrix in (i, j) position.

In Bayesian statistics, it is the conjugate prior of the precision matrix (inverse covariance-matrix) of a multivariate normal distribution. The least informative, proper Wishart prior is obtained by setting $n_0 = p$. The prior mean of $W_p(n_0, \boldsymbol{\Lambda}_0)$ is $n_0\Lambda_0$, suggesting that a reasonable choice for $\boldsymbol{\Lambda}_0^{-1}$ would be $n\boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}_0$ is some prior guess for the covariance matrix.

2.2 Interval Censoring

Censoring is one of the main characteristic that distinguishes survival analysis from other fields of statistics. Basically, a censored observation contains only partial information about the variable of interest. There are different types of censoring, here we consider an interval censoring in the study. We now briefly describe the some types of interval-censored data considered in this section.

"Case 1" interval censoring or current status data.

Let T be the unobservable failure time and suppose that L is an examination time (or observation time). Then suppose that an observation consists of the random vector (Δ, L) where $\Delta = 1_{[T \leq L]}$. In this case, the only knowledge about the "failure time" T is whether it has occurred before L or not.

"Case 2" and "Case k" interval censoring.

In the "Case 2" interval censored data, we only know that the unobservable failure time T has occurred either within some random time interval, or before the left end point of the time interval, or after the right end point of the time interval. More precisely, suppose that there are two examination (or observation) times L and R , the data observed is

$$(L, R, \Delta_1, \Delta_2, \Delta_3) = (L, R, 1_{[T \leq L]}, 1_{[L < T \leq R]}, 1_{[T > R]}),$$

Note that $\Delta_3 = 1 - \Delta_1 - \Delta_2$. In the particular situation that T has occurred either within some random time interval or after the right end point of the time interval, the data observed can be denoted as

$$(L, R, \Delta_1, \Delta_2) = (L, R, 1_{[L<T\leq R]}, 1_{[T>R]}),$$

it also equivalent

$$(L, R, \Delta) = (L, R, 1_{[R<\infty]}).$$

A "Case k" interval censoring arises when there are k examination times per subject, which is a generalization of "Case 2" interval censoring (see Wellner (1995)).

Particular situation of "Case 2" interval censored data with latent competing risks.

In this case, we assume that the event of the interest (failure) occurs due to the several latent (non-observed) competing risks. In practice, there are three most popular situations. The first one is the event of the interest occur if any of latent risk is activated; the second one is the event of interest occur if all of latent risks are activated, and the last one is the event of the interest occur if one random latent risk is activated. Here, we suppose the time to event (failure time) T has occurred either within some random time interval or after the right end point of the time interval, i.e., $L < T \leq R$ or $T > R$. Thus, the data observed in this case is

$$(L, R, \Delta) = (L, R, 1_{[L<T\leq R]}) = (L, R, 1_{[R<\infty]}).$$

2.3 The likelihood functions

We assume that the examination times are independent of the failure time and that their distribution is independent of the distribution function of the failure time. With these conditions, the joint densities and the likelihood functions for the given types of interval-censored data will be presented follow.

"Case 1" interval censoring

Let T be a failure time with distribution F and L be an observation time with distribution G. The joint density of a single observation (l, δ) is given by

$$F(l)^{\delta}(1 - F(l))^{1-\delta} g(l),$$

where $g(l)$ is the density of L.

Proof: For a single observation, we have two cases $\Delta = 1$ and $\Delta = 0$. We first consider $\Delta = 1$

$$\begin{aligned} P(L \leq l, \Delta = 1) &= P(L \leq l, T \leq L) \\ &= \int_{\mathbb{R}} P(L \leq l, T \leq L | L = s) dG(s) \quad \text{(conditioning on } L\text{)} \\ &= \int_{-\infty}^{l} P(T \leq s | L = s) dG(s) \\ &= \int_{-\infty}^{l} P(T \leq s) dG(s) \quad \text{(using independence of } T \text{ and } L\text{)} \\ &= \int_{-\infty}^{l} F(s) dG(s). \end{aligned}$$

We obtain the corresponding density by differentiating with respect to l. Assuming that G had a density g. Using the integration by parts we have

$$\begin{aligned} \int_{-\infty}^{l} F(s) dG(s) &= F(s)G(s)|_{-\infty}^{l} - \int_{-\infty}^{l} f(s)G(s) ds \\ &= F(l)G(l) - \int_{-\infty}^{l} f(s)G(s) ds. \end{aligned}$$

Thus,

$$\begin{aligned} f(l, \Delta = 1) &= \frac{\partial}{\partial l} \int_{-\infty}^{l} F(s) dG(s) \\ &= \frac{\partial}{\partial l} \left(F(l)G(l) - \int_{-\infty}^{l} f(s)G(s) ds \right) \\ &= F(l)g(l) \end{aligned}$$

Similarly, considering $\Delta = 0$ we have

$$\begin{aligned} P(L \leq l, \Delta = 0) &= P(L \leq l, T > L) \\ &= \int_{\mathbb{R}} P(L \leq l, T > L | L = s) dG(s) \quad \text{(conditioning on } L\text{)} \\ &= \int_{-\infty}^{l} P(T > s | L = s) dG(s) \\ &= \int_{-\infty}^{l} P(T > s) dG(s) \quad \text{(using independence of } T \text{ and } L\text{)} \\ &= \int_{-\infty}^{l} (1 - F(s)) dG(s). \end{aligned}$$

Using the Integration by parts, we obtain

$$f(l, \Delta = 0) = \frac{\partial}{\partial l} \int_{-\infty}^{l} 1 - F(s) dG(s) = (1 - F(l)) g(l).$$

Combining the terms for $\Delta = 1$ and $\Delta = 0$, we get the following density for one observation:

$$[F(l)g(l)]^{\delta} [(1 - F(l))g(l)]^{1-\delta} = F(l)^{\delta} (1 - F(l))^{1-\delta} g(l). \qquad \square$$

Note that this density again factors in a part depending on F and a part depending on g. Since G and g do not involve any of the parameters in F, they can be neglected. Hence, the likelihood function L_n of a random sample $(l_1, \delta_1), \ldots, (l_n, \delta_n)$ is given by

$$L_n = \prod_{i=1}^{n} F(l_i)^{\delta_i} (1 - F(l_i))^{1-\delta_i}. \tag{2.13}$$

Note that the likelihood function $L_n(F)$ also can be rewritten in terms of observed sets as

$$L_n = \prod_{i=1}^{n} P_F(R_i),$$

where

$$R_i = \begin{cases} (0, l_i], & \text{if } \delta_i = 1 \\ (l_i, \infty), & \text{if } \delta_i = 0 \end{cases}$$

and $P_F(R_i)$ denotes the probability that $T \in R_i$ under distribution F for $i = 1 \ldots, n$. Now we can derive the likelihood in a slightly simpler way by observing that

$$\begin{aligned} P(\Delta = 1 | L = l) &= P(T \leq L | L = l) = P(T \leq L) = F(l), \\ P(\Delta = 0 | L = l) &= P(T > L | L = l) = P(T > L) = 1 - F(l). \end{aligned}$$

Hence, $\Delta | L$ is a Bernoulli random variable with parameter $F(L)$. It then follows that the density of

one observation is

$$\begin{aligned} P(\Delta = \delta, L = l) &= P(\Delta = \delta | L = l) g(l) \\ &= F(l)^{\delta} (1 - F(l))^{1-\delta} g(l), \end{aligned}$$

exactly as we found before.

2.3.1 "Case 2" interval censoring

Let T be a failure time with distribution F and (L, R) (with $L < R$) be a pair of observation times with the joint distribution G. The joint density of a single observation $(l, r, \delta_1, \delta_2, \delta_3)$ is given by

$$F(l)^{\delta_1} [F(r) - F(l)]^{\delta_2} (1 - F(r))^{1-\delta_3} g(l, r),$$

where $g(l, r)$ is the joint density of (L, R) and $\delta_3 = 1 - \delta_1 - \delta_2$.

Prove: In this case, we have three situations which are $T \leq L$, $L < T \leq R$ and $T > R$ with the probabilities

$$\begin{aligned} p_1 = P(\Delta_1 = 1 | L = l, R = r) &= P(T \leq L | L = l, R = r) \\ &= P(T \leq l | L = l, R = r) \\ &= P(T \leq l) \quad \text{(using independence of } T \text{ and } (L, R)) \\ &= F(l), \end{aligned}$$

$$\begin{aligned} p_2 = P(\Delta_2 = 1 | L = l, R = r) &= P(L < T \leq R | L = l, R = r) \\ &= P(l < T \leq r | L = l, R = r) \\ &= P(l < T \leq r) \quad \text{(using independence of } T \text{ and } (L, R)) \\ &= F(r) - F(l) \end{aligned}$$

and

$$\begin{aligned} p_3 = P(\Delta_3 = 1 | L = l, R = r) &= P(T > R | L = l, R = r) \\ &= P(T > r | L = l, R = r) \\ &= P(T > r) \quad \text{(using independence of } T \text{ and } (L, R)) \\ &= 1 - F(r), \end{aligned}$$

Hence, $\boldsymbol{\Delta}|L,R \sim Multinomial(1,p_1,p_2,p_3)$ with $p_1+p_2+p_3=1$. It then follows that the density of a single observation $(l,r,\delta_1,\delta_2,\delta_3)$ is given

$$\begin{aligned} P(\boldsymbol{\Delta}=(\delta_1,\delta_2,\delta_3),L=l,R=r) &= P(\boldsymbol{\Delta}=(\delta_1,\delta_2,\delta_3)|L=l,R=r)g(l,r) \\ &= F(l)^{\delta_1}(F(r)-F(l))^{\delta_2}(1-F(r))^{\delta_3}g(l,r), \end{aligned}$$

where $\delta_1+\delta_2+\delta_3=1$. $\square$

Since G and g do not involve any of the parameters in F, they can be neglected. Hence, the likelihood function of a random sample $(l_1,r_1,\delta_{11},\delta_{21}),\ldots,(l_n,r_n,\delta_{1n},\delta_{2n})$ is given by

$$L_n=\prod_{i=1}^{n}F(l_i)^{\delta_{1i}}(F(r_i)-F(l_i))^{\delta_{2i}}(1-F(r_i))^{(1-\delta_{1i}-\delta_{2i})}.$$

Note that when we assume that T has occurred after the first examination time L, we just have two situations which are $T\in(L,R]$ and $T\in(L,\infty)$ which is equivalent to verify whether $R<\infty$ or $R=\infty$. Let $\Delta=1_{[L<T\leq R]}=1_{[R<\infty]}$, the probabilities of these situations are given by:

$$\begin{aligned} p_1=P(\Delta=1|L=l,R=r) &= P(L<T\leq R|L=l,R=r) \\ &= P(l<T\leq r|L=l,R=r) \\ &= P(l<T\leq r) \quad \text{(using independence of } T \text{ and } (L,R)) \\ &= F(r)-F(l) \end{aligned}$$

and

$$\begin{aligned} p_2=P(\Delta=0|L=l,R=r) &= P(T>L|L=l,R=r) \\ &= P(T>l|L=l,R=r) \\ &= P(T>l) \quad \text{(using independence of } T \text{ and } (L,R)) \\ &= 1-F(l). \end{aligned}$$

Hence, the density of one observation is

$$\begin{aligned} P(\Delta=\delta,L=l,R=r) &= P(\Delta=\delta|L=l,R=r)g(l,r) \\ &= (F(r)-F(l))^{\delta}(1-F(l))^{1-\delta}g(l,r). \end{aligned}$$

The likelihood function of a random sample $(l_1, r_1, \delta_1), \ldots, (l_n, r_n, \delta_n)$ can be written as

$$L_n = \prod_{i=1}^{n} (F(r_i) - F(l_i))^{\delta_i} (1 - F(l_i))^{(1-\delta_i)}.$$

Interval censored data with latent competing risks

Let M denote the number of latent risks and assume that M has a known discrete distribution with the p.d.f. denoted by $P(M = m)$. Let Y_j for $j = 1 \ldots, M$ denote the failure times due to the jth latent risk and we assume that given $M = m$, Y_j's are i.i.d with a distribution $F(\cdot) = 1 - S(\cdot)$. The time to event of interest (failure time) which is defined by random variable $T = Y_{(R^*)}$, for $M \geq 1$ and $T = \infty$ if $M = 0$ with $P(T = \infty | M = 0) = 1$, where $Y_{(R^*)}$ is the Rth statistic order and R^* can indicate resistance factors of the immune system of the individual in many biological processes. It can be a fixed constant, a function of M or a random variable specified through a conditional distribution on M. In this work, we deal with two specifications for R^*, there are $R^* = 1$ and $R^* = M$.

The survival functions of the random variable T considering $R^* = 1$ (i.e., $T = Y_{(1)} = \min\{Y_j, j = 1, \ldots, M\}$) is given by

$$S_{pop}(t) = P(T > t) = \sum_{m=0}^{\infty} S(t)^m P[M = m]. \tag{2.14}$$

Proof:

$$\begin{aligned}
S_{pop}(t) &= P(T > t) \\
&= \sum_{m=0}^{\infty} P(T > t | M = m) P(M = m) \\
&= P[T > t | M = 0] P(M = 0) + \sum_{m=1}^{\infty} P(T > t | M = m) P(M = m) \\
&= P(M = 0) + \sum_{m=1}^{\infty} P[\min\{Y_j, j = 1, \ldots, m\} > t | M = m] P(M = m) \\
&= P(M = 0) + \sum_{m=1}^{\infty} P[Y_1 > t, \ldots, Y_m > y | M = m] P(M = m) \\
&= P(M = 0) + \sum_{m=1}^{\infty} P[Y_1 > t]^m P[M = m] \\
&= P(M = 0) + \sum_{m=1}^{\infty} S(t)^m P[M = m] \\
&= \sum_{m=0}^{\infty} S(t)^m P[M = m].
\end{aligned}$$

The survival functions of the random variable T considering $R^* = M$ (i.e., $T = Y_{(M)} = \max\{Y_j, j = 1, \ldots, M\}$) is given by

$$S_{pop}(t) = P(T > t) = 1 + P[M = 0] - \sum_{m=0}^{\infty} F(t)^m P[M = m]. \tag{2.15}$$

Proof:

$$\begin{aligned}
S_{pop}(t) &= P(T > t) \\
&= \sum_{m=0}^{\infty} P(T > t | M = m) P(M = m) \\
&= P[T > t | M = 0] P(M = 0) + \sum_{m=1}^{\infty} P(T > t | M = m) P(M = m) \\
&= P(M = 0) + \sum_{m=1}^{\infty} P[\max\{Y_j, j = 1, \ldots, m\} > t | M = m] P(M = m) \\
&= P(M = 0) + \sum_{m=1}^{\infty} \left(1 - P[\max\{Y_j, j = 1, \ldots, m\} \leq t | M = m]\right) P(M = m) \\
&= P(M = 0) + \sum_{m=1}^{\infty} \left(1 - P[Y_1 \leq t, \ldots, Y_m \leq y | M = m]\right) P(M = m) \\
&= P(M = 0) + \sum_{m=1}^{\infty} P(M = m) - \sum_{m=1}^{\infty} P[Y_1 \leq t]^m P[M = m] \\
&= P(M = 0) + 1 - P(M = 0) - \sum_{m=1}^{\infty} P[Y_1 \leq t]^m P[M = m] \\
&= P(M = 0) + 1 - \sum_{m=0}^{\infty} F(t)^m P[M = m] \\
&= 1 + P(M = 0) - \sum_{m=0}^{\infty} F(t)^m P[M = m].
\end{aligned}$$

Assuming T independent on the observed failure times (L, R) with the joint distribution G, the joint density of a single observation (l, r, δ) is given by

$$(S_{pop}(l) - S_{pop}(r))^{\delta} S_{pop}(l)^{1-\delta} g(l, r),$$

where $g(l, r)$ is the joint density of (L, R).

Proof: In this case, we have two situations which are $L < T \leq R$ and $T > R$ with the

probabilities

$$\begin{aligned}
p_1 &= P[\delta = 1 | L = l, R = r] \\
&= \sum_{m=0}^{\infty} P[\delta = 1 | M = m, L = l, R = r] P(M = m) \\
&= \sum_{m=0}^{\infty} P[L < T \leq R | M = m, L = l, R = r] P[M = m] \\
&\quad \text{(using independence of } T \text{ and } (L, R)) \\
&= \sum_{m=0}^{\infty} P[l < T \leq r | M = m] P[M = m] \\
&= \sum_{m=0}^{\infty} P[T \leq r | M = m] P[M = m] - \sum_{m=0}^{\infty} P[T \leq l | M = m] P[M = m] \\
&= \sum_{m=0}^{\infty} P[T > l | M = m] P[M = m] - \sum_{m=0}^{\infty} P[T > r | M = m] P[M = m] \\
&= S_{pop}(l) - S_{pop}(r)
\end{aligned}$$

and

$$\begin{aligned}
p_2 &= P[\delta = 0 | L = l, R = r] \\
&= \sum_{m=1}^{\infty} P[\delta = 0 | M = m, L = l, R = r] P(M = m) \\
&= \sum_{m=0}^{\infty} P[T > L | M = m, L = l, R = r] P(M = m) \\
&\quad \text{(using independence of } T \text{ and } (L, R)) \\
&= \sum_{m=0}^{\infty} P[T > l | M = m] P(M = m) \\
&= S_{pop}(l).
\end{aligned}$$

Hence, the density of one observation is

$$\begin{aligned}
P[\Delta = \delta | L = l, R = r] &= P[\Delta = \delta | L = l, R = r] g(l, r) \\
&= (S_{pop}(l) - S_{pop}(r))^{\delta} (S_{pop}(l))^{1-\delta} g(l, r).
\end{aligned}$$

Since G and g do not involve any of the parameters in S_{pop}, they can be neglected. Thus, the likelihood function of random sample $(l_1, r_1, \delta_1), \ldots, (l_n, r_n, \delta_n)$ is given by

$$\prod_{i=1}^{n} (S_{pop}(l_i) - S_{pop}(r_i))^{\delta_i} (S_{pop}(l_i))^{1-\delta_i}. \tag{2.16}$$

2.4 Cure rate model

With the development of medical technology, in many clinical cancer studies, some patients can return a normal life after a treatment, i.e, there is a percentage of patients will not occur the event of interest after a long follow-up period of the study. In this situation, the usual survival models, assumed that all individuals occur the event of interest after a long follow-up period, are not adequate to fit this kind of data set. Alternatively, the cure rate models (also known as the survival models with cure fraction), which assume that a significant fraction of individuals will not occur the event of interest even after a long follow-up period of the study, can be used. In the literature, a percentage of individuals will not occur the event of interest is known as a cure fraction. There are many cure rate models have been widely developed. The reference papers are Maller & Zhou (1996), Ibrahim *et al.* (2001a), Tsodikov *et al.* (2003), Cooner *et al.* (2007), Tournoud & Ecochard (2007), Lopes *et al.* (2012), Rodrigues *et al.* (2009a), Cancho *et al.* (2009), Cancho *et al.* (2011), Rodrigues *et al.* (2010a) and Rodrigues *et al.* (2010b). In this section, some principal cure rate models will be presented as follow.

Mixture cure rate model

Perhaps the most popular type of cure rate model is the mixture cure model introduced by Boag (1949) and Berkson & Gage (1952). In this distribution, it is assumed that a certain proportion of the individuals are cured. The survival function for the population of the mixture cure model is given by

$$S_{pop}(t) = p_0 + (1 - p_0)S(t), \tag{2.17}$$

where p_0 is the cure fraction (that is, proportion of the cured individuals) and $S(t)$ is the survival function of the non-cured (or susceptible) individuals. Note that this model also can be introduced under structure of latent competing risks, assuming M has a Bernoulli distribution with the success parameter $1 - p_0$. The survival function of the non-cured individuals $S(t)$ can taken different approaches such as parametric, semi-parametric and non-parametric (Maller & Zhou, 1996; Peng, 2003; Lu, 2010).

Although the mixture cure model is widely used in the literature, it has some disadvantages which were commented by Chen *et al.* (1999). First, in the presence of covariates, it does not have the proportional hazard structure; Second, If covariates are included in the cure fraction through a standard binomial regression model with the improper priors for the coefficient parameters, the posterior distributions of the parameters will be improper, i.e., the mixture cure model requires proper priors distributions in the Bayesian inference.

Promotion time cure model

Later, the promotion time cure model has been proposed by Yakovlev & Tsodikov (1996) and Chen *et al.* (1999) as an alternative cure rate model with desirable properties. The cure model was derived in a context in which relapse occurs in patients with cancer. Let M_i denotes the number of latent risk (in the cancer study, M_i denote the number of carcinogenic cells in the beginning of a treatment) for ith individual, and assume that M_i has Poisson distribution with mean θ. Let Y_j for $j = 1, \ldots, M_i$ denote the failure time due to the jth latent cause, that is, the time until jth carcinogenic cell produces a detectable cancer. Suppose that given M_i, the random variables Y_j are independent and identically distributed (i.i.d.) with c.d.f. $F(\cdot) = 1 - S(\cdot)$ and the presence of any of latent risk (i.e., $M_i \geq 1$) will ultimately lead to the occurrence of the event. Thus, the time to event of interest (time to detect cancer) is defined by random variable $T = \min\{Y_j, j = 0, \cdots, M_i\}$ where $P(Y_0 = \infty) = 1$. Using the equation (2.14) and $M_i \sim Poisson(\theta)$, the survival function of promotion time cure model for the population is given by

$$\begin{aligned} S_{pop}(t) &= \sum_{m=0}^{\infty} S(t)^m \frac{e^{-\theta}\theta^m}{m!} \\ &= e^{-\theta} \sum_{m=0}^{\infty} \frac{(\theta S(t))^m}{m!} \\ &= e^{-\theta} e^{\theta S(t)} \\ &= e^{-\theta F(t)}. \end{aligned}$$

The corresponding p.d.f. and hazard function of T are given by

$$f_{pop}(t) = \theta f(t) e^{-\theta F(t)} \qquad \text{and} \qquad h_{pop}(t) = \theta f(t), \tag{2.18}$$

respectively, where $f(t) = \frac{\partial}{\partial t} F(t)$.

A cure fraction of the promotion time cure model is given by

$$S_{pop}(\infty) = \lim_{t \to \infty} S_{pop}(t) = \lim_{t \to \infty} e^{-\theta F(t)} = e^{-\theta}. \tag{2.19}$$

It is easy to note that this model belongs to the family of Cox proportional hazards models (Cox, 1972). Suppose two individuals, say i and j, have the parameters associated with the cure fractions given by θ_i and θ_j, respectively. Thus,

$$\frac{h_{pop}(t|\theta_i)}{h_{pop}(t|\theta_j)} = \frac{\theta_i f(t)}{\theta_j f(t)} = \frac{\theta_i}{\theta_j}$$

does not depend on the time t.

Complementary promotion time cure model

The promotion time cure model assumed that the presence of any of latent risk will lead to the occurrence of the event. Now, we assume the other situation that the occurrence of the event will be occur when all of the latent risk are activated. Using the same definitions for M_i and Y_j as above, now the time to event of interest is defined by random variable $T = \max\{Y_j, j = 1, \cdots, M_i\}$ for $M_i \geq 1$ and $T = \infty$ if $M_i = 0$ with $P(T = \infty | M_i = 0) = 1$. Similarly, using the equation (2.15) and $M_i \sim Poisson(\theta)$, the survival function complementary promotion time cure model for the population is given by

$$\begin{aligned} S_{pop}(t) &= 1 + P[M = 0] - \sum_{m=0}^{\infty} F(t)^m P[M = m] \\ &= 1 + e^{-\theta} - \sum_{m=0}^{\infty} F(t)^m \frac{e^{-\theta}\theta^m}{m!} \\ &= 1 + e^{-\theta} - e^{-\theta} \sum_{m=0}^{\infty} \frac{(\theta F(t))^m}{m!} \\ &= 1 + e^{-\theta} - e^{-\theta} e^{\theta F(t)} \\ &= 1 + e^{-\theta} - e^{-\theta S(t)}. \end{aligned}$$

The corresponding p.d.f. by and hazard function of T are given by

$$f_{pop}(t) = \theta f(t) e^{-\theta S(t)} \qquad \text{and} \qquad h_{pop}(t) = \frac{\theta f(t) e^{-\theta S(t)}}{1 + e^{-\theta} - e^{-\theta S(t)}}, \tag{2.20}$$

respectively, where $f(t) = \frac{\partial}{\partial t} F(t)$.

A cure fraction of the complementary promotion time cure model is given by

$$S_{pop}(\infty) = \lim_{t \to \infty} S_{pop}(t) = \lim_{t \to \infty} 1 + e^{-\theta} - e^{-\theta S(t)} = e^{-\theta}. \tag{2.21}$$

Note that it is the same as the cure fraction of the promotion time cure model.

Geometric cure rate model

Let M_i denote the number of latent risk for ith individual, and assume that M_i has geometric distribution with $1/(1+\theta)$, the probability mass function is given by

$$P(M_i = m) = \frac{\theta^m}{(\theta+1)^{m+1}}, \quad m = 0, 1, 2 \cdots, \tag{2.22}$$

where $\theta > 0$, $E(M_i) = \theta$ and $Var(M_i) = \theta(1+\theta)$. Let Y_j for $j = 1, \ldots, M_i$ denote the failure time due to the jth latent cause, that is, the time until jth carcinogenic cell produces a detectable cancer. Suppose that given M_i, the random variables Y_j are independent and identically distributed (i.i.d.) with c.d.f. $F(\cdot) = 1 - S(\cdot)$ and the presence of any of latent risk (i.e., $M_i \geq 1$) will ultimately lead to the occurrence of the event. Thus, the time to event of interest is defined by random variable $T = \min\{Y_j, j = 0, \cdots, M_i\}$ where $P(Y_0 = \infty) = 1$. From (2.14), the survival function of geometric cure model for the population is given by

$$S_{pop}(t) = \left[1 + \theta F(t)\right]^{-1}. \tag{2.23}$$

Note that this survival function has a proportional odds structure when covariates $\boldsymbol{x}_i$ are modeled via $\theta_i(\boldsymbol{x}_i)$ and the latent survival $F(t_i)$ is free of $\boldsymbol{x}_i$, because

$$\frac{1 - S_{pop}(t_i|\boldsymbol{x}_i)}{S_{pop}(t_i|\boldsymbol{x}_i)} = \theta_i(\boldsymbol{x}_i)F(t_i).$$

Recently, Gu *et al.* (2011) studied the geometric cure model under the proportional odds structure and renamed the geometric cure model as the cure rate proportional odds (CRPO) model. Unlike the model proposed by Chen *et al.* (1999), the ratio of hazards for the CRPO model for two covariante values does not remain over time.

The corresponding p.d.f and the hazard function associated to (4.2) are given by

$$f_{pop}(t) = \theta f(t_{ij}) \left[1 + \theta F(t)\right]^{-2} \quad \text{and} \quad h_{pop}(t) = \theta f(t) \left[1 + \theta F(t)\right]^{-1},$$

respectively, where $f(t) = \frac{\partial}{\partial t} F(t)$.

Note that, the survival function in (4.2) can also be written as a mixture cure model

$$S_{\text{pop}}(t) = (1+\theta)^{-1} + \left(1 - (1+\theta)^{-1}\right) \left\{ \frac{\left[1 + \theta F(t)\right]^{-1} - (1+\theta)^{-1}}{1 - (1+\theta)^{-1}} \right\},$$

Thus, the survival functions of uncured (susceptible) individuals can be expressed by

$$S_{\text{sus}}(t) = \frac{[1+\theta F(t)]^{-1} - (1+\theta)^{-1}}{1-(1+\theta)^{-1}}.$$

If we assume another situation in which the presence of all latent risks will ultimately lead to the occurrence of the event, then the time to the event of interest is defined by the random variable $T = \max\{Y_j, j = 1, \cdots, M_i\}$ for $M_i \geq 1$ and $T = \infty$ if $M_i = 0$ with $P(T = \infty | M_i = 0) = 1$. The survival function for the population is given by

$$S_{pop}(t) = 1 + (1+\theta)^{-1} - [1+\theta S(t)]^{-1}. \tag{2.24}$$

The corresponding p.d.f. and the hazard function are given by

$$f_{\text{pop}}(t) = \theta_{ij} f(t)(1+\theta S(t))^{-2},$$

and

$$h_{\text{pop}}(t) = \frac{\theta f(t)\left[1+\theta S(t_{ij})\right]^{-2}}{1+(1+\theta)^{-1} - (1+\theta S(t))^{-1}},$$

respectively. The survival function (4.4) can also be written as a mixture cure model

$$S_{\text{pop}}(t) = (1+\theta)^{-1} + \left(1-(1+\theta)^{-1}\right)\left\{\frac{1-(1+\theta S(t))^{-1}}{1-(1+\theta)^{-1}}\right\},$$

Thus, the survival functions of susceptible individuals is given by

$$S_{\text{sus}}(t) = \frac{1-(1+\theta S(t))^{-1}}{1-(1+\theta)^{-1}}.$$

2.5 Frailty model

2.5.1 Introduction

Ordinary methods in survival analysis assume the populations are homogenous, that is, assuming the lifetimes of each individual are mutually independent with same distribution, which imply that all individuals have the same risk of death. Although this assumption is valid for many studies, it can be inadequate for others. In many situations, the lifetime data are observed as repeated measurements or collected from several clusters, such times within each cluster cannot be mutually independents. For example, the behavior of the observed lifetimes between members of the same

family displays certain similarities that would not be observed among individuals without family tie. Therefore, in this case, it is reasonable to assume that there is an association between the lifetimes of the same cluster. This association between the lifetimes characterizes multivariate lifetimes data. A frailty model is commonly used in this context to consider this association.

A frailty model also can be used for univariate (independent) lifetimes. In this case, each individual has its own frailty, which has different meaning of the frailty in the multivariate context. Here, the frailty is a heterogenous measure of the individuals, while in multivariate survival is also a measure of association. In this chapter, the univariate and multivariate frailty will be presented in the following sections.

2.5.2 Univariate frailty model

Situations in which each individual has its own frailty component, which could be seen as the special case in which all groups or families have unitary size, characterize univariate survival data. The supposition that individual has its own frailty component is not difficult to justify. In the medicine study, for example, the argument that individuals are inherently different is widely accepted. Two individuals with exactly the same values of the covariates are not expected to experience any medical response at exactly the same time. There are no observable biological variations, such as genetic factors with respect to one disease.

Considering the situation the heterogeneity of individuals affect the observed survival data, a frailty (or random effect) was introduced in the hazard models. Vaupel *et al.* (1979) introduced the concept of frailty to the biostatistical community and applied it to population mortality data. Lancaster (1979) dealt with times of unemployment and introduced the model to the econometric literature, where the model is known as the mixed proportional hazards model. The classical and most frequently applied model assumes a proportional hazards structure that is conditional on the random effect (frailty). To be more specific, the hazard function of an individual depends on an unobservable, time-independent random variable Z. In the multiplicative hazards framework, which has been used in most survival data studies, Z acts multiplicative effect on the baseline hazard function h_0. The frailty model without observed covariates for the individual i $(i = 1, \ldots, n)$ is given by

$$h(t|z_i) = z_i h_0(t), \tag{2.25}$$

where the $z_1, \ldots, z_n$ is a sample of random variables $Z_1, \ldots, Z_n$ i.i.d. with a known distribution with mean one and unknown variance σ^2. The variance (if it exists) is interpretable as a measure of heterogeneity across the population in baseline risk. When σ^2 is small, the values of Z are closely

located around one. If σ^2 is large, then values of Z are more dispersed, inducing greater heterogeneity in the individual hazards $Zh_0(t)$. It is natural to introduce observed covariates into model (2.25) similar to the Cox model by

$$h_i(t|\boldsymbol{x_i}, z_i) = z_i h_0(t) \exp(\boldsymbol{x_i}^\top \boldsymbol{\beta}), \tag{2.26}$$

where $\boldsymbol{\beta}$ is the regression parameters vector associated to covariates $\boldsymbol{x}_i$. Consequently, a frailty model is a generalization of the well-known proportional hazards model. The proportional hazards model is obtained if the frailty distribution degenerates to $Z_i = 1$ for $i = 1, \ldots, n$. In this case, the frailty not only explain the heterogeneity of the individuals but also evaluate the effect of covariates that were not observed at the time of the experiment for some reason, and thus were not included in the analyzes.

Various probability distributions have been proposed in the literature for the frailty variables. Next, we will present the gamma distribution which is used in the work.

The gamma distribution has been widely applied by several authors (Vaupel *et al.* (1979), Lancaster (1979), Hougaard & Hougaard (2000) Duchateau & Janssen (2007)). From a computational and analytical point of view, it fits very well as a mixture distribution to failure data. It is easy to derive the closed-form expressions of unconditional survival, cumulative density, and hazard function, which is due to the simplicity of the Laplace transform. This is also the reason why this distribution has been used in most applications published to date.

Let Z_i, $i = 1 \ldots, n$ be random variables with the gamma distribution presented in Section 5.3.1, i.e., $Z_i \sim \text{Gamma}(\alpha, \beta)$, considering the parameters $\alpha = \beta = \eta^{-1}$, the p.d.f. of Z_i is given by

$$f(z) = \frac{\left(\frac{1}{\eta}\right)^{1/\eta}}{\Gamma\left(\frac{1}{\eta}\right)} z^{\frac{1}{\eta}-1} \exp\left(-\frac{z}{\eta}\right), \quad z \geq 0, \tag{2.27}$$

Note that $E[Z_i] = 1$ and $Var[Z_i] = \eta$. Thus, η can be viewed as a way to quantify this frailty. If $\eta = 0$ (i.e, $Var[Z_i] = 0$) all of the frailty variables $Z_i = 1$, for all $i = 1, \ldots, n$, that is, the gamma distribution is degenerate in 1.

Promotion time cure model with fragility

The promotion time cure model presented in Section 2.4 is assumed that conditional on the number of latent risks $M_i = m$, the random variables $Y_1, \ldots, Y_m$ are mutually independent. This assumption may be unrealistic, because $Y_1, \ldots, Y_m$ are not observed random variables taken on the same subject. One possible relaxation and remedy of this assumption is to introduce a subject-specific frailty Z_i such that conditional on both $M_i = m$ and $Z_i = z$, $Y_1, \ldots, Y_m$ are mutually

independent with distribution function $F(\cdot)$. Moveover, we assume that conditional on $Z_i = z$, M_i has a Poisson distribution with rate $z\theta$, thus Z_i presents the heterogeneity of the Poisson rates in the M_i's. Following the same derivation as before, we then obtain the survival function of time to event, T, is

$$S_{pop}(t) = E_{Z_i}\left[\exp\{-\theta F(t) z\}\right],$$

where E_{Z_i} denotes the expectation with respect to Z_i. Assuming the Z_i has a gamma distribution with its p.d.f. presented in (2.27), then we have

$$\begin{aligned}
S_{pop}(t) &= E_{Z_i}\left[\exp\{-\theta F(t) z\}\right] \\
&= \int_0^\infty \exp\{-\theta F(t) z\} \frac{\left(\frac{1}{\eta}\right)^{1/\eta}}{\Gamma\left(\frac{1}{\eta}\right)} z^{\frac{1}{\eta}-1} \exp\left(-\frac{z}{\eta}\right) dz \\
&= \frac{\left(\frac{1}{\eta}\right)^{1/\eta}}{\Gamma\left(\frac{1}{\eta}\right)} \int_0^\infty z^{\frac{1}{\eta}-1} \exp\left(-\left(\frac{1}{\eta} + \theta F(t)\right) z\right) dz \\
&= \frac{\left(\frac{1}{\eta}\right)^{1/\eta}}{\left(\frac{1}{\eta} + \theta F(t)\right)^{1/\eta}} \int_0^\infty \frac{\left(\frac{1}{\eta} + \theta F(t)\right)^{1/\eta}}{\Gamma\left(\frac{1}{\eta}\right)} z^{\frac{1}{\eta}-1} \exp\left(-\left(\frac{1}{\eta} + \theta F(t)\right) z\right) dz \\
&= (1 + \eta\theta F(t))^{-1/\eta}. \qquad (2.28)
\end{aligned}$$

Note that this cure rate model is the same as the of model which was proposed by Rodrigues *et al.* (2009b). The cure rate model can also called Negative Binomial cure rate model, because it can be proposed similarly to Chen *et al.* (1999). Let M_i denote the number of latent risk for ith individual, and now assume that M_i has negative binomial (NB) distribution with parameters $\theta > 0$ and $\eta > -1/\theta$, i.e., $E[M_i] = \theta$ and $Var[M_i] = \theta(1 + \eta)$. Let Y_j for $j = 1, \ldots, M_i$ denote the failure time due to the jth latent cause. Suppose that given M_i, the random variables Y_j are i.i.d. with c.d.f. $F(\cdot) = 1 - S(\cdot)$ and the presence of any of latent risk (i.e., $M_i \geq 1$) will lead to the occurrence of the event. Thus, the time to event of interest (time to detect cancer) is defined by random variable $T = \min\{Y_j, j = 0, \cdots, M_i\}$ where $P(Y_0 = \infty) = 1$. From (2.14), the survival function for the population can be written as

$$\begin{aligned}
S_{pop}(t) &= \sum_{m=0}^{\infty} S(t)^m \frac{\Gamma(\eta^{-1} + m)}{\Gamma(\eta) m!} \left(\frac{\eta\theta}{1 + \eta\theta}\right)^m (1 + \eta\theta)^{-1/\eta} \\
&= (1 + \eta\theta)^{-1/\eta} \sum_{m=0}^{\infty} \frac{\Gamma(\eta^{-1} + m)}{\Gamma(\eta) m!} \left(\frac{\eta\theta S(t)}{1 + \eta\theta}\right)^m \\
&= (1 + \eta\theta(1 - S(t)))^{-1/\eta} \\
&= (1 + \eta\theta F(t))^{-1/\eta}.
\end{aligned}$$

The corresponding p.d.f. and hazard function of T are given by

$$f_{pop}(t) = \theta f(t)(1+\theta\eta F(t))^{-(1/\eta+1)}$$

and

$$h_{pop}(t) = \theta f(t)(1+\theta\eta F(t))^{-1},$$

respectively, where $f(t) = \frac{\partial}{\partial t}F(t)$. Its cure fraction is given by

$$S_{pop}(\infty) = \lim_{t\to\infty} S_{pop}(t) = \lim_{t\to\infty}(1+\eta\theta S(t))^{-1/\eta} = (1+\eta\theta)^{-1/\eta}.$$

Note that some special sub-models can be obtained if we set some special values for the parameter η. For $\eta \to 0$ we have the promotion time cure model and if $\eta = 1$, we obtain the geometric cure rate model given in Section 2.4.

Complementary promotion time cure model with fragility

Similar to above case, we introduce a frailty Z_i to the complementary promotion time cure model, such that conditional on both $M_i = m$ and $Z_i = z$, $Y_1, \ldots, Y_m$ are mutually independent with distribution function $F(\cdot)$. Assuming conditional on $Z_i = z$, M_i has a Poisson distribution with rate $z\theta$ and Z_i has a gamma distribution with its p.d.f. presented in (2.27), the survival function of time to event, T, is given by

$$\begin{aligned}
S_{pop}(t) &= E_{Z_i}\left[1+e^{-\theta z}-e^{-\theta S(t)z}\right] \\
&= \int_0^\infty \left(1+e^{-\theta z}-e^{-\theta S(t)z}\right)\frac{\left(\frac{1}{\eta}\right)^{1/\eta}}{\Gamma\left(\frac{1}{\eta}\right)} z^{\frac{1}{\eta}-1}\exp\left(-\frac{z}{\eta}\right)dz \\
&= 1+\int_0^\infty \frac{\left(\frac{1}{\eta}\right)^{1/\eta}}{\Gamma\left(\frac{1}{\eta}\right)} z^{\frac{1}{\eta}-1}\exp\left(-(\theta+\eta^{-1})z\right)dz \\
&\quad -\int_0^\infty \frac{\left(\frac{1}{\eta}\right)^{1/\eta}}{\Gamma\left(\frac{1}{\eta}\right)} z^{\frac{1}{\eta}-1}\exp\left(-\left(\eta^{-1}+\theta S(t)\right)z\right)dz \\
&= 1+\frac{\left(\frac{1}{\eta}\right)^{1/\eta}}{\left(\frac{1}{\eta}+\theta\right)^{1/\eta}}\int_0^\infty \frac{\left(\frac{1}{\eta}+\theta\right)^{1/\eta}}{\Gamma\left(\frac{1}{\eta}\right)} z^{\frac{1}{\eta}-1}\exp\left(-\left(\frac{1}{\eta}+\theta\right)z\right)dz \\
&\quad -\frac{\left(\frac{1}{\eta}\right)^{1/\eta}}{\left(\frac{1}{\eta}+\theta S(t)\right)^{1/\eta}}\int_0^\infty \frac{\left(\frac{1}{\eta}+\theta S(t)\right)^{1/\eta}}{\Gamma\left(\frac{1}{\eta}\right)} z^{\frac{1}{\eta}-1}\exp\left(-\left(\frac{1}{\eta}+\theta S(t)\right)z\right)dz \\
&= 1+(1+\eta\theta)^{-1/\eta}-(1+\eta\theta S(t))^{-1/\eta}. \qquad (2.29)
\end{aligned}$$

Similarly, we called this model as complementary negative binomial cure rate model. Assuming the number of latent risk M_i has the negative binomial (NB) distribution with parameters $\theta > 0$ and $\eta > -1/\theta$. Let Y_j for $j = 1, \ldots, M_i$ denote the failure time due to the jth latent cause. Suppose that given M_i, the random variables Y_j are i.i.d. with c.d.f. $F(\cdot) = 1 - S(\cdot)$ and the presence of all latent risks will lead to the occurrence of the event. The time to event of interest is defined by random variable $T = \max\{Y_j, j = 1, \cdots, M_i\}$ for $M_i \geq 1$ and $T = \infty$ if $M_i = 0$ with $P(T = \infty | M_i = 0) = 1$. Using the equation (2.15) and $M_i \sim NB(\theta, \eta)$, the survival function for the population can be written as

$$\begin{aligned} S_{pop}(t) &= 1 + (1+\eta\theta)^{-1/\eta} - \sum_{m=0}^{\infty} F(t)^m \frac{\Gamma(\eta^{-1}+m)}{\Gamma(\eta)m!} \left(\frac{\eta\theta}{1+\eta\theta}\right)^m (1+\eta\theta)^{-1/\eta} \\ &= 1 + (1+\eta\theta)^{-1/\eta} - (1+\eta\theta)^{-1/\eta} \sum_{m=0}^{\infty} \frac{\Gamma(\eta^{-1}+m)}{\Gamma(\eta)m!} \left(\frac{\eta\theta F(t)}{1+\eta\theta}\right)^m \\ &= 1 + (1+\eta\theta)^{-1/\eta} - (1+\eta\theta(1-F(t)))^{-1/\eta} \\ &= 1 + (1+\eta\theta)^{-1/\eta} - (1+\eta\theta S(t))^{-1/\eta}. \end{aligned}$$

The corresponding p.d.f and hazard function of T are given by

$$f_{pop}(t) = \theta f(t)(1+\theta\eta S(t))^{-(1/\eta+1)},$$

and

$$h_{pop}(t) = \frac{\theta f(t)(1+\theta\eta S(t))^{-(1/\eta+1)}}{1 + (1+\theta\eta)^{-1/\eta} - (1+\theta\eta S(t))^{-1/\eta}},$$

respectively, where $f(t) = \frac{\partial}{\partial t}F(t)$. Its cure fraction is given by

$$S_{pop}(\infty) = \lim_{t\to\infty} S_{pop}(t) = \lim_{t\to\infty} 1 + (1+\eta\theta)^{-1/\eta} - (1+\eta\theta S(t))^{-1/\eta} = (1+\eta\theta)^{-1/\eta}.$$

Note that if we set $\eta \to 0$ we have the Complementary Promotion time cure model presented in Section 2.4 and if $\eta = 1$, we obtain the Complementary Geometric cure rate model or Complementary cure rate proportional odds model.

2.5.3 Multivariate frailty models

There are several approaches have been proposed in the literature for analyzing multivariate survival data (Therneau, 2000). Frailty models are classified in the conditional approach, which assume that the lifetimes are conditionally independent given the frailty. This approach is commonly used for modeling the problem of multivariate survival data characterized by the presence of clusters.

One of most popular model in modeling the association between survival times of individuals within each cluster is the shared frailty model. In this case, a frailty is shared in each cluster, that is, the individuals who are in the same cluster have the common frailty. A shared frailty model in survival analysis is defined as follows.

Suppose there are m clusters and that cluster i has n_i observations and associates with the unobserved frailty Z_i for $i = 1, \ldots, m$. The vector $\boldsymbol{X}_{ij}$, $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$ contains the covariate information of the event time T_{ij} of the jth observation in the ith cluster. Conditional on the frailty variance Z_i, the survival times in cluster i are assumed to be independent and their hazard functions to be of the form

$$h_{ij} = z_i h_0(t) \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}), \tag{2.30}$$

where $h_0(t)$ denotes the baseline hazard function, and $\boldsymbol{\beta}$ is the regression parameters vector (fixed effect parameters vector) to be estimated. The frailties Z_i $(i = 1, \ldots, m)$ are assumed to be independently and identically distributed random variables with the known distribution function with mean one and some unknown variances. Note that the equation (2.30) also can be written as

$$h_{ij} = h_0(t) \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + W_i), \tag{2.31}$$

where $W_i = \ln(z_i)$, is assumed to be independently and identically distributed random variables with the a distribution function with mean zero and some unknown variance, so that the proportional hazards model can be obtained if variance has value zero. One of most used model is the normal distribution with mean zero and unknown variance σ^2,i.e., $W_i \sim^{iid} N(0, \sigma^2)$.

In this work, we consider frailty (random effect) corresponding to clusters that are spatially arranged. While it is possible to identify centroid of geographic regions and employ spatial process modeling for the locations, the effects are more naturally associated with areal units. As such we work with conditionally autoregressive (CAR) models and multivariate conditionally autoregressive (MCAR) models for these effects, and these models will be presented as follow.

Conditionally Auto-Regressive (CAR) models: Gaussian case

The CAR models were introduced by Besag (1974), but they have enjoyed a dramatic increase in usage only in the past decade. This resurgence arises from their convenient employment in the context of Gibbs sampling and more general Markov Chain Monte Carlo (MCMC) methods for fitting certain classes of hierarchical spatial models (for more details see Banerjee *et al.* (2004), in Section 5.4.3).

Let $Y_1, \ldots, Y_m$ be m observations with associated with areal units $1, 2, \ldots, m$, and let $W_{m\times m}$ be an adjacent matrix of the map with its elements w_{ij} defined by

$$w_{ij} = \begin{cases} 1, & \text{if region } i \text{ and } j \text{ are adjacent,} \\ 0, & \text{otherwise} \end{cases}$$

and let $w_{i+} = \sum_j w_{ij}$ denoting the number of regions adjacent to region i. For the Gaussian case, we suppose

$$Y_i|y_j, j \neq i \ \sim N\left(\sum_j b_{ij} y_j, \tau_i^2\right), \quad i = 1, \ldots, m. \tag{2.32}$$

These full conditionals are compatible. For obtaining the joint distribution for the Y_i, the Brook's Lemma will be used and it is defined as

Lemma 2.1. ***(Brook's Lemma)*** *Let $\boldsymbol{y}_0 = (y_{10}, \ldots, y_{m0})$ be any fixed point in the support of the joint probability distribution $p(y_1, \ldots, y_m)$.*

$$\begin{aligned} p(y_1, \ldots, y_m) \ &= \ \frac{p(y_1|y_2, \ldots, y_m)}{p(y_{10}|y_2, \ldots, y_m)} \frac{p(y_2|y_{10}, y_3, \ldots, y_m)}{p(y_{20}|y_{10}, y_3 \ldots, y_m)} \times \\ &\qquad \cdots \times \frac{p(y_m|y_{10}, \ldots, y_{m-1,0})}{p(y_{m0}|y_{10}, \ldots, y_{m-1,0})} p(y_{10}, \ldots, y_{m0}). \end{aligned}$$

Banerjee *et al.* (2004) showed that using the Brook's Lemma, the joint probability distribution $p(y_1, \ldots, y_m)$ has expression

$$p(y_1, \ldots, y_m) \propto \exp\left\{-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{D^{-1}(I - B)y}\right\}, \tag{2.33}$$

where $\boldsymbol{I}$ is identity matrix of size n, $\boldsymbol{B} = \{b_{ij}\}$ and $\boldsymbol{D}$ is diagonal with $\boldsymbol{D}_{ii} = \tau_i^2$. The expression (2.33) suggests a joint multivariate normal distribution for $\boldsymbol{Y}$ with mean $\boldsymbol{0}$ and covariance matrix $(\boldsymbol{I - B})^{-1}\boldsymbol{D}$. So let $b_{ij} = \frac{w_{ij}}{w_{i+}}$ and $\tau_i^2 = \frac{\tau^2}{w_{i+}}$, we have the relation

$$\frac{b_{ij}}{\tau_j^2} = \frac{w_{ij}}{w_{i+}}\left(\frac{\tau^2}{w_{j+}}\right)^{-1} = \frac{w_{ji}}{w_{i+}}\left(\frac{\tau^2}{w_{j+}}\right)^{-1} = \frac{w_{ji}}{w_{j+}}\left(\frac{\tau^2}{w_{i+}}\right)^{-1} = \frac{b_{ji}}{\tau_i^2}, \ \text{ for all } i, j,$$

which implies a symmetry of the matrix $\boldsymbol{D^{-1}(I - B)}$. Moreover, the supposition (2.32) is now given by $Y_i|y_j, j \neq i \ \sim N\left(\sum_j w_{ij} y_j / w_{i+}, \tau^2/w_{i+}\right), \ i = 1, \ldots, m$, and the matrix can be written as $\boldsymbol{D^{-1}(I - B)} = \frac{1}{\tau^2}\left(\boldsymbol{D_W - W}\right)$, where $\boldsymbol{D_W}$ is diagonal matrix with $(D_W)_{ii} = w_{i+}$, for $i = 1, \ldots, m$. Thus, the joint probability distribution given in (2.33) becomes

$$p(y_1, \ldots, y_m) \propto \exp\left\{-\frac{1}{2\tau^2}\boldsymbol{y}^\top(\boldsymbol{D_W - W})\boldsymbol{y}\right\},$$

which suggests that it could be a n-dimencional multivariate normal distribution with mean $\mathbf{0}$ and precision matrix $\tau^{-2}(\boldsymbol{D_W} - \boldsymbol{W})$ and we denoted this model by $CAR(\theta)$, where $\theta = \tau^2$. Note that this precision matrix is rank deficient, so it is a non-positive definite matrix and it leads to an improper distribution function. This singularity, while theoretically awkward, creates little problem in a Bayesian implementation, since the identifying sum-to-zero constraint $\sum_{i=1}^{m} Y_i = 0$ is easily imposed in a Gibbs sampler simply by recentering the Y_i draws around zero after every iteration (see Carlin & Louis (2000), pg 263).

Multivariate conditionally autoregressive (MCAR) models

Let $\boldsymbol{Y}^\top = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m)$ where each $\boldsymbol{Y}_i$ is a $p \times 1$, following Mardia (1988), the zero-centered multivariate conditionally autoregressive (MCAR) models sets

$$\boldsymbol{Y}_i | \boldsymbol{Y}_{j \neq i}, \boldsymbol{\Sigma}_i \sim N\left(\sum_j \boldsymbol{B}_{ij}\boldsymbol{Y}_j, \boldsymbol{\Sigma}_i\right), \quad i = 1, \ldots, m, \tag{2.34}$$

where each $\boldsymbol{B}_{ij}$ is $p \times p$, as is each $\boldsymbol{\Sigma}_i$. As in the univariate case, using the Brook's lemma, a joint density for $\boldsymbol{Y}$ of the form

$$p(\boldsymbol{Y} | \{\boldsymbol{\Sigma}_i, i = 1 \ldots, m\}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{Y}^\top \boldsymbol{\Gamma^{-1}}(\boldsymbol{I} - \tilde{\boldsymbol{B}})\boldsymbol{Y}\right\}, \tag{2.35}$$

where $\boldsymbol{\Gamma}$ is block diagonal with blocks $\boldsymbol{\Sigma}_i$ and $\tilde{\boldsymbol{B}}$ is an $np \times np$ with (i, j)th block $\boldsymbol{B}_{ij}$. Similarly the univariate case, we take $b_{ij} = \frac{w_{ij}}{w_{i+}}$ and $\boldsymbol{\Sigma}_i = \frac{\Sigma}{w_{i+}}$, then we have the condition $b_{ij}\Sigma_j = b_{ji}\Sigma_i$ for all i, j, which let the matrix $\boldsymbol{\Gamma^{-1}}(\boldsymbol{I} - \tilde{\boldsymbol{B}})$ be a symmetric. Note that using the Kronecker product notation $\otimes$, the $\boldsymbol{\Gamma^{-1}}(\boldsymbol{I} - \tilde{\boldsymbol{B}})$ matrix can be rewrote as

$$\boldsymbol{\Gamma^{-1}}(\boldsymbol{I} - \tilde{\boldsymbol{B}}) = (\boldsymbol{D_W} \otimes \boldsymbol{\Sigma^{-1}})(\boldsymbol{I} - \boldsymbol{B} \otimes \boldsymbol{I}) = (\boldsymbol{D_W} - \boldsymbol{W}) \otimes \boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma^{-1}}$ and $\boldsymbol{D_W}$, $\boldsymbol{W}$ and $\boldsymbol{B}$ are the same as defined in univariate case.

Again, the singularity of $\boldsymbol{D_W} - \boldsymbol{W}$ implies that $\boldsymbol{\Gamma^{-1}}(\boldsymbol{I} - \tilde{\boldsymbol{B}})$ is singular. An alternative to resolve this problem is insert the smoothness parameter a in $\boldsymbol{D_W} - \boldsymbol{W}$ and standardize $\boldsymbol{W}$ so that each of its rows sum to 1, thus,

$$\boldsymbol{\Gamma^{-1}}(\boldsymbol{I} - \tilde{\boldsymbol{B}}) = (\boldsymbol{D_W} - a\boldsymbol{W}) \otimes \boldsymbol{\Lambda},$$

and the joint probability distribution given in (2.33) becomes

$$p(\boldsymbol{Y}|\{\boldsymbol{\Sigma}_i, i = 1\ldots, m\}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{Y}^\top\left[(\boldsymbol{D_W} - a\boldsymbol{W})\otimes\boldsymbol{\Lambda}\right]\boldsymbol{Y}\right\}, \tag{2.36}$$

which suggests that it could be a multivariate normal distribution with mean $\mathbf{0}$ and the precision matrix

$$(\boldsymbol{D_W} - a\boldsymbol{W})\otimes\boldsymbol{\Lambda}, \tag{2.37}$$

and we denoted this model by $MCAR(a, \boldsymbol{\Lambda})$. The parameter $a \in (0, 1)$ has a spatial smoothness interpretation. Value of a closer to 1 imply greater weight on the adjacency matrix $\boldsymbol{W}$, while a close to 0 implies that the adjacency structure has few role to play in the precision matrix.

Later, Gelfand & Vounatsou (2003) and Carlin & Banerjee (2003) extend the $MCAR(a, \boldsymbol{\Lambda})$ to allow the introduction of a spatial auto-regression coefficient for each component of $\boldsymbol{Y}_i$, and they denoted the extend MCAR model by $MCAR(a_1, \ldots, a_p, \boldsymbol{\Lambda})$.

First, they rearrange the rows of the $np\times 1$ vector $\boldsymbol{Y}$ to block by components, rather than by units, that is, let $\boldsymbol{Y} = (Y_{11}, Y_{21}, \ldots, Y_{m1}, Y_{12}, \ldots, Y_{m2}, \ldots, Y_{1p}, \ldots, Y_{mp})^\top$, thus the precision matrix $\boldsymbol{\Lambda}$ given in (2.37) can be rewritten as $\boldsymbol{\Lambda}\otimes(\boldsymbol{D_W} - a\boldsymbol{W})$.

For the parameter vector $\boldsymbol{a} = (a_1, \ldots, a_p)$, the corresponding positive definite matrix are denoted by $(\boldsymbol{D_W} - a_i\boldsymbol{W})$ for $i = 1, \ldots, p$ and its corresponding the Cholesky factorization are denoted by $\boldsymbol{R_i}^\top\boldsymbol{R_i}$, where $\boldsymbol{R}_i$ is $n\times n$. dimensional matrix. Following the Carlin & Banerjee (2003), the precision matrix can be written as

$$\boldsymbol{\Lambda}\otimes(\boldsymbol{D_W} - \boldsymbol{aW}) = \begin{bmatrix} \lambda_{11}\boldsymbol{R_1}^\top\boldsymbol{R_1} & \lambda_{12}\boldsymbol{R_1}^\top\boldsymbol{R_2} & \ldots & \lambda_{1p}\boldsymbol{R_1}^\top\boldsymbol{R_p} \\ \lambda_{21}\boldsymbol{R_2}^\top\boldsymbol{R_1} & \lambda_{22}\boldsymbol{R_2}^\top\boldsymbol{R_2} & \ldots & \lambda_{2p}\boldsymbol{R_2}^\top\boldsymbol{R_p} \\ \vdots & \vdots & \ldots & \vdots \\ \lambda_{p1}\boldsymbol{R_p}^\top\boldsymbol{R_1} & \lambda_{p2}\boldsymbol{R_p}^\top\boldsymbol{R_2} & \ldots & \lambda_{pp}\boldsymbol{R_p}^\top\boldsymbol{R_p} \end{bmatrix} = \boldsymbol{R}^\top(\boldsymbol{\Lambda}\otimes\boldsymbol{I_m})\boldsymbol{R}, \tag{2.38}$$

where λ_{ij}'s are the elements of the matrix $\boldsymbol{\Lambda}$ and $\boldsymbol{R}$ is a block diagonal matrix with blocks $R_1, \ldots, R_p$. The $\boldsymbol{\Lambda}\otimes(\boldsymbol{D_W} - \boldsymbol{aW})$ is positive definite since $\boldsymbol{\Lambda}$ is positive definite.

2.6 Model comparison criteria

There are several Bayesian criteria to compare competing models for a given data set and to select the one that best fits the data. One of the most used in applied works is the deviance information criterion (DIC), which is based on the posterior mean of deviance. For a model, the

statistic DIC is defined as

$$DIC = \bar{d} + pd.$$

where $\bar{d} = E[D(\boldsymbol{\varphi})]$, $pd = E[D(\boldsymbol{\varphi})] - D[E(\boldsymbol{\varphi})]$ and $D(\boldsymbol{\varphi})$ is the deviance function of the model defined by $-2\log L(\boldsymbol{\varphi})$. L is the likelihood function of the model. Spiegelhalter *et al.* (2002) provide evidences that pd is a suitable measure of model complexity even in hierarchical settings, and thus, DIC is considered as a sensible generalization of the expected Akaike information criterion to hierarchical settings. The model, with the smallest value of DIC, is commonly taken as the preferred model to describe the data set given.

2.7 Bayesian case influence diagnostics

Since regression models are sensitive to underlying model assumptions, generally performing a sensitivity analysis is strongly advisable. One of the most used ways of evaluating the influence of an observation in the fitted model is a case-deletion (Cook & Weisberg, 1982), in which the effects are studied by completely removing cases from the analysis. This reasoning will form the basis of our Bayesian global influence methodology and, in doing so, it will be possible to determine which subjects might influence the analysis. Now, the Bayesian case-deletion influence diagnostic measures for the joint posterior distribution based on the ψ-divergence (Peng & Dey, 1995; Weiss, 1996) will be introduced as follows.

Let $D_\psi(P, P_{(-i)})$ denote the ψ-divergence between P and $P_{(-i)}$, in which P denotes the posterior distribution of $\boldsymbol{\vartheta}$ for full data, and $P_{(-i)}$ denotes the posterior distribution of $\boldsymbol{\vartheta}$ without the ith case. Specifically,

$$D_\psi(P, P_{(-i)}) = \int \psi\left(\frac{\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}^{(-i)})}{\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})}\right)\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})\,d\boldsymbol{\vartheta}. \tag{2.39}$$

where ψ is a convex function with $\psi(1) = 0$. Several choices concerning the ψ are given by Dey & Birmiwal (1994). For example, $\psi(z) = -\log(z)$ defines the Kullback-Leibler (K-L) divergence, $\psi(z) = (z-1)\log(z)$ gives J-distance (or the symmetric version of K-L divergence), $\psi(z) = 0.5|z-1|$ defines the variational distance (or L_1 norm) and $\psi(z) = (z-1)^2$ defines the χ^2-square divergence.

Let $\boldsymbol{\vartheta}^{(1)}, \ldots, \boldsymbol{\vartheta}^{(Q)}$ be a size Q sample of $\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})$, $D_\psi(P, P_{(-i)})$ can be calculated numerically by

$$\widehat{D_\psi}(P, P_{(-i)}) = \frac{1}{Q}\sum_{q=1}^{Q}\psi\left(\frac{\widehat{CPO}_i}{L(y_i|\boldsymbol{\vartheta}^{(q)})}\right), \tag{2.40}$$

where $\widehat{CPO}_i = \left\{\frac{1}{Q}\sum\limits_{q=1}^{Q}\frac{1}{L(y_i|\boldsymbol{\vartheta}^{(q)})}\right\}^{-1}$ is the numerical approximation of the conditional predictive ordinate statistic of i-th observation (Ibrahim *et al.*, 2001a).

Note that $D_\psi(P, P_{(-i)})$ can be interpreted as the ψ-divergence of the effect of deleting the i-th case from the full data on the joint posterior distribution of $\boldsymbol{\vartheta}$. As pointed by Peng & Dey (1995), Weiss (1996) and Cancho *et al.* (2010), it may be difficult for a practitioner to judge the cutoff point of the divergence measure so as to determine whether a small subset of observations is influential or not. In this context, we will use the proposal given by Peng & Dey (1995) and Weiss (1996) by considering a biased coin, which has success probability p. Then the ψ-divergence between the biased and an unbiased coin is

$$D_\psi(f_0, f_1) = \int \psi\left(\frac{f_0(x)}{f_1(x)}\right) f_1(x)dx, \tag{2.41}$$

where $f_0(x) = p^x(1-p)^{1-x}$ and $f_1(x) = 0.5$, $x = 0, 1$. Now if $D_\psi(f_0, f_1) = d_\psi(p)$, then it can be easy to check that d_ψ satisfies the following equation

$$d_\psi(p) = \frac{\psi(2p) + \psi(2(1-p))}{2} \tag{2.42}$$

It is not difficult to see for the divergence measures considered that d_ψ increases as p moves away from 0.5. In addition, $d_\psi(p)$ is symmetric about $p = 0.5$ and d_ψ, achieves its minimum at $p = 0.5$. In this point, $d_\psi(0.5) = 0$, and $f_0 = f_1$. Therefore, if we consider $p > 0.90$ (or $p \leq 0.10$) as a strong bias in a coin, then $d_{\text{K-L}}(0.90) = 0.51$, $d_{\text{J}}(0.90) = 0.88$, $d_{L_1}(0.90) = 0.4$ and $d_{\chi^2}(0.90) = 0.64$. This equation implies that ith case is considered influential when $d_{L_1} > 0.4$ or $d_{\chi^2} > 0.64$. Thus, if we use the Kullback-Leibler divergence, we can consider an influential observation when $d_{\text{K-L}} > 0.51$. Similarly, using the J-distance, an observation which $d_{\text{J}} > 0.88$ can be considered influential.

Chapter 3

Spatial frailty in Cure rate models

In survival analysis, it is common obtain the data set which are collected from different regions, that is the data are clustered by different regions. One of the most used approaches is consider cluster-specific random effects (or frailty) in the modeling. The frailties account for excess heterogeneity in the data, as well as capture similarity across observations within the same cluster. In this section, we will introduce frailties to each spatial cluster in the cure rate models presented in Section 2.5.2 for the interval-censored data.

3.1 Geometric cure rate models with spatial frailties

Supposing that there are I regions and n_i individuals in ith region. Let T_{ij} denotes the random variable for time to the event of the jth individual in the ith region, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, I$. We suppose that the (i, j)th individual is potentially exposed to M_{ij} latent risk, in which M_{ij} denote the initial number of competing causes concerning the occurrence of an event, and assuming M_{ij} has a geometric distribution with parameter $1/(1+\theta_{ij})$, the probability mass function is given by

$$P(M_{ij} = m) = \frac{\theta_{ij}^m}{(\theta_{ij}+1)^{m+1}}, \quad m = 0, 1, 2 \cdots, \tag{3.1}$$

where $\theta_{ij} > 0$, $E(M_{ij}) = \theta_{ij}$ and $Var(M_{ij}) = \theta_{ij}(1+\theta_{ij})$.

Let Y_{cij} denote the lifetime of jth individual in ith region due to the cth ($c = 1, \ldots, M_{ij}$) latent risk. Given $M_{ij} > 0$, $Y_{1ij}, Y_{2ij}, \ldots$ are assumed to be independent and identically distributed with a common distribution function $F(\cdot) = 1 - S(\cdot)$ that does not depend upon M_{ij}. If we assume that the presence of any latent risk will ultimately lead to the occurrence of the event, the time to the event of interest T_{ij} could be defined as $T_{ij} = \min\{Y_{1ij}, \cdots, Y_{M_{ij}ij}\}$ for $M_{ij} \geq 1$. If $M_{ij} = 0$, then

the individual is not at risk of final event and is considered cured. In this case, we define $T_{ij} = \infty$ with $P(T_{ij} = \infty | M_{ij} = 0) = 1$. Thus, the survival function for the population is given by

$$S_{pop}(t_{ij}) = \left[1 + \theta_{ij} F(t_{ij})\right]^{-1}. \tag{3.2}$$

The probability density function (p.d.f) and the hazard function associated to (3.2) are given by

$$f_{pop}(t_{ij}) = \theta_{ij} f(t_{ij}) \left[1 + \theta_{ij} F(t_{ij})\right]^{-2} \quad \text{and} \quad h_{pop}(t_{ij}) = \theta_{ij} f(t_{ij}) \left[1 + \theta_{ij} F(t_{ij})\right]^{-1},$$

respectively, where $f(t_{ij}) = \frac{\partial}{\partial t_{ij}} F(t_{ij})$.

Note that, the survival function in (3.2) can also be written as a mixture cure model

$$S_{\text{pop}}(t_{ij}) = (1+\theta_{ij})^{-1} + \left(1 - (1+\theta_{ij})^{-1}\right) \left\{ \frac{\left[1 + \theta_{ij} F(t_{ij})\right]^{-1} - (1+\theta_{ij})^{-1}}{1 - (1+\theta_{ij})^{-1}} \right\},$$

Thus, the survival functions of uncured (susceptible) individuals can be expressed by

$$S_{\text{sus}}(t_{ij}) = \frac{\left[1 + \theta_{ij} F(t_{ij})\right]^{-1} - (1+\theta_{ij})^{-1}}{1 - (1+\theta_{ij})^{-1}}.$$

If we assume another situation in which the presence of all latent risks will ultimately lead to the occurrence of the event, then the time to the event of interest is defined by the random variable $T_{ij} = \max\{Y_{cij}, c = 1, \cdots, M_{ij}\}$ for $M_{ij} \geq 1$ and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P(T_{ij} = \infty | M_{ij} = 0) = 1$. The survival function for the population is given by

$$S_{pop}(t_{ij}) = 1 + (1+\theta_{ij})^{-1} - \left[1 + \theta_{ij} S(t_{ij})\right]^{-1}. \tag{3.3}$$

The corresponding p.d.f. and the hazard function are given by

$$f_{\text{pop}}(t_{ij}) = \theta_{ij} f(t_{ij}) (1 + \theta_{ij} S(t_{ij}))^{-2},$$

and

$$h_{\text{pop}}(t_{ij}) = \frac{\theta f(t_{ij}) \left[1 + \theta_{ij} S(t_{ij})\right]^{-2}}{1 + (1+\theta_{ij})^{-1} - (1 + \theta_{ij} S(t_{ij}))^{-1}},$$

respectively. The survival function (4.4) can also be written as a mixture cure model

$$S_{\text{pop}}(t_{ij}) = (1+\theta_{ij})^{-1} + \left(1 - (1+\theta_{ij})^{-1}\right) \left\{ \frac{1 - (1 + \theta_{ij} S(t_{ij}))^{-1}}{1 - (1+\theta_{ij})^{-1}} \right\},$$

Thus, the survival functions of susceptible individuals is given by

$$S_{\text{sus}}(t_{ij}) = \frac{1-(1+\theta_{ij}S(t_{ij}))^{-1}}{1-(1+\theta_{ij})^{-1}}.$$

The first situation, also known as the first activation scheme because, in this case, we assume the event of interest occurs when the first possible cause is activated. On the other hand, the second situation is known as the last activation scheme, because the event of interest only takes place after all the latent causes have been activated (see, Cooner *et al.*, 2007). Thus, we denoted the survival functions (3.2) and (3.3) by $S^F_{\text{pop}}(t_{ij})$ and $S^L_{\text{pop}}(t_{ij})$, respectively. There is another kind of situation in which the event of interest occurs: when some of the possible causes are activated and, given the number of latent causes M_{ij}, the number of activated causes is a random variable with discrete uniform distribution in $\{1,\cdots,M_{ij}\}$. This situation is known as random activation scheme. In this case, the survival function for the population is given by

$$S^R_{\text{pop}}(t_{ij}) = (1+\theta_{ij})^{-1} + \left(1-(1+\theta_{ij})^{-1}\right) S(t_{ij}), \tag{3.4}$$

where the superscript R denotes random activation scheme.

Note that whichever the activation scheme, the density and hazard functions of the cure models are improper functions, since the survival functions are not proper. Its cure fraction is the same for these activation schemes and, thus, it can be obtained by $p_{0ij} = \lim_{t_{ij}\to\infty} S_{pop}(t_{ij}) = (1+\theta_{ij})^{-1}$. However, under different activation schemes, the models differ by its surviving, density and hazard functions. Moreover, under the conditions of the models (3.2), (3.3) and (3.4) for any distribution function $F(\cdot)$, we have $S^F_{\text{pop}}(t_{ij}) \leq S^R_{\text{pop}}(t_{ij}) \leq S^L_{\text{pop}}(t_{ij})$ for all $t_{ij} > 0$.

As is well known, the cure fraction plays a key role in the survival models with cure fraction. So we consider the parametrization of the model in cure fraction in expressions. Since $p_{0ij} = (1+\theta_{ij})^{-1}$, we have $\theta_{ij} = p_{0ij}^{-1} - 1$. Moreover, we propose that the cured fraction of an individual (i,j)th be associated with covariates $\boldsymbol{x}_{ij}$. Thus linking p_{0ij} to covariates $\boldsymbol{x}_{ij}$ by

$$p_{0ij} = \frac{\exp(\xi_{ij})}{1+\exp(\xi_{ij})},\ j = 1,\ldots,n_i,\ i = 1,\ldots,I,$$

where ξ_{ij} is a linear function of covariates, $\xi_{ij} = \boldsymbol{x}_{ij}^\top \boldsymbol{b}$ where $\boldsymbol{b}$ is a p_1-dimensional vector which represents the effects of covariates on the cured fraction. Thus, the models in (3.2), (3.3) and (3.4) parameterized in the p_{0ij} can be written as

$$S^F_{pop}(t_{ij}) = \left[1+(p_{0ij}^{-1}-1)F(t_{ij})\right]^{-1}, \tag{3.5}$$

$$S_{pop}^{L}(t_{ij}) = 1 + p_{0ij} - \left[1 + (p_{0ij}^{-1} - 1)S(t_{ij})\right]^{-1} \tag{3.6}$$

and

$$S_{pop}^{R}(t_{ij}) = p_{0ij} + (1 - p_{0ij})S(t_{ij}), \tag{3.7}$$

The model in (3.7) is the same considered by Banerjee & Carlin (2004).

The non-negative random variables Y_{cij}'s can take several distributions, we assume Y_{cij}'s take proportional hazard (PH) model with the baseline hazard function $h_0(t|\boldsymbol{\alpha})$, the conditional hazard function and corresponding survival function are given by

$$h(t|\boldsymbol{\phi}) = h_0(t|\boldsymbol{\alpha})\exp(\lambda_{ij}) \;\text{ or }\; \mathrm{S}(\mathrm{t}|\boldsymbol{\phi}) = \mathrm{S}_0(\mathrm{t}|\boldsymbol{\alpha})^{\exp(\lambda_{\mathrm{ij}})}, \tag{3.8}$$

where $\boldsymbol{\phi} = (\boldsymbol{\alpha}, \lambda_{ij})$, $\lambda_{ij} = \boldsymbol{z}_{ij}'\boldsymbol{\beta}$ is the linear predictor of the covariates, where $\boldsymbol{z}_{ij}$ is covariates of an individual (i, j) and $\boldsymbol{\beta}$ is a p_2-dimensional vector representing the effects of covariates on the survival model component. $S_0(t|\boldsymbol{\alpha})$ is the baseline survival function corresponding to $h_0(t|\boldsymbol{\alpha})$ and $\boldsymbol{\alpha}$ is the parameter vector of the baseline functions. Note that different distributions will be obtained if we take different baseline functions. In this paper, we consider two different distributions for the baseline functions. Firstly, we assume the baseline hazard function $h_0(t|\alpha) = \alpha t^{\alpha-1}$, thus Y_{cij}'s follow a Weibull distribution with its p.d.f. $f(t|\boldsymbol{\phi}) = \alpha t^{\alpha-1}\exp(\lambda_{ij} - t^{\alpha}e^{\lambda_{ij}})$, where $\alpha > 0$ is a shape parameter and e_{ij}^{λ} is a scale parameter. In this case, we called the functions (3.2) and (3.3) by Weibull geometric cure rate (WGCR) model and Complementary Weibull geometric cure rate (CWGCR) model, respectively.

Secondly, we assume that the baseline functions have the piecewise exponential distribution. Let the vector $\boldsymbol{a} = (a_0, a_1, \ldots, a_{Q-1})$ with $0 = a_0 < a_1 < \ldots < a_{Q-1} < \infty$ be a finite partition of time axis and α_q be the hazard rate of qth interval of intervals $(0, a_1], \ldots, (a_{Q-1}, \infty]$, for $q = 1, \ldots, Q$, so the baseline survival function has expression

$$S_0(t|\boldsymbol{\alpha}) = \exp\left\{-\sum_{q=1}^{Q}\alpha_q\Delta_q(t)\right\}, \quad t > 0, \tag{3.9}$$

where

$$\Delta_q(t) = \begin{cases} 0, & \text{if } \mathrm{t} < \mathrm{a_{q-1}}; \\ t - a_{q-1} & \text{if } \mathrm{a_{q-1}} \le \mathrm{t} < \mathrm{a_q}, \quad \mathrm{q} = 1\ldots, \mathrm{Q}. \\ a_q - a_{q-1} & \text{if } \mathrm{t} \ge \mathrm{a_q}. \end{cases}$$

Note that if $\alpha_i = \alpha$ for all $i = 1, \ldots, Q$, we have an exponential distribution with a parameter α as

the particular case. Moreover, if we partition the time axis $0 = a_0 < a_1 < \ldots < a_{Q-1} < \infty$, so that they denote the ordered distinct time points of all observed interval end points, then we have $t_q = a_q$ for $q = 0, \ldots, Q$. Now, the survival function can be written as

$$S_0(t_q|\boldsymbol{\alpha}) = \begin{cases} 1, & q = 0 \\ \exp\left\{-\sum_{k=1}^{q} \alpha_k (a_k - a_{k-1})\right\}, & q = 1, \ldots, Q-1 \\ 0, & q = Q, \end{cases} \tag{3.10}$$

where $\alpha_q(a_q - a_{q-1}) \geq 0$, $q = 1, \ldots, Q$. Here, we called the function (3.2) by PH Geometric cure rate (PHGCR) model and (3.3) by Complementary PH Geometric cure rate (CPHGCR) model.

Now, we will introduce the frailties U_i and V_i to better explain the effect of survival time of susceptible individuals and on the cured probability through linear predictor expression

$$\begin{aligned} \lambda_{ij} &= \boldsymbol{z}'_{ij}\boldsymbol{\beta} + U_i, \\ \xi_{ij} &= \boldsymbol{x}'_{ij}\boldsymbol{b} + V_i, \quad \text{for } j = 1 \ldots, n_i, i = 1, \ldots, I. \end{aligned}$$

Here, the frailties U_i and V_i must be spatially correlated across the regions. In this work we propose two approaches, the first we employ separate independent conditionally auto-regressive (CAR) prior distribution on $(\boldsymbol{U}, \boldsymbol{V})$. The other one we assuming the spatial priors on $(\boldsymbol{U}, \boldsymbol{V})$ are dependent, and they have multivariate conditionally auto-regressive MCAR prior distribution, where the CAR and MCAR distributions were presented in Section 2.5.3 in detail.

3.1.1 Bayesian Inference

Let $\boldsymbol{D}_{obs} = \{(A_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \delta_{ij}); j = 1, \ldots, n_i, i = 1, \ldots M\}$ denote the observed data, where $A_{ij} = (t_{ijL}, t_{ijR}]$ is the interval during which individual j in cluster i occurs the event of interest, $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are the p_1-dimencional and p_2-dimencional vectors of covariates, and δ_{ij} is following interval censoring indicador: $\delta_{ij} = I(t_{ijR} < \infty)$. For the spacial case in which the survival time is right-(left-) censored, $R_{ij} = +\infty (L_{ij} = 0)$, whereas for exact observations, $t_{ijL} = t_{ijR}$. Following Finkelstein(1986), the likelihood function for the general interval-censored cure rate model is given by

$$\begin{aligned} L(\boldsymbol{\varphi}|\boldsymbol{\mathcal{D}}, \boldsymbol{U}, \boldsymbol{V}) &\propto \prod_{i=1}^{I}\prod_{j=1}^{n_i} \left(S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi}) - S_{\text{pop}}(t_{ijR}|\boldsymbol{\varphi})\right)^{\delta_{ij}} S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi})^{1-\delta_{ij}} \\ &\propto \prod_{i=1}^{I}\prod_{j=1}^{n_i} S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi}) \left(1 - \frac{S_{\text{pop}}(t_{ijR}|\boldsymbol{\varphi},)}{S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi})}\right)^{\delta_{ij}}, \end{aligned} \tag{3.11}$$

where $\boldsymbol{\varphi} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\alpha})$, $\boldsymbol{\alpha}$ is the shape parameter of the Weibull distribution for the first model with unitary size and it is the risk parameter vector for the second model with size Q. For a Bayesian analysis, we assume the following prior densities for parameters $\boldsymbol{b}^\top$, $\boldsymbol{\beta}^\top$ and $\boldsymbol{\alpha}$

- $b_j \sim N(\mu_b, \sigma_b^2)$, $j = 0, \ldots, (p1 - 1)$, with μ_b and σ_b known ;
- $\beta_j \sim N(\mu_\beta, \sigma_\beta^2)$, $j = 1, \ldots, p2$, with μ_β and σ_β known;
- $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)\boldsymbol{I}_{(0,\infty)}$, with μ_α and σ_α known, $i = 1$ for first model and $i = 1, \ldots Q$ for second model;

where $N(\mu, \sigma^2)\boldsymbol{I}_{(a,b)}$ denotes the truncated normal distribution which is the probability distribution of a normally distributed random variable whose value lies within the interval $-\infty \leq a < b \leq \infty$. To express vague priors, we consider $\mu_b = \mu_\beta = \mu_\eta = \mu_\alpha = 0$ with large values of σ_b^2, σ_β^2 and σ_α^2. In several areas, special in medicine, it is preferable to use the prior information when they are available. Moreover, it is worth mentioning that using a truncated normal distribution as prior facilitates the insertion of information in certain regions of the parameter space, since the hyperparameters no longer represent the mean and variance but still control the region of higher probability mass.

Independent assumption

For the independent assumption, we employ separate independent CAR priors on the random frailties $\boldsymbol{U} = (U_1, \ldots, U_I)^\top$ and $\boldsymbol{V} = (V_1, \ldots, V_I)^\top$, that is,

- $U_1, \ldots, U_I \sim \text{CAR}(\theta_1)$;
- $V_1, \ldots, V_I \sim \text{CAR}(\theta_2)$;

where θ_1 and θ_2 are positive unknown hyper-parameters, and we assume they have Inverse-Gamma prior with the known shape parameter $a_0 > 0$ and scale parameter $b_0 > 0$. the joint posterior distribution for the parameters is given

$$\pi(\boldsymbol{\varphi}, \theta_1, \theta_2 | \boldsymbol{D}_{\text{obs}}) \propto L(\boldsymbol{\varphi} | \boldsymbol{D}_{obs}, \boldsymbol{U}, \boldsymbol{V}) \pi(\boldsymbol{U} | \theta_1) \pi(\boldsymbol{V} | \theta_2) \pi(\boldsymbol{\varphi}, \theta_1, \theta_2),$$

where $\pi(\boldsymbol{\varphi}, \theta_1, \theta_2) = \pi(\boldsymbol{b})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\alpha})\pi(\theta_1)\pi(\theta_2)$ and $L(\boldsymbol{\varphi}|\boldsymbol{\mathcal{D}})$ is the likelihood function given in (3.11). Note that, this joint posterior density is analytically intractable. So, we based our inference on the Markov chain Monte Carlo (MCMC) simulation methods. We can observed that the full conditional distributions for parameters $\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{U}$ and $\boldsymbol{V}$ have not closed forms, thus we will use the Metropolis-Hastings algorithm to generate posterior samples for these parameter. To avoid range restrictions on

the parameters α_i's, we define $\zeta_i = \log(\alpha_i)$ for $i = 1, 2, \ldots, Q$, to transform all parameters space to real space (necessary to work with Gaussian proposal densities). Let $\boldsymbol{\vartheta} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \theta_1, \theta_2)$, according for the Jacobian of this transformation, the joint prior density $\pi(\boldsymbol{\vartheta})$ has expression

$$\pi(\boldsymbol{\vartheta}) = \pi(\boldsymbol{\varphi}, \theta_1, \theta_2) \times \exp\left(\sum_{i=1}^{Q} \zeta_i\right), \tag{3.12}$$

where $\boldsymbol{\varphi} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}^{-1})$, $\boldsymbol{\zeta}^{-1}$ denote inverse function of $\boldsymbol{\zeta}$, i.e., $\boldsymbol{\zeta}^{-1} = \left\{\zeta_i^{-1} = \exp(\zeta_i), \;\; i = 1 \ldots, Q\right\}$.

On the other hand, the full conditional distributions for parameters θ_i's are given

$$\begin{aligned}
\pi(\theta_i|\boldsymbol{\vartheta}_{-\theta_i}, \boldsymbol{D}_{\text{obs}}) &\propto \pi(\boldsymbol{\psi}_i|\theta_i)\pi(\theta_i) \\
&\propto (\theta_i)^{-k/2} \exp\left(-\frac{1}{2\theta_i}\boldsymbol{\psi}_i^{\top}(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{\psi}_i\right) \theta_i^{-a_0-1} \exp(-b_0\theta_i^{-1}) \\
&\propto \theta_i^{-\left(a_0+\frac{k}{2}\right)-1} \exp\left\{-\left(\frac{\boldsymbol{\psi}_i^{\top}(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{\psi}_i}{2} + b_0\right)\theta_i^{-1}\right\}, \;\; i = 1, 2
\end{aligned}$$

where $\boldsymbol{\psi}_1 = \boldsymbol{U}$, $\boldsymbol{\psi}_2 = \boldsymbol{V}$ and k is the rank of the matrix $\boldsymbol{D_W} - \boldsymbol{W}$. Thus, the full conditional distributions of the parameter θ_i is an Inverse-Gamma distribution with parameters $a_0 + \frac{k}{2}$ e $b_0 + \frac{1}{2}(\boldsymbol{\psi}_i'(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{\psi}_i)$. In this case, the Gibbs sampler algorithm (see Gamerman & Lopes, 2006) is used to generate a posteriori sample.

Thus, the joint posterior density of $\pi(\boldsymbol{\vartheta}|\boldsymbol{D}_{\text{obs}})$ is proportional to

$$\begin{aligned}
&L(\boldsymbol{\varphi}|\boldsymbol{D}_{obs}, \boldsymbol{U}, \boldsymbol{V}) \exp\left\{-\frac{1}{2}\left[\sigma_b^{-2}\sum_{i=0}^{p1-1} b_i^2 + \sigma_\beta^{-2}\sum_{i=1}^{p2}\beta_i^2 + \sum_{i=1}^{Q}\frac{\exp(2\zeta_i)}{\sigma_\alpha^2} + \frac{\boldsymbol{U}^{\top}(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{U}}{\theta_1}\right.\right. \\
&\left.\left. + \frac{\boldsymbol{V}^{\top}(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{V}}{\theta_2}\right] - (a_0 + 1)\left(\log(\theta_1) + \log(\theta_2)\right) - \left(\frac{b_0}{\theta_1} + \frac{b_0}{\theta_2}\right) + \sum_{i=1}^{Q}\zeta_i\right\}.
\end{aligned}$$

Dependent assumption

Now we assume that the spatial priors on the parameters $(\boldsymbol{U}, \boldsymbol{V})$ are dependent on each other. Let $\boldsymbol{\psi} = (\boldsymbol{U}^{\top}, \boldsymbol{V}^{\top})^{\top}$, we first employ of the parameter $\boldsymbol{\psi}$ has a MCAR distribution with a common smoothness parameter a, i.e.,

$$\boldsymbol{\psi} \sim MCAR(a, \boldsymbol{\Lambda}).$$

Further, we assume the parameter $\boldsymbol{\psi}$ has a extended MCAR distribution which assuming the different smoothness parameters for the parameters $\boldsymbol{U}$ and $\boldsymbol{V}$, say a_1 and a_2, that is,

$$\boldsymbol{\psi} \sim MCAR(a_1, a_2, \boldsymbol{\Lambda}).$$

The prior distributions for $\boldsymbol{a}$ and $\boldsymbol{\Lambda}$ are given by

- $a_i \sim$ Uniform$(0,1)$ or $a_i \sim$ Beta$(18,2)$, for i,
- $\boldsymbol{\Lambda} \sim$ Wishart(n_0, Λ_0), with n_0 and Λ_0 known,

where i=1 for $\boldsymbol{\psi} \sim MCAR(a, \boldsymbol{\Lambda})$ and i=1,2 for $\boldsymbol{\psi} \sim MCAR(a_1, a_2, \boldsymbol{\Lambda})$). The prior distributions for the parameter a_i is used by Banerjee & Carlin (2004), in which $a_i \sim$ Uniform$(0,1)$ is a non-informative prior, and $a_i \sim$ Beta$(18,2)$ is an informative prior with $E[a_i] = 0.9$ and $Var[a_i] = 0.004285$; On the other hand, the prior distribution for the parameter $\boldsymbol{\Lambda}$ is used not only by Carlin & Banerjee (2003) but also by Gelfand & Vounatsou (2003) and Banerjee & Carlin (2004). They suggested that n_0 can take value as the dimension of matrix $\boldsymbol{\Lambda}$. However, Gelfand & Vounatsou (2003) and Banerjee & Carlin (2004) considered Λ_0 equals $\boldsymbol{I}$ and $0.01\boldsymbol{I}$ in their papers, respectively, where $\boldsymbol{I}$ denote a identity matrix. Both authors also commented that they had no prior knowledge regarding the nature or extent of dependence for the parameter $\boldsymbol{\Lambda}$. Note that $\boldsymbol{\Lambda}^{-1}$ describe the relative variability and covariance relationship between the different diseases given the neighboring site. Thus, if $\Lambda_0 = 0.01\boldsymbol{I}$, we assumed high relative variability between neighborhoods and we assumed low relative variability between neighborhoods if $\Lambda_0 = \boldsymbol{I}$. Thus, it is necessary to conduct a prior study for the parameter Λ_0 to verify the influence of Λ_0 in the estimation, in order to have a value for appropriate Λ_0.

To avoid range restrictions on the parameters a_i, considering the transformations $\rho_i = \log(a_i/(1-a_i)) \in \mathbb{R}$, then, the joint posterior density is given by

$$\begin{aligned}
\pi(\boldsymbol{\vartheta}|\boldsymbol{D}_{\text{obs}}) \;\propto\; & L(\boldsymbol{\varphi}|\boldsymbol{D}_{obs}, \boldsymbol{\psi}) \exp\left\{ -\frac{1}{2}\left[\sigma_b^{-2} \sum_{i=0}^{p_1} b_i^2 + \sigma_\beta^{-2} \sum_{i=1}^{p_2} \beta_i^2 + \sum_{i=1}^{Q} \frac{\exp(2\zeta_i)}{\sigma_\alpha^2} \right] \right. \\
+ \;& \boldsymbol{\psi}^\top \left[\boldsymbol{\Lambda} \otimes (\boldsymbol{D_W} - \boldsymbol{aW}) \right] \boldsymbol{\psi} + \log|\boldsymbol{\Lambda} \otimes \boldsymbol{aW}| + \frac{n_0 - 4}{2} \log|\boldsymbol{\Lambda}| - \frac{1}{2} tr(\Lambda_0^{-1}\Lambda) \\
+ \;& \left. + \sum_{i=1}^{Q} \zeta_i \right\} \pi(\boldsymbol{\rho}),
\end{aligned}$$

where $\boldsymbol{\varphi} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}^{-1})$ and $\pi(\rho_i) = 1$ if $a_i \sim$ Uniform$(0,1)$ and $\pi(\rho_i) = \frac{1}{B(18,2)} \frac{\exp(17\rho_i)}{(1+\exp(\rho_i))^{18}}$ if $a_i \sim$ Beta$(18,2)$, where $B(18,2) = \frac{17!}{18!} = \frac{1}{18}$.

This joint posterior density is analytically intractable. So, we based our inference on the Markov chain Monte Carlo (MCMC) simulation methods. We can observed that the full conditional

distributions for parameters $\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\psi}$ and $\boldsymbol{\rho}$ have not closed forms, thus we will use the Metropolis-Hastings algorithm to generate a posteriori samples for these parameter. However, the Gibbs sampler algorithm is used to generate a posteriori sample for the parameter $\boldsymbol{\Lambda}$, because its the full conditional distribution has a closed form. The full conditional distribution $\pi(\boldsymbol{\Lambda}|\boldsymbol{\vartheta}_{(-\boldsymbol{\Lambda})}, \boldsymbol{D}_{\text{obs}}))$ is proportional to

$$\begin{aligned} & \pi(\boldsymbol{\psi}|\boldsymbol{\Lambda}, \boldsymbol{a})\pi(\boldsymbol{\Lambda}) \\ \propto\ & |\boldsymbol{\Lambda} \otimes \boldsymbol{D_W} - \boldsymbol{aW}|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{\psi}^\top(\boldsymbol{D_W} - \boldsymbol{aW})\boldsymbol{\psi}\right) |\boldsymbol{\Lambda}|^{(n_0-4)/2} \exp\left(-\frac{1}{2}tr(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda})\right) \\ \propto\ & |\boldsymbol{\Lambda}|^{(I+n_0-4)/2} \exp\left(-\frac{1}{2}tr((\boldsymbol{\Lambda}_0^{-1} + \boldsymbol{B})\boldsymbol{\Lambda})\right), \end{aligned} \tag{3.13}$$

where

$$\boldsymbol{B} = \begin{bmatrix} tr(\boldsymbol{R_1 U}(\boldsymbol{R_1 U})^\top) & tr(\boldsymbol{R_1 U}(\boldsymbol{R_2 V})^\top) \\ tr(\boldsymbol{R_2 V}(\boldsymbol{R_1 U})^\top) & tr(\boldsymbol{R_2 V}(\boldsymbol{R_2 V})^\top) \end{bmatrix}$$

Thus, the full conditional distribution for $\boldsymbol{\Lambda}$ can be taken the Wishart distribution with scala matrix $(\Lambda_0^{-1} + B)^{-1}$ and degrees of freedom $I + n_0$.

3.1.2 Simulation study

In this section we present simulation studies for WGCR model, CWGCR model, PHGCR model and CPHGCR model with the dependent assumption in order to examine the theirs performances. The interval-censored survival times $(t_{ijL}, t_{ijR}, \delta_{ij})$ with the cure fraction under the first and last activations are generated in a manner similar to that employed by Yau & Ng (2001) with some modifications.

First, we generate latent Geometric variable M_{ij}, which denote the initial number of competing causes related to the event, with parameter $p_{0ij} = [1 + \exp(-(b_0 + b_1)x_{ij} + v_i)]^{-1}$ for the jth individual in the ith region, $j = 1, \ldots, n_i$, $i = 1, \ldots, I$, where covariate x_{ij} follows Bernoulli(0.5) distribution. Interval-censored data $(t_{ijL}, t_{ijR}, \delta_{ij})$ are then generated as follows:

(i) If $M_{ij} = 0$, then let $t_{ij} = t_{ijL}$ from the exponential distribution with hazard rate 10 and let censoring indicator $\delta_{ij} = 0$.

(ii) If $M_{ij} > 0$, then

- we generate M_{ij} latent Weibull variables with parameter α and $\lambda_{ij} = (\beta x_{ij} + u_i)$, if Y_{cij}'s has the Weibull distribution;
- or we generate M_{ij} latent Exponential variables with hazard rate $\alpha\lambda_{ij} = \alpha(\beta x_{ij} + u_i))$, if Y_{cij}'s take the PH model.

Let t_{ij} takes lowest generated variable in case of generating the variables of model under first activation and t_{ij} takes largest generated variable in case of generating the variables of model under last activation. The censoring variable c_{ij} is generated from $U(0, cc)$, $cc > 0$ is fixed to control the percentage of censored data. Let $\delta_{ij} = 1$ if $t_{ij} \leq c_{ij}$ and $\delta_{ij} = 0$ otherwise.

(iii) For $\delta_{ij} = 0$, let $0 < t_{ijL} < t_{ijR} = \infty$.

(iv) For $\delta_{ij} = 1$, we create len_{ij} from distribution $U(0.2, 0.7)$ and l_{ij} from $U(0, 0.01)$. Then, from $(0, l_{ij}], (l_{ij}, l_{ij} + len_{ij}], \ldots, (l_{ij} + klen_{ij}, \infty]$, $k = 1, 2, \ldots,$, $(t_{ijL,t_{ijR}}]$ is chosen as that satisfying $t_{ijL} < t_{ij} \leq t_{ijR}$.

In the simulation study, we consider $I = 5$ regions (Zip) with the corresponding adjacent matrix is $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$, the random effects u_i and v_i are generated from Normal distribution with mean $\mathbf{0}$ and precision matrix $\mathbf{\Lambda} \bigotimes (\boldsymbol{D_W} - a\boldsymbol{W})$, where $\boldsymbol{W}$ is standardized adjacent matrix so that each of its rows sum to one, $D_W = \text{Diag}(1, 1, 2, 1, 1)$ is a diagonal matrix and we fixed $a = 0.9$ and $\Lambda = \text{Diag}(4, 4)$, i.e. we fixed $\Lambda_{11} = 4$, $\Lambda_{22} = 4$ and $\Lambda_{12} = \Lambda_{21} = 0$. We consider 100 individuals in the simulation studies. The corresponding Zip codes for each individual was distributed using sample with replace, thus the number of individuals in each region n_i, $i = 1 \ldots, 5$ are varied, that is, these five regions could present different numbers of individuals with $\sum_{i=1}^{5} n_i = 100$. Thus, we have sample size $n = 100$ and we fixed the parameters $b_0 = -1.50$, $b_1 = -0.50$, $\beta = -0.15$, $\alpha = 0.30$ for WGCR and CWFCR models and $\alpha = 1.0$ for PHGCR and CPHGCR models. In simulations, we consider around 40 per cent of the censored data for each generated sample and 500 repeated samples are simulated for each model. The priors for the parameters b_0, b_1, β_1 and α used in the studies, are $b_0 \sim N(0, 3^2)$, $b_1 \sim N(0, 3^2)$, $\beta_1 \sim N(0, 3^2)$, and $\alpha \sim N(0, 10^2)\boldsymbol{I}_{(0,\infty)}$.

For each generated data set we simulate one chain of size 10000 for each parameter, disregarding the first 1000 iterations to eliminate the effect of the initial values and to avoid correlation problems and thinning to every third iteration, thus obtaining a effective sample of size 3000 upon which the posterior is based on. To evaluate the performance of the parameter estimates, the average bias (Bias), standard deviation (SD) of the estimate, average standard deviation (SDs mean) and mean square error (MSE) are calculated for the fitted models, the summaries are presented in Table 3.17 and 3.18. We can note that the bias and MSE of parameter Λ_{12} are lager than others in all fitting models. The estimator of Λ_{12} presents a negative biases for WGCR and PHGCR models and it presents a positive biases for the complementary cure rate models, however its biases and MSEs

are always near zero. Moreover, the simulation results for the cure rate models considering the prior 1 very close to those obtained considering the prior 2.

Table 3.1: Simulation results for WGCR and CWGCR models with depended spatial fragilities

WGCR Model						
Parameter	True value	Estimate mean	SD of the estimate	Bias	MSE	SDs mean
Prior 1: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.9, 1))$						
b_0	-1.50	-1.4812	0.0708	0.0188	0.0054	0.2685
b_1	-0.50	-0.5209	0.1280	-0.0209	0.0168	0.2485
β	-0.15	-0.1360	0.0466	0.0140	0.0024	0.1915
α	0.30	0.1930	0.0521	-0.1070	0.0142	0.0677
Λ_{11}	4.00	4.0062	0.1597	0.0062	0.0255	2.4728
Λ_{22}	4.00	4.0120	0.1918	0.0120	0.0369	2.6151
Λ_{12}	0.00	-0.4541	0.1343	-0.4541	0.2242	1.9196
a	0.90	0.9001	0.0016	0.0001	0.0000	0.0653
Prior 2: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.9, 1))$						
b_0	-1.50	-1.4902	0.0701	0.0098	0.0050	0.2583
b_1	-0.50	-0.5376	0.1330	-0.0376	0.0191	0.2227
β	-0.15	-0.1295	0.0493	0.0205	0.0028	0.1870
α	0.30	0.1863	0.0443	-0.1137	0.0149	0.0536
Λ_{11}	4.00	4.1638	0.1676	0.1638	0.0549	2.5070
Λ_{22}	4.00	4.2657	0.1819	0.2657	0.1036	2.6919
Λ_{12}	0.00	-0.5809	0.1472	-0.5809	0.3591	1.9647
a_1	0.90	0.8999	0.0015	-0.0002	0.0000	0.0655
a_2	0.90	0.9002	0.0015	0.0002	0.0000	0.0654
CWGCR Model						
Parameter	True value	Estimate mean	SD of the estimate	Bias	MSE	SDs mean
Prior 1: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.85, 1))$						
b_0	-1.50	-1.4814	0.0552	0.0186	0.0034	0.2697
b_1	-0.50	-0.4285	0.1003	0.0715	0.0152	0.2726
β	-0.15	-0.1352	0.0918	0.0148	0.0086	0.1438
α	0.30	0.4089	0.0519	0.1089	0.0145	0.0620
Λ_{11}	4.00	4.1665	0.3941	0.1665	0.1827	2.3424
Λ_{22}	4.00	3.9432	0.1946	-0.0568	0.0410	2.6040
Λ_{12}	0.00	0.2061	0.2249	0.2061	0.0929	1.8821
a	0.90	0.8999	0.0016	-0.0001	0.0000	0.0658
Prior 2: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.85, 1))$						
b_0	-1.50	-1.4852	0.0585	0.0148	0.0036	0.2697
b_1	-0.50	-0.4329	0.1063	0.0671	0.0158	0.2728
β_1	-0.15	-0.1340	0.0905	0.0161	0.0084	0.1441
α	0.30	0.4095	0.0530	0.1095	0.0148	0.0619
Λ_{11}	4.00	4.2048	0.4139	0.2048	0.2130	2.3612
Λ_{22}	4.00	3.9361	0.1967	-0.0639	0.0427	2.5979
Λ_{12}	0.00	0.2176	0.2215	0.2176	0.0963	1.8797
a_1	0.90	0.9003	0.0016	0.0003	0.0000	0.0653
a_2	0.90	0.9001	0.0015	0.0001	0.0000	0.0652

Table 3.2: Simulation results for PHGCR model and CPHGCR model with depended spatial fragilities

PHGCR Model						
Parameter	True value	Estimate mean	SD of the estimate	Bias	MSE	SDs mean
Prior 1: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.9, 1))$						
b_0	-1.50	-1.6441	0.0515	-0.1441	0.0234	0.2710
b_1	-0.50	-0.5215	0.1179	-0.0215	0.0143	0.2482
β	-0.15	-0.1538	0.0479	-0.0038	0.0023	0.1877
α	1.00	1.1920	0.0382	0.1920	0.0370	0.1396
Λ_{11}	4.00	4.2224	0.1479	0.2224	0.0713	2.5349
Λ_{22}	4.00	3.9272	0.1801	-0.0728	0.0377	2.5894
Λ_{12}	0.00	-0.4142	0.1420	-0.4142	0.1917	1.9325
a	0.90	0.8999	0.0015	-0.0001	0.0000	0.0655
Prior 2: $\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.9, 1))$						
b_0	-1.50	-1.6418	0.0464	-0.1418	0.0222	0.2710
b_1	-0.50	-0.5146	0.1231	-0.0146	0.0153	0.2491
β	-0.15	-0.1552	0.0502	-0.0052	0.0025	0.1875
α	1.00	0.8980	0.0612	-0.1020	0.1900	0.0100
Λ_{11}	4.00	4.2411	0.1471	0.2411	0.0797	2.5337
Λ_{22}	4.00	3.9259	0.1915	-0.0741	0.0421	2.5853
Λ_{12}	0.00	-0.4153	0.1437	-0.4153	0.1931	1.9307
a_1	0.90	0.9001	0.0016	0.0001	0.0000	0.0655
a_2	0.90	0.9000	0.0016	0.0000	0.0000	0.0653
CPHCRM						
Parameter	True value	Estimate mean	SD of the estimate	Bias	MSE	SDs mean
Prior 1: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.75, 1))$						
b_0	-1.50	-1.6533	0.0852	-0.1533	0.0308	0.2635
b_1	-0.50	-0.5056	0.0998	-0.0056	0.0100	0.2652
β	-0.15	-0.1298	0.0933	0.0202	0.0091	0.1323
α	1.00	0.9090	0.0408	-0.0910	0.1770	0.0850
Λ_{11}	4.00	4.2564	0.2096	0.2564	0.1096	2.3576
Λ_{22}	4.00	3.7852	0.3098	-0.2148	0.1420	2.5456
Λ_{12}	0.00	0.3803	0.1762	0.3803	0.1756	1.7807
a	0.90	0.9001	0.0016	0.0001	0.0000	0.0653
Prior 2: $\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.75, 1))$						
b_0	-1.50	-1.6518	0.0867	-0.1518	0.0306	0.2657
b_1	-0.50	-0.5190	0.1084	-0.0190	0.0121	0.2656
β	-0.15	-0.1373	0.1006	0.0127	0.0103	0.1325
α	1.00	0.9200	0.0407	-0.0800	0.1800	0.0600
Λ_{11}	4.00	4.0064	0.2029	0.0064	0.0411	2.2193
Λ_{22}	4.00	3.7107	0.3010	-0.2893	0.1741	2.4853
Λ_{12}	0.00	0.3619	0.1646	0.3619	0.1580	1.7052
a_1	0.90	0.9002	0.0015	0.0002	0.0000	0.0654
a_2	0.90	0.8996	0.0017	-0.0004	0.0000	0.0655

Influence of outlying observations

One of our main goals in this study is to show the need for robust models to deal with the presence of outliers in the data. Considering the same the parameter values and setup as above and two cases for perturbation, thus eight data sets of size 100 were generated from the WGCR, CWGCR, PHGCR and CPHGCR models with depended spatial fragilities.

We selected cases 18 and 80 for perturbation. To create influential observation in the data set, we choose one or two of these selected cases and perturbed the response variable as follows $\widetilde{t_{kL}} = t_{kL} + 10S_L$ and $\widetilde{t_{kR}} = t_{kR} + 10S_L$, for $k = 1$ and 18, where S_L is the standard deviations of the t_{ijL}'s. Note that using this kind of perturbation, the interval of observed interval time of perturbation candidate observation is not charged. Here, we considere four setups in the study. Setup A: original dataset, without outliers; Setup B: data with outlier 18; Setup C: data with outlier 80; and Setup D: data with outliers 18 and 80. The MCMC computations were made similar to those in the last section and further to monitor the convergence of the Gibbs samples we used the Geweke's convergence diagnostic proposed por Geweke (1992).

Tables 3.3, 3.4, 3.5 and 3.6 reports posterior mean, standard deviation (SD), bias and mean square error (MSE) of the parameters of WGCR, CWGCR, PHGCR and CPHGCR models, respectively. For WGCR model, Table 3.3 shows that the estimative of parameter Λ_{11} creasing in the perturbation cases when prior 1 is used. On the other way, considering prior 2 for the parameters, the estimative of all parameters of cases B, C and D are very closed the case A, which means the parameters are not sensitive to perturbations. It also can be observed on the Table 3.4. For PHGCR model, Table 3.5 shows that parameter Λ_{11} is litter sensitive to perturbations. The estimative of Λ_{11} decreasing in the perturbation cases when considering prior 1 or prior 2 for the parameters and it is more sensitive using prior 1 then prior 2. For CPHGCR model, considering prior 1 Λ_{11} is litter sensitive in cases C and D and Λ_{12} is sensitive in case B; considering prior 2 Λ_{11} is litter sensitive in cases B and C and Λ_{12} is sensitive in cases B and D. This results can be observed on Table 3.6.

For each simulated data set the four divergence measures (d_{KL}, d_J, d_{L_1}, d_{χ^2}) of the perturbed cases and DIC values for the proposed cure rate models were calculated and reported in Table 3.7. We can see that all measures providing larger ψ-divergence measures when compared to the non-perturbed setup (setup A) and the difference between the measures of perturbed case and non-perturbed case is more clearly for PH Geometric cure rate models than Weibull Geometric cure rate models. Furthermore, we can observed that the values of the measures from the cure models wheatear considering the prior 1 or prior 2 for the parameters are similarly. To show better the results, we ploted the ψ-divergence measure from the fitted models. The Figures 3.1 to 3.32 show

the divergence measures before the perturbation (setup A), the model indicate the absence outline observations, and after perturbation observations (setups B, C and D).

For WGCR model with prior 1, we note that the observation 18 cannot be easy detected by all four divergence measure, and observation 80 just be detected by $J-$distance and χ^2 divergence. It also can be observed for the WGCR model with prior 2, moreover, in this case, observation 80 just be detected by χ^2 divergence. For CWGCR model with prior 1, we note that the observation 18 cannot be easy detected when both observations were perturbed (setup D). The both perturbation observations were detected by χ^2 divergence, other three measure only detected observations 80. For CWGCR model with prior 2, we note that the both perturbation observations did not be detected by KL divergence and L_1 norm distance when both observations were perturbed. The $J-$distance was detected only the observation 80, however the both perturbation observations were detected by χ^2 divergence. All perturbation observations selected can be detected by all four divergence measure for PHGCR and CPHGCR models with prior 1 or 2. We also note that the χ^2 divergence is a little bit sensitive for CPHGCR model with both priors, indeed a non-perturbed observation was detected in setup A in Figure 3.31 and 3.32.

Table 3.3: Simulation results of the perturbed cases for WGCR model

Setup	Perturbed case	Prior 1					Prior 2				
		Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	-1.482	0.278	0.018	0.000	b_0	-1.341	0.277	0.159	0.025
		b_1	-0.859	0.259	-0.359	0.129	b_1	-0.554	0.248	-0.054	0.003
		β	-0.036	0.193	0.114	0.013	β	-0.105	0.197	0.045	0.002
		α	0.228	0.068	-0.072	0.005	α	0.112	0.072	-0.188	0.035
		Λ_{11}	4.172	2.533	0.172	0.030	Λ_{11}	4.052	2.486	0.052	0.003
		Λ_{22}	4.152	2.693	0.152	0.023	Λ_{22}	4.012	2.636	0.012	0.000
		Λ_{12}	-0.502	1.945	-0.502	0.252	Λ_{12}	-0.540	1.905	-0.540	0.291
		a	0.902	0.064	0.002	0.000	a_1	0.899	0.068	-0.001	0.000
							a_2	0.899	0.067	-0.001	0.000
B	{18}	b_0	-1.526	0.263	-0.026	0.001	b_0	-1.572	0.275	-0.072	0.005
		b_1	-0.439	0.253	0.061	0.004	b_1	-0.622	0.260	-0.122	0.015
		β	-0.155	0.189	-0.005	0.000	β	-0.099	0.190	0.051	0.003
		α	0.231	0.061	-0.069	0.005	α	0.332	0.073	0.032	0.001
		Λ_{11}	4.057	2.433	0.057	0.003	Λ_{11}	4.056	2.494	0.056	0.003
		Λ_{22}	3.947	2.630	-0.053	0.003	Λ_{22}	3.913	2.561	-0.087	0.008
		Λ_{12}	-0.497	1.904	-0.497	0.247	Λ_{12}	-0.406	1.991	-0.406	0.164
		a	0.900	0.065	0.000	0.000	a_1	0.899	0.065	-0.001	0.000
							a_2	0.900	0.065	0.000	0.000
C	{80}	b_0	-1.510	0.275	-0.010	0.000	b_0	-1.524	0.263	-0.024	0.001
		b_1	-0.586	0.257	-0.086	0.007	b_1	-0.679	0.264	-0.179	0.032
		β	-0.139	0.194	0.011	0.000	β	-0.083	0.193	0.067	0.005
		α	0.255	0.082	-0.045	0.002	α	0.235	0.078	-0.065	0.004
		Λ_{11}	3.832	2.406	-0.168	0.028	Λ_{11}	4.018	2.386	0.018	0.000
		Λ_{22}	3.535	2.423	-0.465	0.216	Λ_{22}	3.919	2.649	-0.081	0.007
		Λ_{12}	-0.115	1.835	-0.115	0.013	Λ_{12}	-0.302	1.969	-0.302	0.091
		a	0.900	0.063	0.000	0.000	a_1	0.899	0.064	-0.001	0.000
							a_2	0.900	0.068	0.000	0.000
D	{18,80}	b_0	-1.460	0.259	0.040	0.002	b_0	-1.599	0.266	-0.099	0.010
		b_1	-0.348	0.248	0.152	0.023	b_1	-0.316	0.246	0.184	0.034
		β	-0.203	0.189	-0.053	0.003	β	-0.208	0.191	-0.058	0.003
		α	0.187	0.059	-0.113	0.013	α	0.253	0.061	-0.047	0.002
		Λ_{11}	4.205	2.569	0.205	0.042	Λ_{11}	4.259	2.541	0.259	0.067
		Λ_{22}	4.086	2.634	0.086	0.007	Λ_{22}	4.002	2.549	0.002	0.000
		Λ_{12}	-0.515	1.971	-0.515	0.265	Λ_{12}	-0.588	1.946	-0.588	0.346
		a	0.900	0.064	0.000	0.000	a_2	0.901	0.065	0.001	0.000
							a_2	0.898	0.068	-0.002	0.000

Table 3.4: Simulation results of the perturbed cases for CWGCR model

Setup	Perturbed case	Prior 1					Prior 2				
		Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	-1.475	0.269	0.025	0.001	b_0	-1.499	0.272	0.001	0.000
		b_1	-0.498	0.255	0.002	0.000	b_1	-0.446	0.260	0.054	0.003
		β	-0.066	0.136	0.084	0.007	β	-0.005	0.139	0.145	0.021
		α	0.323	0.052	0.023	0.001	α	0.384	0.060	0.084	0.007
		Λ_{11}	3.880	2.207	-0.120	0.014	Λ_{11}	3.905	2.214	-0.095	0.009
		Λ_{22}	4.080	2.738	0.080	0.006	Λ_{22}	4.040	2.636	0.040	0.002
		Λ_{12}	0.308	1.884	0.308	0.095	Λ_{12}	0.243	1.908	0.243	0.059
		a	0.902	0.065	0.002	0.000	a_1	0.903	0.064	0.003	0.000
							a_2	0.901	0.064	0.001	0.000
B	{18}	b_0	-1.548	0.254	-0.048	0.002	b_0	-1.445	0.282	0.055	0.003
		b_1	-0.276	0.265	0.224	0.050	b_1	-0.400	0.283	0.100	0.010
		β	-0.019	0.141	0.131	0.017	β	-0.141	0.143	0.009	0.000
		α	0.389	0.056	0.089	0.008	α	0.306	0.050	0.006	0.000
		Λ_{11}	4.611	2.517	0.611	0.373	Λ_{11}	4.125	2.333	0.125	0.016
		Λ_{22}	4.127	2.613	0.127	0.016	Λ_{22}	3.623	2.456	-0.377	0.142
		Λ_{12}	0.451	1.919	0.451	0.204	Λ_{12}	0.354	1.840	0.354	0.125
		a	0.899	0.065	-0.001	0.000	a_1	0.900	0.066	0.000	0.000
							a_2	0.901	0.063	0.001	0.000
C	{80}	b_0	-1.523	0.265	-0.023	0.001	b_0	-1.529	0.279	-0.029	0.001
		b_1	-0.523	0.270	-0.023	0.001	b_1	-0.468	0.272	0.032	0.001
		β	-0.059	0.135	0.091	0.008	β	-0.127	0.138	0.023	0.001
		α	0.375	0.056	0.075	0.006	α	0.388	0.060	0.088	0.008
		Λ_{11}	4.515	2.545	0.515	0.265	Λ_{11}	3.727	2.112	-0.273	0.075
		Λ_{22}	4.139	2.709	0.139	0.019	Λ_{22}	3.765	2.471	-0.235	0.055
		Λ_{12}	0.416	1.980	0.416	0.173	Λ_{12}	0.158	1.837	0.158	0.025
		a	0.901	0.064	0.001	0.000	a_1	0.900	0.065	0.000	0.000
							a_2	0.902	0.063	0.002	0.000
D	{18,80}	b_0	-1.514	0.268	-0.014	0.000	b_0	-1.582	0.271	-0.082	0.007
		b_1	-0.458	0.268	0.042	0.002	b_1	-0.485	0.268	0.015	0.000
		β	-0.032	0.151	0.118	0.014	β	-0.111	0.141	0.039	0.002
		α	0.357	0.049	0.057	0.003	α	0.329	0.048	0.029	0.001
		Λ_{11}	4.750	2.618	0.750	0.563	Λ_{11}	3.964	2.240	-0.036	0.001
		Λ_{22}	4.053	2.627	0.053	0.003	Λ_{22}	4.127	2.698	0.127	0.016
		Λ_{12}	0.306	1.930	0.306	0.094	Λ_{12}	0.417	1.889	0.417	0.174
		a	0.900	0.065	0.000	0.000	a_1	0.900	0.066	0.000	0.000
							a_2	0.902	0.063	0.002	0.000

Table 3.5: Simulation results of the perturbed cases for PHGCR model

Setup	Perturbed case	Prior 1					Prior 2				
		Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	-1.456	0.270	0.044	0.002	b_0	-1.670	0.267	-0.170	0.029
		b_1	-0.431	0.249	0.069	0.005	b_1	-0.611	0.241	-0.111	0.012
		β	-0.176	0.182	-0.026	0.001	β	-0.156	0.177	-0.006	0.000
		α	1.192	0.137	0.192	0.037	α	0.898	0.190	-0.102	0.010
		Λ_{11}	4.275	2.500	0.275	0.075	Λ_{11}	3.840	2.348	-0.160	0.026
		Λ_{22}	3.927	2.568	-0.073	0.005	Λ_{22}	4.058	2.688	0.058	0.003
		Λ_{12}	-0.363	1.951	-0.363	0.132	Λ_{12}	-0.637	1.917	-0.637	0.405
		a	0.898	0.070	-0.002	0.000	a_1	0.901	0.064	0.001	0.000
							a_2	0.902	0.064	0.002	0.000
B	{18}	b_0	-1.348	0.278	0.152	0.023	b_0	-1.387	0.250	0.113	0.013
		b_1	-0.436	0.233	0.064	0.004	b_1	-0.337	0.251	0.163	0.026
		β	-0.207	0.184	-0.057	0.003	β	-0.193	0.191	-0.043	0.002
		α	0.912	0.095	-0.088	0.008	α	0.797	0.205	-0.203	0.041
		Λ_{11}	3.137	2.021	-0.863	0.745	Λ_{11}	3.240	2.181	-0.760	0.577
		Λ_{22}	4.050	2.661	0.050	0.003	Λ_{22}	4.302	2.828	0.302	0.091
		Λ_{12}	-0.300	1.776	-0.300	0.090	Λ_{12}	-0.161	1.867	-0.161	0.026
		a	0.903	0.065	0.003	0.000	a_1	0.899	0.066	-0.001	0.000
							a_2	0.899	0.065	-0.001	0.000
C	{80}	b_0	-1.494	0.253	0.006	0.000	b_0	-1.336	0.273	0.164	0.027
		b_1	-0.643	0.248	-0.143	0.020	b_1	-0.193	0.247	0.307	0.094
		β	-0.183	0.186	-0.033	0.001	β	-0.148	0.193	0.002	0.000
		α	1.112	0.112	0.112	0.012	α	0.866	0.198	-0.134	0.018
		Λ_{11}	3.101	2.049	-0.899	0.809	Λ_{11}	3.419	2.174	-0.581	0.337
		Λ_{22}	4.221	2.637	0.221	0.049	Λ_{22}	3.600	2.468	-0.400	0.160
		Λ_{12}	-0.532	1.831	-0.532	0.283	Λ_{12}	-0.263	1.756	-0.263	0.069
		a	0.900	0.065	0.000	0.000	a_1	0.899	0.067	-0.001	0.000
							a_2	0.900	0.065	0.000	0.000
D	{18,80}	b_0	-1.498	0.276	0.002	0.000	b_0	-1.526	0.277	-0.026	0.001
		b_1	-0.551	0.249	-0.051	0.003	b_1	-0.343	0.243	0.157	0.025
		β	-0.136	0.188	0.014	0.000	β	-0.302	0.185	-0.152	0.023
		α	1.029	0.099	0.029	0.001	α	0.761	0.216	-0.239	0.057
		Λ_{11}	2.821	1.858	-1.179	1.389	Λ_{11}	3.190	2.045	-0.810	0.656
		Λ_{22}	4.276	2.734	0.276	0.076	Λ_{22}	4.160	2.698	0.160	0.026
		Λ_{12}	-0.428	1.791	-0.428	0.183	Λ_{12}	-0.661	1.799	-0.661	0.437
		a	0.905	0.067	0.005	0.000	a_1	0.900	0.064	0.000	0.000
							a_2	0.898	0.067	-0.002	0.000

Table 3.6: Simulation results of the perturbed cases for CPHGCR model

Setup	Perturbed case	Prior 1					Prior 2				
		Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	-1.679	0.268	-0.179	0.032	b_0	-1.840	0.290	-0.340	0.116
		b_1	-0.492	0.271	0.008	0.000	b_1	-0.600	0.263	-0.100	0.010
		β	-0.226	0.132	-0.076	0.006	β	-0.177	0.119	-0.027	0.001
		α	0.909	0.177	-0.091	0.008	α	0.920	0.180	-0.080	0.006
		Λ_{11}	4.258	2.370	0.258	0.067	Λ_{11}	4.351	2.449	0.351	0.123
		Λ_{22}	3.880	2.560	-0.120	0.014	Λ_{22}	2.980	2.239	-1.020	1.040
		Λ_{12}	0.002	1.774	0.002	0.000	Λ_{12}	-0.054	1.594	-0.054	0.003
		a	0.902	0.066	0.002	0.000	a_1	0.901	0.066	0.001	0.000
							a_2	0.902	0.065	0.002	0.000
B	{18}	b_0	-1.678	0.267	-0.178	0.032	b_0	-1.646	0.263	-0.146	0.021
		b_1	-0.599	0.252	-0.099	0.010	b_1	-0.369	0.266	0.131	0.017
		β	-0.384	0.124	-0.234	0.055	β	-0.230	0.138	-0.080	0.006
		α	0.918	0.179	-0.082	0.007	α	0.789	0.194	-0.211	0.045
		Λ_{11}	4.124	2.242	0.124	0.015	Λ_{11}	3.180	1.937	-0.820	0.672
		Λ_{22}	3.743	2.440	-0.257	0.066	Λ_{22}	3.798	2.552	-0.202	0.041
		Λ_{12}	0.259	1.742	0.259	0.067	Λ_{12}	0.420	1.745	0.420	0.176
		a	0.901	0.064	0.001	0.000	a_1	0.899	0.068	-0.001	0.000
							a_2	0.900	0.066	0.000	0.000
C	{80}	b_0	-1.634	0.262	-0.134	0.018	b_0	-1.716	0.271	-0.216	0.047
		b_1	-0.521	0.260	-0.021	0.000	b_1	-0.401	0.263	0.099	0.010
		β	-0.231	0.123	-0.081	0.007	β	-0.040	0.132	0.110	0.012
		α	0.903	0.180	-0.097	0.009	α	0.850	0.184	-0.150	0.023
		Λ_{11}	3.972	2.205	-0.028	0.001	Λ_{11}	3.464	1.970	-0.536	0.287
		Λ_{22}	3.928	2.642	-0.072	0.005	Λ_{22}	3.779	2.531	-0.221	0.049
		Λ_{12}	0.130	1.793	0.130	0.017	Λ_{12}	-0.040	1.623	-0.040	0.002
		a	0.901	0.066	0.001	0.000	a_1	0.900	0.066	0.000	0.000
							a_2	0.902	0.063	0.002	0.000
D	{18,80}	b_0	-1.643	0.275	-0.143	0.021	b_0	-1.515	0.254	-0.015	0.000
		b_1	-0.444	0.270	0.056	0.003	b_1	-0.305	0.246	0.195	0.038
		β	-0.041	0.134	0.109	0.012	β	-0.161	0.137	-0.011	0.000
		α	0.770	0.216	-0.230	0.053	α	0.806	0.192	-0.194	0.037
		Λ_{11}	2.547	1.553	-1.453	2.111	Λ_{11}	3.708	2.101	-0.292	0.085
		Λ_{22}	3.958	2.529	-0.042	0.002	Λ_{22}	4.303	2.689	0.303	0.092
		Λ_{12}	-0.038	1.707	-0.038	0.001	Λ_{12}	0.350	1.774	0.350	0.122
		a	0.901	0.065	0.001	0.000	a_1	0.899	0.067	-0.001	0.000
							a_2	0.901	0.064	0.001	0.000

Table 3.7: Divergence measures of the perturbed cases and DIC values for the simulated data sets.

Model	Prior	Setup	Case number	d_{KL}	d_J	d_{L1}	d_{χ^2}	DIC
WGCR	1	A	18	0.006	0.011	0.042	0.012	142.754
			80	0.030	0.060	0.097	0.067	
		B	18	0.080	0.171	0.159	0.238	164.476
		C	80	0.246	0.602	0.277	2.128	153.082
		D	18	0.069	0.143	0.147	0.181	185.709
			80	0.282	0.677	0.304	1.742	
	2	A	18	0.007	0.014	0.046	0.014	140.186
			80	0.033	0.067	0.102	0.075	
		B	18	0.036	0.075	0.106	0.084	164.446
		C	80	0.294	0.940	0.288	14.544	149.760
		D	18	0.062	0.131	0.138	0.176	184.934
			80	0.120	0.269	0.190	0.498	
CWGCR	1	A	18	0.038	0.079	0.108	0.096	401.704
			80	0.001	0.003	0.019	0.003	
		B	18	0.369	0.890	0.347	2.987	431.013
		C	80	0.605	1.381	0.440	4.905	436.193
		D	18	0.243	0.555	0.283	1.227	447.962
			80	0.723	1.729	0.481	8.789	
	2	A	18	0.066	0.137	0.145	0.161	404.322
			80	0.045	0.092	0.119	0.103	
		B	18	0.677	1.720	0.486	8.654	408.839
		C	80	0.288	0.672	0.305	1.844	416.505
		D	18	0.131	0.291	0.206	0.501	425.875
			80	0.256	0.624	0.288	2.408	
PHGCR	1	A	18	0.016	0.032	0.071	0.034	228.007
			80	0.036	0.073	0.106	0.081	
		B	18	1.644	4.288	0.693	168.408	263.195
		C	80	1.265	4.761	0.649	584.824	241.465
		D	18	1.750	4.472	0.718	84.141	240.257
			80	0.479	1.248	0.402	8.804	
	2	A	18	0.002	0.004	0.026	0.004	211.243
			80	0.013	0.027	0.065	0.029	
		B	18	1.455	3.501	0.662	35.776	278.356
		C	80	1.438	3.271	0.660	22.489	284.756
		D	18	0.462	1.097	0.390	3.959	290.520
			80	0.460	1.162	0.390	5.777	
CPHGCR	1	A	18	0.018	0.037	0.076	0.040	387.037
			80	0.025	0.050	0.086	0.057	
		B	18	2.727	5.539	0.788	73.127	427.090
		C	80	3.659	7.131	0.845	110.592	452.871
		D	18	2.087	4.366	0.726	34.092	483.157
			80	4.358	9.412	0.900	648.576	
	2	A	18	0.011	0.023	0.059	0.023	423.219
			80	0.025	0.052	0.090	0.057	
		B	18	3.790	7.791	0.870	199.173	446.544
		C	80	3.347	6.929	0.829	164.743	470.571
		D	18	3.880	8.238	0.870	365.590	477.539
			80	3.829	8.759	0.883	595.102	

WGCR model

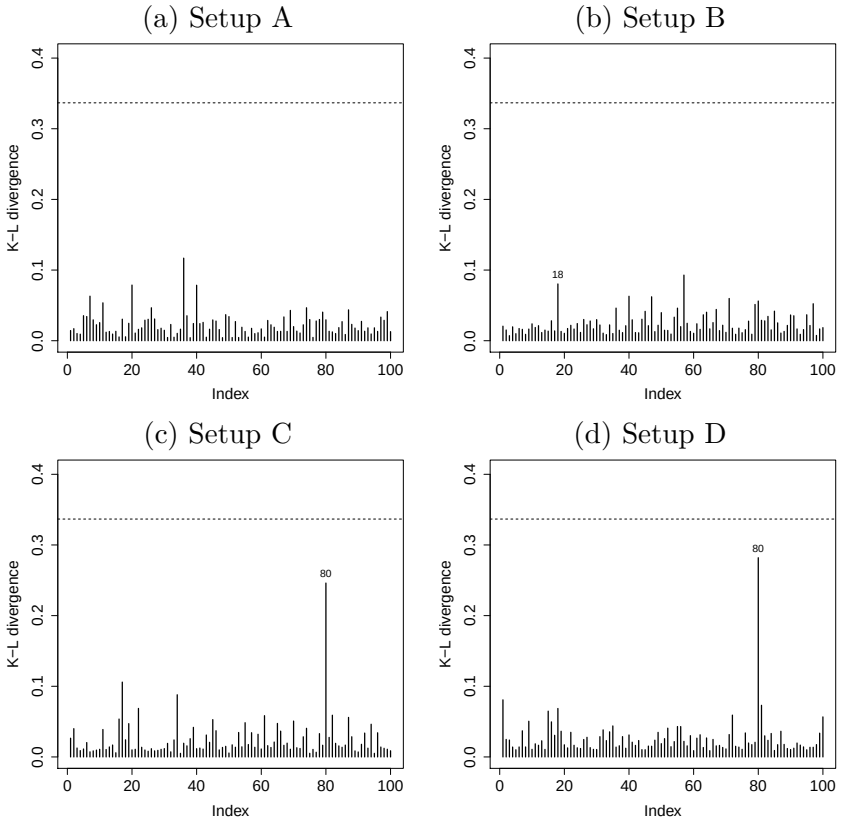

Figure 3.1: Index plots of Kullback-Leibler divergence measure from the fitted of the WGCR model considering prior 1.

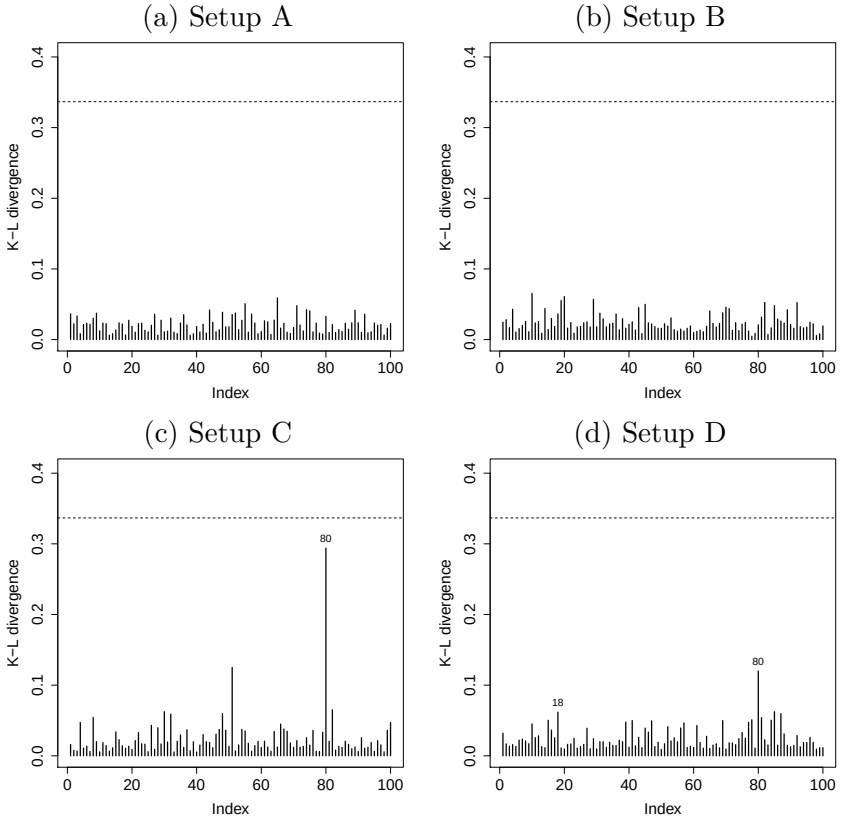

Figure 3.2: Index plots of Kullback-Leibler divergence measure from the fitted WGCR model considering prior 2.

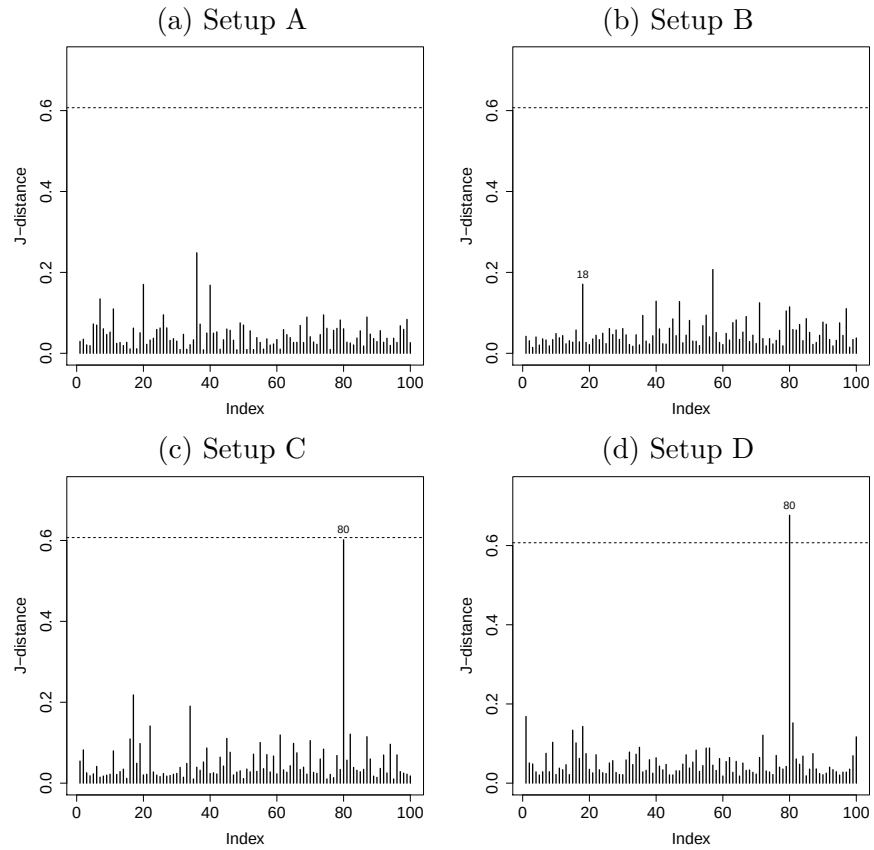

Figure 3.3: Index plots of J-distance from the fitted WGCR model considering prior 1.

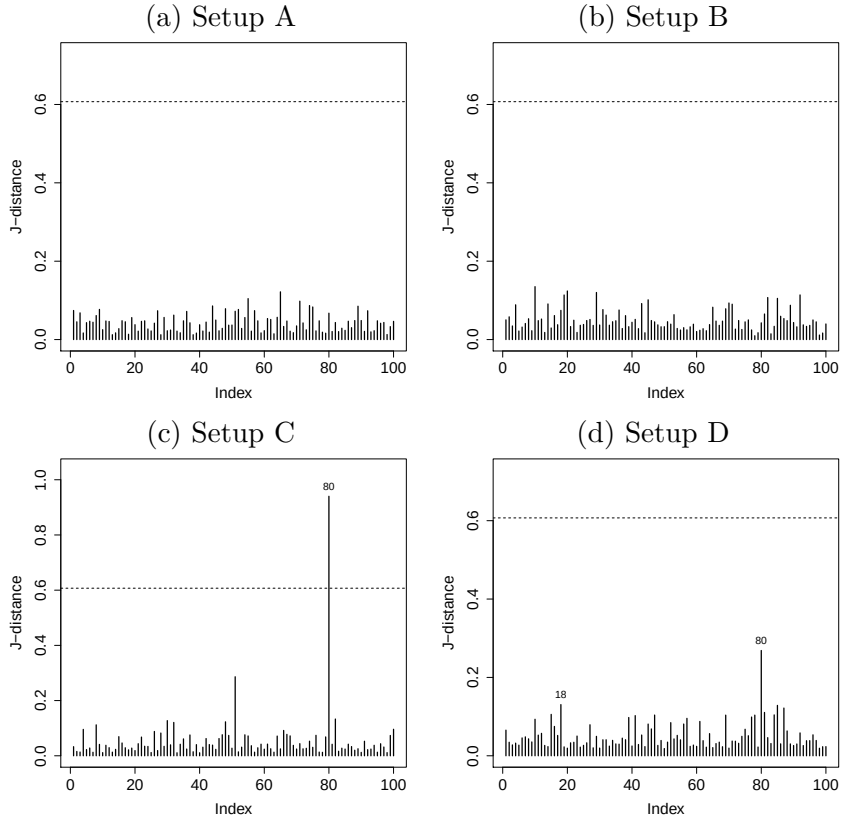

Figure 3.4: Index plots of J-distance from the fitted WGCR model considering prior 2.

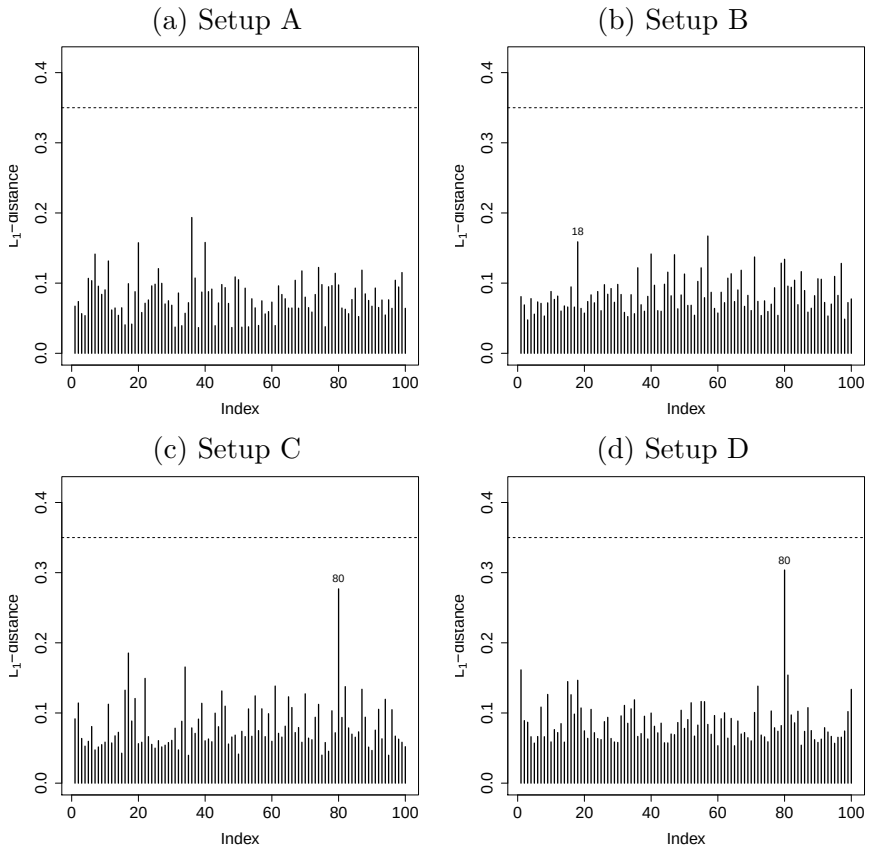

Figure 3.5: Index plots of L_1 norm distance from the fitted WGCR model considering prior 1.

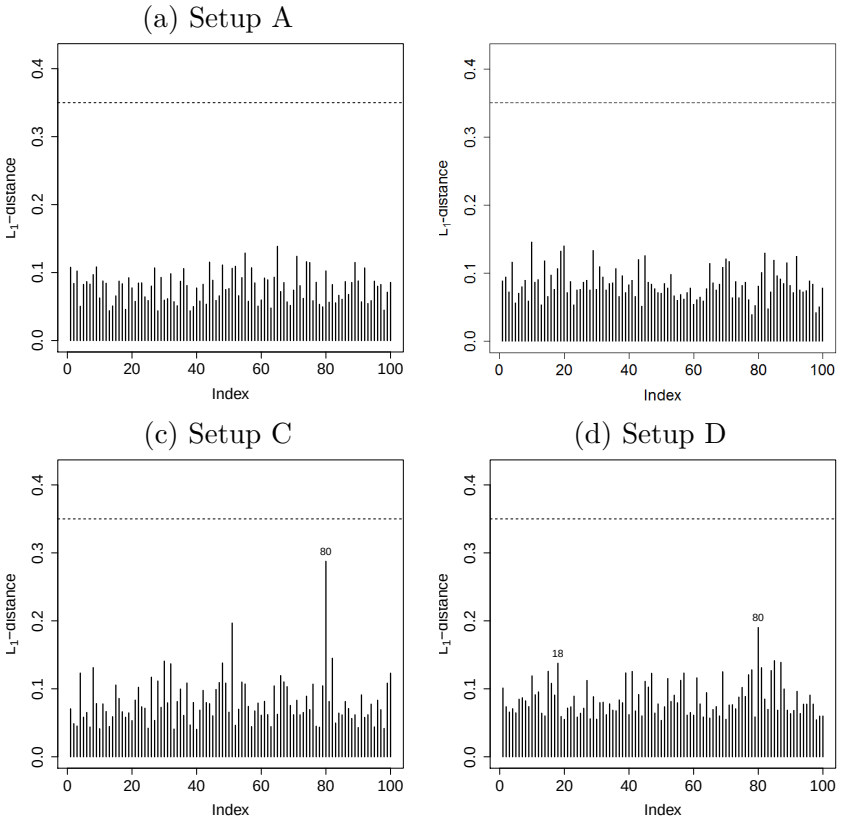

Figure 3.6: Index plots of L_1 norm distance from the fitted WGCR model considering prior 2.

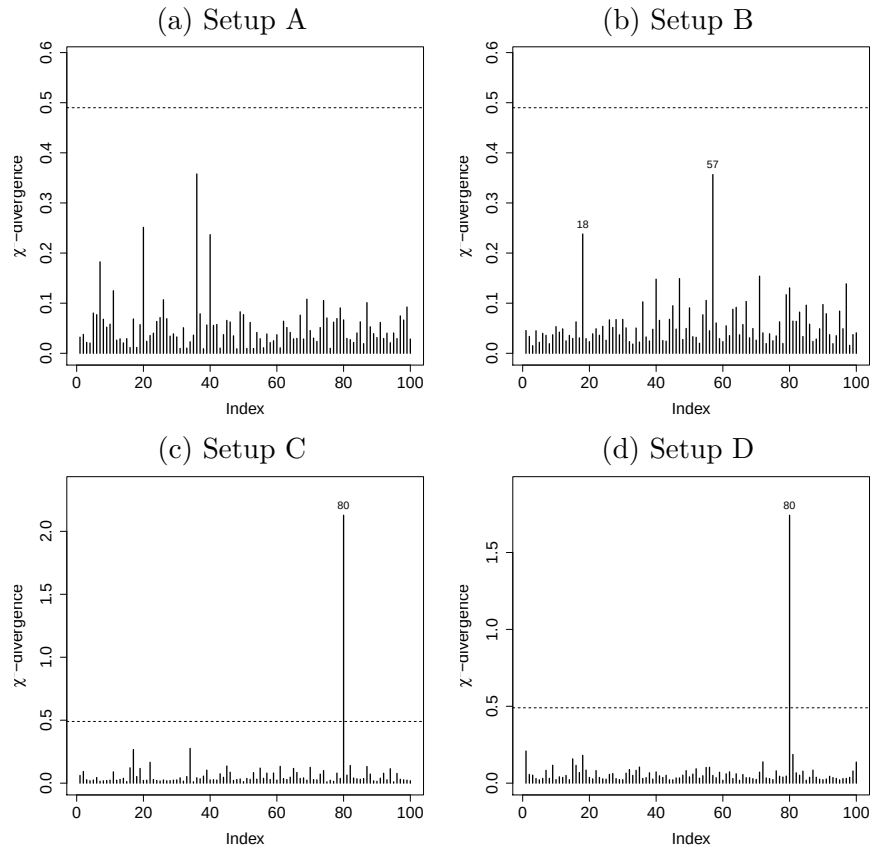

Figure 3.7: Index plots of χ^2-square divergence from the fitted WGCR model considering prior 1.

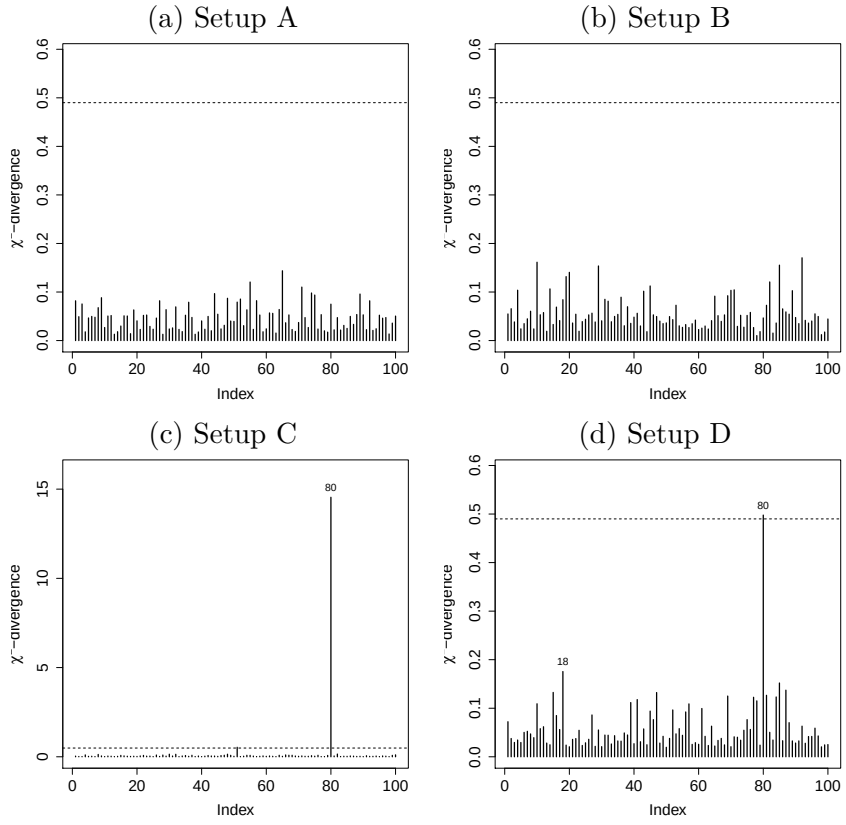

Figure 3.8: Index plots of χ^2-square divergence from the fitted WGCR model considering prior 2.

CWGCR model

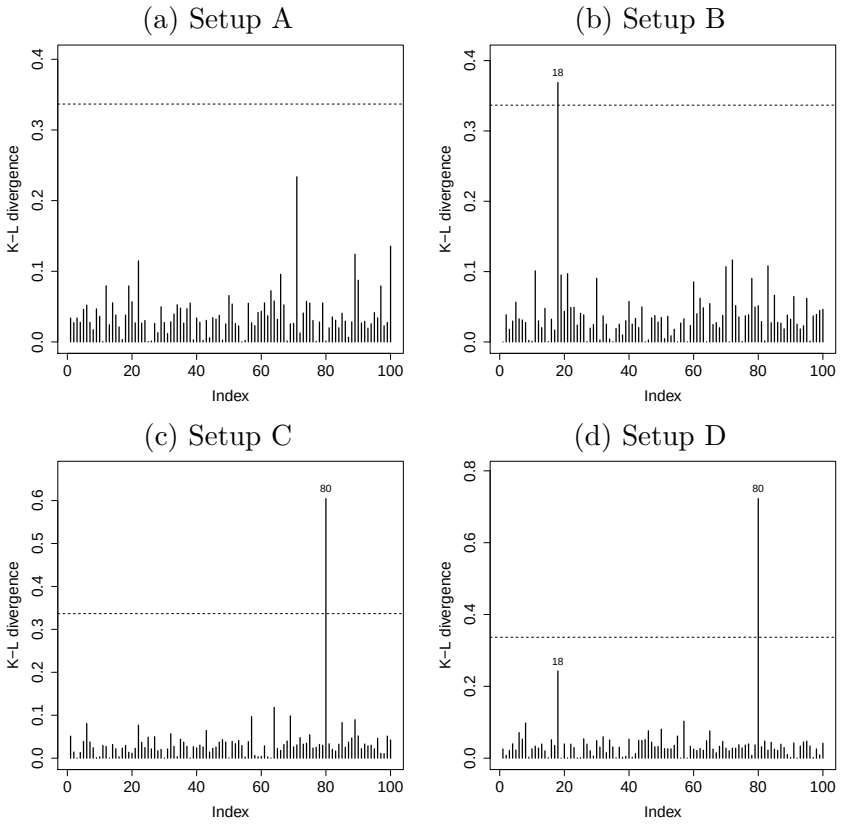

Figure 3.9: Index plots of Kullback-Leibler divergence measure from the fitted CWGCR model considering prior 1.

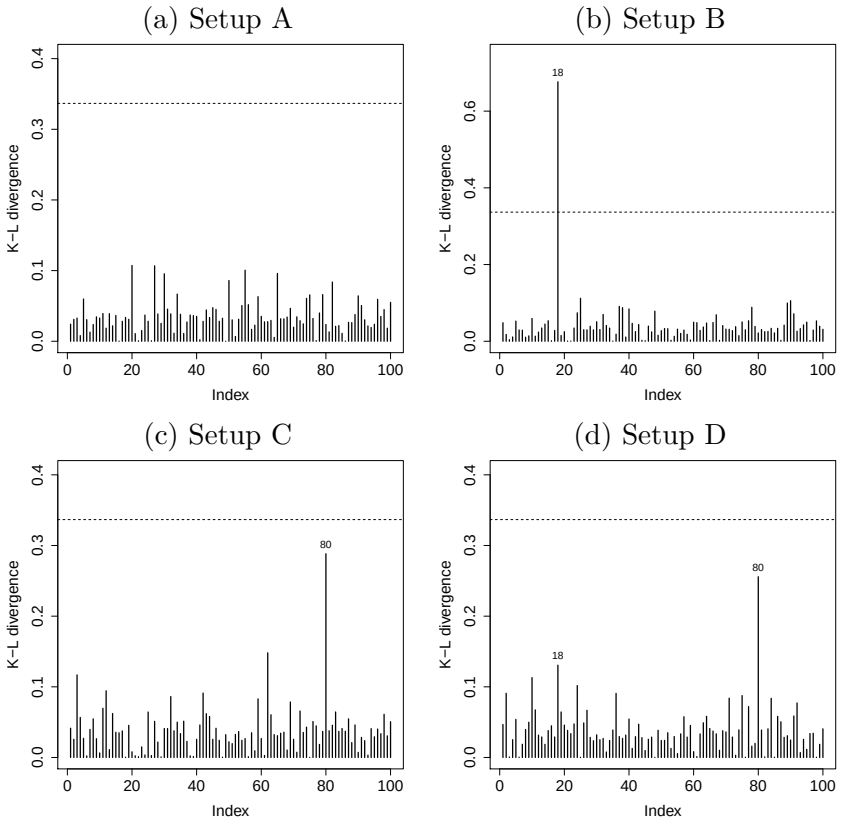

Figure 3.10: Index plots of Kullback-Leibler divergence measure from the fitted CWGCR model considering prior 2.

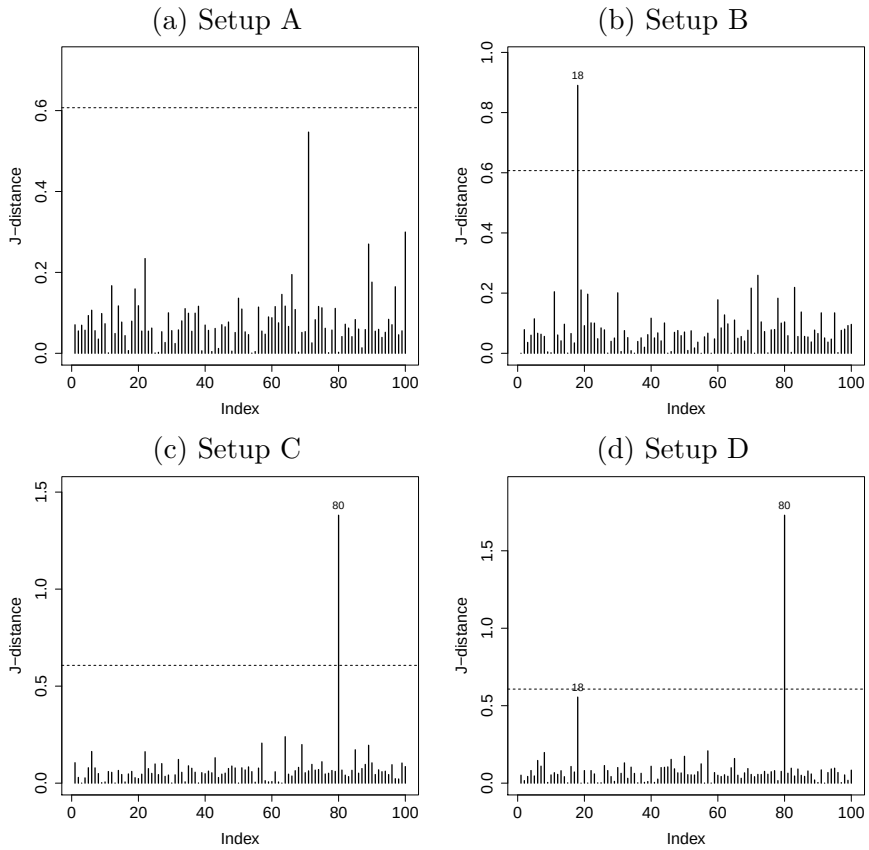

Figure 3.11: Index plots of J-distance from the fitted CWGCR model considering prior 1.

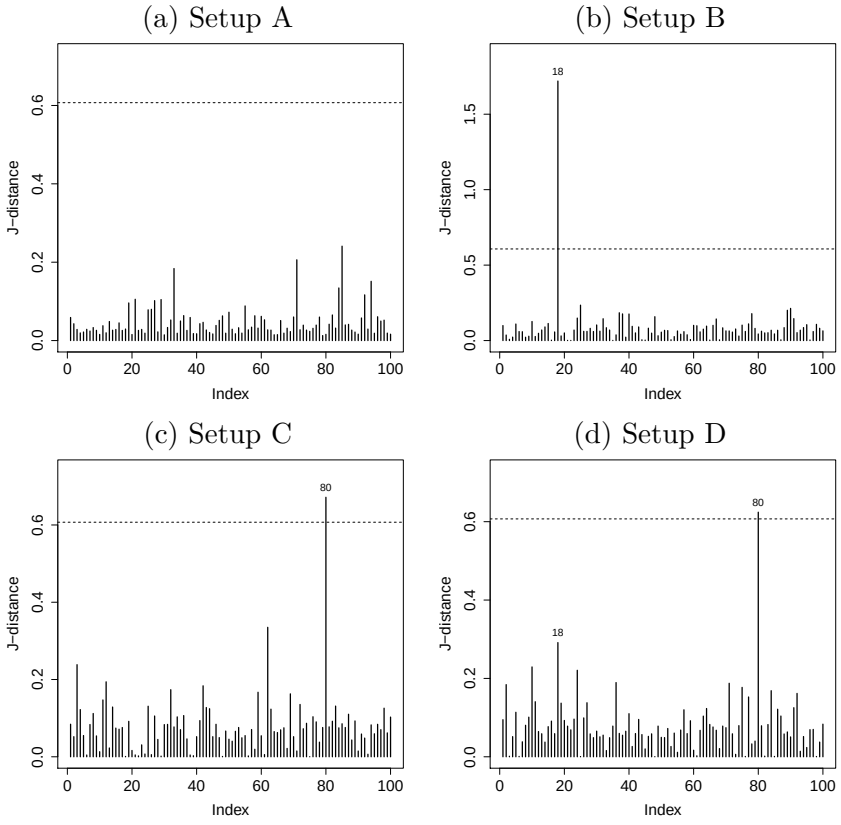

Figure 3.12: Index plots of J-distance from the fitted CWGCR model considering prior 2.

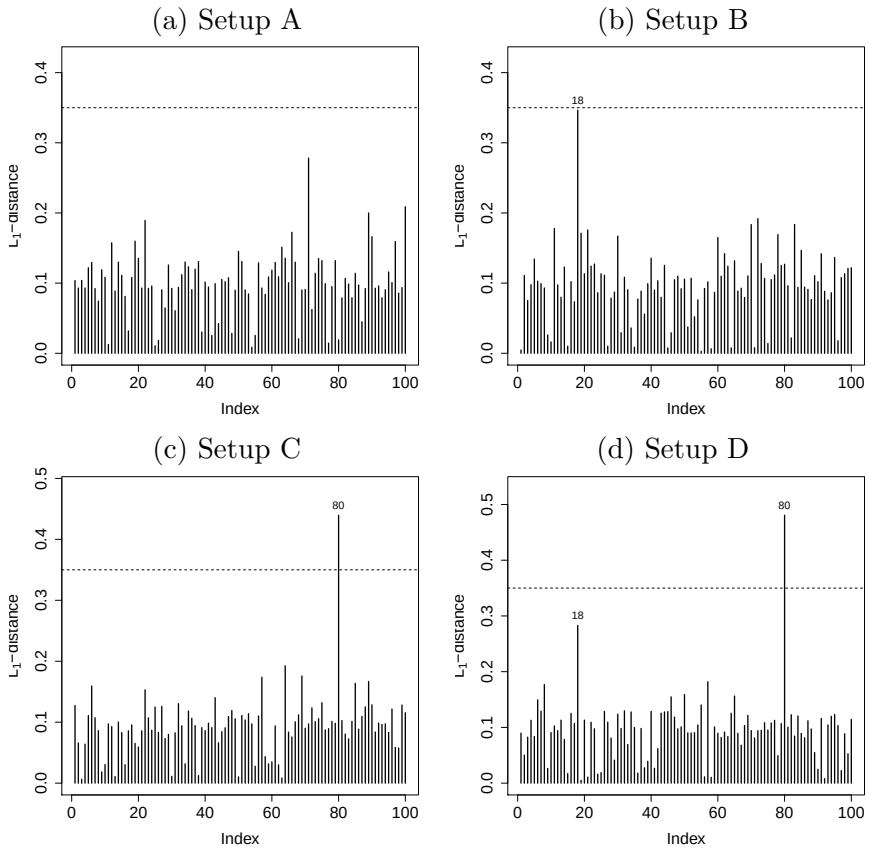

Figure 3.13: Index plots of L_1 norm distance from the fitted CWGCR model considering prior 1.

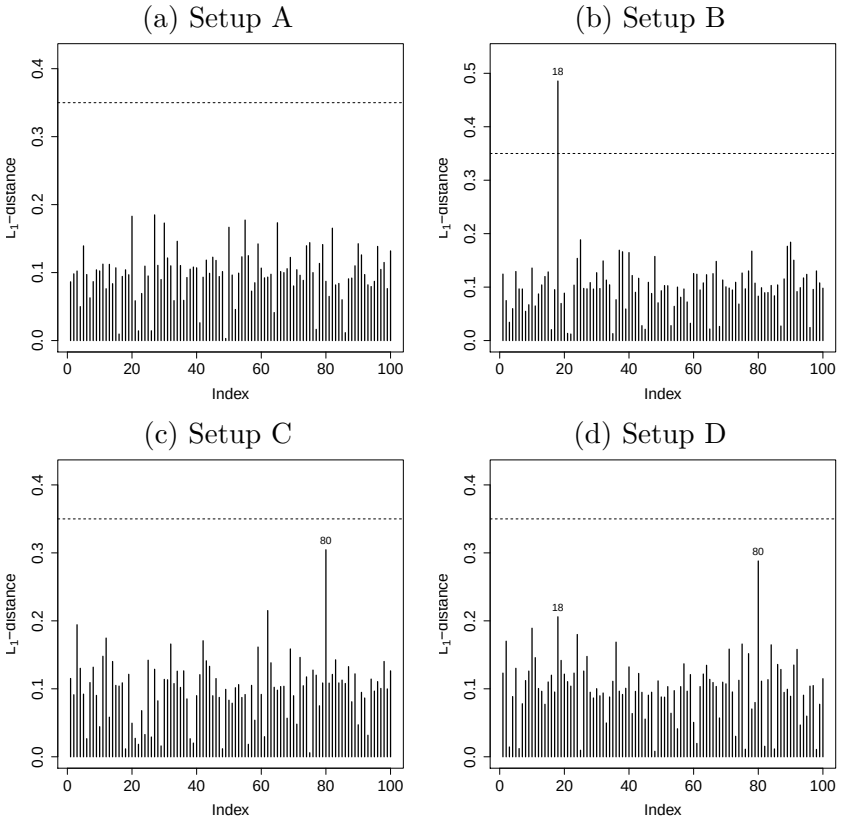

Figure 3.14: Index plots of L_1 norm distance from the fitted CWGCR model considering prior 2.

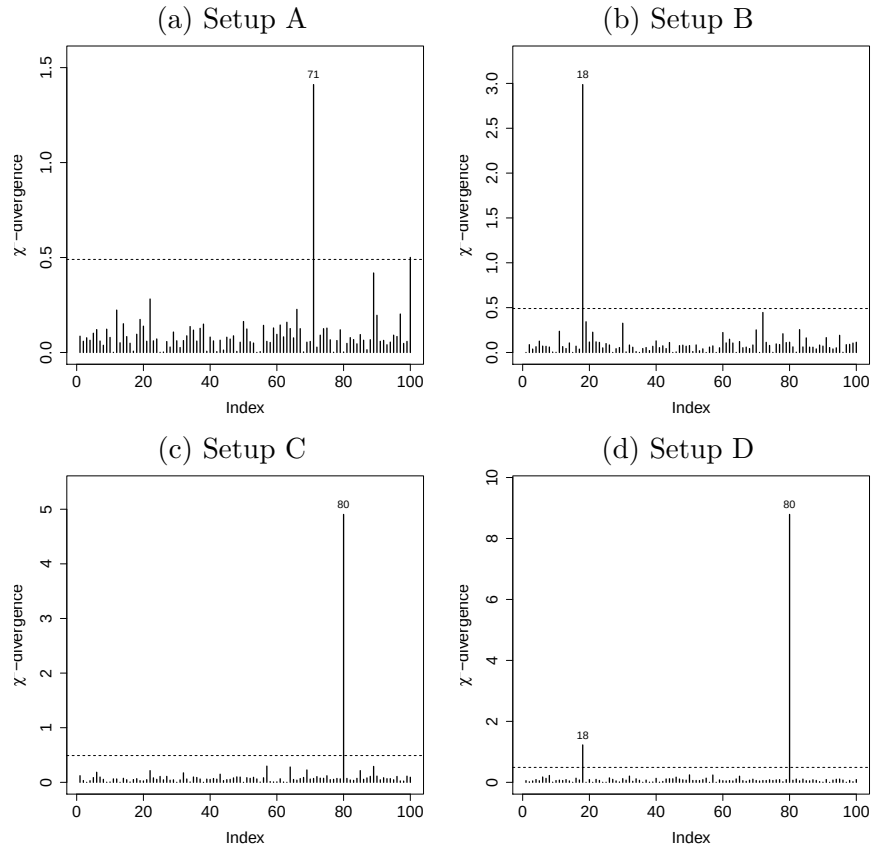

Figure 3.15: Index plots of χ^2-square divergence from the fitted CWGCR model considering prior 1.

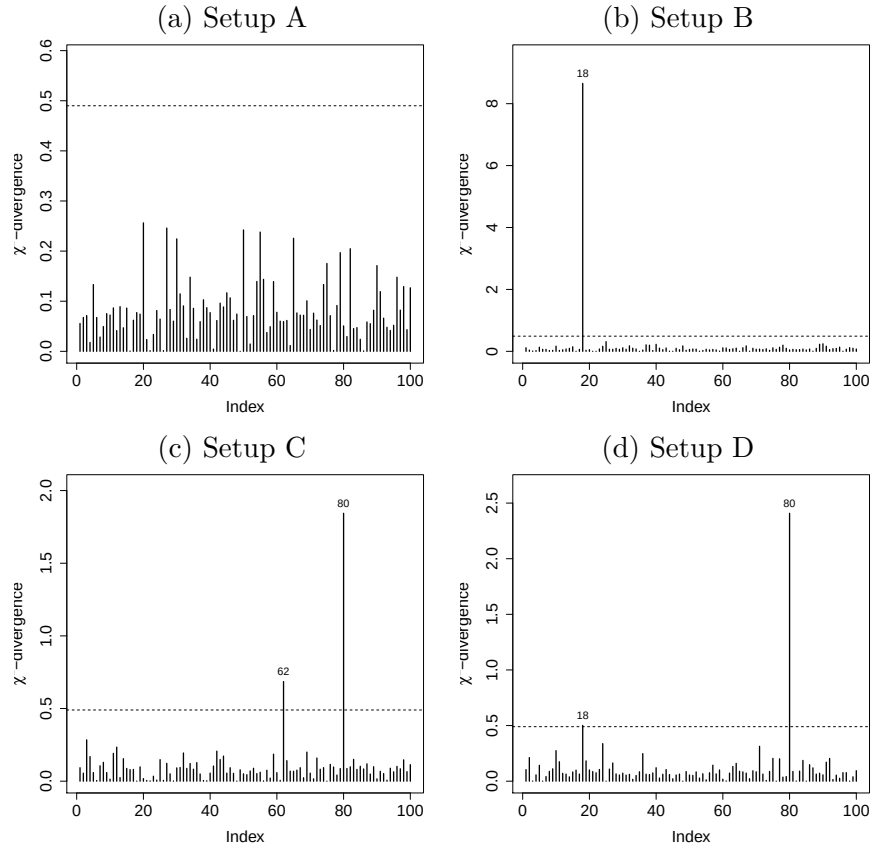

Figure 3.16: Index plots of χ^2-square divergence from the fitted CWGCR model considering prior 2.

PHGCR model

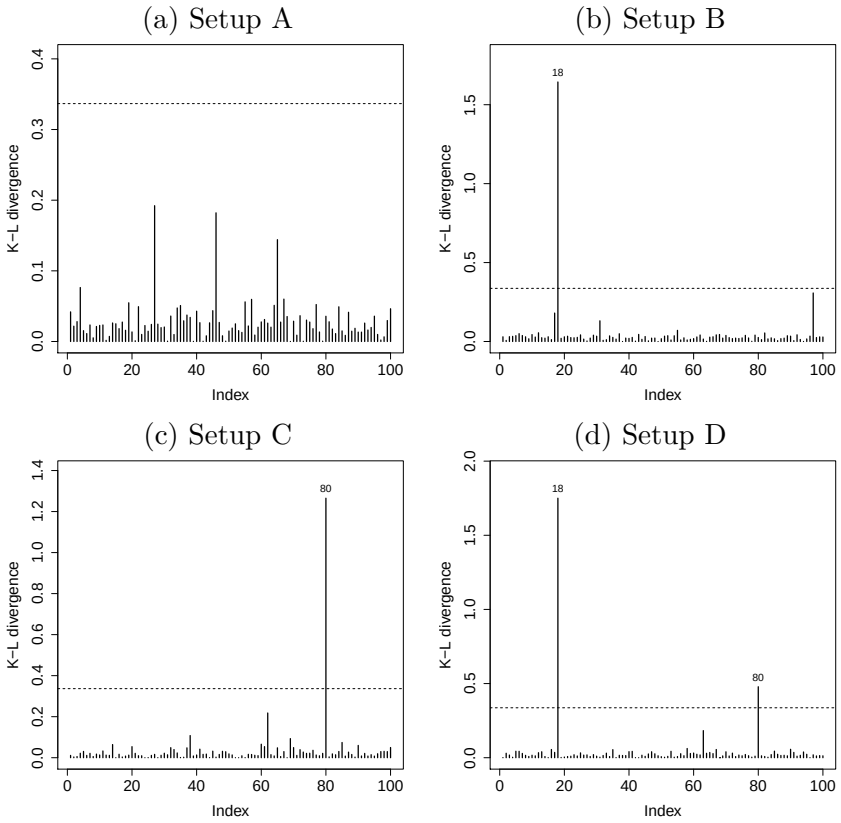

Figure 3.17: Index plots of Kullback-Leibler divergence measure from the fitted PHGCR model considering prior 1.

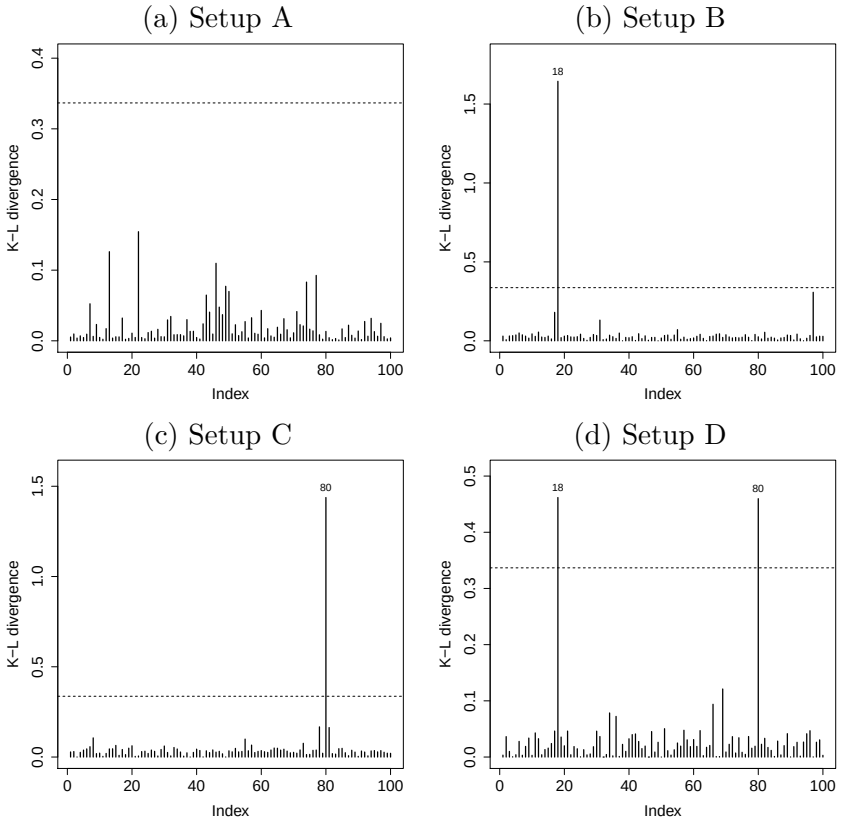

Figure 3.18: Index plots of Kullback-Leibler divergence measure from the fitted PHGCR model considering prior 2.

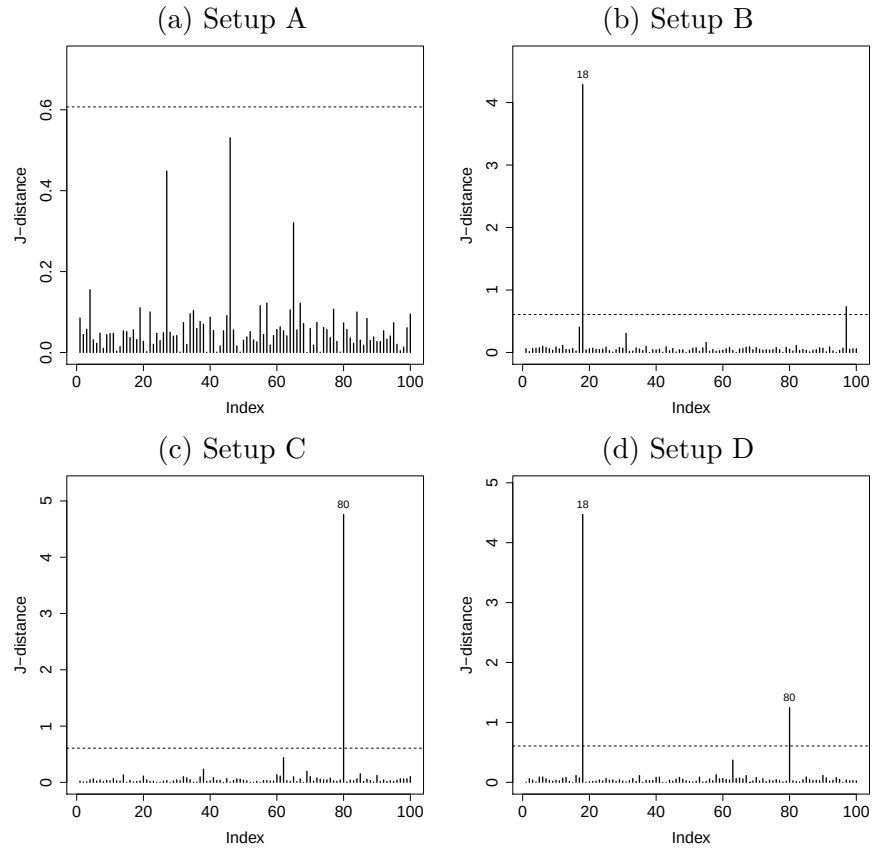

Figure 3.19: Index plots of J-distance from the fitted PHGCR model considering prior 1.

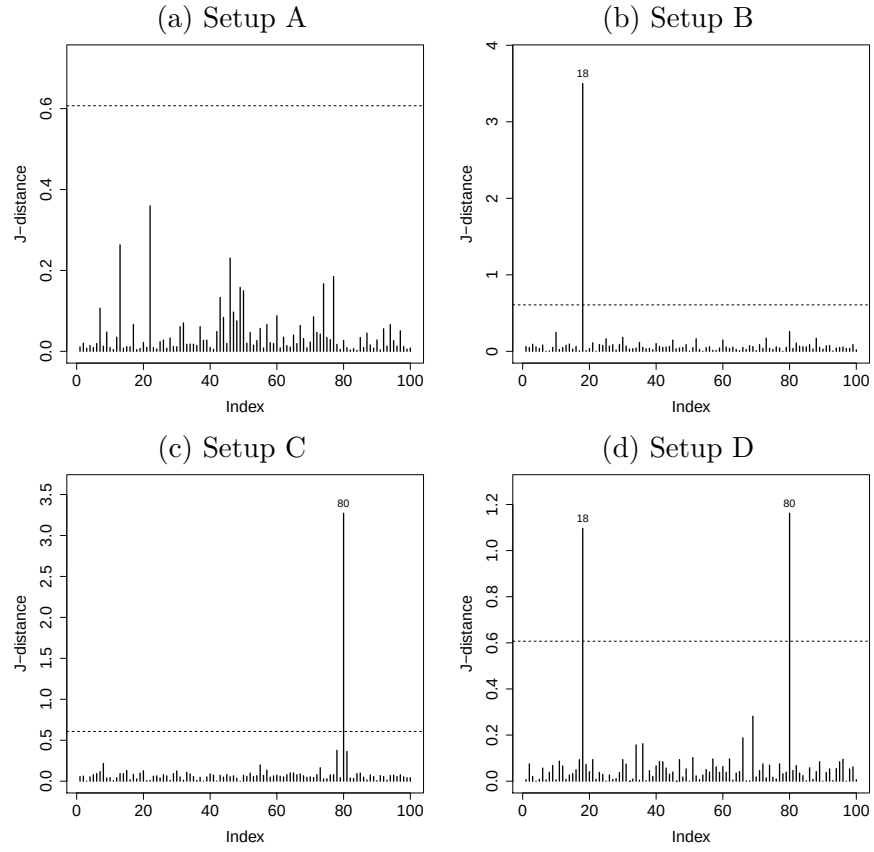

Figure 3.20: Index plots of J-distance from the fitted PHGCR model considering prior 2.

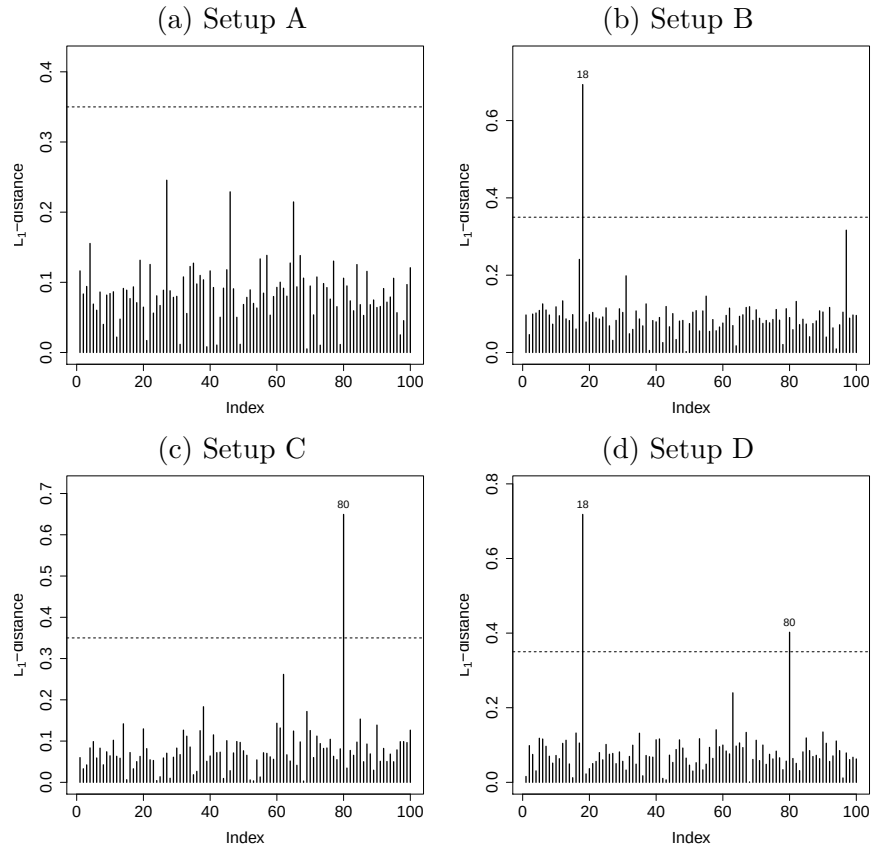

Figure 3.21: Index plots of L_1 norm distance from the fitted PHGCR model considering prior 1.

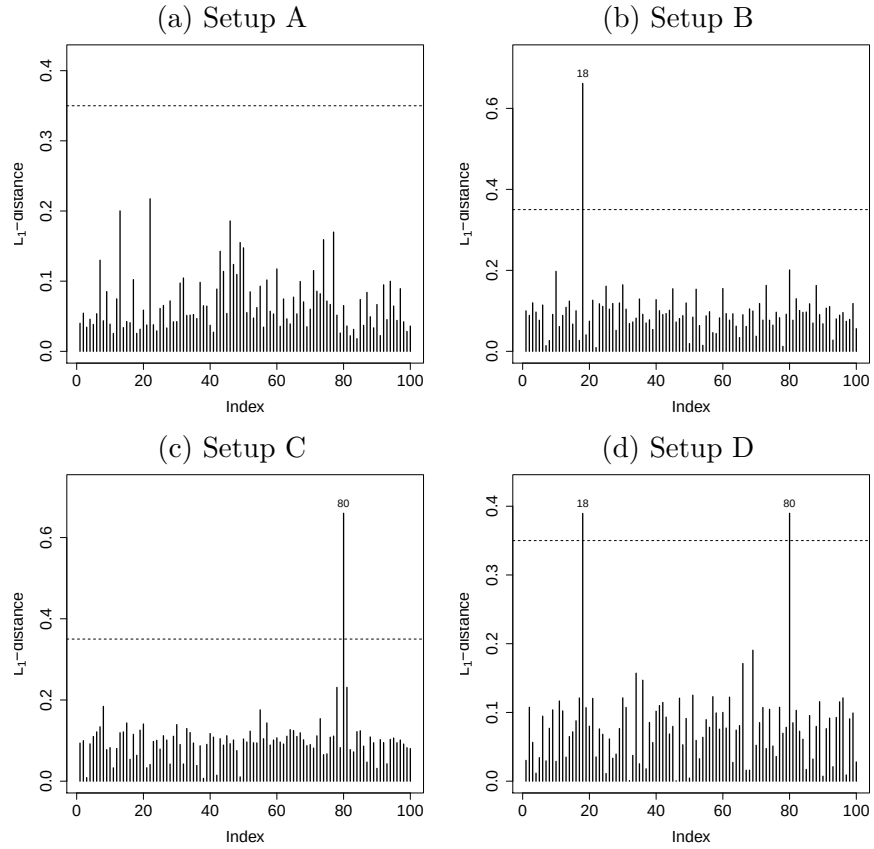

Figure 3.22: Index plots of L_1 norm distance from the fitted PHGCR model considering prior 2.

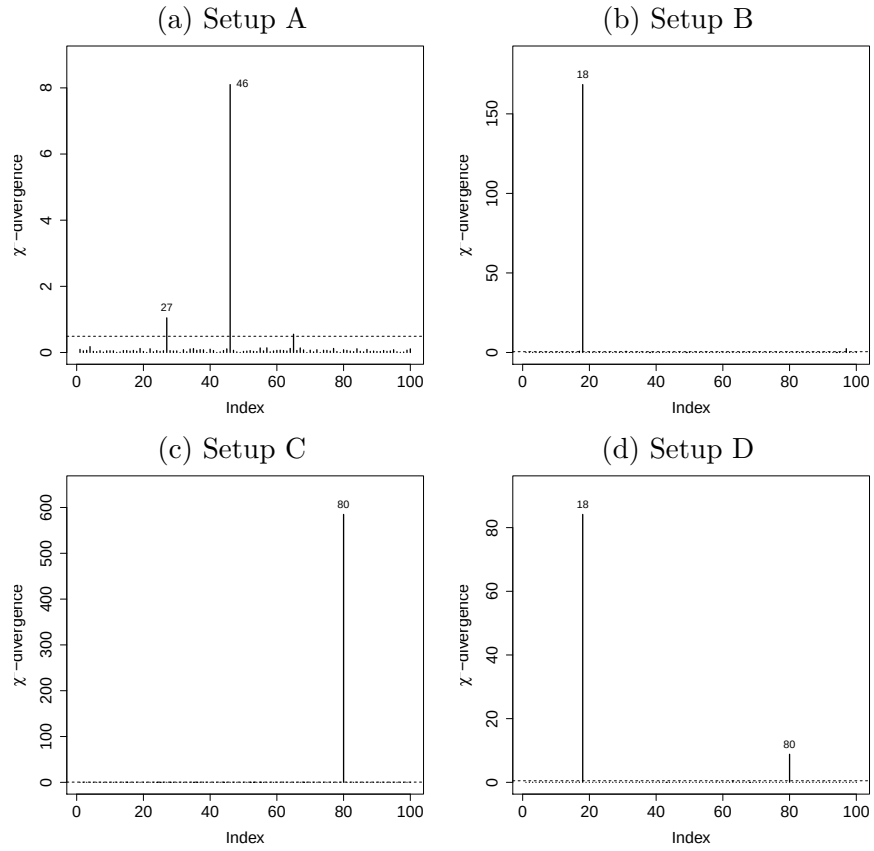

Figure 3.23: Index plots of χ^2-square divergence from the fitted PHGCR model considering prior 1.

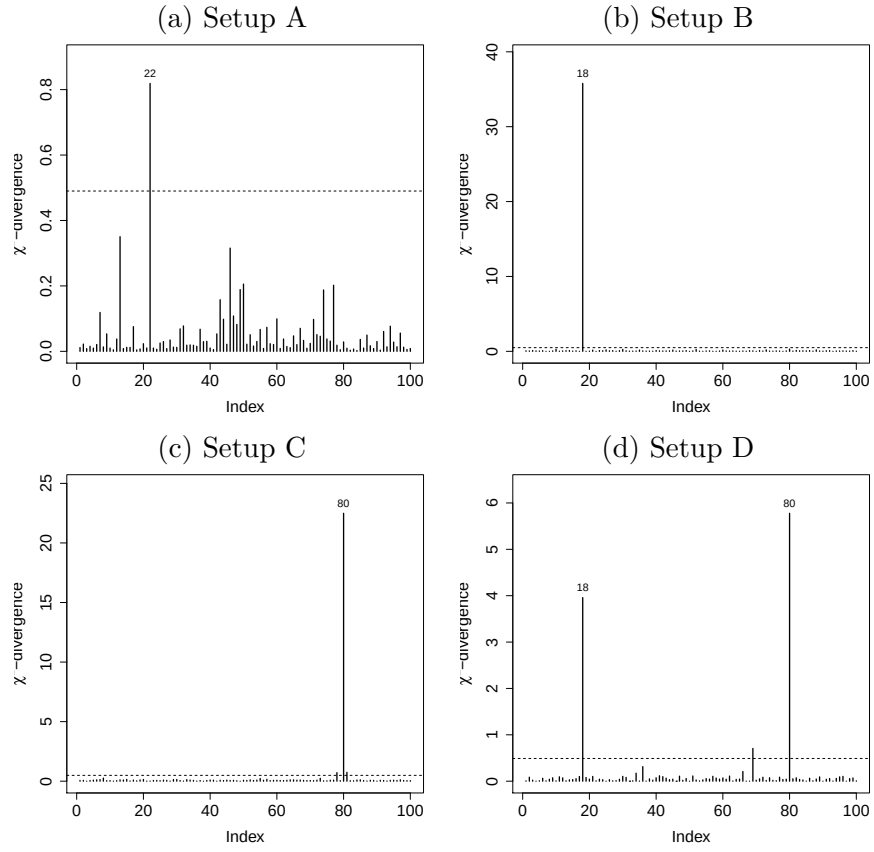

Figure 3.24: Index plots of χ^2-square divergence from the fitted PHGCR model considering prior 2.

CPHGCR model

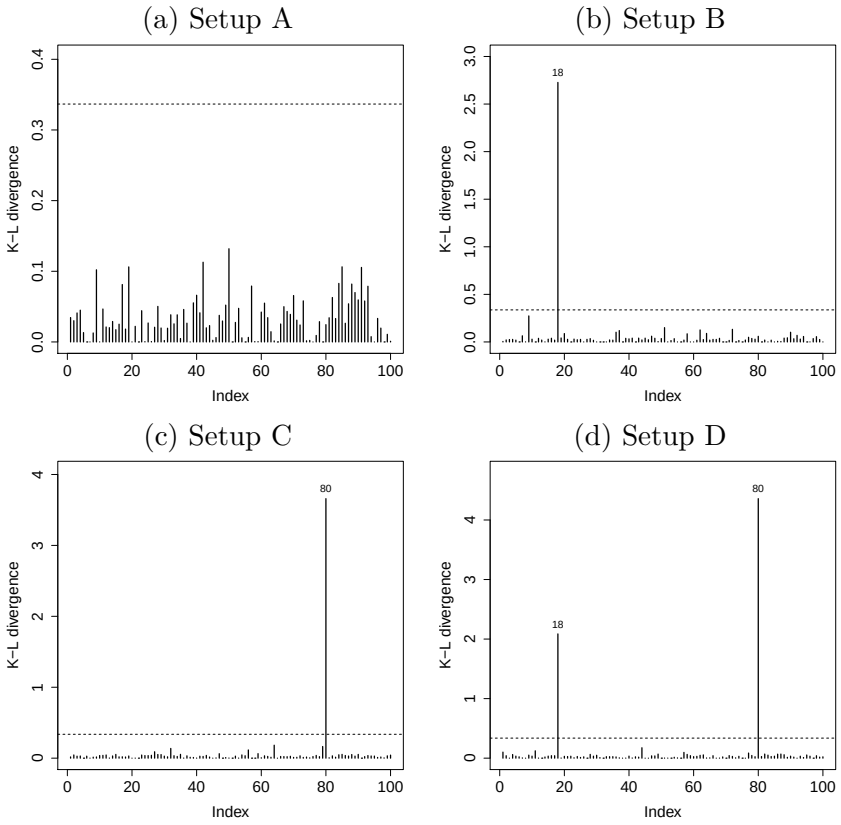

Figure 3.25: Index plots of Kullback-Leibler divergence measure from the fitted CPHGCR model considering prior 1.

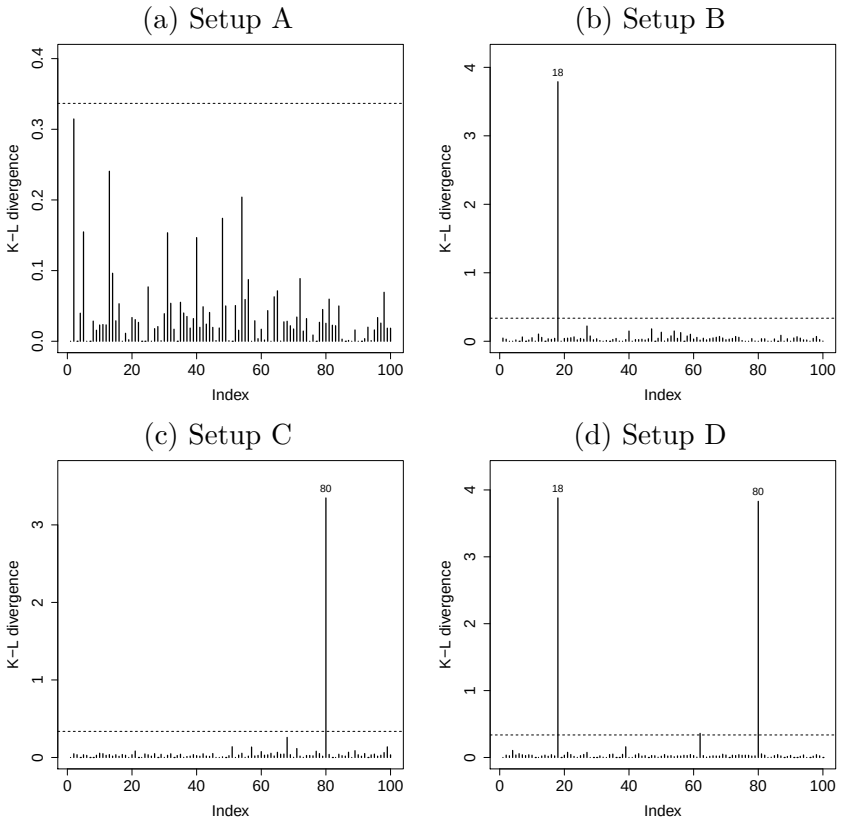

Figure 3.26: Index plots of Kullback-Leibler divergence measure from the fitted CPHGCR model considering prior 2.

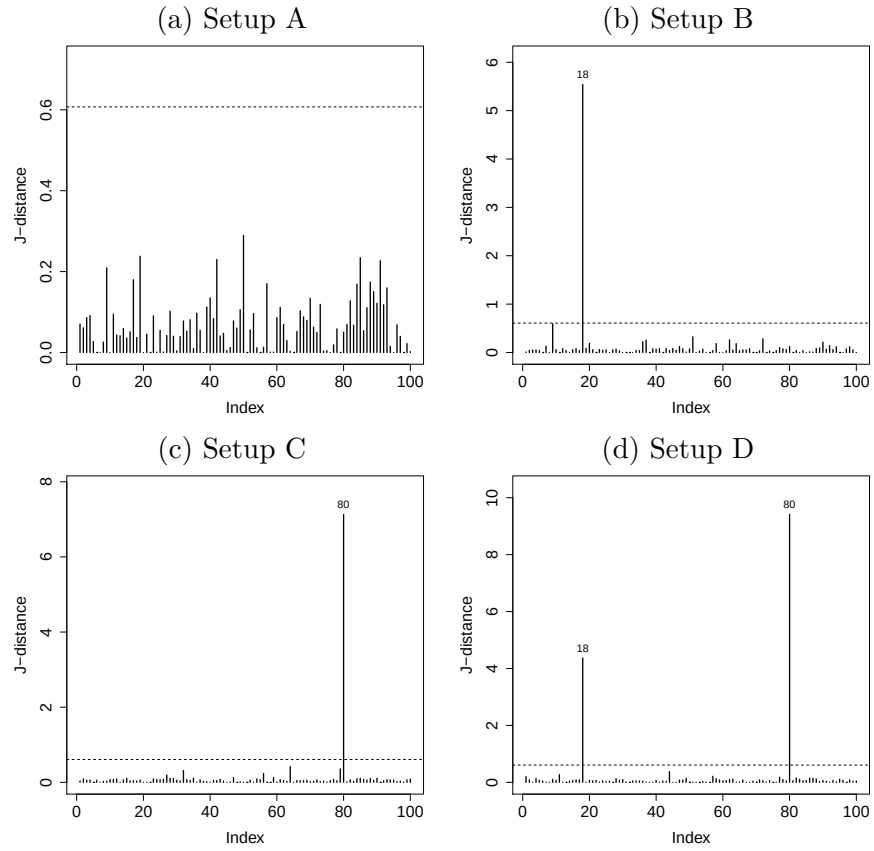

Figure 3.27: Index plots of J-distance from the fitted CPHGCR model considering prior 1.

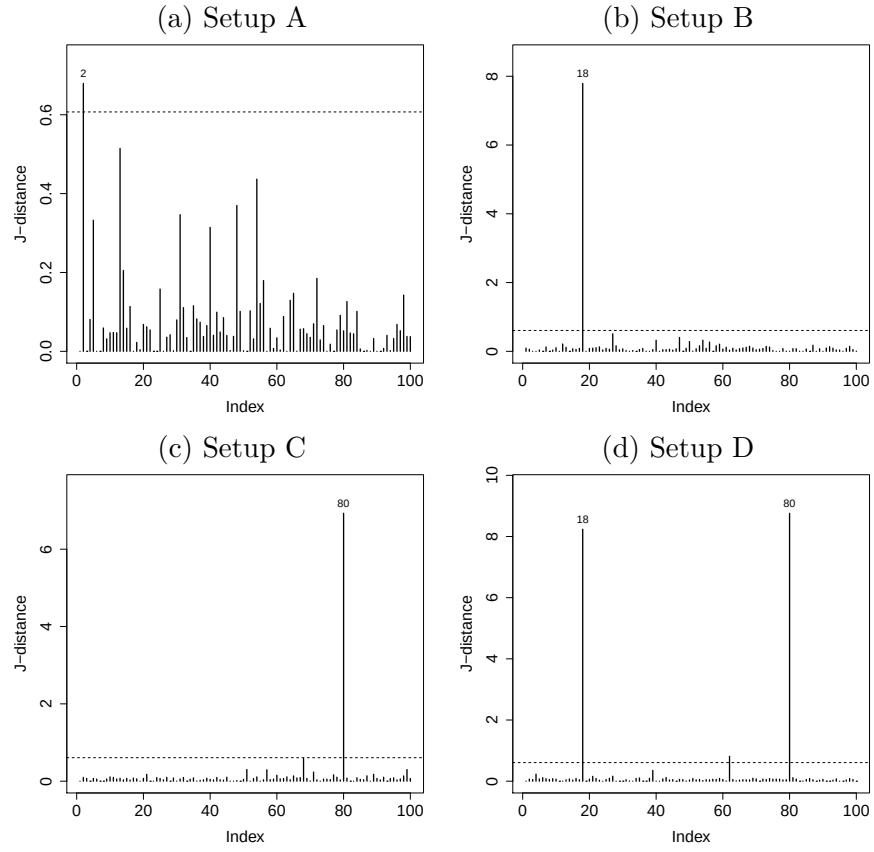

Figure 3.28: Index plots of J-distance from the fitted CPHGCR mode considering prior 2.

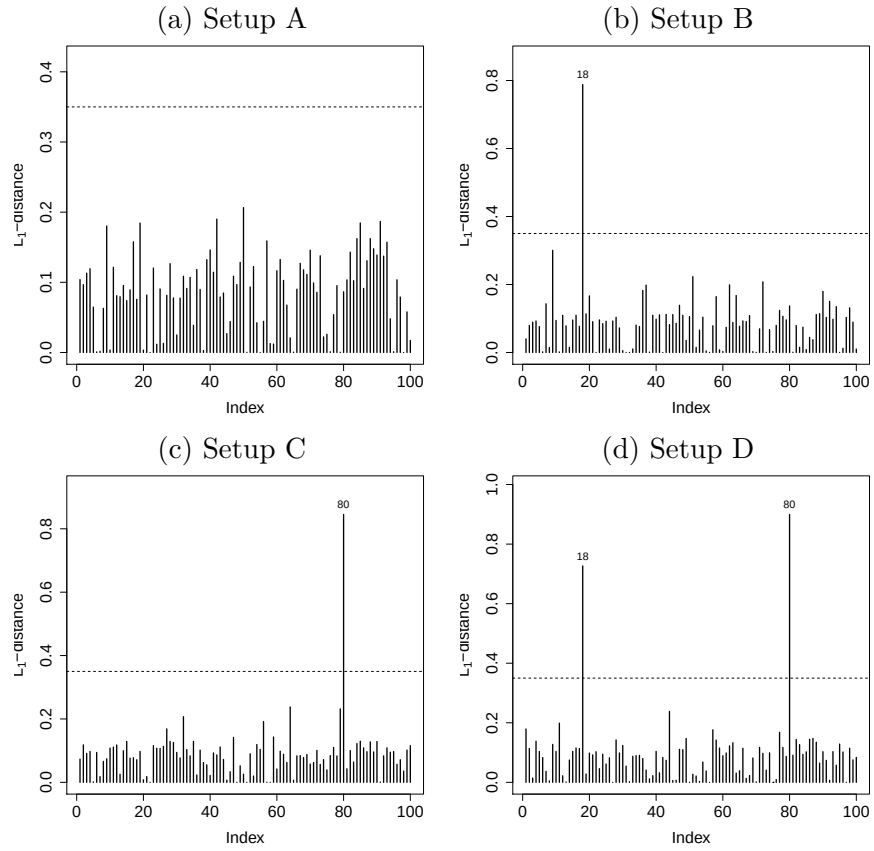

Figure 3.29: Index plots of L_1 norm distance from the fitted CPHGCR model considering prior 1.

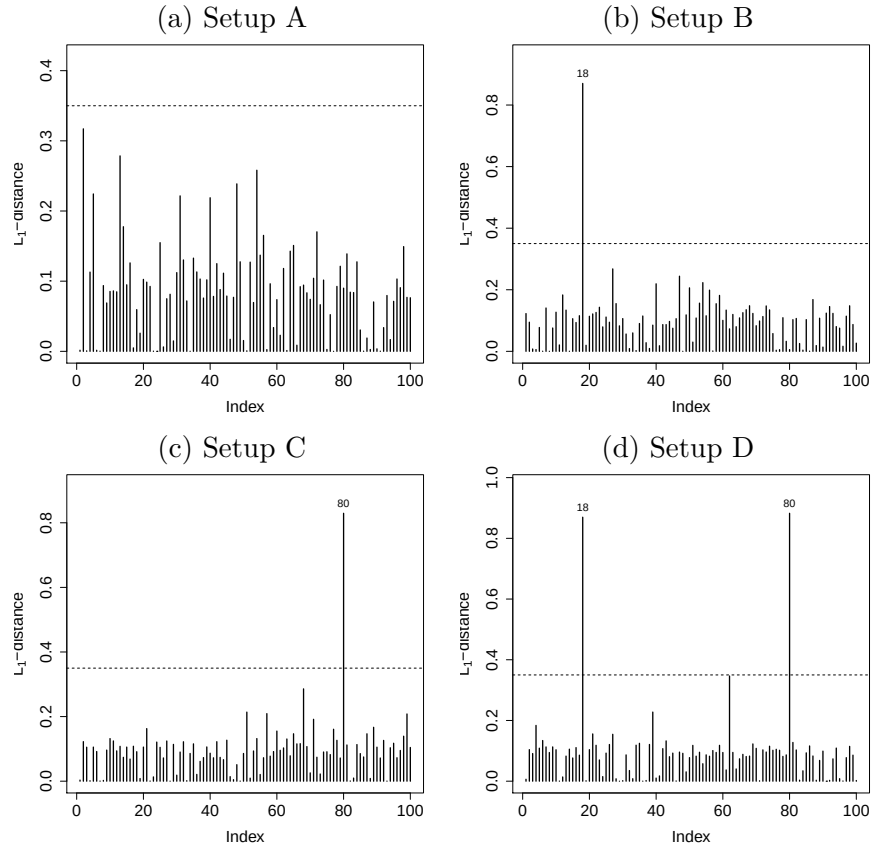

Figure 3.30: Index plots of L_1 norm distance from the fitted CPHGCR model considering prior 2.

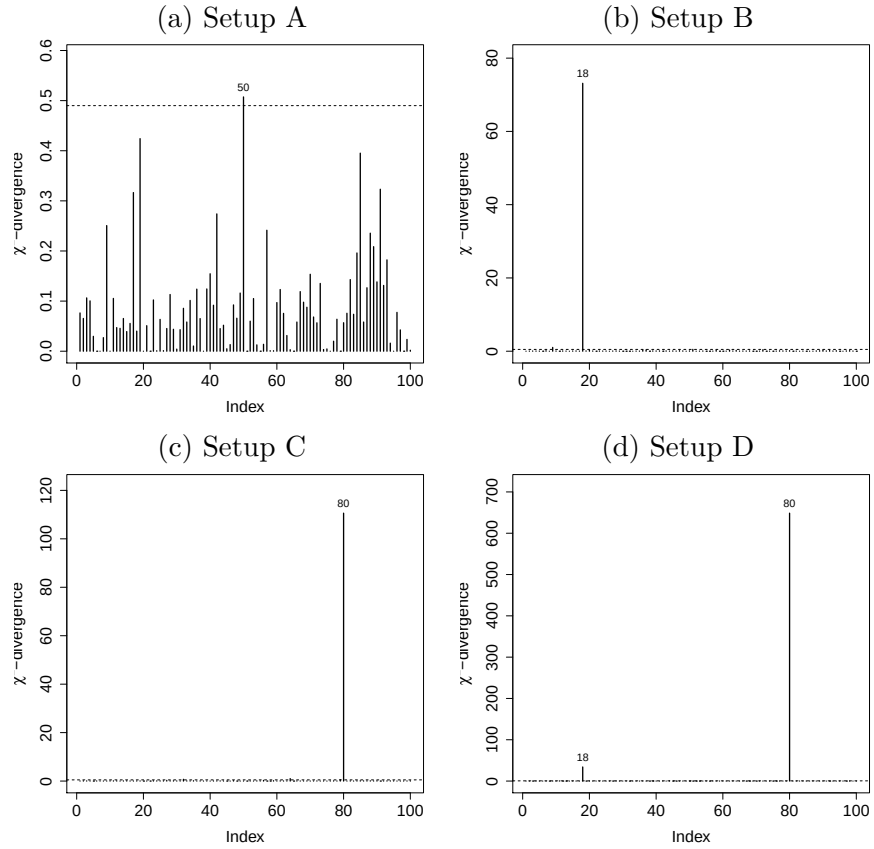

Figure 3.31: Index plots of χ^2-square divergence from the fitted CPHGCR model considering prior 1.

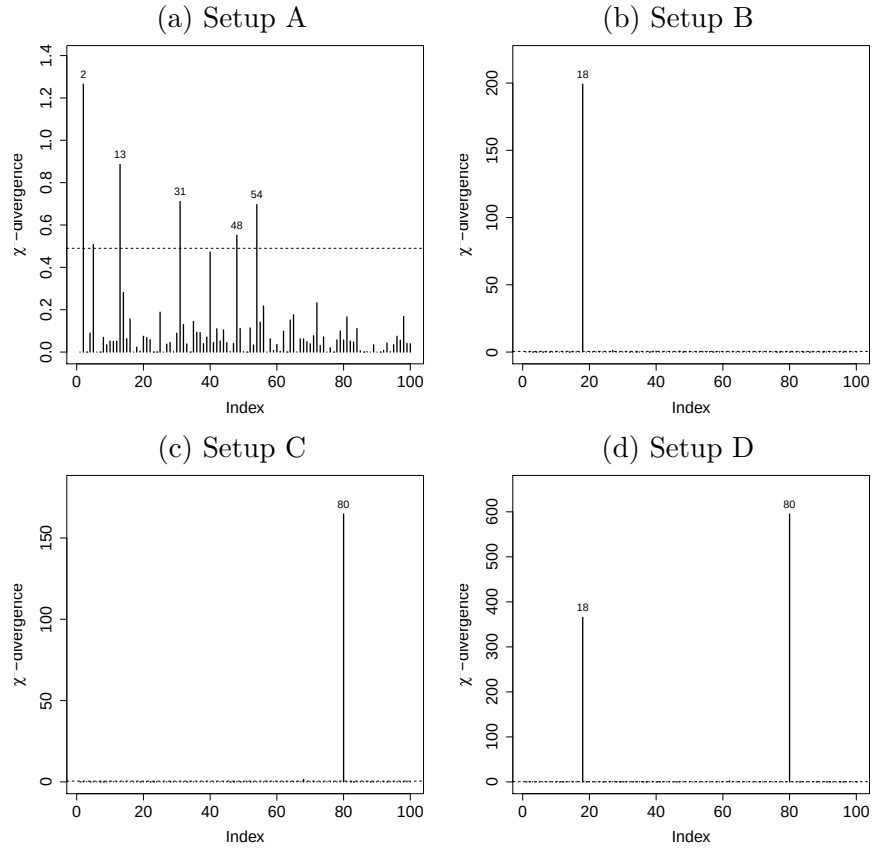

Figure 3.32: Index plots of χ^2-square divergence from the fitted CPHGCR model considering prior 2.

3.1.3 Application

We illustrate the proposed method for the interval-censored smoking cessation data presented in Section 1.1. Firstly, we fitted WGCR and CWGCR models considering the different spatial frailties in the models to the data set. The a priori distributions for the parameters $\boldsymbol{b}$, $\boldsymbol{\beta}$, and α are $b_j \sim N(0, 100)$, $j = 0, \ldots, 4$, $\beta_j \sim N(0, 100)$, $j = 1, \ldots, 4$, and $\alpha \sim N(0, 100)I_{(0,\infty)}$.

Because of the high computational cost, we implement the MCMC algorithms in the C programming language and the results were analyzed in the R language (R Development Core Team (2010)) through the "coda" package (Plummer *et al.*, 2005). All of our MCMC algoritms ran a total of 60,000 iterations discarding the first 20,000 realizations as burn-in and thinning to every fifth iteration. Posterior results are then based on 8,000 realizations of the Markov chain. Our Meteropolis acceptance rate for these parameters ranged from 25% to 50%. The convergence was checked using the Geweke diagnostic which did not indicate lack of convergence. The models are compared using DIC criterion.

Table 3.8 provides the DIC scores for a variety of effects of the fitted cure models. The DIC scores of the model 1 and 5 stand out as the best models in spite of the fact that the DIC values are close to each other. We also can note that the WGCR model are more adequate than the CWGCR model considering different priors to the parameters.

Table 3.8: Bayesian criteria for the fitted models.

WGCR model		Criteria	
Model	Priors	DIC	*pd*
1	$\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	416.4	11.5
2	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	416.9	11.6
3	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	417.5	12.1
4	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	417.2	12.7
5	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	416.7	11.7
CWGCR model		Criteria	
Model	Priors	DIC	*pd*
6	$\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	419.3	11.9
7	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	419.2	12.3
8	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	419.5	12.7
9	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	417.9	13.6
10	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	418.6	13.9

For the comparison with the models proposed by Carlin & Banerjee (2003), we consider the same prior distributions for the parameters $\boldsymbol{b}$ and $\boldsymbol{\beta}$ as considered by these authors. Table 3.9 reports the DIC scores for a variety of effects the Weibull Geometric cure rate model. We observe that the DIC scores in Table 3.9 are smaller than the values in Table 3.8. However, the DIC scores for WGCR models are very close each other, which indicates that these models are equivalent. Moveover, similarly to the previous case, the DIC values of WGCR models are smaller than CWGCR models.

Comparing the obtained DIC scores in Table 3.8 and 3.9 with the DIC scores presented in the study carried out by Carlin & Banerjee (2003), in which they proposed the mixture cure model with the spatial fragility, we can conclude that all our models are more adequate since all our DIC scores are smaller. Here, we select the model 15 as our working model.

Table 3.9: Bayesian criteria for the fitted models.

WGCR model		Criteria	
Model	Priors	DIC	*pd*
11	$\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	414.4	8.2
12	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	414.8	10.9
13	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$ 1	414.7	10.8
14	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	414.3	10.9
15	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	414.5	10.9
CWGCR model		Criteria	
Model	Priors	DIC	*pd*
16	$\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	418.1	9.4
17	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	417.3	11.7
18	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	416.8	11.5
19	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	416.9	11.6
20	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	416.8	11.6

Table 3.10: Posterior summaries of the parameter of the model 15 for the smoking cessation data.

	Survival Model					**Cure rate**			
Parameter		Mean	SD	2.5%	97.5%		Mean	SD	2.5%	97.5%
Intercept						b_0	1.3736	0.5789	0.2600	2.5442
Sex (male=0)	β_1	-0.1562	0.4551	-1.0626	0.6992	b_1	-0.5096	0.3718	-1.2711	0.2052
SI/UC (UC=0)	β_2	0.8427	0.5191	-0.1659	1.8703	b_2	0.8601	0.4310	0.0671	1.7192
Cigarettes per day	β_3	-0.1148	0.0378	-0.1809	-0.0322	b_3	-0.0728	0.0290	-0.1345	-0.0201
Duration as smoker	β_4	-0.0246	0.0343	-0.1003	0.0345	b_4	0.0197	0.0230	-0.0264	0.0651
α		2.4097	0.3073	1.8113	3.0065					
$a_1(a_u)$		0.8968	0.0676	0.7261	0.9874					
$a_2(a_v)$							0.8994	0.0670	0.7370	0.9881
Λ_{11}		2.6769	0.6377	1.5543	4.0673					
Λ_{22}							2.5743	0.6413	1.4924	3.9718
Λ_{12}		-0.0104	0.4625	-0.9321	0.8919					
Σ_{11}		0.4113	0.1075	0.2531	0.6754					
Σ_{22}							0.4298	0.1236	0.2561	0.7298
$\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$		0.0035	0.1828	-0.3655	0.3559					

where Λ_{ij} is the element of precision matrix $\boldsymbol{\Lambda}$ in position (i, j), and Σ_{ij} is the element of matrix $\boldsymbol{\Sigma} = \Lambda^{-1}$ in position (i, j), this Σ_{11} is the spatial variance component of $\boldsymbol{U}$ and Σ_{22} is the spatial variance component of $\boldsymbol{V}$, $\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$ denote their correlation.

The posterior summary of the parameters of the model 15 are presented in the Table 3.10. We note that only the parameters b_0, b_2, b_3 and β_3 are significant. In the cure rate, the negative value of b_3 means that the individuals with higher levels of cigarette consumption have lower probability to quit smoking, while the positive value of b_2 implies the individuals with special intervention have higher probability to quit smoking than those with usual care.

In the survival model, it is shows that the special intervention and the number of cigarettes smoked per day have negative effects on the hazard rate of the relapse time, that is, individuals with special intervention do not present lower hazard rates for the relapse time when compared to those who attend to usual care, while the individuals with a higher level of cigarette consumption do not present high hazard rates.

The estimated standard deviation $\Sigma_{11}^{1/2}$ of random spatial effects in the survival model is 0.4113, and the estimated standard deviation $\Sigma_{22}^{1/2}$ of the random spatial effects in the cure rate is 0.4268, which indicates that there is a considerable heterogeneity among the clusters. Moreover, it is observed that there are no correlations between the spatial effects $\boldsymbol{U}$ and $\boldsymbol{V}$.

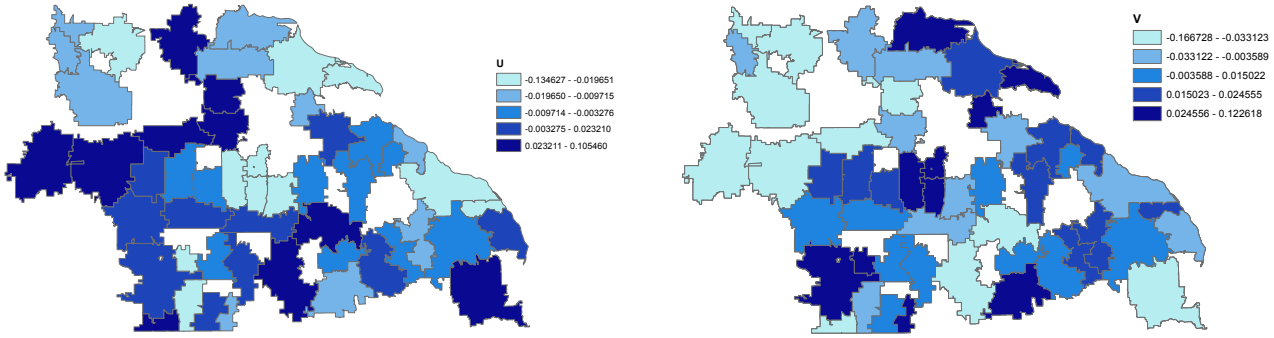

Figure 3.33: Maps of posterior means for frailties $\boldsymbol{U}$ (left panel) and $\boldsymbol{V}$ (right panel) in model 15.

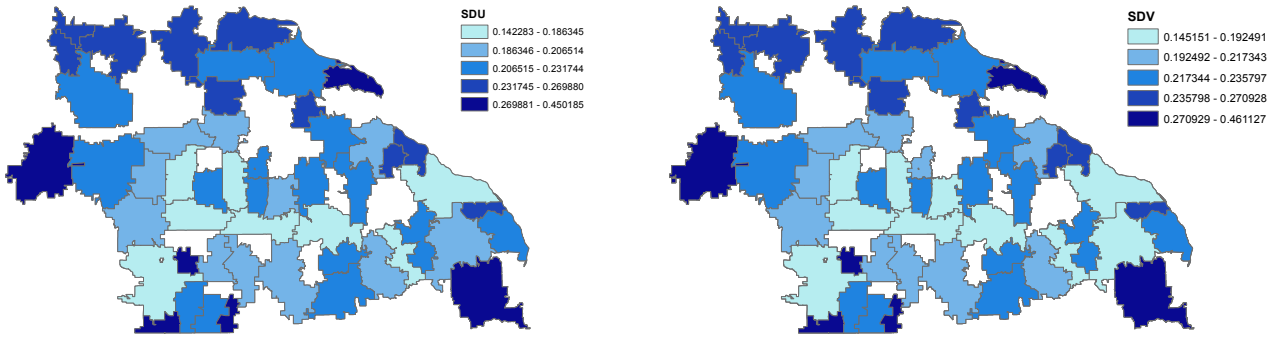

Figure 3.34: Maps of posterior standard derivations for frailties $\boldsymbol{U}$ (left panel) and $\boldsymbol{V}$ (right panel) in model 15.

The Figure 3.33 shows the posterior means of the frailties $\boldsymbol{U}$ and $\boldsymbol{V}$ in the model 15. For the frailties $\boldsymbol{U}$ of which the high value presents the high relapse rate, we note that the city of Owatonna and some cities of South have higher values, that is, the individuals in these regions have higher relapse rates than others. On the other hand, the city of Rochester (in the central regions of the

map) suggests slightly better than avenge cessation behavior, which can also be observed by the frailties $\boldsymbol{V}$. Note that the high value of $\boldsymbol{V}$ presents the high cure probability.

The Figure 3.34 maps the posterior standard deviations of the frailties $\boldsymbol{U}$ and $\boldsymbol{V}$ corresponding to the posterior means mapped in Figure 3.33. We note the posterior standard deviations of the frailties $\boldsymbol{U}$ and $\boldsymbol{V}$ have approximated values. In both maps show that the cities round the central region have lower values and the some periphery cites have higher values.

In order to detect possible influential observations in the posterior distribution of the parameters of model 15, the estimates of ψ-divergence measures, which were obtained by the posteriori sample of the parameters of the model, are presented in Figure 3.35. It shows that there are some individuals which can be influential observations were detected by divergence measures. Here, we will analyze the individuals 72, 138, 151, and 199 were detected by the J-distance and the χ^2-divergence. Table 3.11 presents information on them, so that we note that these four individuals had special interventions, not consuming high amounts of cigarettes per day (27 cigarettes per day on average), while the individuals 72, 138, and 151 had relapse, but the individual 199 did not. To reveal the impact of these possible influential observations on the parameter estimates and inferences, we removed such observations, refitting the models. We also calculated the relative variations (RV) for the posterior mean of the parameters. The RV is defined by $RV = (\hat{\vartheta}_{d,-I} - \hat{\vartheta}_d)/\hat{\vartheta}_d$, for all d, where I denotes a set of influential observations, d is the index of the parameters, $\hat{\vartheta}_{d,-I}$ denotes the posterior mean of $\vartheta_{d,-I}$, after the set of observations I was removed. In this case, we have $I = \{72, 138, 151, 199\}$.

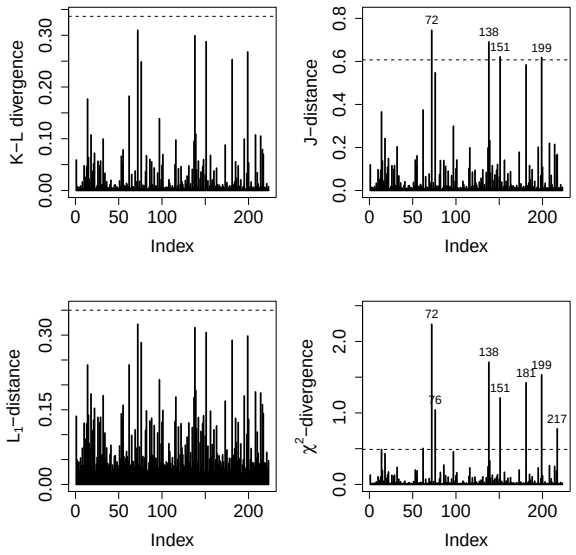

Figure 3.35: Estimates of ψ-divergence measures for Model 15

The posterior summaries of the parameters for the adjusted model 15 and RV for the posterior means of the parameters are presented in Table 3.12. We can note that only the values of RV for the posterior means of the parameters Λ_{12} and Σ_{12} are more than one, but they still have posterior

Table 3.11: Possible influential observations are detected by four divergence measures

Obs.	Sex	Duration	Intervention	Num. cigarettes	Relapse	Time interval	Zip
72	0	20	1	25	1	(3.159, 3.929)	55987
138	1	25	1	20	1	(2.998, 3.992)	55021
151	0	39	1	10	1	(0.923, 3.962)	55057
199	0	22	0	20	1	(3.885, 5.013)	55904

means near zero, and others parameters have the posterior means near the obtained values for the completed data set. In this case, there are not inferential changes after removing the observations.

Table 3.12: Posterior summaries of the parameter of the model 15 and relative variations adjusted for the smoking cessation data without detected individuals 72, 138, 151 and 199.

	Survival Model					**Cure Rate**			
Parameter		Mean	SD	2.5%	97.5%		Mean	SD	2.5%	97.5%
Intercept						b_0	1.3697 (-0.0028)	0.5981	0.2041	2.5265
Sex (male=0)	β_1	-0.2101 (-0.3455)	0.4326	-1.0848	0.6344	b_1	-0.5657 (0.1101)	0.3425	-1.2511	0.0902
SI/UC (UC=0)	β_2	1.0138 (0.2031)	0.5071	0.0486	2.0473	b_2	0.9023 (0.0491)	0.4002	0.1406	1.7522
Cigarettes per day	β_3	-0.1048 (-0.0873)	0.0377	-0.1800	-0.0304	b_3	-0.0611 (-0.1602)	0.0257	-0.1221	-0.0183
Duration as smoker	β_4	-0.0308 (0.2518)	0.0361	-0.1073	0.0375	b_4	0.0184 (-0.0656)	0.0230	-0.0280	0.0629
α		2.6474 (0.0987)	0.3515	1.9792	3.3582					
a_1 (a_u)		0.9006 (0.0043)	0.0665	0.7388	0.9880					
a_2 (a_v)							0.9002 (0.0009)	0.0650	0.7374	0.9864
Λ_{11}		2.6749 (-0.0008)	0.6497	1.5769	4.0686					
Λ_{22}							2.5717 (-0.0010)	0.6364	1.5178	3.9480
Λ_{12}		0.0113 (-2.0812)	0.4709	-0.9257	0.9130					
Σ_{11}		0.4122 (0.0023)	0.1086	0.2515	0.6701					
Σ_{22}							0.4307 (0.0021)	0.1208	0.2575	0.7207
Σ_{12}		-0.0054 (-2.5601)	0.1839	-0.3637	0.3568					

Now, we fitted the PHGCR and CPHGCR models, considering the different spatial frailties in the models to the data set. Since the piecewise exponential distribution has better approximation to any unknown function when the length of each interval becomes smaller, we partition the time axis so that they denoted the ordered distinct time points of all observed interval end points. Therefore, we have 178 parameters need to be estimated. In several areas, especial in medicine, the available prior information is also importance to be considered in the analysis. Therefore, we specify priori distributions for the parameters $\boldsymbol{b}$, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{178})$ to ensure weakly prior information following the analysis results obtained by Carlin & Banerjee (2003), that is let $b_j \sim N(0, 1^2)$, $j = 0, \ldots, 4$, $\beta_j \sim N(0, 1^2)$, $j = 1, \ldots, 4$, and $\alpha_i \sim N(0, 2^2)I_{(0,\infty)}$, $i = 1, \ldots, 178$.

Table 3.13 provides the DIC scores for a variety of effects PHGCR and CPHGCR models. The DIC scores of the fitted models are closely, this indicated these models are almost equivalent. In that follows we present results for Model 9 having low DIC scores, but emphasize that virtually any of the models in Table 3.13 could be used with equal confidence.

Table 3.13: Bayesian criteria for the PHGCR and CPHGCR models.

PHGCR model		Criteria	
Model	Priors	DIC	*pd*
1	$\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	395.8	9.55
2	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	396.2	10.63
3	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	395.8	12.12
4	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	394.8	11.57
5	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	397.8	11.65
CPHGCR model		Criteria	
Model	Priors	DIC	*pd*
6	$\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	395.9	10.29
7	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	395.3	12.22
8	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	395.3	12.16
9	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	394.5	11.70
10	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	394.8	11.71

To compare the proposed cure rate models, we observed the scores of criterion for fitted models presented in Table 3.9 and 3.13. We note that none of the models in Table 3.9 is better than the models presented in Table 3.13. Comparing the obtained DIC scores with the DIC values presented in the paper of Pan *et al.* (2014), where they proposed Bayesian semi-parametric model with the spatial fragility, it is shown that both PHGCR and CPHGCR models have DIC values smaller. Here, we select Model 9 which has the smallest DIC as our working model.

Table 3.14: Posterior summaries of the parameter of Model 9 for the smoking cessation data.

	Survival Model					**Cure rate**			
Parameter		Mean	SD	2.5%	97.5%		Mean	SD	2.5%	97.5%
Intercept						b_0	0.0072	0.8279	-1.6658	1.5794
Sex (male=0)	β_1	0.1748	0.2694	-0.3421	0.7128	b_1	-0.3696	0.4800	-1.2913	0.5993
SI/UC (UC=0)	β_2	-0.1392	0.3107	-0.7249	0.4876	b_2	0.8275	0.5500	-0.2217	1.9833
Cigarettes per day	β_3	-0.0168	0.0234	-0.0602	0.0267	b_3	-0.0499	0.0470	-0.1241	0.0606
Duration as smoker	β_4	-0.0292	0.0259	-0.0787	0.0178	b_4	0.0306	0.0600	-0.1055	0.1202
a		0.8981	0.0668	0.7349	0.9879					
Λ_{11}		2.6680	0.6501	1.5594	4.0684					
Λ_{22}							2.5805	0.6427	1.4840	4.0155
Λ_{12}		-0.0085	0.4670	-0.9244	0.9317					
Σ_{11}		0.4130	0.1084	0.2511	0.6668					
Σ_{22}							0.4289	0.1192	0.2537	0.7187
$\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$		0.0028	0.1820	-0.3645	0.3524					

where Λ_{ij} is the element of precision matrix $\boldsymbol{\Lambda}$ in position (i, j), and Σ_{ij} is the element of matrix $\boldsymbol{\Sigma} = \Lambda^{-1}$ in position (i, j), this Σ_{11} is the spatial variance component of $\boldsymbol{U}$ and Σ_{22} is the spatial variance component of $\boldsymbol{V}$, $\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$ denote their correlation.

Table 3.27 presents posterior means, standard deviations and 95% highest posterior density (HPD) intervals of the parameter of Model 9. We can note that all covariates are not significant when we consider the (95%) credibility interval, but the covariate of the "intervention type SI/UC" in the cure rate and in the survival model will become significant since we consider the lower credibility interval. In cure rate, the sign of the parameters b_2 and b_3 are the same as the results above, which means that the individuals with a higher level of cigarette consumption have lower probability of quit smoking and the individuals had especial with special intervention have higher probability of quit smoking than those with usual care. In the survival function, the negative value of β_2 implies that individuals with special intervention have lower hazard rate of the relapse time than those with usual care.

The estimates of the spatial variance component of $\boldsymbol{U}$ in the survival model (Σ_{11}) is 0.4130, the spatial variance component of $\boldsymbol{V}$ in the cure rate (Σ_{22}) is 0.4289, which indicate that there is considerable heterogeneity among the clusters. Moreover, it is observed that there are not correlations between the spatial effects $\boldsymbol{U}$ and $\boldsymbol{V}$.

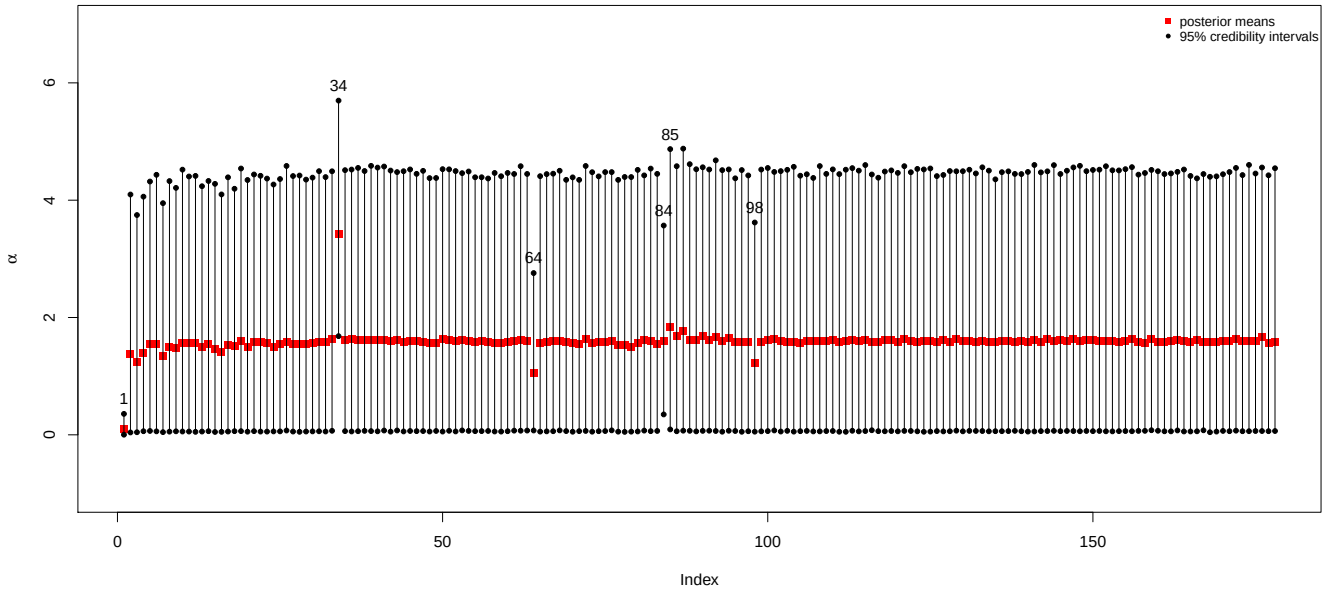

Figure 3.36: Posterior means and credibility intervals of α_i's

The posterior means and 95% HPD intervals of α_i's are presented in the Figure 3.73, it is showed that there are some values of α's, which are indicated, have different values of others. According this Figure, we can partition the time axis so that we consider just risk parameters (α_1, α_2, α_{34}, α_{35}, α_{64}, α_{65}, α_{84}, α_{85}, α_{88}, α_{98}, α_{99}), thus just 11 parameters need to be estimated, it will lower computational time cost.

The Figure 3.37 maps the posterior means of frailties $\boldsymbol{U}$ and $\boldsymbol{V}$ in the CPHGCR model. For

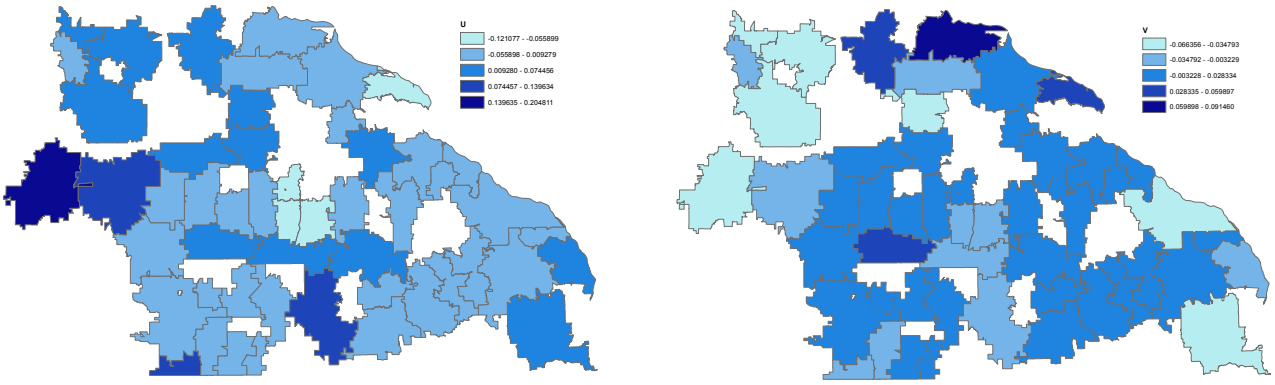

Figure 3.37: Maps of posterior means for frailties $\boldsymbol{U}$ (left panel) and $\boldsymbol{V}$ (right panel).

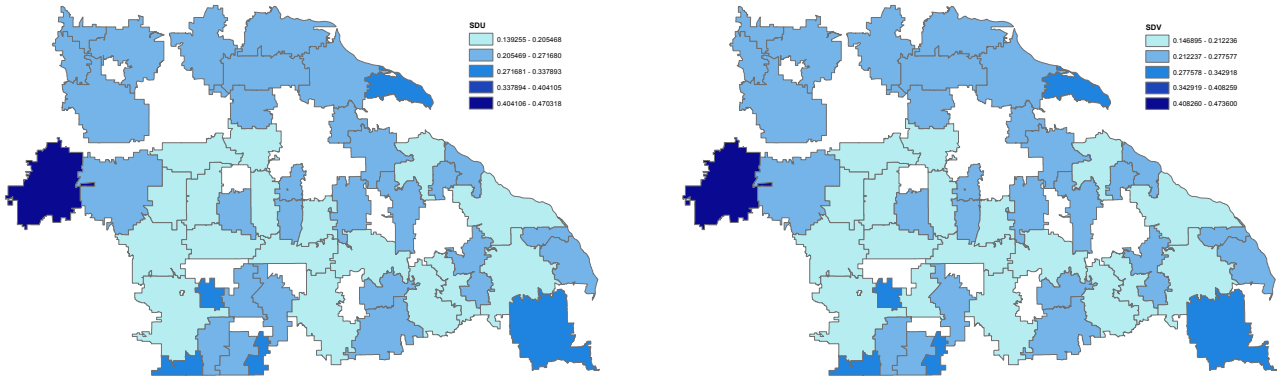

Figure 3.38: Maps of posterior standard derivations for frailties $\boldsymbol{U}$ (left panel) and $\boldsymbol{V}$ (right panel).

the frailties $\boldsymbol{U}$, the result is closer which obtained in model 15, but the frailties $\boldsymbol{V}$ show that almost all of regions have closed cure probabilities. The corresponding posterior standard deviations are presented in the Figure 3.38 have results similar in model 15 as well.

Considering the samples of posterior distribution of the parameters of the Model 9, the ψ-divergence measures are computed to detect possible influential observations in the posterior distribution of the parameters of the Model 9 and presented in the Figure 3.75. It shows that there are some possible influential observations were detected by divergence measures, but they are different from the observations which were detected previously. Here, we will just analyze individuals 14 and 86 which were detected by both J-distance and χ^2-divergence measure. In the Table 3.15, we can note that both individuals had special interventions but occurred relapse. In order to reveal the impact of this possible influent observation on the parameter estimates and inference, we removed this observation and readjust the model. Note that, in the piecewise exponential model, the time axis are partitioned by the ordered distinct time points of all observed interval end points, thus we have different and less risk parameter α's after removed the observations.

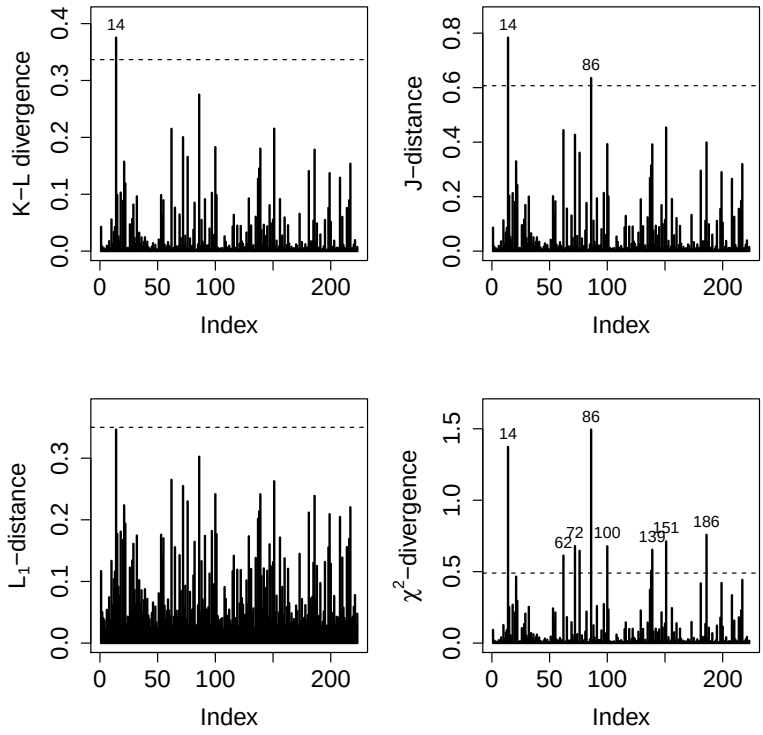

Figure 3.39: Estimates of ψ-divergence measures for Model 9

Table 3.15: Possible influential observations are detected the divergence measures

Obs.	Sex	Duration	Intervention	Num. cigarettes	Relapse	Time interval	Zip
14	0	32	1	60	1	(1.035, 4.211)	55904
86	1	24	1	40	1	(3.885, 5.073)	55987

The posterior summaries of the parameters for the readjust Model 9 and RV for the posterior mean of the parameters are presented in the Table 3.16. We can note that just the posterior means of the parameters b_0 and β_4 have relative variations more higher, but both have values keep on closed zero. All parameters had not sign change except b_0. The posterior means and 95% credibility intervals of α_i's are presented in the Figure 3.40, it is showed that the values of α's is similar as the estimates in Figure 3.73. In this case, we do not have inferential changes after removing the observations. So this model is not sensitive with influent observations. The values of DIC for fitted models is 385.0749 that is lower than Model 9 for the data without removing the detected observations.

3.1.4 Conclusions

In this work, we have described an approach to extend proportional odds cure models to allow for spatial correlations by including spatial fragility for the interval-censored data setting. We use the MCMC methods in Bayesian inference approach for our models and some used Bayesian comparison criterions were used. The results of the application show that WGCR model with fragilities has better fittings, but the PHGCR and CPHGCR models stand out better. Comparing the proposed models

Table 3.16: Posterior summaries of the parameter of Model 9 and RV adjusted for the smoking cessation data without detected individuals 14 and 86.

		Survival Model					**Cure Rate**			
Parameter		Mean	SD	2.5%	97.5%		Mean	SD	2.5%	97.5%
Intercept						b_0	-0.0358 (-5.9465)	0.8369	-1.6263	1.5895
Sex (male=0)	β_1	0.1934 (0.1064)	0.2708	-0.3352	0.7389	b_1	-0.3168 (-0.1428)	0.4536	-1.1974	0.5985
SI/UC (UC=0)	β_2	-0.2241 (0.6097)	0.3002	-0.7793	0.4069	b_2	0.6060 (-0.2677)	0.4968	-0.3488	1.5578
Cigarettes per day	β_3	-0.0334 (0.9819)	0.0193	-0.0709	0.0024	b_3	-0.0862 (0.7276)	0.0298	-0.1425	-0.0232
Duration as smoker	β_4	-0.0109 (-0.6254)	0.0206	-0.0479	0.0332	b_4	0.0708 (1.3100)	0.0350	0.0000	0.1313
a		0.9038 (0.0063)	0.0619	0.7568	0.9864					
Λ_{11}		2.6738 (0.0021)	0.6573	1.5785	4.1271					
Λ_{22}							2.5637 (-0.0065)	0.6365	1.4845	3.9550
Λ_{12}		-0.0063 (-0.2618)	0.4685	-0.9134	0.9258					
Σ_{11}		0.4126 (-0.0011)	0.1085	0.2483	0.6684					
Σ_{22}							0.4320 (0.0073)	0.1221	0.2569	0.7200
Σ_{12}		0.0029 (0.0349)	0.1830	-0.3552	0.3599					

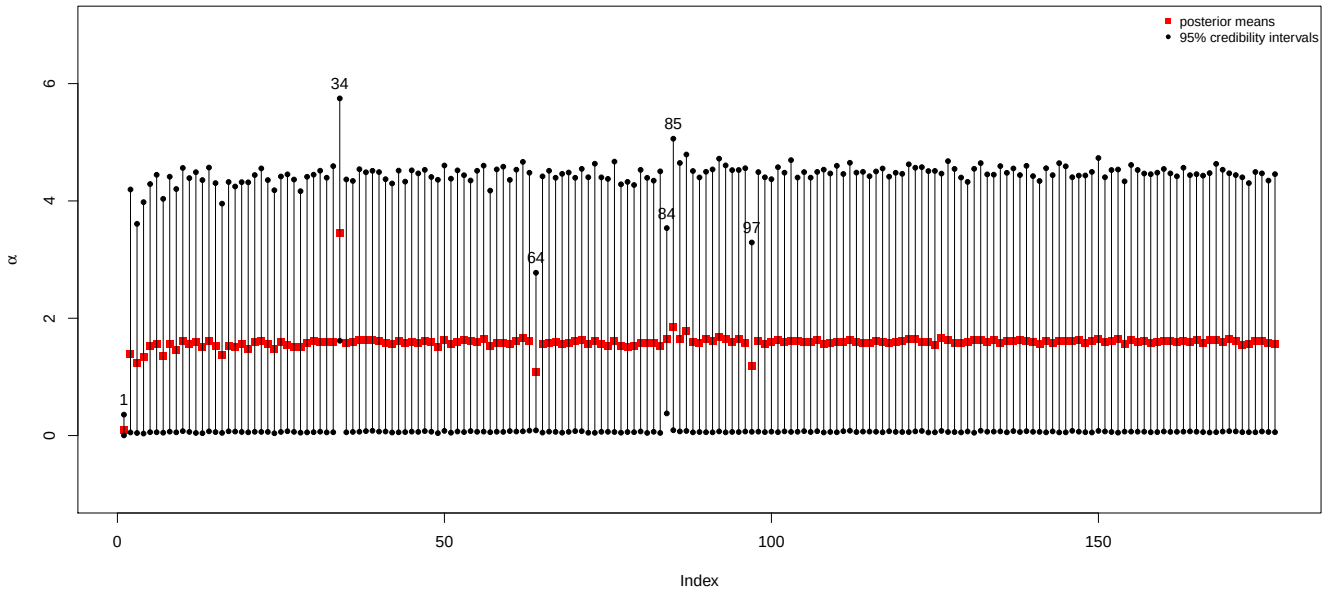

Figure 3.40: Posterior means and credibility intervals of α_i's for the data set without individuals 14 and 86.

with models introduced Carlin & Banerjee (2003) and Pan *et al.* (2014), it is showed that the PHGCR and CPHGCR models are more adequate. Moreover, The proposed model is not sensible with influent observation, which can be observed by the influence diagnostic analysis in the application. The interpretation of the covariates is easy due to the parametrization in the cure proportion. Moveover, the MCAR prior can be used even if fragilities have low or not correlations.

3.2 Negative-Binomial cure rate models with spatial frailties

Suppose there are I regions and n_i individuals in the ith region. We denote by T_{ij} the random variable for the observed time to event of the jth individual in the ith region, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, I$. Suppose that the (i, j)th individual is potentially exposed to latent risk M_{ij}, where M_{ij} denotes the initial number of competing causes related to the occurrence of an event and assuming M_{ij} has a negative binomial (NB) distribution with parameters θ_{ij} and η (Piegorsch, 1990), with the probability mass function

$$P(M_{ij} = m) = \frac{\Gamma(\eta^{-1} + m)}{\Gamma(\eta^{-1})m!}\left(\frac{\eta\theta_{ij}}{1 + \eta\theta_{ij}}\right)^m (1 + \eta\theta_{ij})^{-1/\eta}, \quad m = 0, 1, 2\cdots . \tag{3.14}$$

where $\theta_{ij} > 0$, $\eta > -1/\theta_{ij}$, so that $E(M_{ij}) = \theta_{ij}$ and $Var(M_{ij}) = \theta_{ij}\,(1 + \eta\theta_{ij})$. Here, η is a dispersion parameter (Saha & Paul, 2005), and values of $\eta > 0$ ($\eta < 0$) correspond to over (under) dispersion relative to the Poisson distribution. Particularly, when $\eta \to 0$, the NB approaches the Poisson distribution and the geometric distribution with parameter $1/(1+\theta_{ij})$ can be obtained when $\eta = 1$.

Let Y_{cij} for $c = 1, \ldots, M_{ij}$ denote the failure time of the jth individual in the ith region due to the cth latent risk. Suppose that given M_{ij}, the random variables Y_{cij} are mutually independent with distribution function $F(\cdot) = 1 - S(\cdot)$. If we assume that the presence of any latent risk will ultimately lead to the occurrence of the event, the time to event of interest can be defined by the random variable $T_{ij} = \min\{Y_{cij}, c = 1, \cdots, M_{ij}\}$ for $M_{ij} \geq 1$ and $T_{ij} = \infty$ if $M_{ij} = 0$, with $P(T_{ij} = \infty | M_{ij} = 0) = 1$. Thus, the survival function for the population is given by:

$$S_{pop}(t_{ij}) = \left[1 + \eta\theta_{ij}F(t_{ij})\right]^{-1/\eta}. \tag{3.15}$$

Let η take some different values. We have the mixture model (Berkson & Gage, 1952): $S_{pop}(t_{ij}) = (1 - \theta_{ij}) + \theta_{ij}S(t_{ij})$, if $\eta = -1$; promotion time cure model(Chen *et al.*, 1999): $S_{pop}(t_{ij}) = \exp\{-\theta_{ij}F(t_{ij})\}$, if $\eta \to 0$ and cure rate proportional odds model (Gu *et al.*, 2011): $S_{pop}(t_{ij}) = [1 + \theta_{ij}F(t_{ij})]^{-1}$, if $\eta = 1$.

The probability density function (p.d.f) and hazard function associated with (3.15) are given by

$$f_{pop}(t_{ij}) = \theta_{ij}f(t_{ij})\left[1 + \eta\theta_{ij}F(t_{ij})\right]^{-(1+1/\eta)} \quad \text{and} \quad \mathrm{h_{pop}(t_{ij})} = \theta_{\mathrm{ij}}\mathrm{f(t_{ij})}\left[1 + \eta\theta_{\mathrm{ij}}\mathrm{F(t_{ij})}\right]^{-1}$$

respectively, where $f(t_{ij}) = \frac{\partial}{\partial t_{ij}}F(t_{ij})$.

Note that the survival function in (3.15) can also be written as a mixture cure model

$$S_{pop}(t_{ij}) = (1+\eta\theta_{ij})^{-1/\eta} + \left(1-(1+\eta\theta_{ij})^{-1/\eta}\right)\left\{\frac{[1+\eta\theta_{ij}F(t_{ij})]^{-1/\eta} - (1+\eta\theta_{ij})^{-1/\eta}}{1-(1+\eta\theta_{ij})^{-1/\eta}}\right\}.$$

Thus, the survival function of uncured (susceptible) individuals has the expression

$$S_{sus}(t_{ij}) = \frac{[1+\eta\theta_{ij}F(t_{ij})]^{-1/\eta} - (1+\eta\theta_{ij})^{-1/\eta}}{1-(1+\eta\theta_{ij})^{-1/\eta}}.$$

Now, we assume another situation where the presence of all latent risks will ultimately lead to the occurrence of the event. Thus, the time to event of interest is defined by the random variable $T_{ij} = \max\{Y_{cij}, c = 1, \cdots, M_{ij}\}$ for $M_{ij} \geq 1$ and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P(T_{ij} = \infty | M_{ij} = 0) = 1$. The survival function for the population is then given by

$$S_{pop}(t_{ij}) = 1 + (1+\eta\theta_{ij})^{-1/\eta} - (1+\eta\theta_{ij}S(t_{ij}))^{-1/\eta}. \tag{3.16}$$

Similarly, let η take some different values. Then, the mixture model, complementary cure rate proportional odds model and complementary promotion time cure model can be obtained with the following survival functions respectively

$$\begin{aligned}
S_{pop}(t_{ij}) &= (1-\theta_{ij}) + \theta_{ij}S(t_{ij}), \quad \text{if} \quad \eta = -1;\\
S_{pop}(t_{ij}) &= 1 + \exp\{-\theta_{ij}\} - \exp\left\{-\theta_{ij}S(t_{ij})\right\}, \quad \text{if} \quad \eta = 1;\\
S_{pop}(t_{ij}) &= 1 + (1+\theta_{ij})^{-1} - \left[1+\theta_{ij}S(t_{ij})\right]^{-1}, \quad \text{if} \quad \eta \to 0.
\end{aligned}$$

The corresponding p.d.f. and hazard function are given respectively by

$$f_{pop}(t_{ij}) = \theta_{ij}f(t_{ij})(1+\eta\theta_{ij}S(t_{ij}))^{-(1/\eta+1)},$$

and

$$h_{pop}(t_{ij}) = \frac{\theta_{ij}f(t_{ij})(1+\eta\theta_{ij}S(t_{ij}))^{-(1/\eta+1)}}{1+(1+\eta\theta_{ij})^{-1/\eta} - (1+\eta\theta_{ij}S(t_{ij}))^{-1/\eta}}.$$

The survival function (3.16) also can be written as a mixture cure model

$$S_{pop}(t_{ij}) = (1+\eta\theta_{ij})^{-1/\eta} + \left(1-(1+\eta\theta_{ij})^{-1/\eta}\right)\left\{\frac{1-(1+\eta\theta_{ij}S(t_{ij}))^{-1/\eta}}{1-(1+\eta\theta_{ij})^{-1/\eta}}\right\}.$$

Thus, the survival functions of susceptible individuals is given by

$$S_{sus}(t_{ij}) = \frac{1-(1+\eta\theta_{ij}S(t_{ij}))^{-1/\eta}}{1-(1+\eta\theta_{ij})^{-1/\eta}}.$$

The first situation is also known as the first activation scheme, because in this case we assume that the event of interest occurs when the first possible cause is activated. On the other hand, the second situation is known as last activation scheme, because the event of interest only takes place after all the latent causes have occurred. Thus, we denote the survival functions (3.15) and (3.16) by $S^F_{pop}(t_{ij})$ and $S^L_{pop}(t_{ij})$, respectively. There is another kind of situation where the event of interest occurs when some of the possible causes are activated, and given the number of latent causes M_{ij}, the number of activated causes is a random variable with a discrete uniform distribution on $\{1, \cdots, M\}$. This situation is known as the random activation scheme. In this case, the survival function for the population has the expression

$$S_{pop}(t_{ij}) = (1+\eta\theta_{ij})^{-1/\eta} + (1-(1+\eta\theta_{ij})^{-1/\eta})S(t_{ij}), \tag{3.17}$$

and is denoted by $S^R_{pop}(t_{ij})$. Note that whichever the activation scheme, the density and hazard functions of the cure models are improper functions, since the survival functions are not proper. The cure fraction is the same for these activation schemes, and can be obtained by: $p_{0ij} = \lim_{t_{ij}\to\infty} S_{pop}(t_{ij}) = (1+\eta\theta_{ij})^{-1/\eta}$. However, under the different activation schemes, the models differ by their survival, density and hazard functions. Moreover, under conditions of models (3.15), (3.16) and (3.17) for any distribution function $F(\cdot)$, we have $S^F_{pop}(t_{ij}) \leq S^R_{pop}(t_{ij}) \leq S^L_{pop}(t_{ij})$ for all $t_{ij} > 0$.

As is well known, the cure fraction plays a key role in survival models with cure fraction. So we consider the parameterization of the model in the expressions of the cure fraction. Since $p_{0ij} = (1+\eta\theta_{ij})^{-1/\eta}$, we have $\theta_{ij} = (p_{0ij}^{-\eta} - 1)/\eta$. Moreover, we propose that the cure probability of an individual (i, j) is associated with covariates $\boldsymbol{x}_{ij}$ and it can be modeled by a logistic regression

$$p_{0ij} = \frac{\exp(\xi_{ij})}{1+\exp(\xi_{ij})},$$

where ξ_{ij} is a linear form of covariates, $\xi_{ij} = \boldsymbol{x}_{ij}^\top \boldsymbol{b}$ and $\boldsymbol{b}$ is a p_1-dimensional vector representing the effects of covariates on the cured probability.

Using p_{0ij} as a parameter, the improper survival functions can be written as

$$\begin{aligned} S^F_{pop}(t_{ij}) &= \left[1+(p_{0ij}^{-\eta}-1)F(t_{ij})\right]^{-1/\eta}, \\ S^L_{pop}(t_{ij}) &= 1+p_{0ij}-\left[1+(p_{0ij}^{-\eta}-1)S(t_{ij})\right]^{-1/\eta}, \\ S^R_{pop}(t_{ij}) &= p_{0ij}+(1-p_{0ij})S(t_{ij}). \end{aligned}$$

Note that $S^R_{pop}(t_{ij})$ is a mixture rate model with the cure fraction p_{0ij}.

Assuming Y_{cij}'s take proportional hazard (PH) model with the baseline hazard function $h_0(t|\boldsymbol{\alpha})$, the conditional hazard function $h(t|\boldsymbol{\phi}) = h_0(t|\boldsymbol{\alpha})\exp(\lambda_{ij})$, where $\boldsymbol{\phi} = (\boldsymbol{\alpha}, \lambda_{ij})$, $\lambda_{ij} = \boldsymbol{z}'_{ij}\boldsymbol{\beta}$ is the linear predictor of the covariates, where $\boldsymbol{z}_{ij}$ is covariates of an individual (i,j) and $\boldsymbol{\beta}$ is a p_2-dimensional vector representing the effects of covariates on the survival model component and $\boldsymbol{\alpha}$ is the parameter vector of the baseline functions. Considering the baseline functions the same as the geometric cure rate models presented in section 3.1. Therefore, firstly, the baseline hazard function is $h_0(t|\alpha) = \alpha t^{\alpha-1}$, thus Y_{cij}'s follow a Weibull distribution with the shape parameter $\alpha > 0$ and scale parameter λ_{ij}. We called the functions (3.15) and (3.16) by Weibull negative binomial cure rate (WNBCR) model and complementary Weibull negative binomial cure rate (CWNBCR) model, respectively. Secondly, the baseline function has the piecewise exponential distribution with the vector of parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)$. In this case, we called the functions (3.15) and (3.16) by proportional hazard negative binomial cure rate (PHNBCR) model and complementary proportional hazard negative binomial cure rate (CPHNBCR) model, respectively.

Similarly, we introduce the frailties U_i and V_i to better explain the effect of survival time of susceptible individuals and on the cure probability through a linear predictor expression $\lambda_{ij} = \boldsymbol{z}'_{ij}\boldsymbol{\beta} + U_i$, and $\xi_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{b} + V_i$, for $j = 1\ldots, n_i, i = 1, \ldots, I$.

Here, the frailties U_i and V_i are spatially correlated across the regions. In this work, we propose two approaches. In the first we employ a separate independent conditionally autoregressive (CAR) prior distribution on $(\boldsymbol{U}, \boldsymbol{V})$. Second, we assuming the spatial priors on $(\boldsymbol{U}, \boldsymbol{V})$ are dependent, and they have multivariate conditionally auto-regressive MCAR prior distribution, where the CAR and MCAR distributions were presented in Section 2.5.3 in detail.

3.2.1 Bayesian Inference

Let $\boldsymbol{D}_{obs} = \{(A_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \delta_{ij}); j = 1, \ldots, n_i, i = 1, \ldots M\}$ denote the observed data, where $A_{ij} = (t_{ijL}, t_{ijR}]$ is the interval during which individual j in cluster i occur the event of interest, $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are the p_1-dimencional and p_2-dimencional vectors of covariates, and δ_{ij} is following

interval censoring indicador: $\delta_{ij} = I(t_{ijR} < \infty)$. For the spacial case in which the survival time is right-(left-) censored, $R_{ij} = +\infty (L_{ij} = 0)$, whereas for exact observations, $t_{ijL} = t_{ijR}$. Given the frailties $\boldsymbol{U}$ and $\boldsymbol{V}$, the likelihood function for the general interval-censored cure rate model following Finkelstein(1986), is given by

$$L\{\boldsymbol{\varphi}|\boldsymbol{D}_{obs}, \boldsymbol{U}, \boldsymbol{V}\} \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i} \left(S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi}) - S_{\text{pop}}(t_{ijR}|\boldsymbol{\varphi})\right)^{\delta_{ij}} S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi})^{1-\delta_{ij}}, \tag{3.18}$$

where $\boldsymbol{\varphi} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \eta)$, $\boldsymbol{\alpha}$ is the shape parameter of the Weibull distribution for the first model with unitary size and it is the risk parameter vector for the second model with size Q. For a Bayesian analysis, we assume the prior densities for parameters are $b_j \sim N(\mu_b, \sigma_b^2)$ for $j = 0, \ldots, (p_1 - 1)$; $\beta_j \sim N(\mu_\beta, \sigma_\beta^2)$ for $j = 1, \ldots, p_2$; $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)\boldsymbol{I}_{(0,\infty)}$, $i = 1$ for Weibull distribution and $i = 1, \ldots Q$ for piecewise exponential distribution; $\eta \sim N(\mu_\eta, \sigma_\eta^2)\boldsymbol{I}_{(0,\infty)}$, where μ_b, μ_β, μ_α, μ_η , σ_b, σ_β, σ_α. σ_η are known hyperparameters. To express vague priors, we consider $\mu_b = \mu_\beta = \mu_\alpha = \mu_\eta = 0$ with large values of σ_b^2, σ_β^2, σ_η^2 and σ_α^2. Here, $N(\mu, \sigma^2)\boldsymbol{I}_{(a,b)}$ denotes the truncated normal distribution which is the probability distribution of a normally distributed random variable whose values lies within the interval $-\infty \leq a < b \leq \infty$. In several areas, specially in medicine, it is preferable to use the prior information when they are available, moreover it is worth mentioning that using a truncated normal distribution as prior facilitates the insertion of information in certain regions of the parameter space, since the hyperparameters no longer represent the mean and variance but still control the region of higher probability mass.

Independent assumption

For the independent assumption, we employ separate independent CAR prior on the random frailties $\boldsymbol{U} = (U_1, \ldots, U_I)^\top$ and $\boldsymbol{V} = (V_1, \ldots, V_I)^\top$, that is,

$$U_1, \ldots, U_I \sim \text{CAR}(\theta_1) \quad \text{and} \quad V_1, \ldots, V_I \sim \text{CAR}(\theta_2),$$

where θ_1 and θ_2 are positive unknown hyper-parameters, and we assume they have Inverse-Gamma prior with the known shape parameter $a_0 > 0$ and scale parameter $b_0 > 0$. In this paper, we assume $a_0 = b_0 = 0.01$ to consider vague priors for the parameters θ_1 and θ_2.

Assuming the independence of the parameters $\boldsymbol{b}$, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, η, θ_1 and θ_2 and combining the likelihood function (3.18), the joint posterior distribution for the parameters is given

$$\pi(\boldsymbol{\varphi}, \theta_1, \theta_2|\boldsymbol{D}_{\text{obs}}) \propto L(\boldsymbol{\varphi}|\boldsymbol{D}_{obs}, \boldsymbol{U}, \boldsymbol{V})\pi(\boldsymbol{U}, \boldsymbol{V}|\theta_1, \theta_2)\pi(\boldsymbol{\varphi}, \theta_1, \theta_2),$$

where $\pi(\boldsymbol{\varphi}, \theta_1, \theta_2) = \pi(\boldsymbol{b})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\alpha})\pi(\eta)\pi(\theta_1)\pi(\theta_2)$.

This joint posterior density is analytically intractable. So, we based our inference on the Markov chain Monte Carlo (MCMC) simulation methods. We can observe that the full conditional distributions for parameters $\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\alpha}$ and η have not closed forms, thus we will use the Metropolis-Hastings algorithm to generate a posteriori samples for these parameter. To avoid range restrictions on the parameters α_i's and η, we define $\zeta_i = \log(\alpha_i)$ for $i = 1, 2, \ldots, Q$ and $\kappa = \log(\eta)$ to transform all parameters space to real space (necessary to work with Gaussian proposal densities). Let $\boldsymbol{\vartheta} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \kappa, \theta_1, \theta_2)$, according for the Jacobian of this transformation, the joint posterior density of $\pi(\boldsymbol{\vartheta}|\boldsymbol{D}_{\text{obs}})$ is proportional to

$$L(\boldsymbol{\varphi}|\boldsymbol{D}_{obs}, \boldsymbol{U}, \boldsymbol{V}) \exp\left\{ -\frac{1}{2}\left[\sigma_b^{-2} \sum_{i=0}^{p1-1} b_i^2 + \sigma_\beta^{-2} \sum_{i=1}^{p2} \beta_i^2 + \sum_{i=1}^{Q} \frac{\exp(2\zeta_i)}{\sigma_\alpha^2} + \frac{\exp(2\kappa)}{\sigma_\eta^2} + \frac{\boldsymbol{U}^\top(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{U}}{\theta_1} \right.\right.$$
$$\left.\left. + \frac{\boldsymbol{V}^\top(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{V}}{\theta_2} \right] - (a_0 + 1)\left(\log(\theta_1) + \log(\theta_2)\right) - \left(\frac{b_0}{\theta_1} + \frac{b_0}{\theta_2}\right) + \sum_{i=1}^{Q} \zeta_i + \kappa \right\},$$

where $\boldsymbol{\varphi} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}^{-1}, \kappa^{-1})$, $\boldsymbol{\zeta}^{-1} = \left\{\zeta_i^{-1} = \exp(\zeta_i) = \alpha_i, \;\; i = 1 \ldots, Q\right\}$ denotes inverse function of $\boldsymbol{\zeta}$, and $\kappa^{-1} = \exp(\kappa) = \eta$ denotes inverse function of η.

On the other hand, the full conditional distributions for parameters θ_i's are given by

$$\begin{aligned}
\pi(\theta_i|\boldsymbol{\vartheta}_{-\theta_i}, \boldsymbol{D}_{\text{obs}}) &\propto \pi(\boldsymbol{\psi}_i|\theta_i)\pi(\theta_i) \\
&\propto (\theta_i)^{-k/2} \exp\left(-\frac{1}{2\theta_i}\boldsymbol{\psi}_i^\top(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{\psi}_i\right) \theta_i^{-a_0-1} \exp(-b_0\theta_i^{-1}) \\
&\propto \theta_i^{-\left(a_0 + \frac{k}{2}\right)-1} \exp\left\{ -\left(\frac{\boldsymbol{\psi}_i^\top(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{\psi}_i}{2} + b_0 \right) \theta_i^{-1} \right\}, \;\; i = 1, 2,
\end{aligned}$$

where $\boldsymbol{\psi}_1 = \boldsymbol{U}$, $\boldsymbol{\psi}_2 = \boldsymbol{V}$ and k is the rank of the matrix $\boldsymbol{D_W} - \boldsymbol{W}$. Thus, the full conditional distributions of the parameter θ_i is an Inverse-Gamma distribution with parameters $a_0 + \frac{k}{2}$ e $b_0 + \frac{1}{2}\left(\boldsymbol{\psi}_i'(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{\psi}_i\right)$. In this case, the Gibbs sampler algorithm (see Gamerman & Lopes, 2006) is used to generate a posteriori sample.

Dependent assumption

Now we assume that the spatial priors on the parameters $(\boldsymbol{U}, \boldsymbol{V})$ are dependent on each other. Let $\boldsymbol{\psi} = (\boldsymbol{U}^\top, \boldsymbol{V}^\top)^\top$, we first employ the parameter $\boldsymbol{\psi}$ has a MCAR distribution with a common smoothness parameter a, i.e.,

$$\boldsymbol{\psi} \sim MCAR(a, \boldsymbol{\Lambda}).$$

Further, we employ the parameter $\boldsymbol{\psi}$ has an extend MCAR distribution with assuming the different smoothness parameters for the parameters $\boldsymbol{U}$ and $\boldsymbol{V}$, say a_1 and a_2, that is,

$$\boldsymbol{\psi} \sim MCAR(a_1, a_2, \boldsymbol{\Lambda}).$$

The prior distributions for $\boldsymbol{a}$ and $\boldsymbol{\Lambda}$ are given by

- $a_i \sim$ Uniform$(0,1)$ or $a_i \sim$ Beta$(18,2)$, for i,
- $\boldsymbol{\Lambda} \sim$ Wishart(n_0, Λ_0), with n_0 and Λ_0 known,

where i=1 for $\boldsymbol{\psi} \sim MCAR(a, \boldsymbol{\Lambda})$ and i=1,2 for $\boldsymbol{\psi} \sim MCAR(a_1, a_2, \boldsymbol{\Lambda})$). The prior distributions for the parameter a_i is used by Banerjee & Carlin (2004), in which $a_i \sim$ Uniform$(0,1)$ is a non-informative prior, and $a_i \sim$ Beta$(18,2)$ is an informative prior with $E[a_i] = 0.9$ and $Var[a_i] = 0.004285$; On the other hand, the prior distribution for the parameter $\boldsymbol{\Lambda}$ is used not only by Carlin & Banerjee (2003) but also by Gelfand & Vounatsou (2003) and Banerjee & Carlin (2004). They suggested that n_0 can take value as the dimension of matrix $\boldsymbol{\Lambda}$. However, Gelfand & Vounatsou (2003) and Banerjee & Carlin (2004) considered Λ_0 equals $\boldsymbol{I}$ and $0.01\boldsymbol{I}$ in their papers, respectively, where $\boldsymbol{I}$ denote a identity matrix. Both authors also commented that they had no prior knowledge regarding the nature or extent of dependence for the parameter $\boldsymbol{\Lambda}$. Note that $\boldsymbol{\Lambda}^{-1}$ describe the relative variability and covariance relationship between the different diseases given the neighboring site. Thus, if $\Lambda_0 = 0.01\boldsymbol{I}$, we assumed high relative variability between neighborhood and we assumed low relative variability between neighborhood if $\Lambda_0 = \boldsymbol{I}$. Thus, it is necessary to conduct a prior study for the parameter Λ_0 to verify the influence of Λ_0 in the estimation, in order to have a value for appropriate Λ_0.

To avoid range restrictions on the parameters a_i, we consider the transformations $\rho_i = \log(a_i/(1-a_i)) \in \mathbb{R}$, then, the joint posterior density is given by

$$\begin{aligned}
\pi(\boldsymbol{\vartheta}|\boldsymbol{D}_{\text{obs}}) \;\propto\; & L(\boldsymbol{\varphi}|\boldsymbol{D}_{obs}, \boldsymbol{\psi}) \exp\left\{ -\frac{1}{2}\left[\sigma_b^{-2} \sum_{i=0}^{p_1} b_i^2 + \sigma_\beta^{-2} \sum_{i=1}^{p_2} \beta_i^2 + \sum_{i=1}^{Q} \frac{\exp(2\zeta_i)}{\sigma_\alpha^2} + \frac{\exp(2\kappa)}{\sigma_\eta^2} \right] \right. \\
+ \;& \boldsymbol{\psi}^\top \left[\boldsymbol{\Lambda} \otimes (\boldsymbol{D_W} - \boldsymbol{aW}) \right] \boldsymbol{\psi} + \log|\boldsymbol{\Lambda} \otimes \boldsymbol{aW}| + \frac{n_0 - 4}{2} \log|\boldsymbol{\Lambda}| - \frac{1}{2} tr(\Lambda_0^{-1}\Lambda) \\
+ \;& \left. \sum_{i=1}^{Q} \zeta_i + \kappa \right\} \pi(\boldsymbol{\rho}),
\end{aligned}$$

where $\boldsymbol{\varphi} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}^{-1}, \kappa^{-1})$ and $\pi(\rho_i) = 1$ if $a_i \sim$ Uniform$(0,1)$ and $\pi(\rho_i) = \frac{1}{B(18,2)} \frac{\exp(17\rho_i)}{(1+\exp(\rho_i))^{18}}$ if $a_i \sim$ Beta$(18,2)$, where $B(18,2) = \frac{17!}{18!} = \frac{1}{18}$.

This joint posterior density is analytically intractable. Thus, we again based our inference on the Markov Chain Monte Carlo (MCMC) simulation methods. We observe that the full conditional

distributions for parameters $\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}$ and $\boldsymbol{\rho}$ do not have closed forms, thus we will use the Metropolis-Hastings algorithm to generate a posteriori samples for these parameter. However, the Gibbs sampler algorithm is used to generate a posteriori sample for the parameter $\boldsymbol{\Lambda}$, because the full conditional distribution has a closed form. The full conditional distribution $\pi(\boldsymbol{\Lambda}|\boldsymbol{\vartheta}_{(-\boldsymbol{\Lambda})}, \boldsymbol{D}_{\text{obs}}))$ is proportional to

$$\begin{aligned} & \pi(\boldsymbol{\psi}|\boldsymbol{\Lambda}, \boldsymbol{a})\pi(\boldsymbol{\Lambda}) \\ \propto\ & |\boldsymbol{\Lambda} \otimes \boldsymbol{D_W} - \boldsymbol{aW}|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{\psi}^\top(\boldsymbol{D_W} - \boldsymbol{aW})\boldsymbol{\psi}\right) |\boldsymbol{\Lambda}|^{(n_0-4)/2} \exp\left(-\frac{1}{2}tr(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda})\right) \\ \propto\ & |\boldsymbol{\Lambda}|^{(I+n_0-4)/2} \exp\left(-\frac{1}{2}tr((\boldsymbol{\Lambda}_0^{-1} + \boldsymbol{B})\boldsymbol{\Lambda})\right), \end{aligned} \tag{3.19}$$

where

$$\boldsymbol{B} = \begin{bmatrix} tr(\boldsymbol{R_1 U}(\boldsymbol{R_1 U})^\top) & tr(\boldsymbol{R_1 U}(\boldsymbol{R_2 V})^\top) \\ tr(\boldsymbol{R_2 V}(\boldsymbol{R_1 U})^\top) & tr(\boldsymbol{R_2 V}(\boldsymbol{R_2 V})^\top) \end{bmatrix}$$

Thus, it follows that the full conditional distribution for $\boldsymbol{\Lambda}$ has the Wishart distribution with scale matrix $(\Lambda_0^{-1} + B)^{-1}$ and degrees of freedom $I + n_0$.

3.2.2 Simulation study

In this section we present some simulation studies for the proposed models with the dependent assumption in order to examine the theirs performances. The interval-censored survival times $(t_{ijL}, t_{ijR}, \delta_{ij})$ with the cure fraction under the first and last activations are generated in a manner similar to that employed by Yau & Ng (2001) with some modifications. First, we generate latent NB variable M_{ij}, which denote the initial number of competing causes related to the event, with parameter $p_{0ij} = [1 + \exp(-b_0 + b_1 x_{ij} + v_i)]^{-1}$ for the jth individual in the ith region, $j = 1, \ldots, n_i$, $i = 1, \ldots, I$, where covariate x_{ij} follows Bernoulli(0.5) distribution. Interval-censored data $(t_{ijL}, t_{ijR}, \delta_{ij})$ are then generated as follows:

(i) If $M_{ij} = 0$, then let $t_{ij} = t_{ijL}$ from the exponential distribution with hazard rate 10 and let censoring indicator $\delta_{ij} = 0$.

(ii) If $M_{ij} > 0$, then

- we generate M_{ij} latent Weibull variables with parameter α and $\lambda_{ij} = (\beta x_{ij} + u_i)$, if Y_{cij}'s have the Weibull distribution;
- or we generate M_{ij} latent Exponential variables with hazard rate $\alpha\lambda_{ij} = \alpha(\beta x_{ij} + u_i))$, if Y_{cij}'s take the PH model.

Let t_{ij} takes lowest generated variable in case of generating the variables of model under first activation and t_{ij} takes largest generated variable in case of generating the variables of model

under last activation. The censoring variable c_{ij} is generated from $U(0, cc)$, $cc > 0$ is fixed to control the percentage of censored data. Let $\delta_{ij} = 1$ if $t_{ij} \leq c_{ij}$ and $\delta_{ij} = 0$ otherwise.

(iii) For $\delta_{ij} = 0$, let $0 < t_{ijL} < t_{ijR} = \infty$.

(iv) For $\delta_{ij} = 1$, we create len_{ij} from distribution $U(0.2, 0.7)$ and l_{ij} from $U(0, 0.01)$. Then, from $(0, l_{ij}], (l_{ij}, l_{ij} + len_{ij}], \ldots, (l_{ij} + klen_{ij}, \infty]$, $k = 1, 2, \ldots,$, $(t_{ijL,t_{ijR}}]$ is chosen as that satisfying $t_{ijL} < t_{ij} \leq t_{ijR}$.

In the simulation study, we consider $I = 5$ regions (Zip) with the corresponding adjacent matrix is $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$, the random effects u_i and v_i are generated from Normal distribution with mean $\mathbf{0}$ and precision matrix $\boldsymbol{\Lambda} \bigotimes (\boldsymbol{D_W} - a\boldsymbol{W})$, where $\boldsymbol{W}$ is standardized adjacent matrix so that each of its rows sum to one, $D_W = \text{Diag}(1, 1, 2, 1, 1)$ is a diagonal matrix and we fixed $a = 0.9$ and $\Lambda = \text{Diag}(4, 4)$, i.e. we fixed $\Lambda_{11} = 4$, $\Lambda_{22} = 4$ and $\Lambda_{12} = \Lambda_{21} = 0$. We set $N = 100$ and the number of Zip was distributed for each individual using sample with replace, thus the number of individuals in each region n_i, $i = 1 \ldots, 5$ are varied, that is this five regions could have different number of individuals.

We fixed parameters $b_0 = -1.50$, $b_1 = -0.50$ and $\beta = -0.15$. For Weibull cure rate models, we fixed forma parameter $\alpha = 0.30$ and for the PH cure rate models, we fixed risk parameter $\alpha = 1.00$. We consider around 40% censored data for each generated sample and 500 repeated samples are simulated for each model. In the simulations, the vague priors for the parameters are used. For each generated data set we simulate one chain of size 10000 for each parameter, disregarding the first 1000 iterations to eliminate the effect of the initial values and to avoid correlation problems and thinning to every third iteration, thus obtaining a effective sample of size 3000 upon which the posterior is based on. To evaluate the performance of the parameter estimates, the average bias (Bias), standard deviation (SD) of the estimate, average standard deviation (SDs mean) and mean square error (MSE) are calculated for WNBCR, CWNBCR, PHNBCR and CPHNBCR models, the summaries are presented in Table 3.17 and 3.18. We note that the bias and MSE of parameter Λ_{12} are lager than others in all fitting models. The estimator of Λ_{12} present a negative biases for the WNBCR, PHNBCR and CPHNBCR models and it present a positive biases for the CWNBCR model, however its biases and MSEs are always near zero. Moreover, for both models, the simulation results for the models considering the prior 1 are very close to those obtained using the prior 2.

Table 3.17: Simulation results for WNBCR and CWNBCR models with depended spatial fragilities

WNBCR model						
Parameter	True value	Estimate mean	SD of the estimate	Bias	MSE	SDs mean
Prior 1: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.9, 1))$						
b_0	-1.5	-1.5688	0.0654	-0.0688	0.0090	0.2770
b_1	-0.5	-0.5362	0.1096	-0.0362	0.0133	0.2637
β	-0.2	-0.1306	0.0606	0.0194	0.0040	0.1852
α	0.3	0.2105	0.0465	-0.0895	0.0102	0.0667
Λ_{11}	4.0	4.1621	0.1860	0.1621	0.0608	2.5225
Λ_{22}	4.0	3.8009	0.1730	-0.1991	0.0695	2.5500
Λ_{12}	0.0	-0.3843	0.1649	-0.3843	0.1748	1.9049
a	0.9	0.8998	0.0017	-0.0002	0.0000	0.0656
η	0.4	0.5048	0.0362	0.1048	0.0123	0.2571
Prior 2: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.9, 1))$						
b_0	-1.5	-1.5725	0.0653	-0.0725	0.0095	0.2767
b_1	-0.5	-0.5367	0.1024	-0.0367	0.0118	0.2645
β	-0.2	-0.1291	0.0567	0.0209	0.0036	0.1850
α	0.3	0.2122	0.0475	-0.0878	0.0100	0.0671
Λ_{11}	4.0	4.1815	0.1850	0.1815	0.0671	2.5383
Λ_{22}	4.0	3.8016	0.1744	-0.1984	0.0697	2.5529
Λ_{12}	0.0	-0.3764	0.1593	-0.3764	0.1670	1.8995
a_1	0.9	0.9000	0.0015	0.0000	0.0000	0.0655
a_2	0.9	0.8999	0.0015	-0.0001	0.0000	0.0656
η	0.4	0.5037	0.0335	0.1037	0.0119	0.2569
CWNBCR model						
Parameter	True value	Estimate mean	SD of the estimate	Bias	MSE	SDs mean
Prior 1: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.75, 1))$						
b_0	-1.5	-1.5271	0.0598	-0.0271	0.0043	0.2750
b_1	-0.5	-0.4567	0.0883	0.0433	0.0097	0.2818
β	-0.2	-0.1242	0.0913	0.0258	0.0090	0.1478
α	0.3	0.3941	0.0534	0.0941	0.0117	0.0593
Λ_{11}	4.0	3.9625	0.2996	-0.0375	0.0910	2.2131
Λ_{22}	4.0	3.8841	0.1857	-0.1159	0.0479	2.6307
Λ_{12}	0.0	0.1555	0.1799	0.1555	0.0565	1.7815
a	0.9	0.9000	0.0015	0.0000	0.0000	0.0655
η	0.4	0.3177	0.0347	-0.0823	0.0080	0.2096
Prior 2: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.75, 1))$						
b_0	-1.5	-1.5269	0.0607	-0.0269	0.0044	0.2741
b_1	-0.5	-0.4565	0.0874	0.0435	0.0095	0.2809
β	-0.2	-0.1235	0.0938	0.0265	0.0095	0.1484
α	0.3	0.3939	0.0500	0.0939	0.0113	0.0597
Λ_{11}	4.0	3.9828	0.2942	-0.0172	0.0867	2.2170
Λ_{22}	4.0	4.0973	0.2033	0.0973	0.0507	2.7874
Λ_{12}	0.0	0.1524	0.1836	0.1524	0.0569	1.8314
a_1	0.9	0.9000	0.0015	0.0000	0.0000	0.0654
a_2	0.9	0.9000	0.0015	0.0000	0.0000	0.0652
η	0.4	0.3213	0.0347	-0.0787	0.0074	0.2114

Table 3.18: Simulation results for PHNBCR and CPHNBCR models with depended spatial fragilities

PHNBCR model						
Parameter	True value	Estimate mean	SD of the estimate	Bias	MSE	SDs mean
Prior 1: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.85, 1.00))$						
b_0	-1.50	-1.5807	0.0593	-0.0807	0.0100	0.2720
b_1	-0.50	-0.5219	0.0951	-0.0219	0.0095	0.2695
β	-0.15	-0.1705	0.0642	-0.0205	0.0045	0.1801
α	1.00	0.9571	0.0338	-0.0429	0.0030	0.1881
Λ_{11}	4.00	4.1890	0.1848	0.1890	0.0698	2.4506
Λ_{22}	4.00	3.7513	0.1797	-0.2487	0.0941	2.5440
Λ_{12}	0.00	-0.4464	0.1262	-0.4464	0.2152	1.8658
a	0.90	0.8999	0.0016	-0.0001	0.0000	0.0653
η	0.40	0.3023	0.0553	-0.0977	0.0126	0.1930
Prior 2: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.85, 1.00))$						
b_0	-1.50	-1.5796	0.0574	-0.0796	0.0096	0.2731
b_1	-0.50	-0.5049	0.0984	-0.0049	0.0097	0.2688
β	-0.15	-0.1810	0.0624	-0.0310	0.0048	0.1789
α	0.30	0.9573	0.0340	-0.0427	0.0030	0.1879
Λ_{11}	4.00	3.9937	0.1613	-0.0063	0.0260	2.3200
Λ_{22}	4.00	3.7484	0.1787	-0.2516	0.0952	2.5440
Λ_{12}	0.00	-0.4434	0.1199	-0.4434	0.2109	1.8120
a_1	0.90	0.8999	0.0015	-0.0001	0.0000	0.0655
a_2	0.90	0.8999	0.0016	-0.0001	0.0000	0.0653
η	0.40	0.2999	0.0573	-0.1001	0.0133	0.1921
CPHNBCR model						
Parameter	True value	Estimate mean	SD of the estimate	Bias	MSE	SDs mean
Prior 1: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.85, 1.00))$						
b_0	-1.50	-1.6574	0.0910	-0.1574	0.0330	0.2736
b_1	-0.50	-0.4931	0.0977	0.0069	0.0096	0.2757
β	-0.15	-0.1880	0.1041	-0.0380	0.0123	0.1386
α	0.30	0.8189	0.0567	-0.1811	0.0360	0.1800
Λ_{11}	4.00	4.9581	0.4269	0.9581	1.0999	2.8623
Λ_{22}	4.00	3.4462	0.3012	-0.5538	0.3972	2.3753
Λ_{12}	0.00	-0.2046	0.2298	-0.2046	0.0946	1.9200
a	0.90	0.8999	0.0017	-0.0001	0.0000	0.0657
η	0.40	0.4420	0.0985	0.0420	0.0115	0.2305
Prior 2: $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.85, 1.00))$						
b_0	-1.50	-1.6572	0.0942	-0.1572	0.0336	0.2725
b_1	-0.50	-0.5047	0.1013	-0.0047	0.0103	0.2756
β	-0.15	-0.1890	0.0998	-0.0390	0.0115	0.1391
α	0.30	0.8346	0.0510	-0.1655	0.0300	0.1840
Λ_{11}	4.00	4.0689	0.3224	0.0689	0.1084	2.3206
Λ_{22}	4.00	3.6089	0.3077	-0.3911	0.2474	2.5019
Λ_{12}	0.00	-0.1823	0.1932	-0.1823	0.0705	1.7578
a_1	0.90	0.9000	0.0018	0.0000	0.0000	0.0655
a_2	0.90	0.8998	0.0017	-0.0002	0.0000	0.0654
η	0.40	0.4373	0.0921	0.0373	0.0099	0.2304

Influence of outlying observations

One of our main goals in this study is to show the need for robust models to deal with the presence of outliers in the data. Considering the same the parameter values and setup as above and two cases for perturbation, thus eight data sets of size 100 were generated from the WNBCR, CWNBCR, PHNBCR and CPHNBCR models with depended spatial fragilities.

We selected cases 18 and 80 for perturbation. To create influential observation in the data set, we choose one or two of these selected cases and perturbed the response variable as follows $\widetilde{t_{kL}} = t_{kL} + 10S_L$ and $\widetilde{t_{kR}} = t_{kR} + 10S_L$, for $k = 1$ and 18, where S_L is the standard deviations of the t_{ijL}'s. Note that using this kind of perturbation, the interval of observed interval time of perturbation candidate observation is not charged. Here, we considere four setups in the study. Setup A: original dataset, without outliers; Setup B: data with outlier 18; Setup C: data with outlier 80; and Setup D: data with outliers 18 and 80. The MCMC computations were made similar to those in the last section and further to monitor the convergence of the Gibbs samples we used the Geweke's convergence diagnostic proposed por Geweke (1992).

Tables 3.19, 3.20, 3.21 and 3.22 reports posterior mean, standard deviation (SD), bias and mean square error (MSE) of the parameters of WNBCR, CWNBCR, PHNBCR and CPHNBCR models, respectively. For WNBCR model, Table 3.19 shows that the absolute values of bias of estimates creasing little bit in the perturbation cases when prior 1 is used for the parameters. On the other way, considering prior 2 for the parameters, the estimates of all parameters of cases B, C and D are very closed the case A, which means the parameters are not sensitive to perturbations. It also can be observed on the Table 3.20. For PHNBCR model, Table 3.21 shows that parameter Λ_{11} is little sensitive to perturbations. The estimates of Λ_{11} decreasing in the perturbation cases when considering prior 1 or prior 2 for the parameters and we obtained similarly simulation results considering both priors.

For CPHNBCR model, considering prior 1, the parameter η is little sensitive in perturbation cases and Λ_{11} is sensitive in case B; considering prior 2, the parameters η and Λ_{12} is litter sensitive in perturbation cases, which can be observed on Table 3.22.

For each simulated data set the four divergence measures (d_{KL}, d_J, d_{L_1}, d_{χ^2}) of the perturbed cases and DIC values for four cure rate models were calculated and reported in Table 3.23. We can see that all measures providing larger ψ-divergence measures when compared to the non-perturbed setup (setup A) and the difference between the measures of perturbed case and non-perturbed case is more clearly for CWNBCR, PHNBCR and CPHNBCR models than WNBCR model. we can note that, the obtained measures values from WNBCR model for data with outlier 18 (setup B) are close

to the setup A. Furthermore, we observed that the obtained measures values from the CWNBCR, PHNBCR and CPHNBCR models wheatear considering the prior 1 or prior 2 are similarly. However, the measures values from the WNBCR model for setup C considering the prior 2 for the parameters are much larger than values obtained using the prior 1. To show better the results, we plot the J-distance divergence measure from the WNBCR model considering the prior 1 and 2 and J-distance divergence measure from the CWNBCR, PHNBCR and CPHNBCR models considering only prior 1.

The Figures 3.41 to 3.72 show the divergence measures before the perturbation (setup A) and after perturbation observations (setups B, C and D). The Figures 3.41 to 3.48 show that outline observation 18 can not be detected by four divergence measures, the outline observation 80 just can be detected by four divergence measures when parameters have prior 2, and the χ^2-squre divergence more sensible than other three measures for the WNBCR model considering prior 2 for the parameters. Moreover, estimates of all parameters of perturbed case are very closed the non-perturbed case, we can conclude that the WNBCR model is not sensitive with this kind of perturbation. The Figures 3.49 to 3.56 show that all perturbation observations selected can be detected by four divergence measures for the CWNBCR model considering prior 1 for the parameters. When the parameters have prior 2, only the KL and χ^2-square divergences can detect the perturbation observations, L_1 norm and $J-$distance cannot detected the perturbation observations when the both observations were perturbed (Setup D). The Figures 3.55(a), 3.56(a), and 3.56(c) also show that the χ^2-square divergence is more sensible than other measures. The Figures 3.57 to 3.64 show that all perturbation observations selected can be detected by all divergence measures for the PHNBCR model considering prior 1 for the parameters. However, the perturbation observation 18 cannot be detected by χ^2-square divergence when the parameters have prior 2, and the Figures 3.65 to 3.72 show that all perturbation observations selected can be detected by all divergence measures for the CPHNBCR model considering prior 1 or 2 for the parameters.

Table 3.19: Simulation results of the perturbed cases for WNBCR model

WNBCR model										
Setup	Perturbed case	Prior 1					Prior 2			
		Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	-1.566	0.280	-0.066	0.004	b_0	-1.557	0.276	-0.057	0.003
		b_1	-0.443	0.271	0.057	0.003	b_1	-0.530	0.254	-0.030	0.001
		β	-0.196	0.187	-0.046	0.002	β	-0.138	0.182	0.012	<0.001
		α	0.206	0.065	-0.094	0.009	α	0.203	0.073	-0.097	0.009
		Λ_{11}	4.435	2.547	0.435	0.189	Λ_{11}	4.319	2.600	0.319	0.102
		Λ_{22}	3.815	2.662	-0.185	0.034	Λ_{22}	3.875	2.586	-0.125	0.016
		Λ_{12}	-0.352	1.919	-0.352	0.124	Λ_{12}	-0.395	1.980	-0.395	0.156
		a	0.900	0.063	<0.001	<0.001	a_1	0.902	0.063	0.002	<0.001
		η	0.469	0.246	0.069	0.005	a_2	0.902	0.064	0.002	<0.001
							η	0.513	0.252	0.113	0.013
B	{18}	b_0	-1.609	0.278	-0.109	0.012	b_0	-1.560	0.283	-0.060	0.004
		b_1	-0.591	0.267	-0.091	0.008	b_1	-0.565	0.258	-0.065	0.004
		β	-0.097	0.184	0.053	0.003	β	-0.114	0.189	0.036	0.001
		α	0.240	0.068	-0.060	0.004	α	0.264	0.070	-0.036	0.001
		Λ_{11}	3.874	2.393	-0.126	0.016	Λ_{11}	4.280	2.536	0.280	0.079
		Λ_{22}	3.499	2.460	-0.501	0.251	Λ_{22}	3.817	2.541	-0.183	0.033
		Λ_{12}	-0.132	1.866	-0.132	0.017	Λ_{12}	-0.439	1.879	-0.439	0.192
		a	0.900	0.065	<0.001	<0.001	a_1	0.902	0.064	0.002	<0.001
			0.537	0.257	0.137	0.019	a_2	0.902	0.064	0.002	<0.001
							η	0.512	0.242	0.112	0.013
C	{80}	b_0	-1.643	0.267	-0.143	0.020	b_0	-1.491	0.275	0.009	<0.001
		b_1	-0.358	0.268	0.142	0.020	b_1	-0.479	0.262	0.021	<0.001
		β	-0.258	0.179	-0.108	0.012	β	-0.171	0.187	-0.021	<0.001
		α	0.245	0.061	-0.055	0.003	α	0.163	0.059	-0.137	0.019
		Λ_{11}	4.182	2.477	0.182	0.033	Λ_{11}	4.517	2.654	0.517	0.267
		Λ_{22}	3.749	2.614	-0.251	0.063	Λ_{22}	3.876	2.526	-0.124	0.015
		Λ_{12}	-0.309	1.908	-0.309	0.096	Λ_{12}	-0.623	2.004	-0.623	0.389
		a	0.901	0.065	0.001	<0.001	a_1	0.903	0.064	0.003	<0.001
		η	0.481	0.261	0.081	0.007	a_2	0.901	0.066	0.001	<0.001
							η	0.449	0.241	0.049	0.002
D	{18,80}	b_0	-1.611	0.267	-0.111	0.012	b_0	-1.562	0.286	-0.062	0.004
		b_1	-0.746	0.261	-0.246	0.061	b_1	-0.641	0.259	-0.141	0.020
		β	-0.019	0.186	0.131	0.017	β	-0.082	0.185	0.068	0.005
		α	0.196	0.066	-0.104	0.011	α	0.210	0.065	-0.090	0.008
		Λ_{11}	4.465	2.659	0.465	0.216	Λ_{11}	4.243	2.506	0.243	0.059
		Λ_{22}	3.981	2.617	-0.019	<0.001	Λ_{22}	3.782	2.610	-0.218	0.047
		Λ_{12}	-0.527	1.955	-0.527	0.277	Λ_{12}	-0.411	1.900	-0.411	0.169
		a	0.903	0.061	0.003	<0.001	a_1	0.900	0.065	<0.001	<0.001
		η	0.559	0.264	0.159	0.025	a_2	0.899	0.064	-0.001	<0.001
							η	0.509	0.253	0.109	0.012

Table 3.20: Simulation results of the perturbed cases for CWNBCR model

CWNBCR model										
Setup	Perturbed case	Prior 1					Prior 2			
		Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	-1.533	0.269	-0.033	0.001	b_0	-1.464	0.275	0.036	0.001
		b_1	-0.425	0.281	0.075	0.006	b_1	-0.459	0.295	0.041	0.002
		β	-0.129	0.149	0.021	<0.001	β	-0.102	0.163	0.048	0.002
		α	0.403	0.063	0.103	0.011	α	0.370	0.059	0.070	0.005
		Λ_{11}	3.877	2.134	-0.123	0.015	Λ_{11}	4.057	2.275	0.057	0.003
		Λ_{22}	4.074	2.712	0.074	0.005	Λ_{22}	3.992	2.691	-0.008	<0.001
		Λ_{12}	0.260	1.801	0.260	0.068	Λ_{12}	0.009	1.827	0.009	<0.001
		a	0.901	0.065	0.001	<0.001	a_1	0.901	0.062	0.001	<0.001
		η	0.319	0.209	-0.081	0.007	a_2	0.900	0.063	<0.001	<0.001
							η	0.312	0.208	-0.088	0.008
B	{18}	b_0	-1.622	0.271	-0.122	0.015	b_0	-1.536	0.270	-0.036	0.001
		b_1	-0.469	0.284	0.031	0.001	b_1	-0.385	0.272	0.115	0.013
		β	-0.122	0.145	0.028	0.001	β	0.070	0.147	0.220	0.048
		α	0.402	0.053	0.102	0.010	α	0.356	0.050	0.056	0.003
		Λ_{11}	4.343	2.422	0.343	0.117	Λ_{11}	3.988	2.212	-0.012	<0.001
		Λ_{22}	4.041	2.740	0.041	0.002	Λ_{22}	3.989	2.629	-0.011	<0.001
		Λ_{12}	0.223	1.809	0.223	0.050	Λ_{12}	0.309	1.854	0.309	0.095
		a	0.899	0.067	-0.001	<0.001	a_1	0.901	0.063	0.001	<0.001
		η	0.342	0.210	-0.058	0.003	a_2	0.902	0.065	0.002	<0.001
							η	0.300	0.216	-0.100	0.010
C	{80}	b_0	-1.503	0.270	-0.003	<0.001	b_0	-1.439	0.273	0.061	0.004
		b_1	-0.387	0.298	0.113	0.013	b_1	-0.361	0.269	0.139	0.019
		β	-0.065	0.148	0.085	0.007	β	0.051	0.151	0.201	0.040
		α	0.426	0.065	0.126	0.016	α	0.413	0.066	0.113	0.013
		Λ_{11}	4.324	2.387	0.324	0.105	Λ_{11}	3.944	2.270	-0.056	0.003
		Λ_{22}	3.839	2.587	-0.161	0.026	Λ_{22}	3.835	2.634	-0.165	0.027
		Λ_{12}	0.043	1.852	0.043	0.002	Λ_{12}	-0.239	1.809	-0.239	0.057
		a	0.899	0.068	-0.001	<0.001	a_1	0.901	0.063	0.001	<0.001
		η	0.296	0.201	-0.104	0.011	a_2	0.901	0.066	0.001	<0.001
							η	0.334	0.212	-0.066	0.004
D	{18,80}	b_0	-1.526	0.276	-0.026	0.001	b_0	-1.599	0.266	-0.099	0.010
		b_1	-0.479	0.279	0.021	<0.001	b_1	-0.405	0.275	0.095	0.009
		β	-0.124	0.140	0.026	0.001	β	0.115	0.142	0.265	0.070
		α	0.354	0.049	0.054	0.003	α	0.362	0.051	0.062	0.004
		Λ_{11}	3.918	2.203	-0.082	0.007	Λ_{11}	4.395	2.444	0.395	0.156
		Λ_{22}	3.755	2.525	-0.245	0.060	Λ_{22}	3.826	2.582	-0.174	0.030
		Λ_{12}	0.419	1.782	0.419	0.175	Λ_{12}	0.033	1.810	0.033	0.001
		a	0.899	0.065	-0.001	<0.001	a_1	0.901	0.065	0.001	<0.001
		η	0.307	0.203	-0.093	0.009	a_2	0.902	0.064	0.002	<0.001
							η	0.288	0.200	-0.112	0.013

Table 3.21: Simulation results of the perturbed cases for PHNBCR model

PHNBCR model										
Setup	Perturbed case	Prior 1					Prior 2			
		Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	-1.595	0.259	-0.095	0.009	b_0	-1.568	0.271	-0.068	0.005
		b_1	-0.420	0.266	0.080	0.006	b_1	-0.358	0.271	0.142	0.020
		β	-0.243	0.177	-0.093	0.009	β	-0.351	0.181	-0.201	0.041
		α	0.871	0.195	-0.129	0.017	α	0.913	0.200	-0.087	0.008
		Λ_{11}	3.813	2.342	-0.187	0.035	Λ_{11}	3.707	2.155	-0.293	0.086
		Λ_{22}	3.882	2.534	-0.118	0.014	Λ_{22}	3.855	2.551	-0.145	0.021
		Λ_{12}	-0.532	1.899	-0.532	0.283	Λ_{12}	-0.718	1.781	-0.718	0.516
		a	0.901	0.065	0.001	<0.001	a_1	0.902	0.064	0.002	<0.001
		η	0.230	0.167	-0.170	0.029	a_2	0.902	0.064	0.002	<0.001
							η	0.274	0.187	-0.126	0.016
B	{18}	b_0	-1.407	0.268	0.093	0.009	b_0	-1.396	0.265	0.104	0.011
		b_1	-0.430	0.261	0.070	0.005	b_1	-0.363	0.262	0.137	0.019
		β	-0.273	0.176	-0.123	0.015	β	-0.298	0.183	-0.148	0.022
		α	0.836	0.199	-0.164	0.027	α	0.824	0.201	-0.176	0.031
		Λ_{11}	3.713	2.306	-0.287	0.082	Λ_{11}	3.629	2.230	-0.371	0.138
		Λ_{22}	3.882	2.461	-0.118	0.014	Λ_{22}	4.066	2.705	0.066	0.004
		Λ_{12}	-0.415	1.824	-0.415	0.173	Λ_{12}	-0.184	1.844	-0.184	0.034
		a	0.900	0.067	0.000	<0.001	a_1	0.898	0.067	-0.002	<0.001
		η	0.401	0.219	0.001	<0.001	a_2	0.901	0.065	0.001	<0.001
							η	0.483	0.253	0.083	0.007
C	{80}	b_0	-1.503	0.270	-0.003	<0.001	b_0	-1.399	0.274	0.101	0.010
		b_1	-0.407	0.267	0.093	0.009	b_1	-0.407	0.259	0.093	0.009
		β	-0.296	0.182	-0.146	0.021	β	-0.161	0.181	-0.011	<0.001
		α	0.902	0.198	-0.098	0.010	α	0.792	0.202	-0.208	0.043
		Λ_{11}	3.202	2.053	-0.798	0.637	Λ_{11}	3.033	1.939	-0.967	0.935
		Λ_{22}	3.953	2.562	-0.047	0.002	Λ_{22}	4.072	2.666	0.072	0.005
		Λ_{12}	-0.403	1.799	-0.403	0.163	Λ_{12}	-0.243	1.750	-0.243	0.059
		a	0.902	0.063	0.002	<0.001	a_1	0.901	0.067	0.001	<0.001
		η	0.405	0.216	0.005	<0.001	a_2	0.903	0.063	0.003	<0.001
							η	0.382	0.225	-0.018	<0.001
D	{18,80}	b_0	-1.488	0.262	0.012	<0.001	b_0	-1.509	0.271	-0.009	<0.001
		b_1	-0.488	0.252	0.012	<0.001	b_1	-0.622	0.249	-0.122	0.015
		β	-0.279	0.177	-0.129	0.017	β	-0.279	0.176	-0.129	0.017
		α	0.822	0.200	-0.178	0.032	α	0.815	0.203	-0.185	0.034
		Λ_{11}	3.183	1.941	-0.817	0.667	Λ_{11}	3.239	1.987	-0.761	0.579
		Λ_{22}	3.814	2.565	-0.186	0.035	Λ_{22}	3.989	2.678	-0.011	<0.001
		Λ_{12}	-0.459	1.749	-0.459	0.210	Λ_{12}	-0.802	1.778	-0.802	0.644
		a	0.902	0.064	0.002	<0.001	a_1	0.900	0.068	0.000	<0.001
		η	0.551	0.230	0.151	0.023	a_2	0.902	0.064	0.002	<0.001
							η	0.425	0.222	0.025	0.001

Table 3.22: Simulation results of the perturbed cases for CPHNBCR model

CPHNBCR model										
Setup	Perturbed case	Prior 1					Prior 2			
		Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	-1.633	0.272	-0.133	0.018	b_0	-1.553	0.265	-0.053	0.003
		b_1	-0.383	0.269	0.117	0.014	b_1	-0.438	0.280	0.062	0.004
		β	-0.071	0.141	0.079	0.006	β	-0.143	0.152	0.007	<0.001
		α	0.773	0.199	-0.227	0.052	α	0.858	0.199	-0.142	0.020
		Λ_{11}	4.439	2.695	0.439	0.192	Λ_{11}	3.208	1.926	-0.792	0.627
		Λ_{22}	3.397	2.349	-0.603	0.363	Λ_{22}	3.800	2.486	-0.200	0.040
		Λ_{12}	-0.250	1.909	-0.250	0.063	Λ_{12}	-0.033	1.776	-0.033	0.001
		a	0.902	0.066	0.002	<0.001	a_1	0.898	0.065	-0.002	<0.001
		η	0.319	0.208	-0.081	0.007	a_2	0.900	0.065	<0.001	<0.001
							η	0.340	0.217	-0.060	0.004
B	{18}	b_0	-1.576	0.280	-0.076	0.006	b_0	-1.666	0.262	-0.166	0.028
		b_1	-0.491	0.286	0.009	<0.001	b_1	-0.483	0.287	0.017	<0.001
		β	-0.415	0.139	-0.265	0.070	β	-0.094	0.143	0.056	0.003
		α	0.685	0.202	-0.315	0.099	α	0.879	0.196	-0.121	0.015
		Λ_{11}	4.721	2.799	0.721	0.519	Λ_{11}	3.447	2.050	-0.553	0.306
		Λ_{22}	3.371	2.313	-0.629	0.396	Λ_{22}	3.595	2.419	-0.405	0.164
		Λ_{12}	0.186	1.946	0.186	0.035	Λ_{12}	-0.390	1.756	-0.390	0.152
		a	0.902	0.063	0.002	<0.001	a_1	0.898	0.066	-0.002	<0.001
		η	0.347	0.223	-0.053	0.003	a_2	0.901	0.065	0.001	<0.001
							η	0.256	0.181	-0.144	0.021
C	{80}	b_0	-1.533	0.268	-0.033	0.001	b_0	-1.688	0.269	-0.188	0.035
		b_1	-0.628	0.270	-0.128	0.017	b_1	-0.444	0.264	0.056	0.003
		β	-0.349	0.153	-0.199	0.040	β	-0.187	0.142	-0.037	0.001
		α	0.749	0.198	-0.251	0.063	α	0.855	0.201	-0.145	0.021
		Λ_{11}	4.086	2.502	0.086	0.007	Λ_{11}	4.000	2.335	<0.001	<0.001
		Λ_{22}	3.707	2.378	-0.293	0.086	Λ_{22}	3.443	2.407	-0.557	0.311
		Λ_{12}	-0.100	1.906	-0.100	0.010	Λ_{12}	-0.416	1.803	-0.416	0.173
		a	0.901	0.066	0.001	<0.001	a_1	0.899	0.066	-0.001	<0.001
		η	0.294	0.198	-0.106	0.011	a_2	0.900	0.065	<0.001	<0.001
							η	0.298	0.200	-0.102	0.010
D	{18,80}	b_0	-1.542	0.266	-0.042	0.002	b_0	-1.656	0.258	-0.156	0.024
		b_1	-0.303	0.283	0.197	0.039	b_1	-0.609	0.278	-0.109	0.012
		β	-0.246	0.149	-0.096	0.009	β	-0.434	0.151	-0.284	0.080
		α	0.638	0.205	-0.362	0.131	α	0.877	0.189	-0.123	0.015
		Λ_{11}	4.109	2.602	0.109	0.012	Λ_{11}	3.498	2.031	-0.502	0.252
		Λ_{22}	3.872	2.504	-0.128	0.016	Λ_{22}	3.645	2.486	-0.355	0.126
		Λ_{12}	0.002	1.947	0.002	<0.001	Λ_{12}	-0.127	1.727	-0.127	0.016
		a	0.900	0.066	<0.001	<0.001	a_1	0.898	0.067	-0.002	<0.001
		η	0.171	0.144	-0.229	0.053	a_2	0.902	0.064	0.002	<0.001
							η	0.253	0.180	-0.147	0.022

Table 3.23: Divergence measures of the perturbed cases and DIC values for the simulated data sets.

Model	Prior	Setup	Case number	d_{KL}	d_J	d_{L1}	d_{χ^2}	DIC
WNBCR	1	A	18	0.019	0.039	0.077	0.042	155.900
			80	0.059	0.125	0.138	0.160	
		B	18	0.049	0.103	0.124	0.124	162.537
		C	80	0.040	0.083	0.112	0.097	168.190
		D	18	0.118	0.256	0.192	0.417	174.948
			80	0.097	0.210	0.174	0.306	
	2	A	18	0.011	0.023	0.060	0.025	144.903
			80	0.009	0.019	0.053	0.021	
		B	18	0.044	0.090	0.118	0.102	160.820
		C	80	0.442	1.690	0.364	69.616	179.784
		D	18	0.090	0.196	0.164	0.313	170.610
			80	0.089	0.194	0.165	0.300	
CWNBCR	1	A	18	0.017	0.035	0.075	0.036	397.683
			80	0.037	0.075	0.110	0.082	
		B	18	0.416	0.963	0.369	3.040	460.719
		C	80	0.585	1.429	0.440	7.355	408.683
		D	18	0.426	1.051	0.376	4.596	428.512
			80	0.417	0.930	0.371	2.187	
	2	A	18	0.038	0.078	0.109	0.089	366.214
			80	0.038	0.078	0.107	0.091	
		B	18	0.616	1.457	0.451	6.008	396.018
		C	80	0.400	0.914	0.365	2.364	399.191
		D	18	0.221	0.495	0.266	1.002	410.241
			80	0.231	0.528	0.274	1.163	
PHNBCR	1	A	18	0.002	0.004	0.025	0.004	262.038
			80	<0.001	<0.001	0.003	<0.001	
		B	18	1.249	3.728	0.624	229.515	274.767
		C	80	0.466	1.140	0.391	5.040	272.823
		D	18	1.281	3.433	0.629	63.147	306.894
			80	1.500	4.096	0.695	105.593	
	2	A	18	0.048	0.099	0.125	0.114	274.876
			80	0.029	0.058	0.096	0.062	
		B	18	1.208	3.045	0.610	37.911	286.468
		C	80	1.095	2.787	0.594	25.740	294.388
		D	18	1.151	2.638	0.596	15.903	281.866
			80	0.407	0.941	0.364	2.671	
CPHNBCR	1	A	18	0.337	0.754	0.333	1.759	427.835
			80	0.025	0.050	0.089	0.053	
		B	18	2.407	5.164	0.765	60.061	445.979
		C	80	2.619	5.240	0.781	38.913	437.124
		D	18	3.202	7.434	0.845	331.339	492.476
			80	2.138	4.935	0.733	114.932	
	2	A	18	0.047	0.097	0.123	0.111	412.991
			80	0.041	0.084	0.114	0.097	
		B	18	1.412	2.873	0.620	10.929	417.466
		C	80	1.412	2.978	0.621	14.312	424.754
		D	18	2.165	5.646	0.749	327.954	459.821
			80	1.117	2.791	0.592	28.511	

WNBCR model

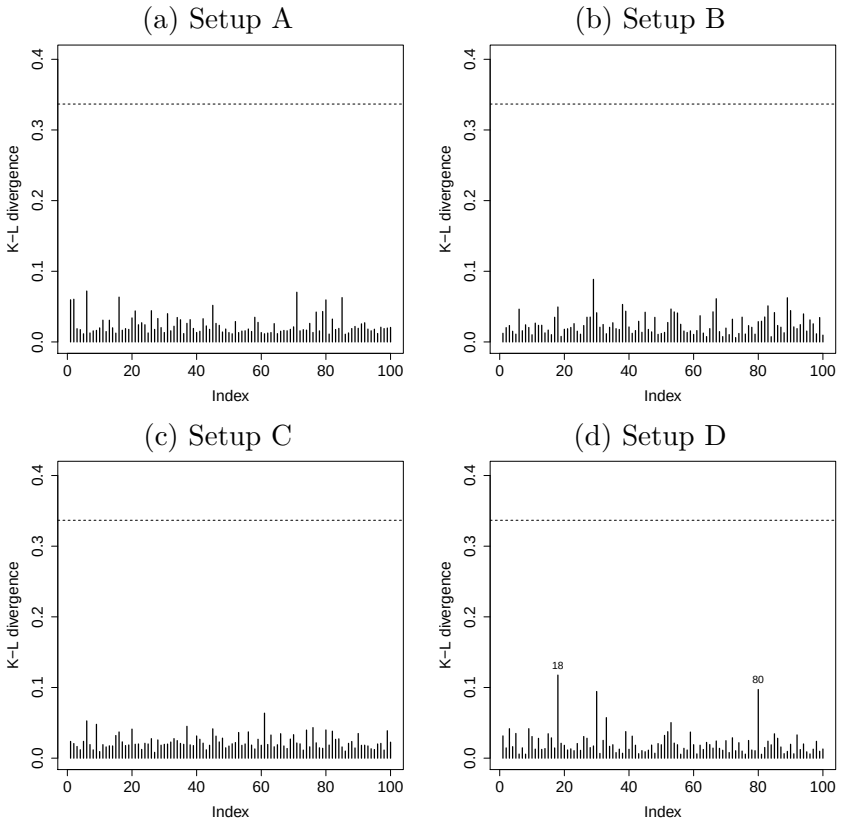

Figure 3.41: Index plots of Kullback-Leibler divergence measure from the fitted WNBCR model considering prior 1.

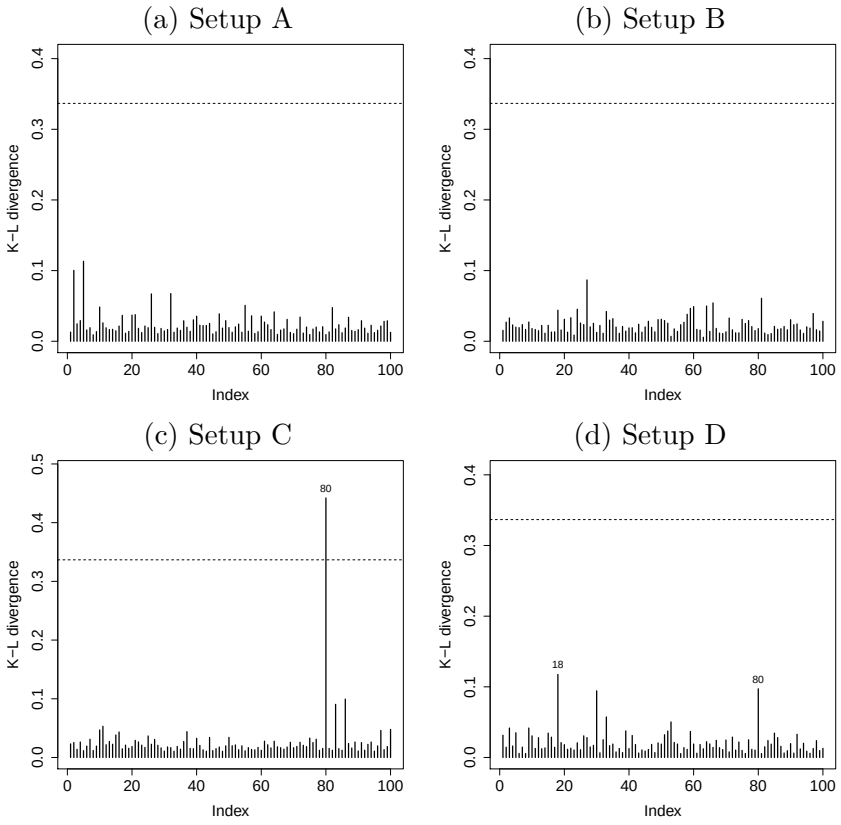

Figure 3.42: Index plots of Kullback-Leibler divergence measure from the fitted WNBCR model considering prior 2.

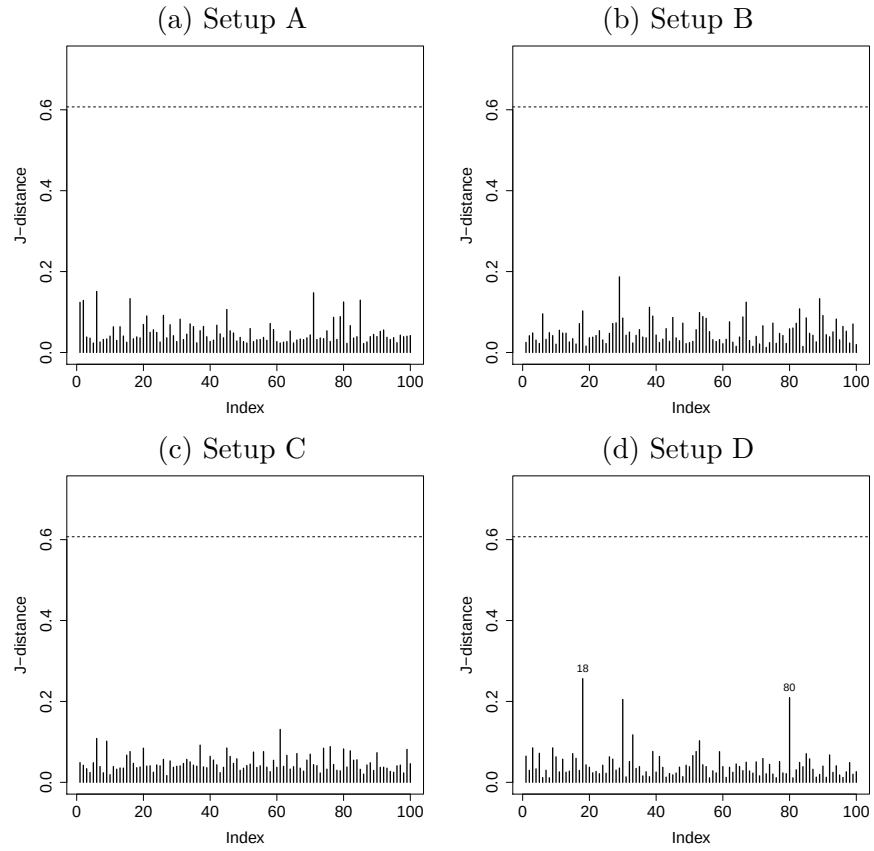

Figure 3.43: Index plots of J-distance from the fitted WNBCR model considering prior 1.

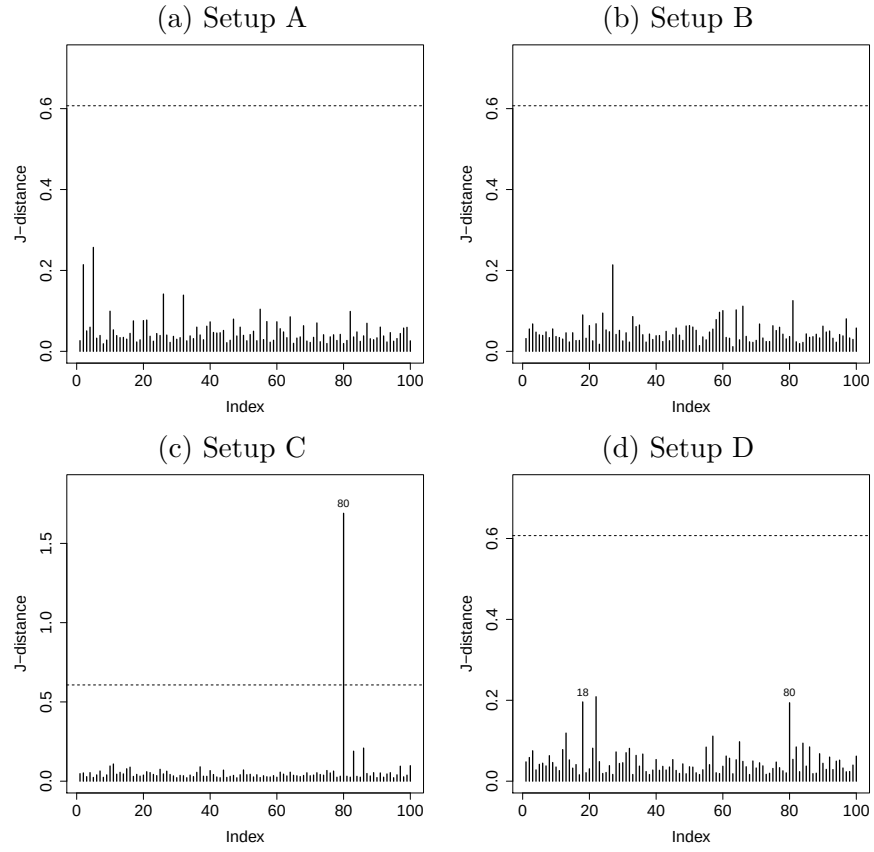

Figure 3.44: Index plots of J-distance from the fitted WNBCR model considering prior 2.

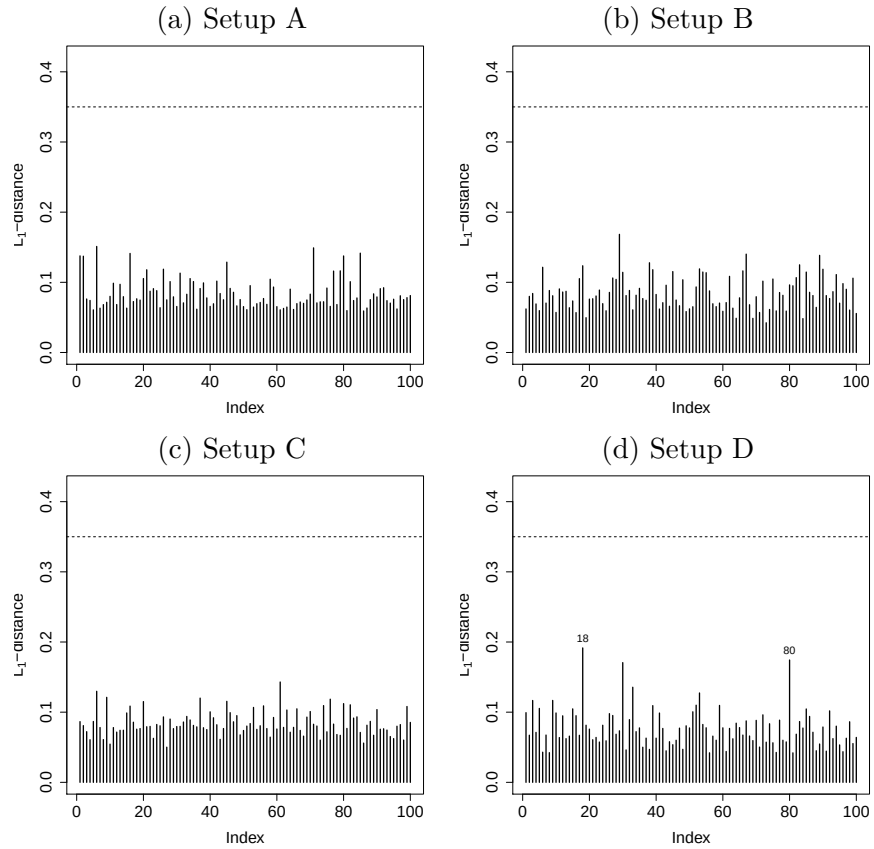

Figure 3.45: Index plots of L_1 norm distance from the fitted WNBCR model considering prior 1.

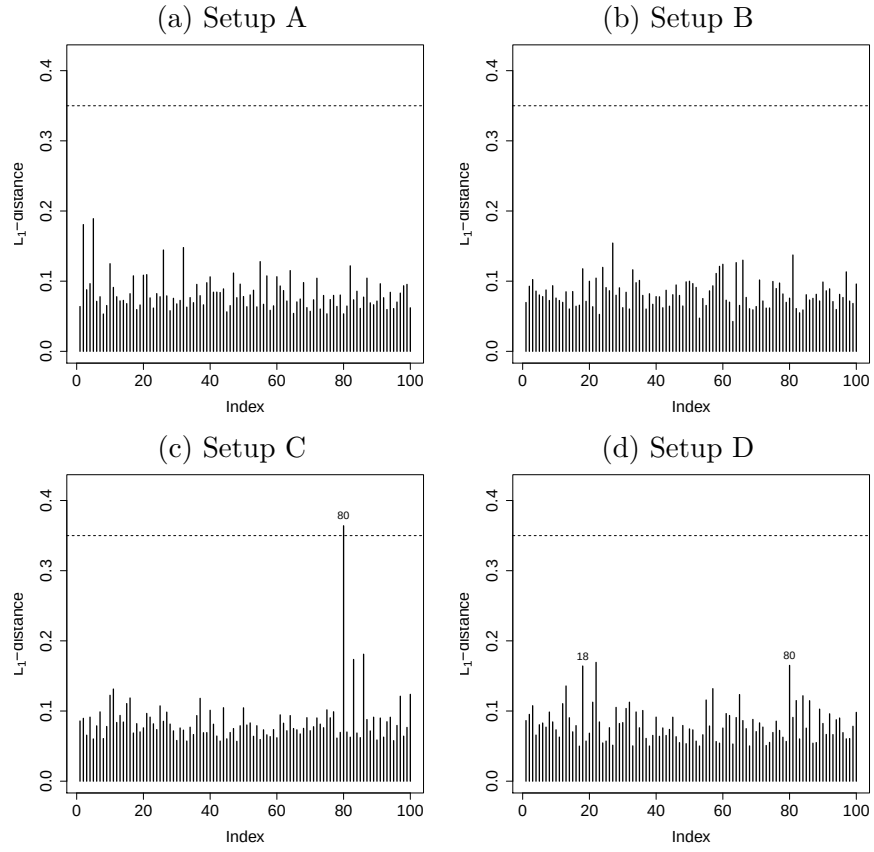

Figure 3.46: Index plots of L_1 norm distance from the fitted WNBCR model considering prior 2.

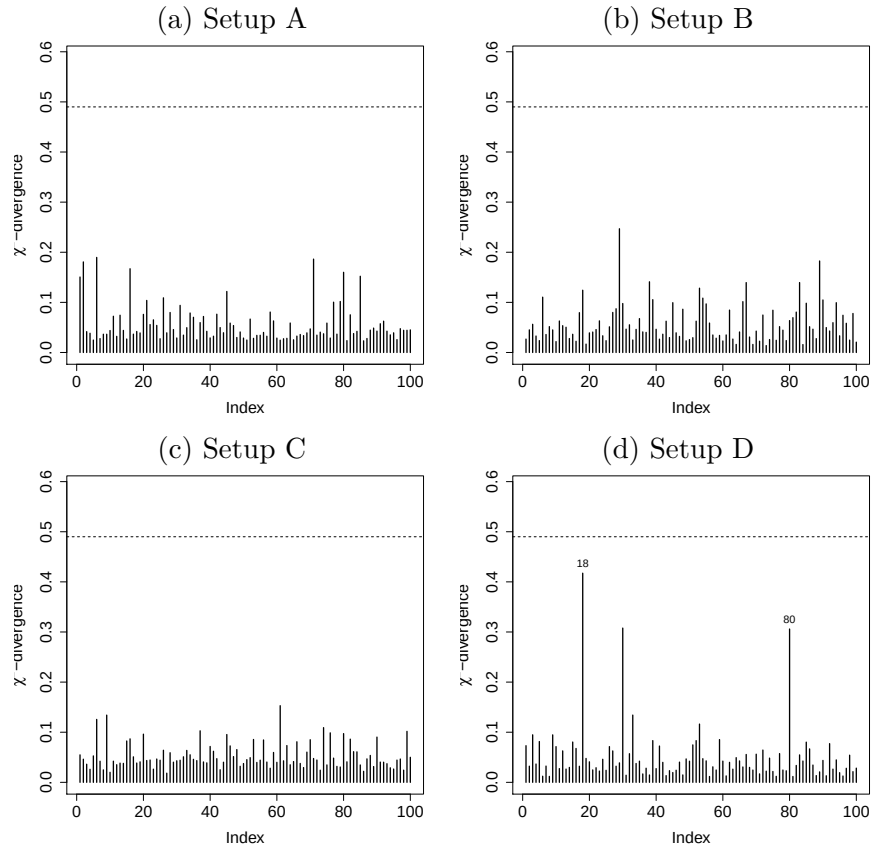

Figure 3.47: Index plots of χ^2-square divergence from the fitted WNBCR model considering prior 1.

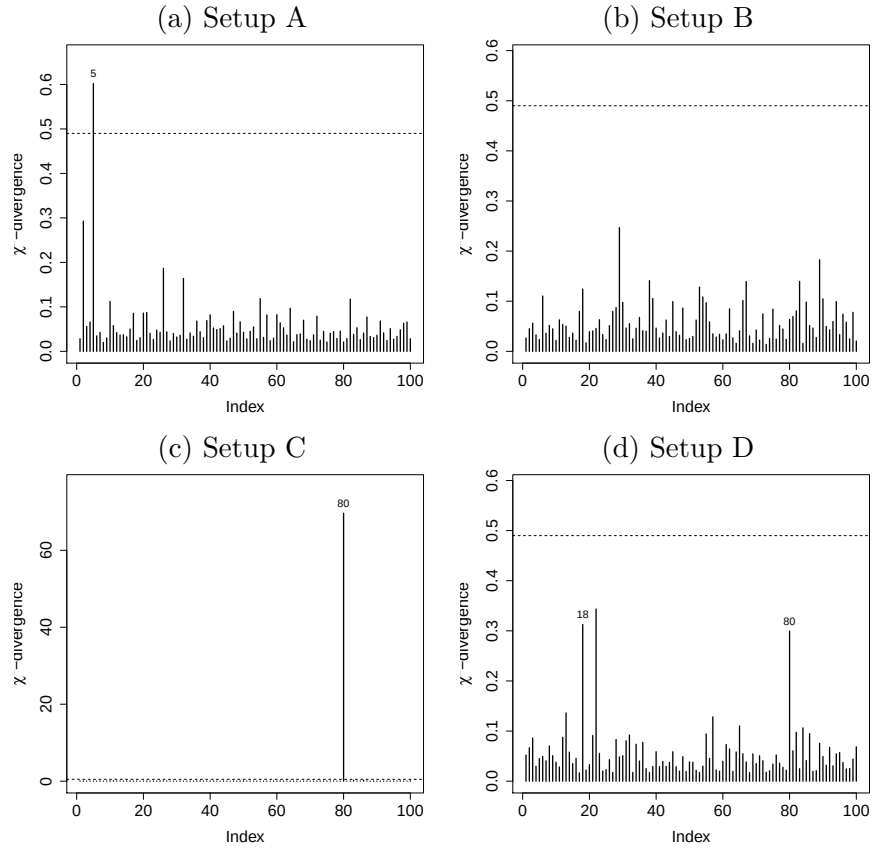

Figure 3.48: Index plots of χ^2-square divergence from the fitted WNBCR model considering prior 2.

CWNBCR model

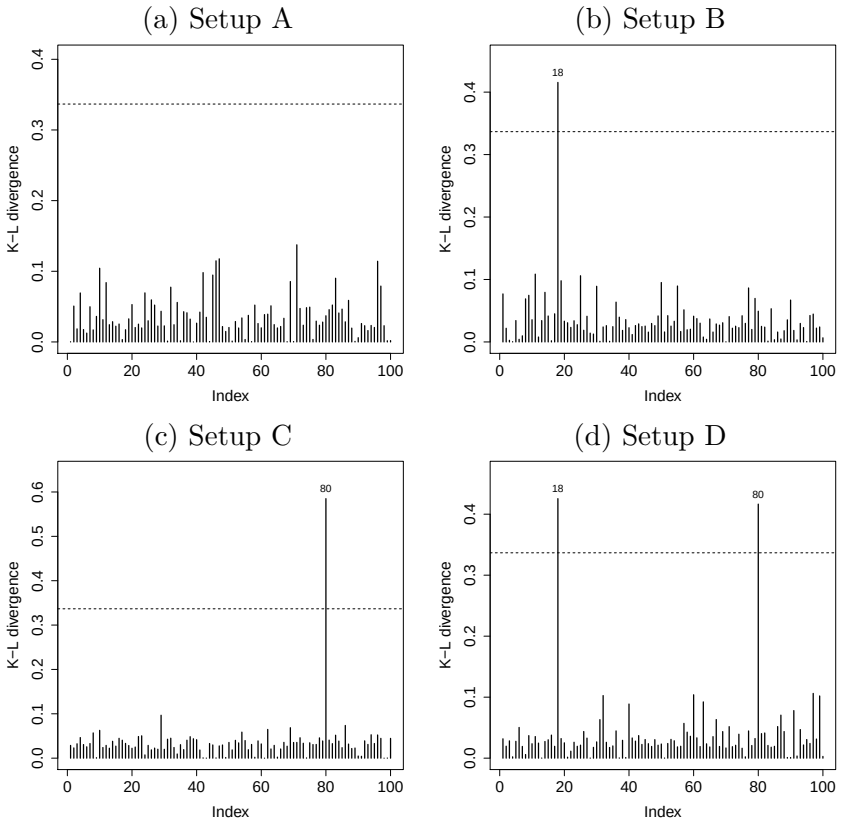

Figure 3.49: Index plots of Kullback-Leibler divergence measure from the fitted CWNBCR model considering prior 1.

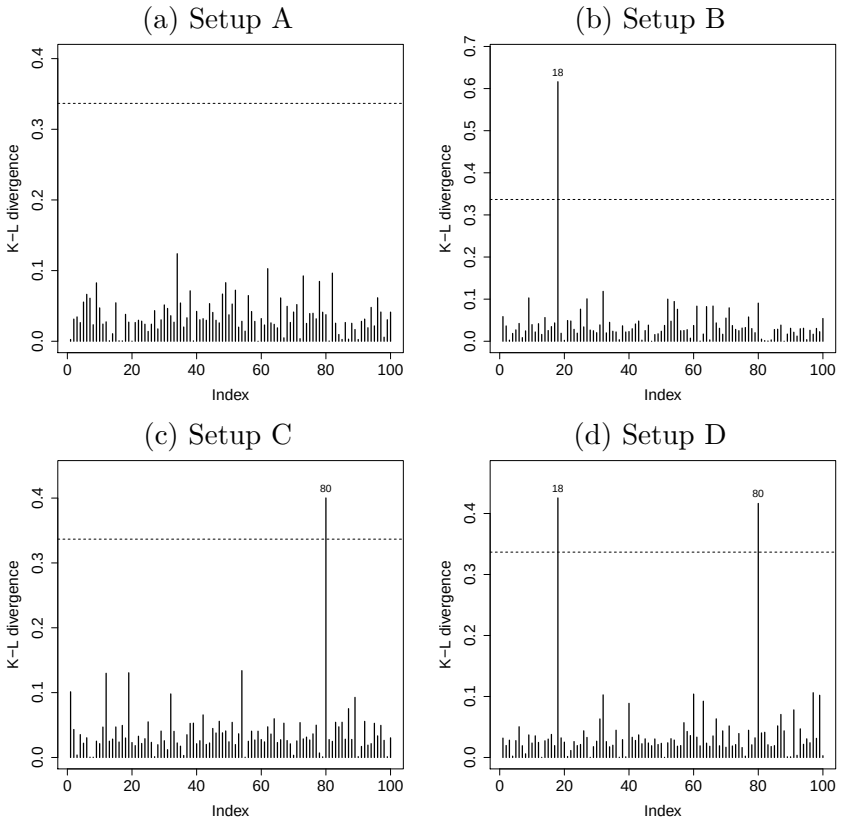

Figure 3.50: Index plots of Kullback-Leibler divergence measure from the fitted CWNBCR model considering prior 2.

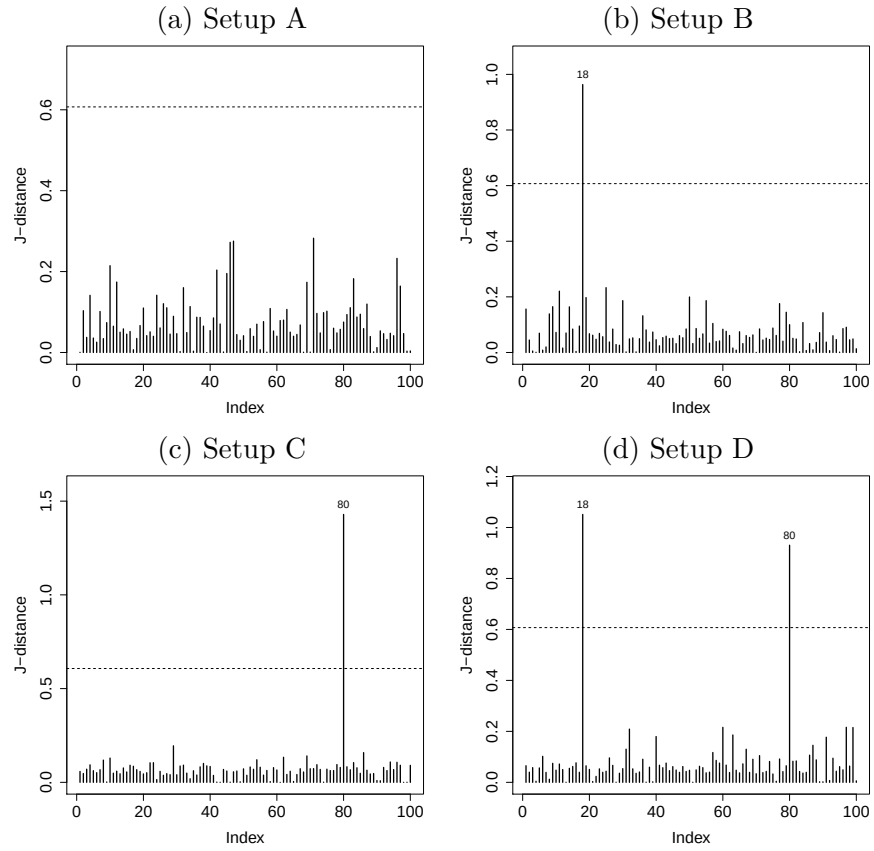

Figure 3.51: Index plots of J-distance from the fitted CWNBCR model considering prior 1.

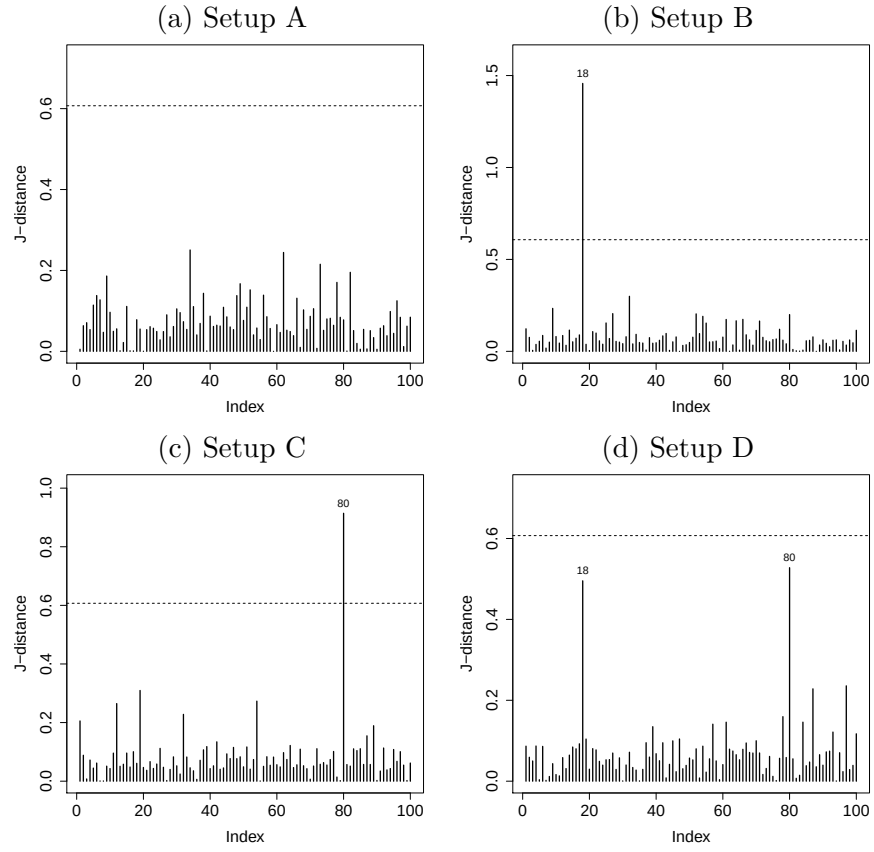

Figure 3.52: Index plots of J-distance from the fitted CWNBCR model considering prior 2.

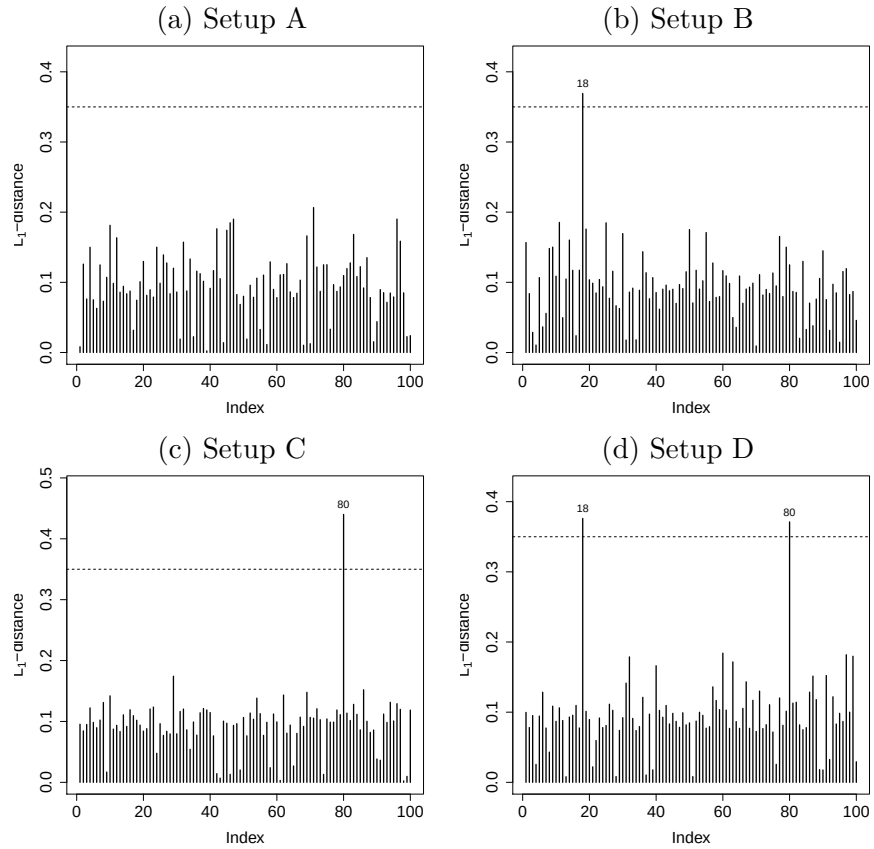

Figure 3.53: Index plots of L_1 norm distance from the fitted CWNBCR model considering prior 1.

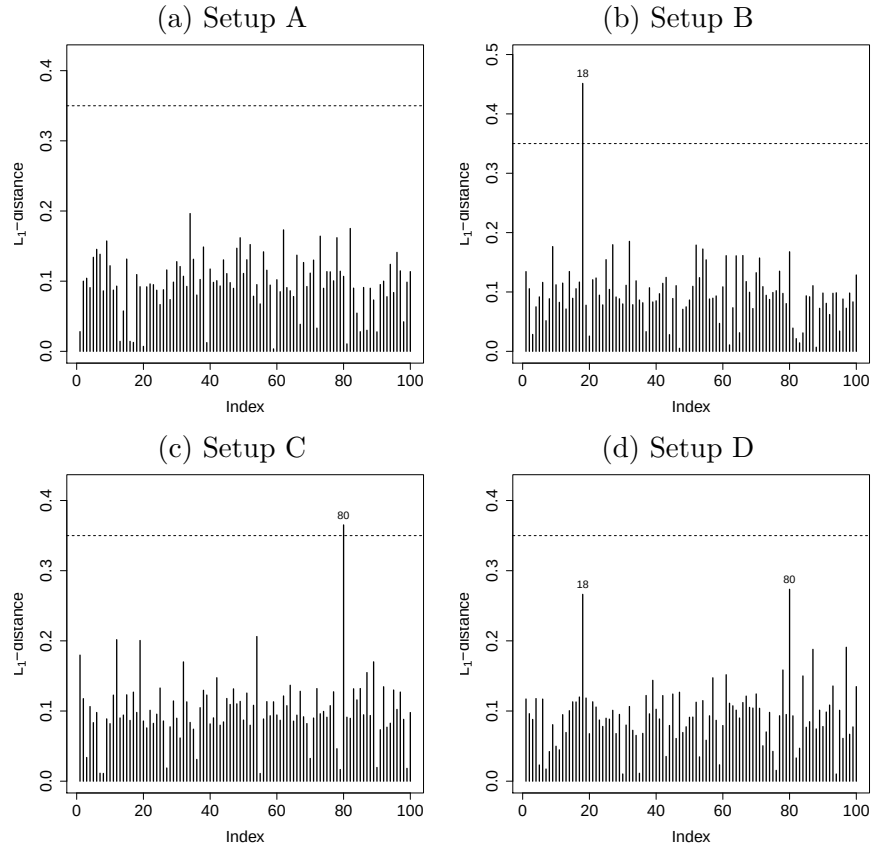

Figure 3.54: Index plots of L_1 norm distance from the fitted CWNBCR model considering prior 2.

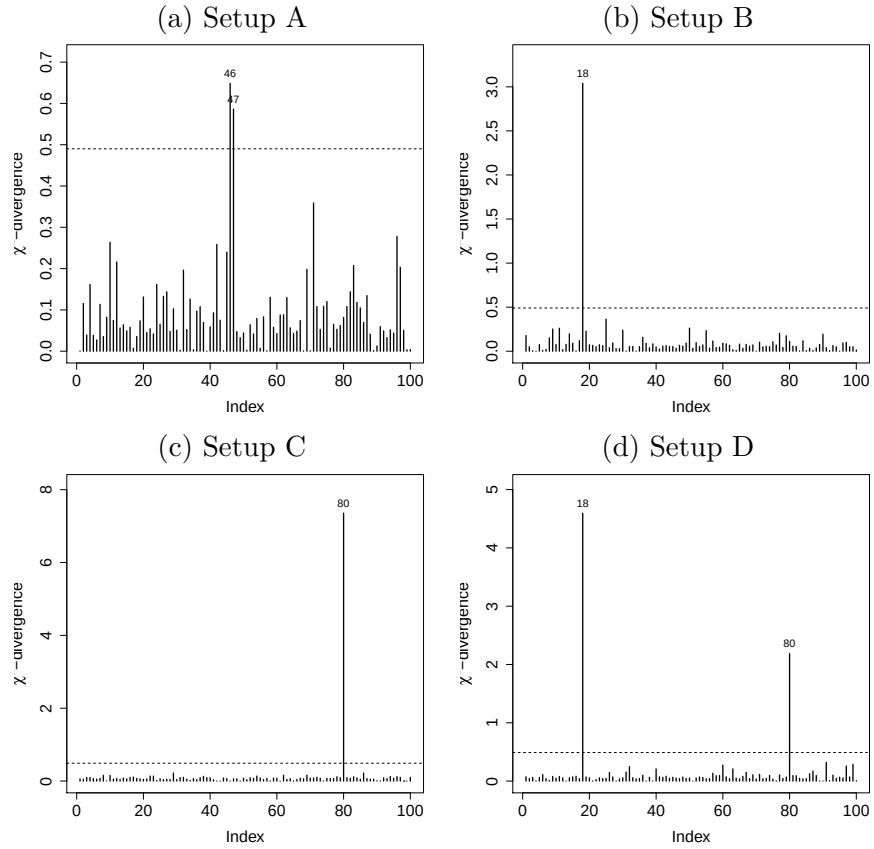

Figure 3.55: Index plots of χ^2-square divergence from the fitted CWNBCR model considering prior 1.

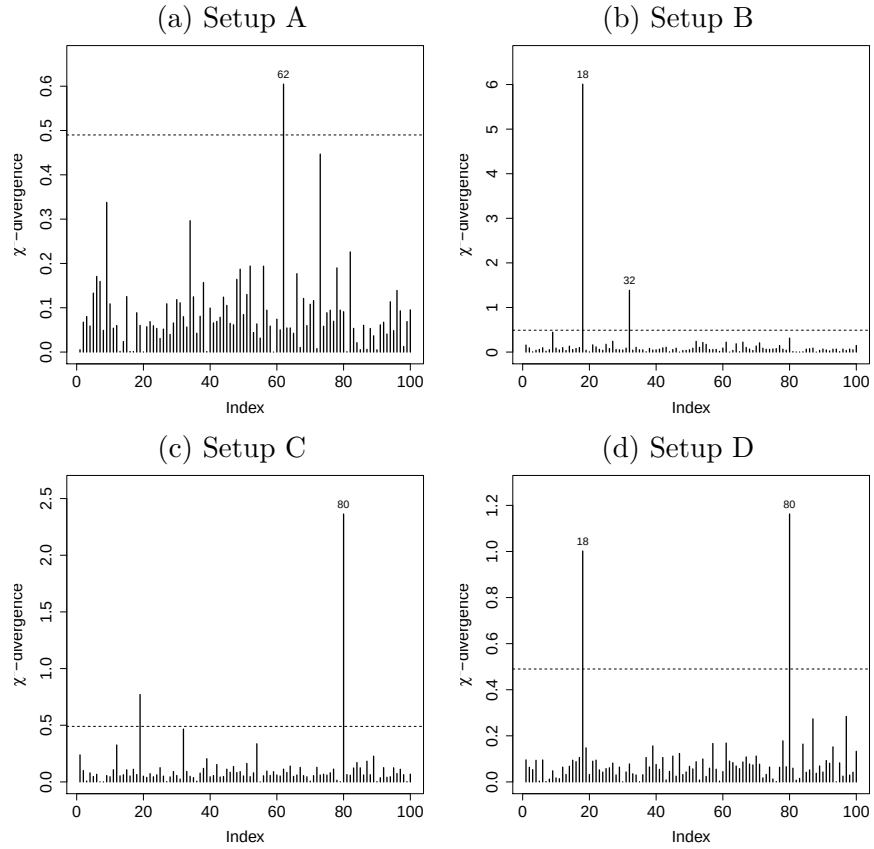

Figure 3.56: Index plots of χ^2-square divergence from the fitted CWNBCR model considering prior 2.

PHNBCR model

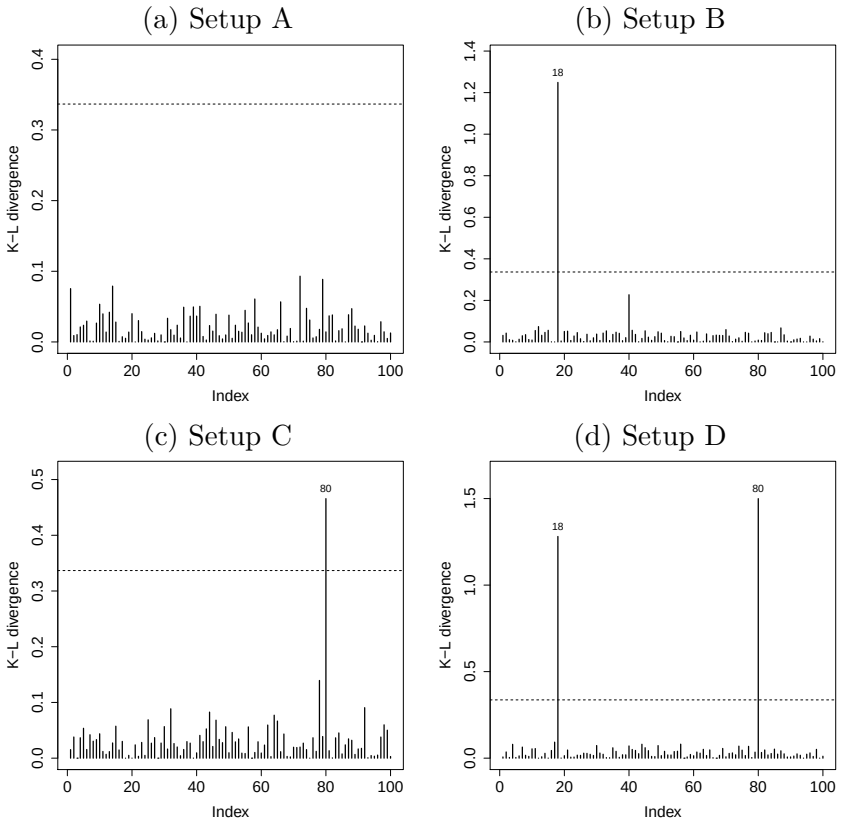

Figure 3.57: Index plots of Kullback-Leibler divergence measure from the fitted PHNBCR model considering prior 1.

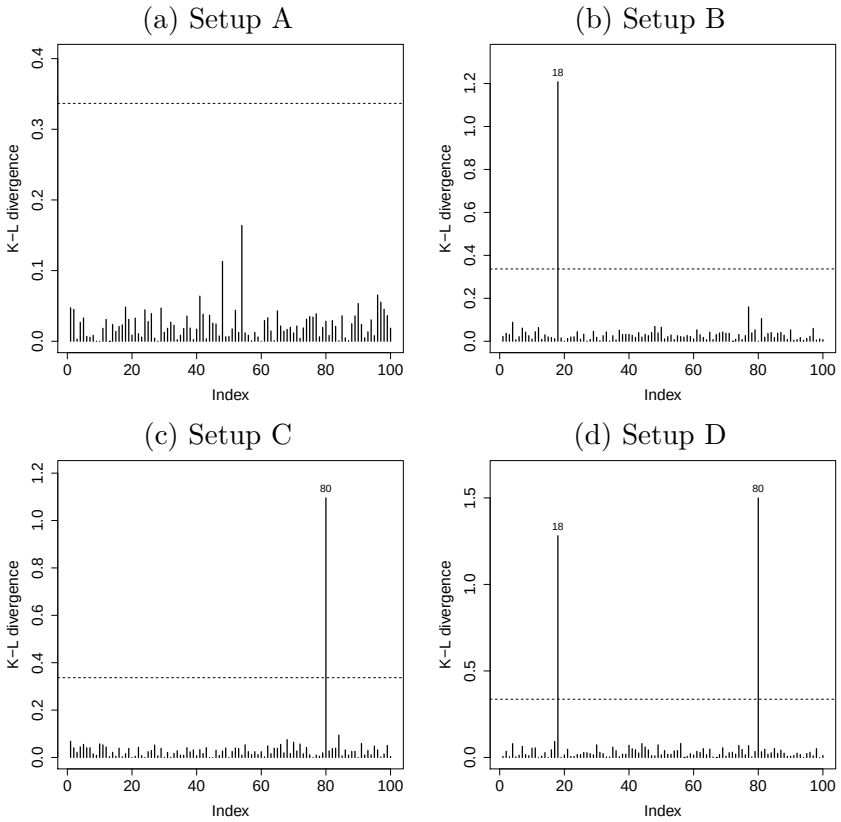

Figure 3.58: Index plots of Kullback-Leibler divergence measure from the fitted PHNBCR model considering prior 2.

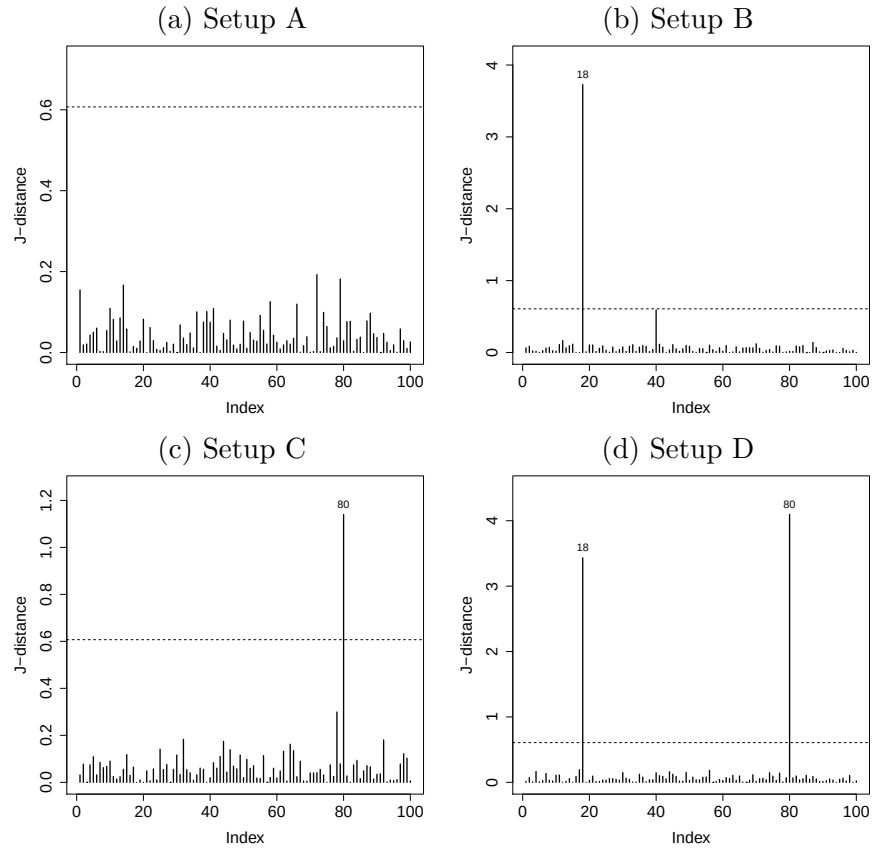

Figure 3.59: Index plots of J-distance from the fitted PHNBCR model considering prior 1.

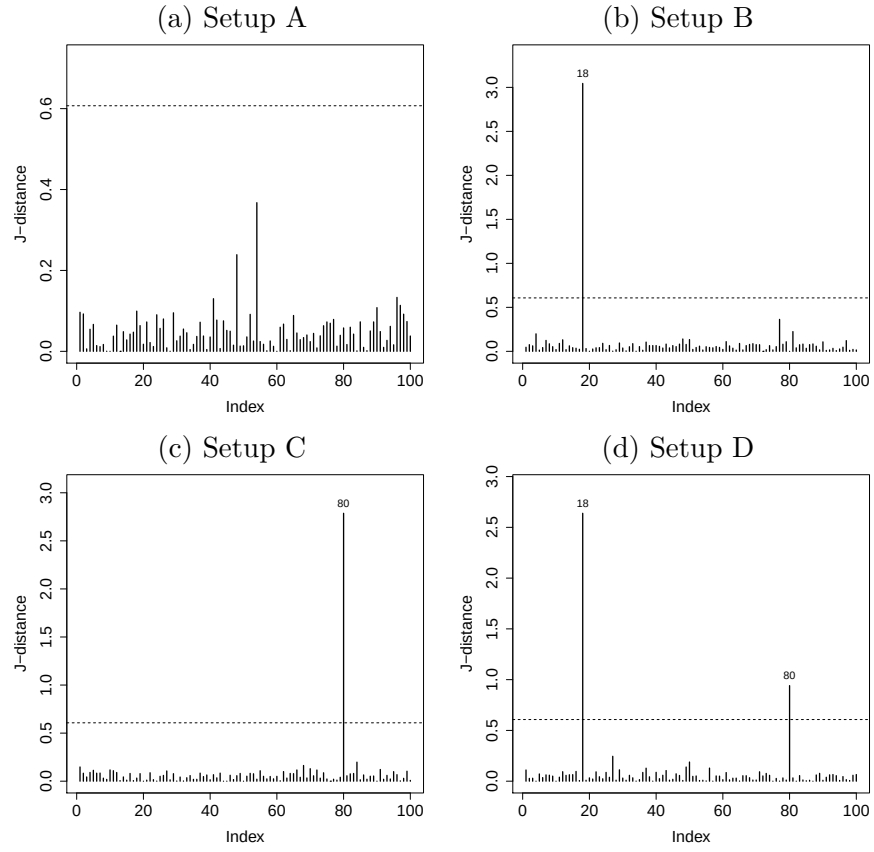

Figure 3.60: Index plots of J-distance from the fitted PHNBCR model considering prior 2.

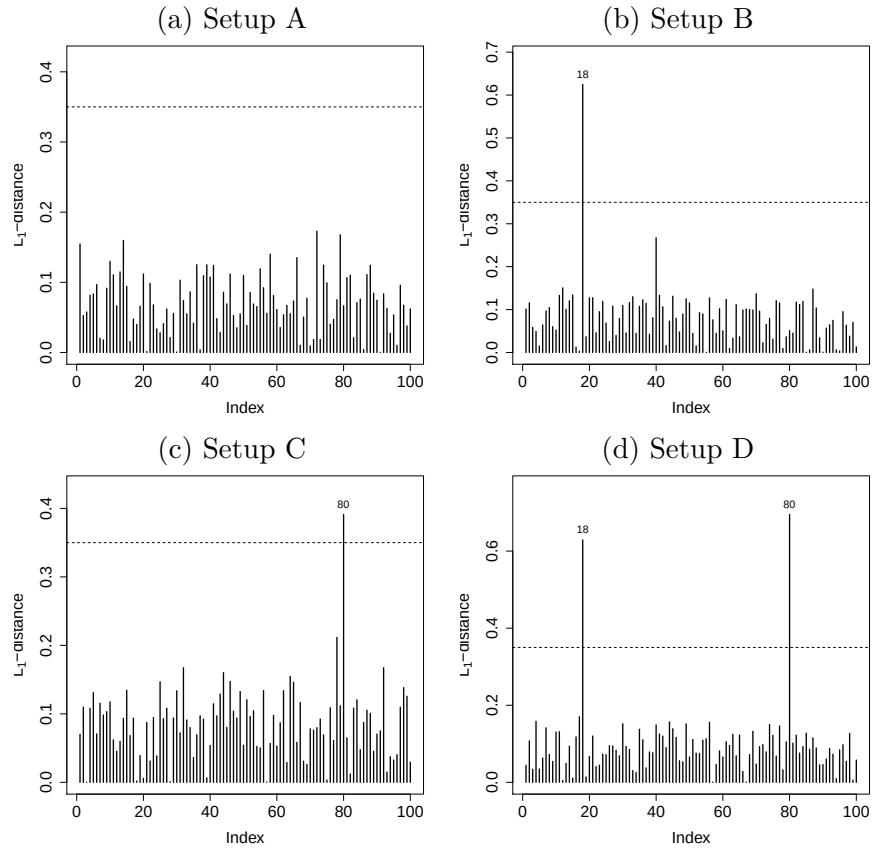

Figure 3.61: Index plots of L_1 norm distance from the fitted PHNBCR model considering prior 1.

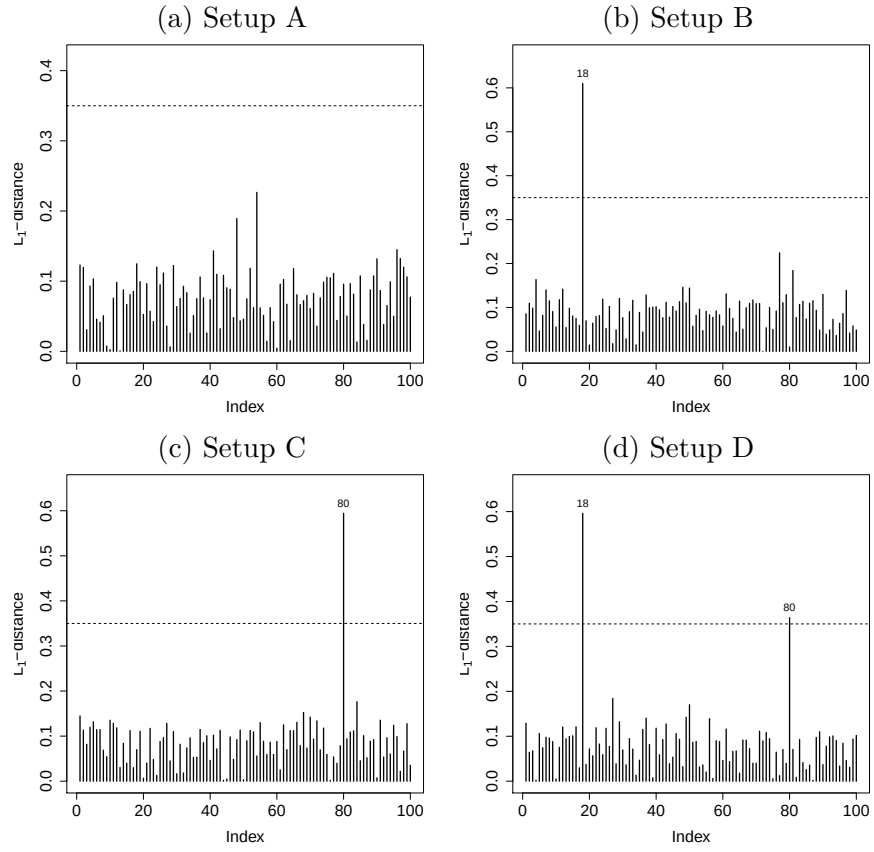

Figure 3.62: Index plots of L_1 norm distance from the fitted PHNBCR model considering prior 2.

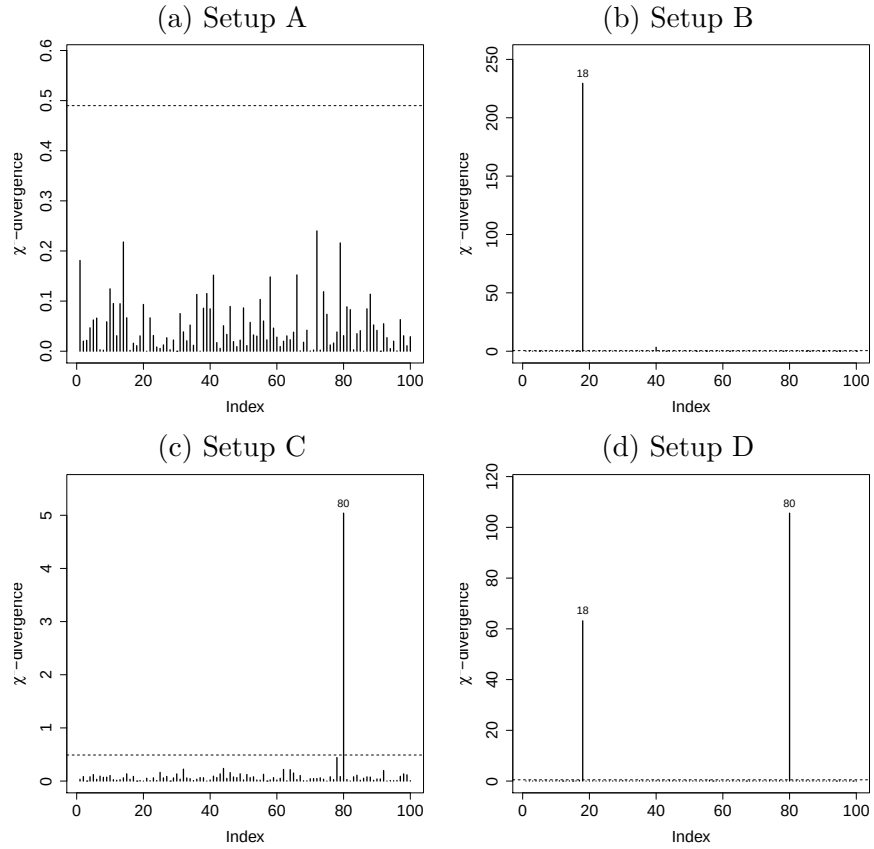

Figure 3.63: Index plots of χ^2-square divergence from the fitted PHNBCR model considering prior 1.

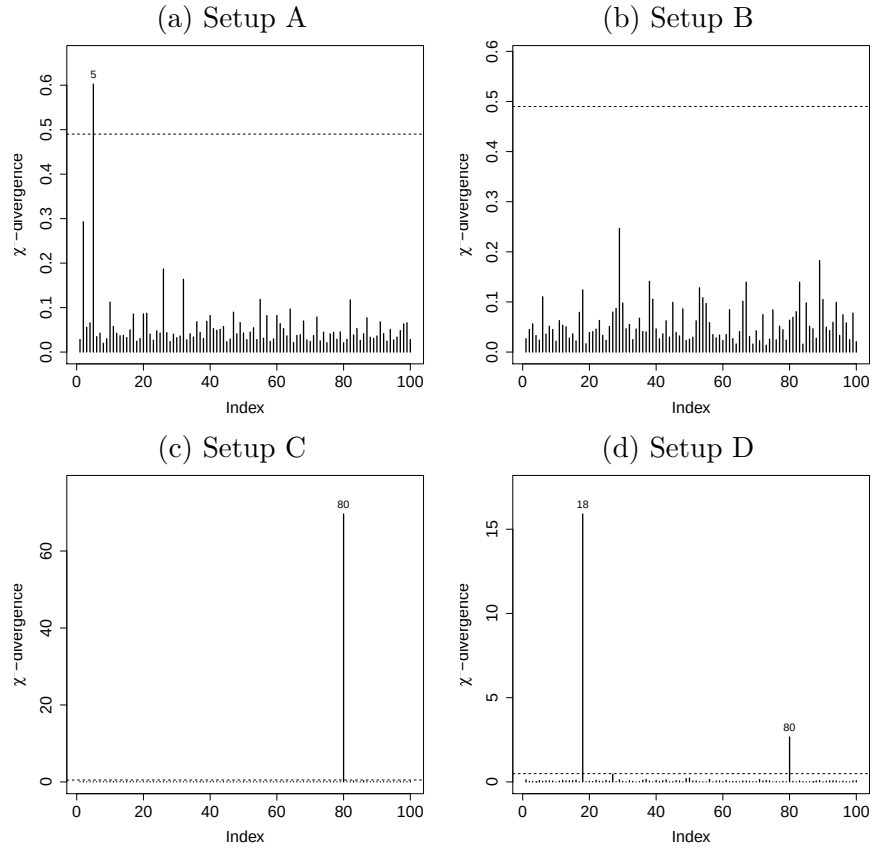

Figure 3.64: Index plots of χ^2-square divergence from the fitted PHNBCR model considering prior 2.

CPHNBCR model

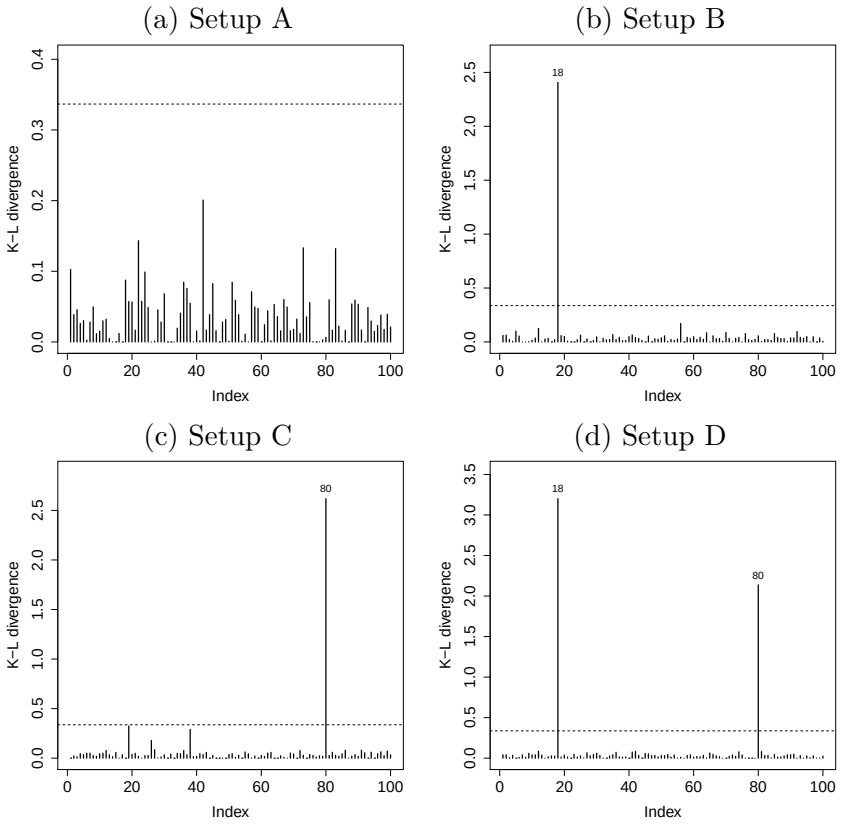

Figure 3.65: Index plots of Kullback-Leibler divergence measure from the fitted CPHNBCR model considering prior 1.

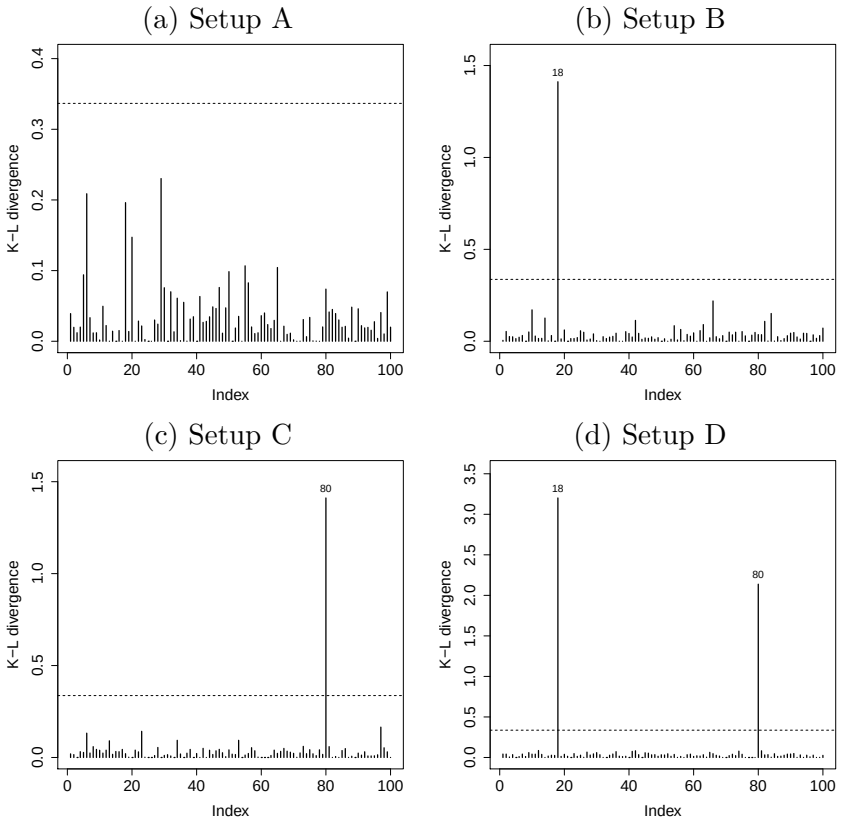

Figure 3.66: Index plots of Kullback-Leibler divergence measure from the fitted PHNBCR model considering prior 2.

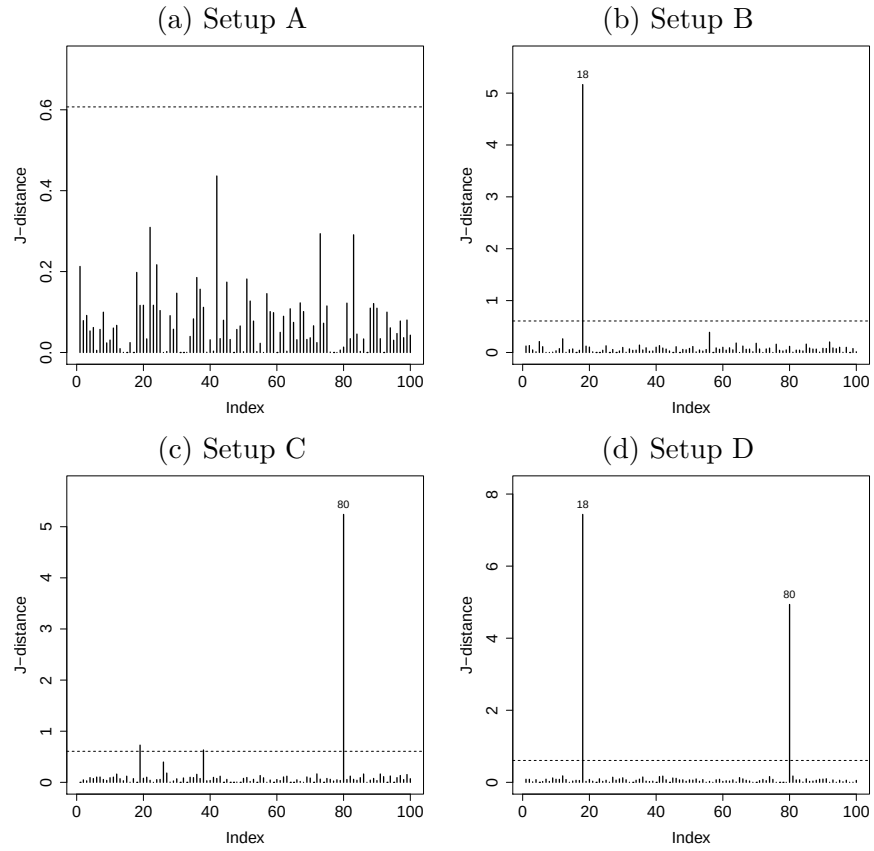

Figure 3.67: Index plots of J-distance from the fitted CPHNBCR model considering prior 1.

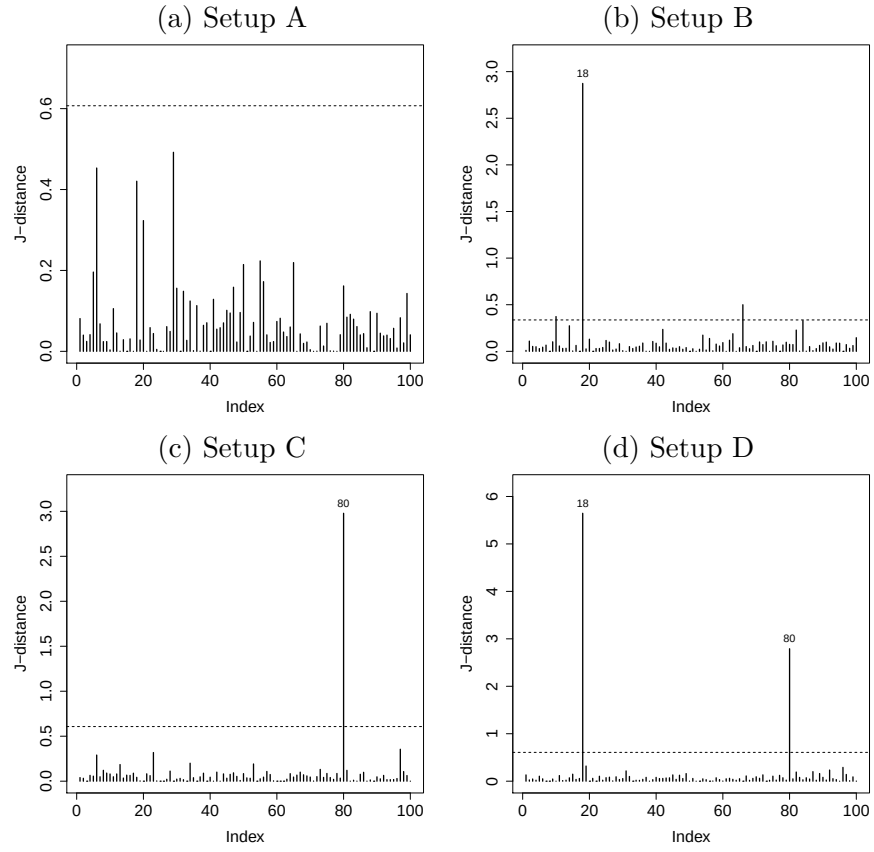

Figure 3.68: Index plots of J-distance from the fitted CPHNBCR model considering prior 2.

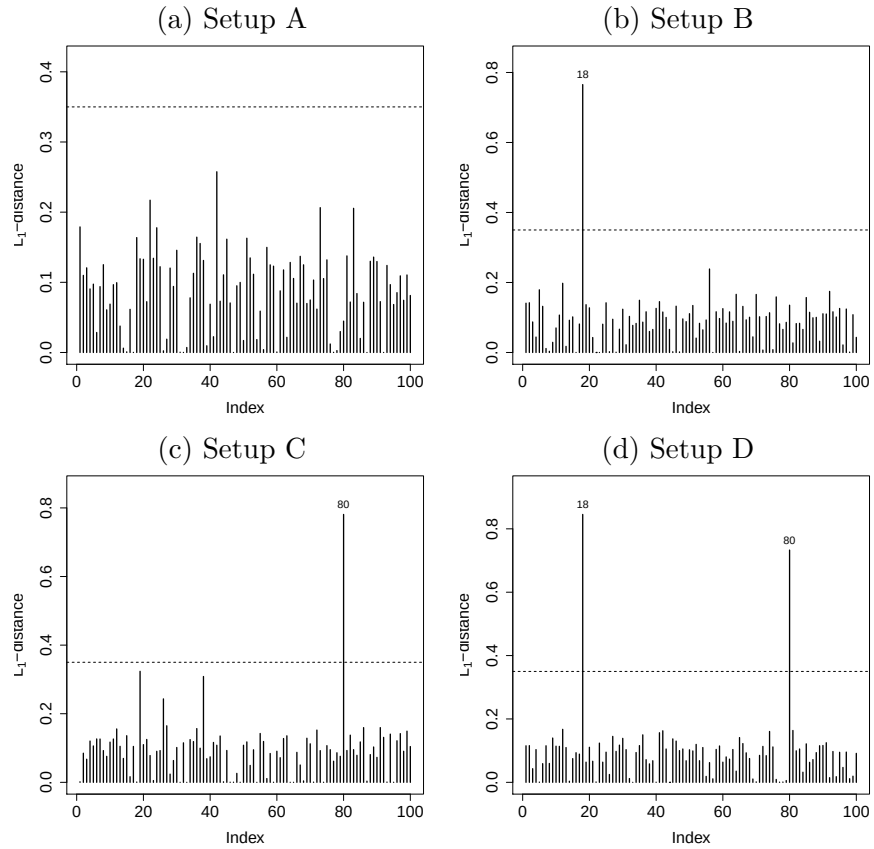

Figure 3.69: Index plots of L_1 norm distance from the fitted CPHNBCR model considering prior 1.

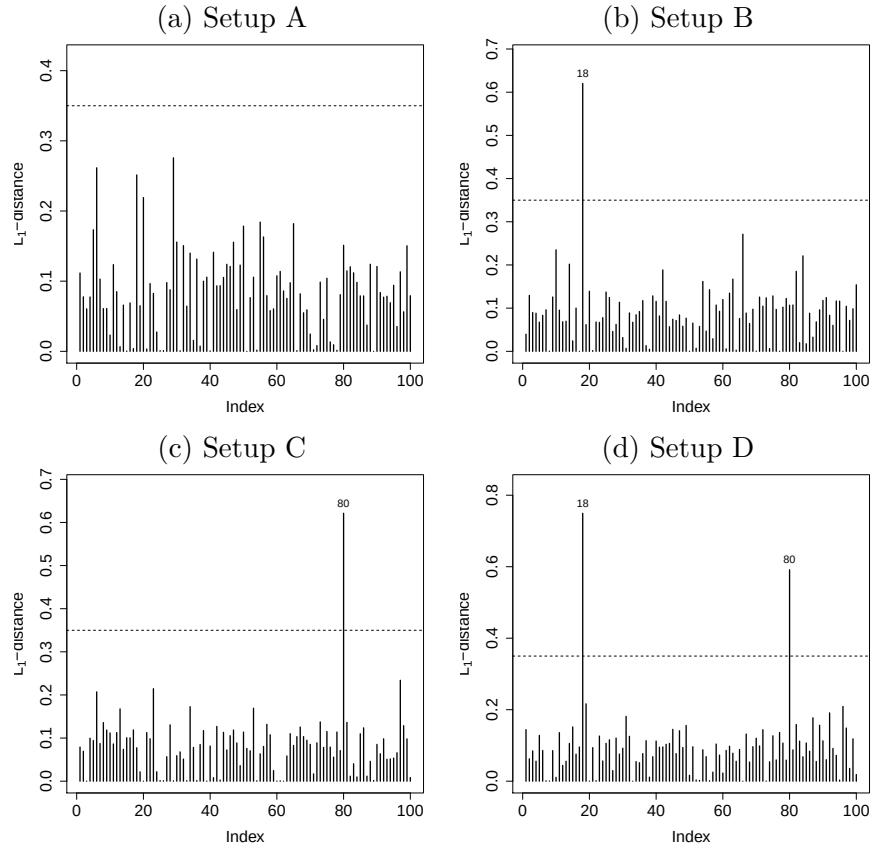

Figure 3.70: Index plots of L_1 norm distance from the fitted CPHNBCR model considering prior 2.

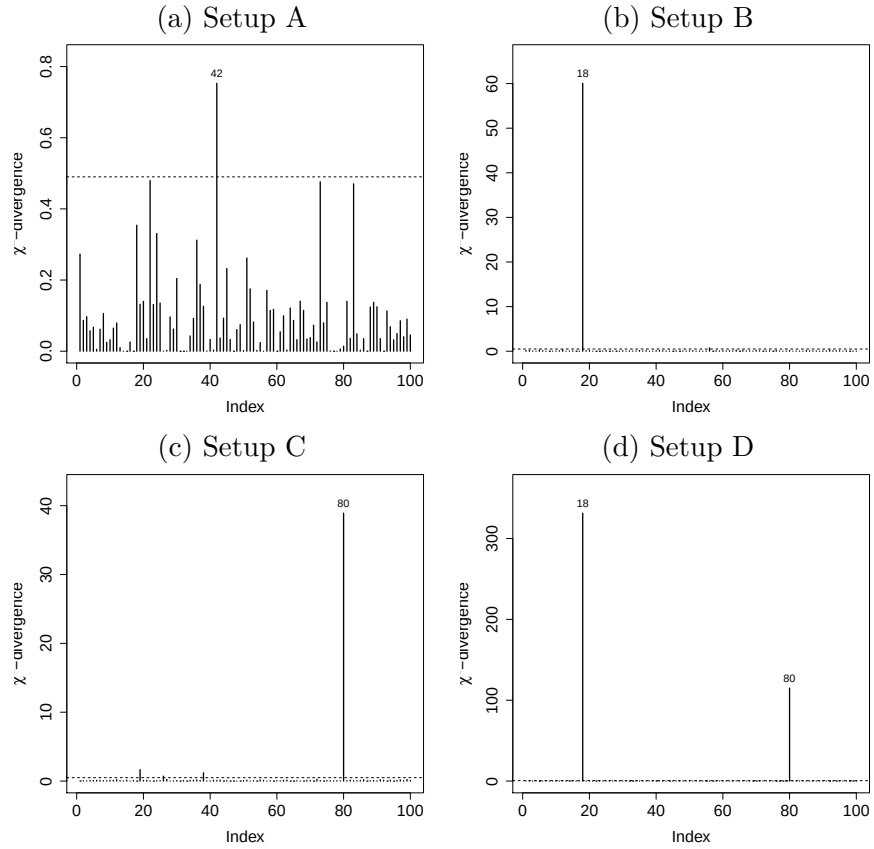

Figure 3.71: Index plots of χ^2-square divergence from the fitted CPHNBCR model considering prior 1.

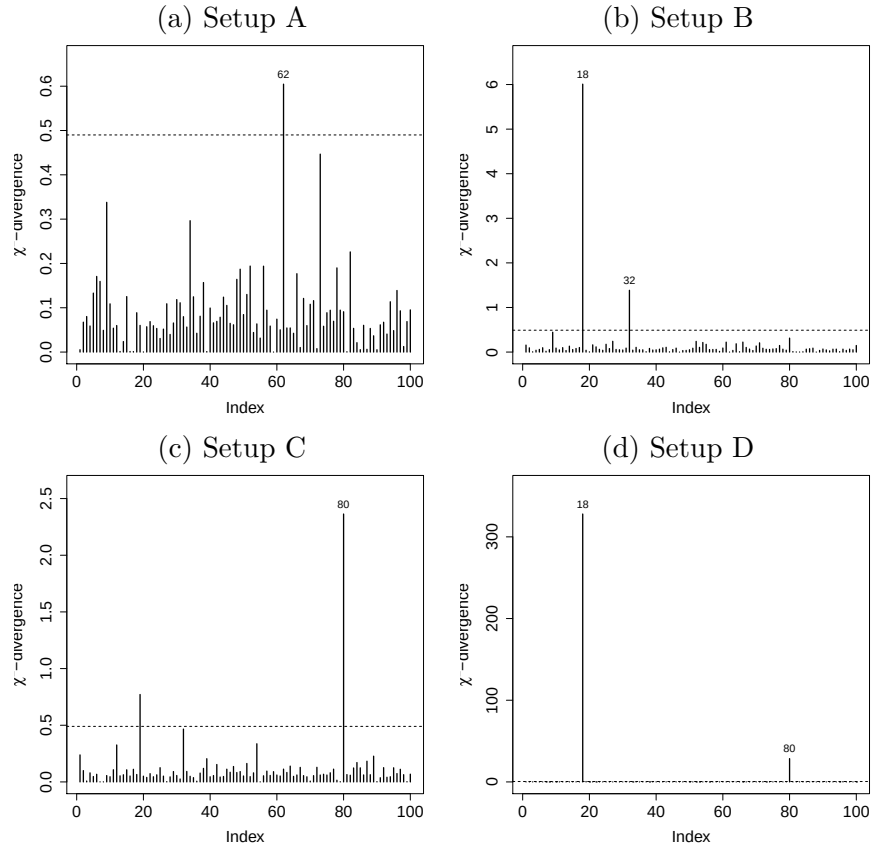

Figure 3.72: Index plots of χ^2-square divergence from the fitted CPHNBCR model considering prior 2.

3.2.3 Application

We now apply the proposed method to the interval-censored smoking cessation data presented in Section 1.1.

Because of the high computational cost, we implement the MCMC algorithms in C language and the results are analyzed in R language (R Development Core Team (2010)) through the "coda" package (Plummer *et al.* (2005)). All of our MCMC algoritms ran a total of 60,000 iterations discarding the first 20,000 realizations as burn-in and thinning to every fifth iteration. Posterior results are then based on 8,000 realizations of the Markov chain. Our Meteropolis acceptance rate for these parameters ranged from 25% to 50%. The convergence was checked using the Geweke diagnostic which did not indicate lack of convergence. The models are compared using DIC and the new proposed measure. Moreover, the case deletion influence diagnostics are also computed to detect possible influential observations.

Firstly, we fit the proposed models considering the different spatial frailties in the models to the data set. Prior distributions for the parameters $\boldsymbol{b}$, $\boldsymbol{\beta}$ and η are $b_j \sim N(0, 100)$, $j = 0, \ldots, 4$, $\beta_j \sim N(0, 100)$, $j = 1, \ldots, 4$, and $\eta \sim N(0, 100)I_{(0,\infty)}$ and a prior distribution for the shape parameter of WNBCR and CWNBCR models is $\alpha \sim N(0, 100)I_{(0,\infty)}$. As know, the piecewise exponential distribution has better approximation to any unknown function when the length of each interval becomes smaller. Therefore, we partition the time axis so that they denoted the ordered distinct time points of all observed interval end points. Thus, we have 178 risk parameters need to estimate. Prior distributions for the risk parameters are $\alpha_i \sim N(0, 100)I_{(0,\infty)}$, $\quad i = 1, \ldots, 178$. For the sub-models of the PHNBCR and CPHNBCR, we used the informative prior distributions for the parameters $\boldsymbol{b}$ and $\boldsymbol{\beta}$, where the priors are based on the posterior distributions of these parameters of PHNBCR and CPHNBCR models, i.e., $b_j \sim N(0, 1)$, $j = 0, \ldots, 4$, and $\beta_j \sim N(0, 0.6)$, $j = 1, \ldots, 4$.

For the assumption that the random frailties are independent, the prior distribution for parameter θ_i is given by $\theta_i \sim \text{InvGamma}(0.01, 0.01)$ for $i = 1, 2$. On the other hand, assuming the dependence of the random frailties, a prior distribution for parameter a_i can be taken $a_i \sim \text{Uniform}(0, 1)$ or $a_i \sim \text{Beta}(18, 2)$ and let $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \boldsymbol{\Lambda}_0)$, $i = 1$ if the random frailties take the traditional MCAR distribution (MCAR$(a, \boldsymbol{\Lambda})$) and $i = 1, 2$ if the random frailties take the extended MCAR distribution (MCAR$(a_1, a_2, \boldsymbol{\Lambda})$). Since Gelfand & Vounatsou (2003) and Banerjee & Carlin (2004) considered different values for $\boldsymbol{\Lambda}_0$, we fixe $\boldsymbol{\Lambda}_0$ equals to $\boldsymbol{I}_{2\times2}$, $0.1\boldsymbol{I}_{2\times2}$, $0.01\boldsymbol{I}_{2\times2}$ and $0.001\boldsymbol{I}_{2\times2}$ for the WNB model. We note that the estimative of parameters $\boldsymbol{\beta}$, $\boldsymbol{b}$, α, η and a are not influenced by Λ_0. However, the Table 3.24 shows that $\boldsymbol{\Lambda}_0$ restrict the posterior estimate of $\boldsymbol{\Lambda}$, i.e., if we assume the small values for the diagonal elements of $\boldsymbol{\Lambda}_0$, the posterior estimative of the elements of $\boldsymbol{\Lambda}$ have

small values and the posterior estimative of the elements of covariance matrix $\boldsymbol{\Sigma}$ have large values. We observed that the estimative of Σ_{11}, Σ_{22} and Σ_{12} have reasonable values when $\Lambda_0 = 0.1\boldsymbol{I}_{2\times2}$, thus, we consider $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$ in the application, where $\text{Diag}(0.1, 0.1) = 0.1\boldsymbol{I}_{2\times2}$.

Table 3.24: Posterior estimate of the elements of matrix $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$

Λ_0	Λ_{11}	Λ_{22}	Λ_{12}	Σ_{11}	Σ_{22}	Σ_{12}
$0.001\boldsymbol{I}_{2\times2}$	0.0342	0.0275	-0.0026	29.4836	36.5596	2.7352
$0.01\boldsymbol{I}_{2\times2}$	0.2774	0.2772	-0.0046	3.6062	3.6079	0.0604
$0.1\boldsymbol{I}_{2\times2}$	2.6616	2.5635	-0.0163	0.3757	0.3901	0.0024
$1\boldsymbol{I}_{2\times2}$	25.9818	25.4708	0.0407	0.0385	0.0393	-0.0001

where Λ_{ij} is the element of precision matrix $\boldsymbol{\Lambda}$ in position (i, j), and Σ_{ij} is the element of matrix $\boldsymbol{\Sigma} = \Lambda^{-1}$ in position (i, j), this Σ_{11} is the spatial variance component of $\boldsymbol{U}$ and Σ_{22} is the spatial variance component of $\boldsymbol{V}$, $\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$ denote their correlation.

The values of the Bayesian model selection criterion for fitted cure rate models are presented in Table 5.4. According to the DIC, the PHNBCR and CPHNBCR models are better than WNBCR and CWNBCR models for all prior distributions considered for the parameters. Although the DIC value of CPHNBCR model is lower than that of PHNBCR model, the DIC values of both models are very close, thus, we conclude that the both model are equivalent. Inasmuch as the proposed cure models are very flexile and encompasses several well-known cure model as its special cases, its sub-models also have been fitted, considering the dependent random frailties with priors (iv) given in Table 5.4 for the parameters.

The criterion values for fitted cure rate models are presented in Table 5.5. We observe that the WNBCR and CPHNBCR models have the smallest DIC and Pd values among all cure models with Weibull distribution and piecewise exponential distribution. Moreover, comparing the obtained DIC values with the values presented in the paper of Carlin & Banerjee (2003), where they proposed the mixture cure model with the spatial frailty, we conclude that all fitted models are more adequate since all their DIC values are smaller. Here, we select the CPHNB model as our working model.

The posterior summary of the parameters except α_i's is presented in the Table 3.27. The posterior means and 95% credible intervals of α_i's are presented in the Figure 3.73, it is showed that there are some values of α's, which are indicated, have different values than others. Thus, we can repartition the time axis so that we consider just risk parameters (α_1, α_2, α_{34}, α_{35}, α_{64}, α_{65}, α_{84}, α_{85}, α_{88}, α_{98}, α_{99}), thus only 11 parameters need to be estimated, which will lower computational time cost.

We note that expect the intercept, all covariates are not significant when we consider the (95%) credible interval, but the covariate of "the intervention type SI/UC", "number of cigarettes

Table 3.25: DIC values for the fitted proposed cure rate models considering different priors for the parameters.

Model	Priors	Criteria DIC	*Pd*
WNB	(i) : $\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	408	4.4
	(ii) : $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	408	8.2
	(iii) : $\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	401	1.1
	(iv) : $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	401	1.7
	(v) : $\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	408	8.6
CWNB	(i) : $\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	420	17.2
	(ii) : $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	418	12.6
	(iii) : $\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	418	12.4
	(iv) : $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	417	13.6
	(v) : $\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	417	13.7
PHNB	(iv) : $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	387	5.7
CPHNB	(iv) : $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	382	5.3

Table 3.26: DIC values for the fitted cure rate models considering prior (iv) for the parameters.

Model	Criteria DIC	*Pd*
Weibull negative binomial cure rate model	401	1.7
Weibull geometric cure rate model	417	12.7
Weibull promotion time cure model	417	10.4
Complementary Weibull negative binomial cure rate model	417	13.6
Complementary Weibull geometric cure rate model	418	13.6
Complementary Weibull promotion time cure model	419	12.4
PH negative binomial cure rate model	387	5.7
PH geometric cure rate model	395	11.6
PH promotion time cure model	404	12.9
Complementary PH negative binomial cure rate model	382	5.3
Complementary PH geometric cure rate model	394	11.7
Complementary PH promotion time cure model	404	13.6

Table 3.27: Posterior summaries of the parameter of the CPHNB model for the smoking cessation data.

		CPHNB model							
		Survival Model					**Cure Rate**		
Parameter		Mean	SD	2.5%	97.5%		Mean	SD	2.5%	97.5%
Intercept						b_0	-1.7523	1.7780	-5.9692	0.8523
Sex (male=0)	β_1	0.1661	0.3411	-0.5797	0.7904	b_1	-0.0368	0.7668	-1.3124	1.7907
SI/UC (UC=0)	β_2	-0.6882	0.4260	-1.6565	0.0501	b_2	1.3076	1.0626	-0.3685	3.9041
Cigarettes per day	β_3	0.0049	0.0157	-0.0269	0.0368	b_3	-0.0313	0.0492	-0.1168	0.0646
Duration as smoker	β_4	-0.0404	0.0222	-0.0800	0.0086	b_4	0.0518	0.0820	-0.1158	0.2083
η		0.8998	0.0660	0.7347	0.9861					
a		13.8192	5.9519	4.2498	26.3673					
Λ_{11}		2.6928	0.6559	1.5703	4.1202					
Λ_{22}							2.5671	0.6478	1.4625	3.9913
Λ_{12}		0.0205	0.4664	-0.8912	0.9630					
Σ_{11}		0.4098	0.1107	0.2485	0.6690					
Σ_{22}							0.4324	0.1265	0.2555	0.7310
$\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$		-0.0076	0.1819	-0.3732	0.3508					

where Λ_{ij} is the element of precision matrix $\boldsymbol{\Lambda}$ in position (i, j), and Σ_{ij} is the element of matrix $\boldsymbol{\Sigma} = \Lambda^{-1}$ in position (i, j), this Σ_{11} is the spatial variance component of $\boldsymbol{U}$ and Σ_{22} is the spatial variance component of $\boldsymbol{V}$, $\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$ denote their correlation.

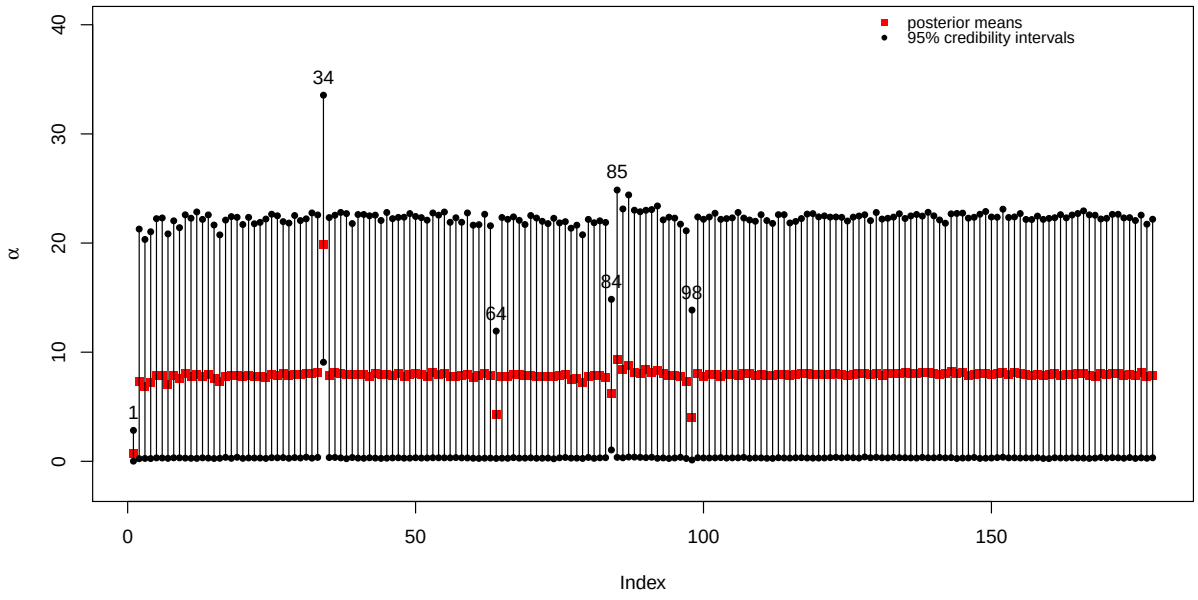

Figure 3.73: Posterior means and credible intervals of α_i's

smoked per day", and "duration of smoking habit" could be significant since we consider the lower credible interval. In cure rate the positive value of b_2 implies that individuals with special intervention have higher probability of quitting smoking than those with usual care and the negative value of b_3 means that the individuals with a higher level of cigarette consumption have lower probability of quitting. But the duration of smoking habit does not have a positive effect on quitting. In the survival function, this shows that individuals with special intervention have lower hazard rate of relapse time than those with usual care. On the other hand, the number of cigarettes smoked per day and the duration of smoking habit have no positive effects on the hazard rate of the relapse, that is, individuals with a higher level of cigarette consumption per day and longer habit do not have a higher hazard rate.

The standard deviation, $\Sigma_{11}^{1/2}$ of random spatial effects in the survival model is 0.6401, and the standard deviation, $\Sigma_{22}^{1/2}$ of random spatial effects in cure rate is 0.6575 which indicates considerable heterogeneity among the clusters. Moreover, there is no linear correlation between the spatial effects $\boldsymbol{U}$ and $\boldsymbol{V}$.

Figure 3.74 maps the posterior means and standard deviations of frailties $\boldsymbol{U}$ and $\boldsymbol{V}$ in the CPHNBCR model. For the frailties $\boldsymbol{U}$ for which the high value represents a high relapse rate, we note that the northwest regions and some cities of the south region have higher values, that is, the individuals in these regions have higher relapse rates than those in other areas. By contrast, the center region (Rochester city) and some northeast cities suggest slightly better than avenge cessation behavior. The frailties $\boldsymbol{V}$ show that the eastern region has lower cure probability and the other regions, which have close probabilities. Note that the posterior standard deviations of frailties $\boldsymbol{U}$ and $\boldsymbol{V}$ have approximate values. Both maps show that the cities round the central region have lower values and the city of Waseca has the highest value.

In order to detect possible influential observations in the posterior distribution of the parameters of CPHNB model, the estimates of ψ-divergence measures, which were obtained from the posterior sample of the parameters of the model, are presented in Figure 3.75. It shows that there are some possible influential observations which were detected by divergence measures. Here, we only analyze the individual 138, who was detected as an influential observation by all four divergence measures. This individual is a women living in Faribault. She received the anti-smoking intervention and her observed interval is $(2.998, 3.992)$. Her average cigarette consumption is 20 per day and smoking duration is 25 years.

In order to reveal the impact of this possible influential observation on the parameter estimates and inference, we removed observation 138 and readjusted the model and calculated the relative variations (RV) for the posterior mean of the parameters. The RV are defined by

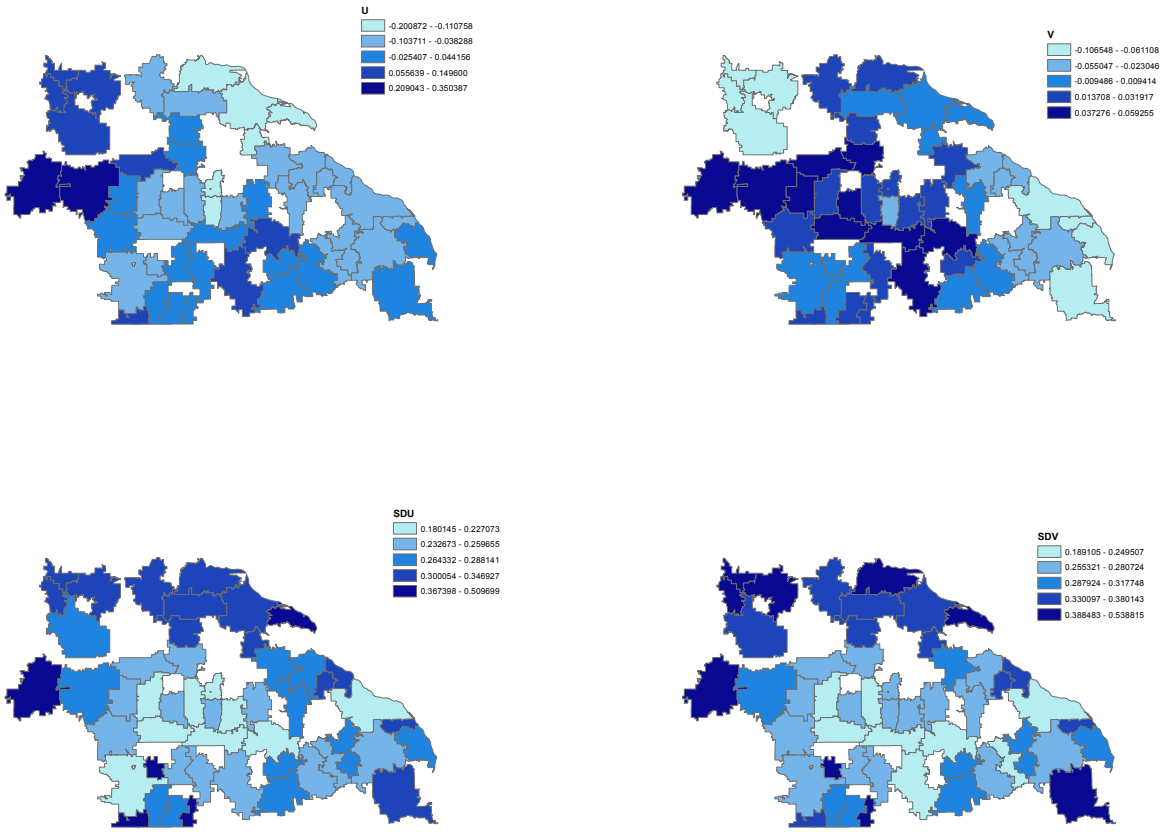

Figure 3.74: Maps of posterior means for frailties $\boldsymbol{U}$ (upper-left panel) and $\boldsymbol{V}$ (upper-right panel) and posterior standard derivations for frailties $\boldsymbol{U}$ (lower-left panel) and $\boldsymbol{V}$ (lower-right panel).

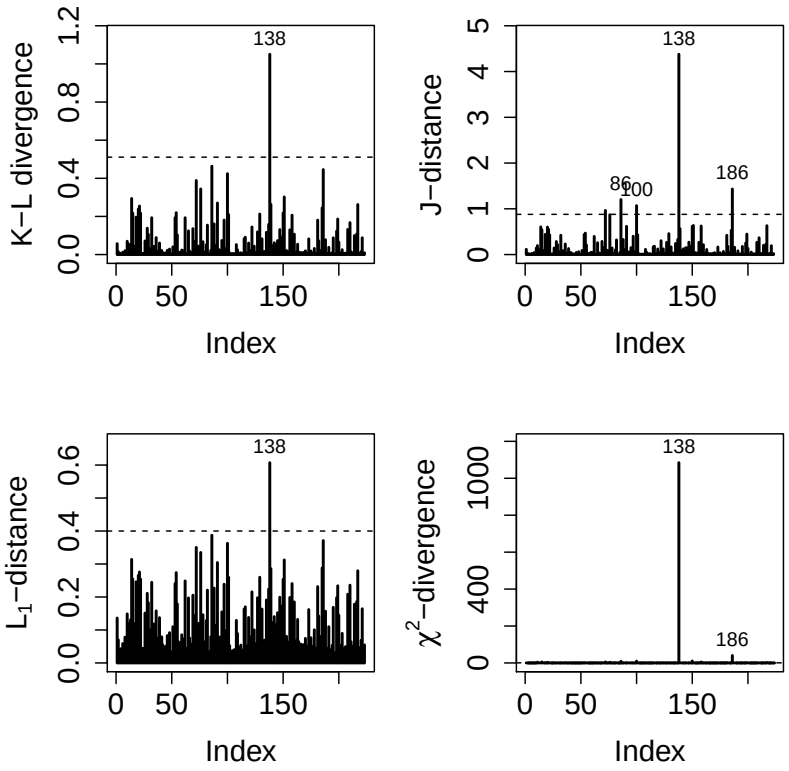

Figure 3.75: Estimates of ψ-divergence measures for CPHNBCR model

$RV = (\hat{\vartheta}_{d,-\{138\}} - \hat{\vartheta}_d)/\hat{\vartheta}_d$, for all d, where d is the index of the parameters, $\hat{\vartheta}_{d,-\{138\}}$ denotes the posterior mean of $\vartheta_{d,-\{138\}}$, after removel of the set of observations $\{138\}$. Note that, in the piecewise exponential model, the time axis is partitioned by the ordered distinct time points of all observed interval end points, so we have different and fewer risk parameter α's after removed observation 138. In this case we have 176 risk parameters to be estimated. The posterior summaries of the parameters for the refitted CPHNB model and RV for the posterior mean of the parameters are presented in Table 3.28. We note that only the parameters b_1 and σ_{12} have larger RV values, but still close to the obtained estimates without removing the detected individual. In this case, all parameters of the CPHNB model are not sensitive under deletion of the outlying observations and we do not have inferential changes after removing the observations. The DIC values and p_d for fitted models are 378 and 6.2, respectively. They are lower than CPHNBCR model for the data without removing the detected observation.

Table 3.28: Posterior summaries of the parameter of CPHNBCR model and RV adjusted for the smoking cessation data without detected individual 138.

		Survival Model					**Cure Rate**			
Parameter		Mean	SD	2.5%	97.5%		Mean	SD	2.5%	97.5%
Intercept						b_0	-1.2963 (-0.2602)	1.424	-4.2843	1.1732
Sex (male=0)	β_1	0.2191 (-0.3191)	0.3350	-0.5109	0.8301	b_1	0.0353 **(-1.9592)**	0.6972	-1.0915	1.7375
SI/UC (UC=0)	β_2	-0.5183 (-0.2469)	0.3807	-1.3837	0.1839	b_2	1.0211 (-0.2191)	0.8139	-0.4098	3.049
Cigarettes per day	β_3	0.0020 (-0.5918)	0.0172	-0.0329	0.0355	b_3	-0.0448 (0.4313)	0.0391	-0.1152	0.0491
Duration as smoker	β_4	-0.0437 (0.0817)	0.0189	-0.0777	0.0024	b_4	0.0633 (0.2220)	0.0569	-0.066	0.1611
η		13.9922 (0.0125)	5.6025	4.4720	26.2489					
a		0.9008 (0.0011)	0.0654	0.7408	0.9863					
Λ_{11}		2.6605 (-0.0120)	0.6542	1.5514	4.1072					
Λ_{22}							2.5590 (-0.0032)	0.6345	1.4802	3.9449
Λ_{12}		-0.0170 (-1.8293)	0.4725	-0.9683	0.8960					
Σ_{11}		0.4149 (0.0126)	0.1092	0.2501	0.6753					
Σ_{22}							0.4334 (0.0022)	0.1228	0.2589	0.7200
Σ_{12}		0.0067 **(-1.8816)**	0.1848	-0.3526	0.3742					

3.2.4 Conclusions

In this work, we described an approach to extend the cure rate model (Cancho *et al.*, 2011) and its complementary model to allow for spatial correlations by including spatial frailty for the interval-censored data setting. The proposed cure rate models with frailty are very flexible because they encompass several known cure rate models as its particular cases. We use the MCMC methods in Bayesian inference approach to fit our models and some Bayesian model comparison criteria were used. The results from the application show that WNBCR model with frailties has better fit than CWNBCR model with frailties, but the proportional hazard cure models with frailties (PHNBCR and CPHNBCR models) stand out better. Comparing the proposed models with models introduced by Carlin & Banerjee (2003) and Pan *et al.* (2014), it is shown that the proportional hazard cure models with frailties are more adequate. Moreover, the proposed models are not sensitive with influential observations, which can be observed through the influence diagnostic in the simulation study as well as in the application. The interpretation of the covariates is easy due to the parametrization of the models considered in the cure rate. Moveover, the MCAR prior can be used even if frailties effects are low or they are not correlated.

Chapter 4

The Power Series Cure Rate Model for Spatially Correlated Interval-Censored Data based on Generalized Extreme Value Distribution

4.1 Introduction

In this section, we propose a new cure rate survival model for spatially correlated interval-censored data based on generalized extreme value distribution. This cure rate model is much more general than the cure models proposed in Sections 3.1 and 3.2. Here, we assume the number of competing causes related to the occurrence of an event is modeled by an exponential composed by discrete power series (PS) distribution (Noack, 1950). Therefore, the cure rate model with the PS distribution is very flexible. Because it can be seen as a general model encompassing several well known cure models such as, bernoulli (the mixture cure model), geometric, logarithmic, and Poisson (the promotion time cure model), among others ones, which can be tested for the best fitting on a straightforwardly way. The MCMC method is used in Bayesian inference approach and some Bayesian model selection criteria are used for model comparison. Moreover, we conduct an influence diagnostic in order to detect possible influential or extreme observations that can cause distortions on the results of the analysis. Finally, the proposed models are applied to analysis of a real data set on a smoking cessation study.

4.2 Power Series cure rate Model

Suppose that there are I regions and n_i individuals in ith region. We denote by T_{ij} the random variable for the observed time to event of jth individual in ith region, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, I$. Suppose that the (i, j)th individual is potentially exposed to M_{ij} latent risk, where M_{ij} denotes the initial number of competing causes related to the occurrence of an event and assuming M_{ij} has a class of random variables with discrete distributions proposed by Noack (1950), with the probability mass function

$$P(M_{ij} = m) = \frac{a_m \theta_{ij}^m}{A(\theta_{ij})}, \quad m = 0, 1, 2 \cdots . \tag{4.1}$$

where $a_m > 0$, $A(\theta_{ij}) = \sum_{m=0}^{\infty} a_m \theta_{ij}^m$ and $\theta_{ij} \in (0, s)$ is chosen such that $A(\theta_{ij})$ is finite and and its first, second and third derivatives are defined. The parameter θ_{ij} is called the power parameter of the distribution and $A(\theta_{ij})$ is the series function. Some important and well-known distributions belong to this class. For example, if $A(\theta_{ij}) = (1 + \theta_{ij})^k$ and $a_m = \binom{k}{m}$ with $\theta_{ij} > 0$ and k is positive integer, then (4.1) defines the binomial distribution. If $A(\theta_{ij}) = \exp(\theta_{ij})$ and $a_m = \frac{1}{m}$, $\theta_{ij} > 0$ then (4.1) defines a Poisson distribution. If $A(\theta_{ij}) = (1-\theta_{ij})^{-k}$ and $a_m = \binom{m+k-1}{m}$, $k > 0$ and $0 < \theta_{ij} < 1$, then (4.1) defines the negative binomial distribution. If $A(\theta_{ij}) = -\log(1 - \theta_{ij})/\theta_{ij}$ and $a_m = \frac{1}{m+1}$, with $0 < \theta_{ij} < 1$, then logarithmic distribution is obtained from (4.1).

Let Y_{cij} for $c = 1, \ldots, M_{ij}$ denote the failure times of jth individual in ith region due to the cth latent risk. We Suppose that, given M_{ij}, the random variables Y_{cij}'s are mutually independent with distribution function $F(\cdot) = 1 - S(\cdot)$. If we assume the presence of any of latent risk will lead to the occurrence of the event, the time to event of interest can be defined by random variable $T_{ij} = \min\{Y_{cij}, c = 1, \cdots, M_{ij}\}$ for $M_{ij} \geq 1$ and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P(T_{ij} = \infty | M_{ij} = 0) = 1$. Note that any survival distribution can be considered to represent our uncertainty about the values of random variables $Y_{cij}, c = 1, \ldots$. The assumption of independence and identical distribution to $Y_{1ij}, Y_{2ij}, \ldots$ is surely a strong one, favoring simplicity and analytical tractability at the expense of a more general formulation, as remarked by Yakovlev and Tsodikov (1996). Despite this shortcoming, these models have proven to be useful in many real-world applications.

Under this setup, the survival function for the population is given by

$$S_{pop}^F(t_{ij}) = P(T_{ij} > t_{ij}) = \sum_{m=0}^{\infty} S(t_{ij})^m \frac{a_m \theta_{ij}^m}{A(\theta_{ij})} = \frac{A\left(\theta_{ij} S(t_{ij})\right)}{A(\theta_{ij})}, \quad t_{ij} > 0. \tag{4.2}$$

This situation is also known as first activation scheme, because in this case we assume that the event of interest occurs when the first possible cause is activated. Here, we called the model in (4.2) as

power series (PS) cure rate model under first activation. The cure fraction can be obtained from (4.2) is given by

$$p_{0ij} = \lim_{t\to\infty} S_{pop}(t_{ij}) = \frac{A(0)}{A(\theta_{ij})} = \frac{a_0}{A(\theta_{ij})} > 0.$$

The corresponding density and hazard functions to (4.2) are given by

$$f_{pop}(t_{ij}) = \frac{A'(\theta_{ij}S(t_{ij}))}{A(\theta_{ij})}\theta_{ij}f(t_{ij})$$

and

$$h_{pop}(t_{ij}) = \frac{A'(\theta_{ij}S(t_{ij}))}{A(\theta_{ij}S(t_{ij}))}\theta_{ij}f(t_{ij}), \tag{4.3}$$

respectively. Where $A'(\theta_{ij}S(t_{ij})) = dA(\theta_{ij}S(t_{ij}))/dt$ and $f(t_{ij}) = -dS(t_{ij})/dt$ denotes the (proper) density function of the time to event Y_{cij} in (4.2). Note that the $f_{pop}(t_{ij})$ is not proper probability density function, since $S_{pop}(t_{ij})$ is not proper survival function. The hazard function in (4.3) satisfies the proportional hazards property if, and only if, $A(\theta_{ij}) = \exp(\theta_{ij})$.

On the other hand, if we assume the presence of all latent risks will ultimately lead to the occurrence of the event. Thus, the time to event of interest is defined by random variable $T_{ij} = \max\{Y_{cij}, c = 1, \cdots, M_{ij}\}$ for $M_{ij} \geq 1$ and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P(T_{ij} = \infty | M_{ij} = 0) = 1$. The survival function for the population is given by

$$S^L_{pop}(t_{ij}) = P(T_{ij} > t_{ij}) = 1 + \frac{A(0)}{A(\theta_{ij})} - \frac{A\left(\theta_{ij}F(t_{ij})\right)}{A(\theta_{ij})}, \quad t_{ij} > 0. \tag{4.4}$$

This situation is known as last activation scheme, because the event of interest only takes place after all the latent causes have been occurred. We called the model in (4.4) as power series cure rate model under last activation. The cure fraction can be obtained from (4.4) is given by $p_{0ij} = \lim_{t\to\infty} S_{pop}(t_{ij}) = \frac{A(0)}{A(\theta_{ij})} > 0$, which has the same expression as the cure fraction obtained from (4.2). The corresponding density and hazard functions to (4.4) are given by

$$f_{pop}(t_{ij}) = \frac{A'(\theta_{ij}F(t_{ij}))}{A(\theta_{ij})}\theta_{ij}f(t_{ij})$$

and

$$h_{pop}(t_{ij}) = \frac{A'(\theta_{ij}F(t_{ij}))\theta_{ij}f(t_{ij})}{A(0) + A(\theta_{ij}) - A(\theta_{ij}F(t_{ij}))},$$

where $A'(\theta_{ij}F(t_{ij})) = dA(\theta_{ij}F(t_{ij}))/dt$ and $f(t_{ij}) = -dS(t_{ij})/dt$ denotes the (proper) density function of the time to event Y_{cij} in (4.4).

There are another situation where the event of interest occurs when the some of the possible causes are activated, and given the number of latent causes M_{ij}, the number of activated causes is a

random variable with the discrete uniform distribution on $\{1, \cdots, M_{ij}\}$. This situation is known as random activation scheme. The survival function for the population is given by

$$S^R_{pop}(t_{ij}) = P(T_{ij} > t_{ij}) = \frac{A(0)}{A(\theta_{ij})} + \left(1 - \frac{A(0)}{A(\theta_{ij})}\right) S(t_{ij}), \quad t_{ij} > 0. \tag{4.5}$$

The corresponding density and hazard functions to (4.5) are given by

$$f_{pop}(t_{ij}) = \left(1 - \frac{A(0)}{A(\theta_{ij})}\right) f(t_{ij})$$

and

$$h_{pop}(t_{ij}) = \frac{(A(\theta_{ij}) - A(0)) f(t_{ij})}{A(0) + (A(\theta_{ij}) - A(0)) S(t_{ij})}.$$

Note that under conditions of the models (4.2), (4.4) and (4.5) for any distribution function $F(\cdot)$, the relationship among the first, last and random activation schemes is $S^F_{\text{pop}}(t_{ij}) \leq S^R_{\text{pop}}(t_{ij}) \leq S^L_{\text{pop}}(t_{ij})$ for all $t_{ij} > 0$.

4.3 Special cases of the PS cure rate model under first/last activation

In this section, we present several important cure model can be obtained directly from our general formulations given in (4.2) and (4.4).

4.3.1 Mixture cure model

Let $A(\theta_{ij}) = (1 + \theta_{ij})$, from the general formulation (4.2) or (4.4), we obtain the classical mixture model (Boag (1949); Berkson & Gage (1952)),

$$S_{pop}(t_{ij}) = \frac{1}{1 + \theta_{ij}} + \frac{\theta_{ij}}{1 + \theta_{ij}} S(t_{ij}).$$

In this case, M_{ij} follows a Bernoulli distribution with parameter $\theta_{ij}/1 + \theta_{ij}$ and the cure rate is given by $p_{0ij} = (1 + \theta_{ij})^{-1}$. The corresponding density function has expression

$$f_{pop}(t_{ij}) = \frac{\theta_{ij}}{1 + \theta_{ij}} f(t_{ij}).$$

4.3.2 Promotion time cure model and complementary promotion time cure model

If $A(\theta_{ij}) = \exp(\theta_{ij})$, from (4.2), then we obtain the promotion time cure model proposed by Chen *et al.* (1999),

$$S_{pop}(t_{ij}) = \exp\{-\theta_{ij}F(t_{ij})\}.$$

Here, M_{ij} follows the Poisson distribution with parameter θ_{ij}. The cure fraction given by $p_{0ij} = \exp\{\theta_{ij}\}$. The corresponding density function is given by

$$f_{pop}(t_{ij}) = \theta_{ij}f(t_{ij})\exp\{-\theta_{ij}F(t_{ij})\}.$$

Moreover, from the formulation (4.4), we obtain the complementary promotion time cure model. The corresponding survival function is given by

$$S_{pop}(t_{ij}) = 1 + \exp\{-\theta_{ij}\} - \exp\left\{-\theta_{ij}S(t_{ij})\right\},$$

and the corresponding density function is given by

$$f_{pop}(t_{ij}) = \theta_{ij}f(t_{ij})\exp\{-\theta_{ij}S(t_{ij})\}.$$

4.3.3 Geometric cure rate model and complementary geometric cure rate model

If $A(\theta_{ij}) = (1-\theta_{ij})^{-1}$, then M_{ij} follows the geometric distribution. From the formulation (4.2) and (4.4), we obtain the geometric and complementary geometric cure models. The corresponding survival functions are given by

$$S_{pop}(t_{ij}) = \frac{1-\theta_{ij}}{1-\theta_{ij}S(t_{ij})},$$

and

$$S_{pop}(t_{ij}) = 1 + (1-\theta_{ij}) - \frac{1-\theta_{ij}}{1-\theta_{ij}F(t_{ij})},$$

respectively. The cure fraction given by $p_{0ij} = 1-\theta_{ij}$. The corresponding density functions are given by

$$f_{pop}(t_{ij}) = \theta_{ij}(1-\theta_{ij})f(t_{ij})\left[1-\theta_{ij}S(t_{ij})\right]^{-2}$$

and

$$f_{pop}(t_{ij}) = \theta_{ij}(1-\theta_{ij})f(t_{ij})\left[1-\theta_{ij}F(t_{ij})\right]^{-2},$$

respectively.

4.3.4 Logarithmic cure rate model and complementary logarithmic cure rate model

Let $A(\theta_{ij}) = -\log(1-\theta_{ij})/\theta_{ij}$, then M_{ij} follows the logarithmic distribution, then from the formulation (4.2) the corresponding survival function is given by

$$S_{pop}(t_{ij}) = \frac{\log(1-\theta_{ij}S(t_{ij}))}{S(t_{ij})\log(1-\theta_{ij})},$$

and the corresponding density function is given by

$$f_{pop}(t_{ij}) = -\frac{f(t_{ij})}{S(t_{ij})\log(1-\theta_{ij})}\left[\frac{\log(1-\theta_{ij}S(t_{ij}))}{S(t_{ij})} + \frac{\theta_{ij}}{1-\theta_{ij}S(t_{ij})}\right].$$

From the formulation (4.4), the survival function of complementary logarithmic cure model has expression

$$S_{pop}(t_{ij}) = 1 - \frac{\theta_{ij}}{\log(1-\theta_{ij})} - \frac{\log(1-\theta_{ij}F(t_{ij}))}{F(t_{ij})\log(1-\theta_{ij})},$$

and the corresponding density function is given by

$$f_{pop}(t_{ij}) = -\frac{f(t_{ij})}{F(t_{ij})\log(1-\theta_{ij})}\left[\frac{\log(1-\theta_{ij}F(t_{ij}))}{F(t_{ij})} + \frac{\theta_{ij}}{1-\theta_{ij}F(t_{ij})}\right].$$

The cure fraction of both models is given by $p_{0ij} = -\theta_{ij}/\log(1-\theta_{ij})$. Other survival models with cure fraction can be obtained in a similar way.

As is well known, the cure fraction plays a key role in the survival models with a cure fraction. Thus, it is important to study the effect of covariates on the cure fraction. Since the cure fraction is in the θ_{ij}'s function, the effect of covariates can be obtained by associate covariates with the parameter θ_{ij}. In this paper, we propose that for the models whose M_{ij} follows Bernoulli or Poisson distribution, the parameter θ_{ij} of an individual (i,j) is associated with covariates $\boldsymbol{x}_{ij}$ and it is modeled by

$$\theta_{ij} = \exp\{\xi_{ij}\},$$

and for the models whose M_{ij} follows geometric or logarithmic distribution, the parameter θ_{ij} of an

individual (i, j) is associated with covariates $\boldsymbol{x}_{ij}$ and it is modeled by a logistic regression

$$\theta_{ij} = \frac{\exp(\xi_{ij})}{1 + \exp(\xi_{ij})},$$

where ξ_{ij} is a linear form of covariates, $\xi_{ij} = \boldsymbol{x}_{ij}^{\top}\boldsymbol{b}$ and $\boldsymbol{b}$ is a p_1-dimensional vector representing the effects of covariates on θ_{ij} which associated the cured probability p_{0ij}.

The non-negative random variables Y_{cij}'s can take several distributions. In this work, we assume they follow proportional hazard (PH) model with the baseline hazard function $h_0(t|\cdot)$, the conditional hazard function and corresponding survival function are given by

$$h(t|\cdot) = h_0(t|\bullet)\exp(\lambda_{ij}) \;\text{ or }\; S(t|\cdot) = S_0(t|\bullet)^{\exp(\lambda_{ij})} \tag{4.6}$$

where $\lambda_{ij} = \boldsymbol{z}_{ij}'\boldsymbol{\beta}$, $\boldsymbol{z}_{ij}$ and $\boldsymbol{\beta}$ is a p_2-dimensional vector representing the effects of covariates on the survival model component, $S_0(t|\cdot)$ is the baseline survival function corresponding to $h_0(t|\cdot)$. Here, we specify the baseline function using a logGEV(μ, σ, ς) distribution given in Section 2.1.3 instead of the commonly used Weibull or Gamma distribution, where $\mu \in \mathbb{R}$, $\sigma > 0$ and $\varsigma \in \mathbb{R}$ are the location, scale and shape parameters respectively. The main reason for choosing this distribution is that the hazard function of the logGEV distribution can take severely different shapes, so it is extremely flexible in modeling survival data.

Now, we will introduce the frailties U_i and V_i to better explain the effect of survival time of susceptible individuals and on the parameter θ_{ij} which related cured probability through linear predictor expression

$$\begin{aligned} \lambda_{ij} &= \boldsymbol{z}_{ij}'\boldsymbol{\beta} + U_i, \\ \xi_{ij} &= \boldsymbol{x}_{ij}'\boldsymbol{b} + V_i, \quad \text{for } j = 1\ldots, n_i, i = 1, \ldots, I. \end{aligned}$$

Here, the frailties U_i and V_i are spatially correlated across the regions. In this work, we assume the spatial priors on $(\boldsymbol{U}, \boldsymbol{V})$ are dependent, and they have multivariate conditionally auto-regressive extend MCAR prior distribution which was studied by Gelfand and Vounatsou (2003) and Carlin and Benerjee (2003). The details to extend the MCAR distribution can be found in Section 2.5.3.

4.4 Bayesian Inference

Let $\boldsymbol{D}_{obs} = \{(A_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \delta_{ij}); j = 1, \ldots, n_i, i = 1, \ldots M\}$ denote the observed data, where $A_{ij} = (t_{ijL}, t_{ijR}]$ is the interval during which individual j in cluster i occur the event of interest,

$\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are the p_1-dimencional and p_2-dimencional vectors of covariates, and δ_{ij} is following interval censoring indicator: $\delta_{ij} = I(t_{ijR} < \infty)$. For the spacial case in which the survival time is right-(left-) censored, $R_{ij} = +\infty (L_{ij} = 0)$, whereas for exact observations, $t_{ijL} = t_{ijR}$. Following Finkelstein(1986), the likelihood function for the general interval-censored cure rate model is given by

$$\begin{aligned} L\{\boldsymbol{\varphi}|\boldsymbol{D}_{obs}, \boldsymbol{U}, \boldsymbol{V}\} &\propto \prod_{i=1}^{I}\prod_{j=1}^{n_i} \left(S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi}) - S_{\text{pop}}(t_{ijR}|\boldsymbol{\varphi})\right)^{\delta_{ij}} S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi})^{1-\delta_{ij}} \\ &\propto \prod_{i=1}^{I}\prod_{j=1}^{n_i} S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi}) \left(1 - \frac{S_{\text{pop}}(t_{ijR}|\boldsymbol{\varphi},)}{S_{\text{pop}}(t_{ijL}|\boldsymbol{\varphi})}\right)^{\delta_{ij}}, \end{aligned} \tag{4.7}$$

where $\boldsymbol{\varphi} = (\boldsymbol{b}, \boldsymbol{\beta}, \mu, \sigma, \varsigma)$. For a Bayesian analysis, we assume the prior densities for parameters are $b_j \sim N(0, \sigma_b^2)$ for $j = 0, \ldots, (p_1 - 1)$; $\beta_j \sim N(0, \sigma_\beta^2)$ for $j = 1, \ldots, p_2$; $\mu \sim N(0, \sigma_\mu^2)$, $\varsigma \sim N(0, \sigma_\varsigma^2)$ and $\sigma^2 \sim IG(a_\sigma, b_\sigma)$, where $IG(a, b)$ is an inverse-gamma distribution with mean $b/(a-1)$ and variance $b^2/\{(a-1)^2(a-2)\}$ and σ_b, σ_β, σ_μ, σ_ς, a_σ and b_σ are known hyperparameters. In several areas, especial in medicine, the available prior information is also importance to be considered in the analysis. Therefore, we specify the hyperparameters to ensure vague prior information following the analysis results obtained by Carlin & Banerjee (2003), that is let $\sigma_b^2 = 1$, $\sigma_\beta^2 = 1$, $\sigma_\mu^2 = 10^2$, $\sigma_\varsigma^2 = 10^2$, $a_\sigma = 2$ and $b_\sigma = 1$. For the parameters of MCAR distribution a_1, a_2 and $\boldsymbol{\Lambda}$, the informative prior distributions are considered following (Carlin & Banerjee, 2003), that is let $a_i \sim \text{Beta}(18, 2)$, for $i = 1, 2$, and $\boldsymbol{\Lambda} \sim \text{Wishart}(n_0, \Lambda_0)$, with $n_0 = 2$ and $\Lambda_0 = 0.1\mathbf{I_2}$ where $\mathbf{I_2}$ is a unit matrix of size 2.

To avoid range restrictions on the parameters a_i and σ^2, considering the transformations $\upsilon = \log(\sigma^2) \in \mathbb{R}$ and $\rho_i = \log(a_i/(1-a_i)) \in \mathbb{R}$, then, the joint posterior density is given by

$$\begin{aligned} \pi(\boldsymbol{\vartheta}|\boldsymbol{D}_{\text{obs}}) &\propto L(\boldsymbol{\varphi}|\boldsymbol{D}_{obs}) \exp\left\{-\frac{1}{2}\left[\sigma_b^{-2}\sum_{i=0}^{p_1} b_i^2 + \sigma_\beta^{-2}\sum_{i=1}^{p_2}\beta_i^2 + \frac{\mu^2}{\sigma_\mu^2} + \frac{\varsigma^2}{\sigma_\varsigma^2} + \frac{\exp(\upsilon)}{\sigma_\sigma^2}\right]\right. \\ &+ \boldsymbol{\psi}^\top \left[\boldsymbol{\Lambda} \otimes (\boldsymbol{D_W} - \boldsymbol{aW})\right] \boldsymbol{\psi} + \log|\boldsymbol{\Lambda} \otimes \boldsymbol{aW}| + \frac{n_0 - 4}{2}\log|\boldsymbol{\Lambda}| - \frac{1}{2}tr(\Lambda_0^{-1}\Lambda) \\ &+ \left.\upsilon\right\} \frac{\exp(17\rho_i)}{(1+\exp(\rho_i))^{18}}, \end{aligned}$$

where $\boldsymbol{\varphi} = (\boldsymbol{b}, \boldsymbol{\beta}, \mu, \varsigma, \upsilon^{-1})$ with $\upsilon^{-1} = \exp\left(\frac{1}{2}\upsilon\right) = \sigma$.

This joint posterior density is analytically intractable. So, we based our inference on the Markov chain Monte Carlo (MCMC) simulation methods. We can observed that the full conditional distributions for parameters $\boldsymbol{b}$, $\boldsymbol{\beta}$, μ, ς, σ^2 and $\boldsymbol{\rho}$ have not closed forms, thus we will use the Metropolis-Hastings algorithm to generate a posteriori samples for these parameter. However, the Gibbs sampler algorithm is used to generate a posteriori sample for the parameter $\boldsymbol{\Lambda}$, because its the full conditional distribution has a closed form. The full conditional distribution $\pi(\boldsymbol{\Lambda}|\boldsymbol{\vartheta}_{(-\boldsymbol{\Lambda})}, \boldsymbol{D}_{\text{obs}}))$ is

proportional to

$$\begin{aligned} & \pi(\boldsymbol{\psi}|\boldsymbol{\Lambda},\boldsymbol{a})\pi(\boldsymbol{\Lambda}) \\ \propto\ & |\boldsymbol{\Lambda}\otimes\boldsymbol{D_W}-\boldsymbol{aW}|^{1/2}\exp\left(-\frac{1}{2}\boldsymbol{\psi}^\top(\boldsymbol{D_W}-\boldsymbol{aW})\boldsymbol{\psi}\right)|\boldsymbol{\Lambda}|^{(n_0-4)/2}\exp\left(-\frac{1}{2}tr(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda})\right) \\ \propto\ & |\boldsymbol{\Lambda}|^{(I+n_0-4)/2}\exp\left(-\frac{1}{2}tr((\boldsymbol{\Lambda}_0^{-1}+\boldsymbol{B})\boldsymbol{\Lambda})\right), \end{aligned} \tag{4.8}$$

where

$$\boldsymbol{B}=\begin{bmatrix} tr(\boldsymbol{R_1U}(\boldsymbol{R_1U})^\top) & tr(\boldsymbol{R_1U}(\boldsymbol{R_2V})^\top) \\ tr(\boldsymbol{R_2V}(\boldsymbol{R_1U})^\top) & tr(\boldsymbol{R_2V}(\boldsymbol{R_2V})^\top) \end{bmatrix}$$

Thus, the full conditional distribution for $\boldsymbol{\Lambda}$ can be taken the Wishart distribution with scala matrix $(\Lambda_0^{-1}+B)^{-1}$ and degrees of freedom $I+n_0$.

4.5 Application

We now apply the proposed method to the interval-censored smoking cessation data presented in Section 1.1. Because of the high computational cost, we implement the MCMC in algorithms C language and the results were analyzed in R language (R Development Core Team (2010)) through the "coda" package (Plummer *et al.* (2005)). All of our MCMC algoritms ran a total of 100,000 iterations discarding the first 40,000 realizations as burn-in and thinning to every fifth iteration. Posterior results are then based on 7,500 realizations of the Markov chain. Our Meteropolis acceptance rate for these parameters ranged from 25% to 50%. The convergence was checked using the Geweke diagnostic which did not indicate lack of convergence. The models are compared using DIC criterion.

We fitted some particular case of PS cure rate model which described in Section 4.3. The values of Bayesian criteria for fitted models are presented in Table 4.1, according to the DIC and *pd* value the complementary promotion time cure model stand outs as the best models and all of the cure rate models under last activations are better than the models under first activations. Comparing the obtained DIC values with the values presented in the paper of Carlin & Banerjee (2003), where they proposed the mixture cure model with the spatial frailty, assuming Y_{cij} has Weibull or gamma distributions. we can conclude that all models in Table 4.1 are more adequate since all our DIC values are smaller.

Here we select four cure models, which have lower DIC values, as our working models. There are cure models under the random and last activation. Since the selected cure models are obtained considering their initial number of competing causes related to the occurrence of event

Table 4.1: DIC values for the fitted cure rate models

Activation	Distribution of M_{ij}	Model	DIC	pd
Random	Bernoulli	Mixture cure model	388.5	5.03
First	Poisson	Promotion time cure model	393.1	9.69
	Geometric	Geometric cure rate model	392.9	8.38
	Logarithmic	Logarithmic cure rate model	393.8	7.20
Last	Poisson	Complementary promotion time cure model	383.5	4.06
	Geometric	Complementary geometric cure rate model	386.8	4.62
	Logarithmic	Complementary logarithmic cure rate model	390.8	5.95

M_{ij} has Bernoulli, Poisson, geometric, logarithmic distributions, to simply the notation we call the mixture model, Complementary promotion time cure model, Complementary geometric cure model, and Complementary logarithmic cure model by the Bernoulli, Poisson, geometric and Logarithmic, respectively. The posterior summaries of the parameter for the selected models are presented in the Table 4.2. We note that the signs of the regression coefficients $b_{\text{intercept}}$, b_{sex}, $b_{\text{treatment}}$, $b_{\text{consumption}}$, b_{duration}, β_{sex} and $\beta_{\text{treatment}}$ are the same for all selected models. However, the sign of coefficients $\beta_{\text{consumption}}$ and β_{duration} are negative for mixture model and positive for other three models. The mean of the parameters μ, σ^2, ς, a_1, a_2, Λ_{11}, Λ_{22} and Λ_{12} have close values for all models. Note that the β parameters are related to the cure fraction, thus the interpolation of cure rate can be obtained. The data shows that women smokers have lower probability of quitting than men smokers, individuals with special intervention have higher probability of quitting smoking than those with usual care and the individuals with a higher level of cigarette consumption have lower probability of quitting than others. On the other hand, we also note that women smokers have high hazard rate of relapse time than the men and individuals with special intervention have lower hazard rate than those with usual care. Here, the cigarette consumption has little effect on hazard rate of relapse time and the duration of smoking habit has little effect on not only on hazard rate of relapse time but also on probability of quitting.

In order to detect possible influential observations in the posterior distribution of the parameters of the fitted models, the estimates of K-L divergence and L_1 distance (two particular cases of ψ-divergence measures), which were obtained from the posteriori sample of the models'parameters, are presented in Figure 4.1. We note that the individual 14 was detected by both divergence measures for the complementary promotion time cure model and no influential individual was detected in the other three models. The detected individual is a male patient who had a 32-year smoking habit, consumed 60 cigarettes per day, lived in Rochester city and relapsed during the treatment.

Table 4.2: Posterior summaries of the parameter of the selected models for the smoking cessation data.

	Bernoulli				Poisson				Geometric				Logarithmic			
Parameter	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%
$b_{\text{intercept}}$	0.395	0.839	-1.195	2.141	0.259	0.744	-1.246	1.693	0.232	0.808	-1.393	1.838	0.507	0.849	-1.182	2.156
b_{sex}	0.238	0.791	-1.264	1.880	0.082	0.512	-0.998	1.079	0.072	0.591	-1.154	1.220	0.192	0.750	-1.337	1.711
$b_{\text{treatment}}$	-0.242	0.875	-1.885	1.558	-0.500	0.536	-1.620	0.497	-0.484	0.612	-1.730	0.732	-0.328	0.787	-1.789	1.348
$b_{\text{consumption}}$	0.047	0.045	-0.039	0.134	0.030	0.032	-0.042	0.080	0.036	0.038	-0.044	0.103	0.052	0.043	-0.037	0.133
b_{duration}	-0.001	0.051	-0.088	0.104	-0.024	0.046	-0.095	0.078	-0.016	0.051	-0.096	0.098	-0.004	0.052	-0.084	0.113
β_{sex}	0.361	0.375	-0.371	1.138	0.371	0.308	-0.198	1.041	0.389	0.301	-0.175	1.025	0.433	0.281	-0.084	1.018
$\beta_{\text{treatment}}$	-0.237	0.429	-0.973	0.746	-0.267	0.338	-0.889	0.466	-0.209	0.315	-0.773	0.473	-0.165	0.296	-0.701	0.462
$\beta_{\text{consumption}}$	-0.004	0.021	-0.052	0.033	0.002	0.019	-0.039	0.035	0.005	0.018	-0.037	0.033	0.009	0.016	-0.029	0.035
β_{duration}	-0.014	0.027	-0.054	0.051	0.005	0.021	-0.029	0.055	0.009	0.020	-0.024	0.056	0.007	0.019	-0.023	0.052
μ	0.751	0.199	0.463	1.288	0.676	0.185	0.382	1.118	0.657	0.176	0.377	1.078	0.653	0.175	0.384	1.071
σ^2	0.841	0.363	0.429	1.835	0.786	0.307	0.410	1.589	0.741	0.288	0.397	1.475	0.711	0.266	0.390	1.410
ς	1.379	0.668	0.348	2.981	1.580	0.852	0.375	3.704	1.456	0.776	0.278	3.337	1.318	0.689	0.268	2.960
a_1	0.901	0.065	0.744	0.987	0.900	0.064	0.744	0.987	0.900	0.064	0.743	0.986	0.899	0.067	0.739	0.987
a_2	0.900	0.066	0.742	0.987	0.900	0.065	0.740	0.987	0.901	0.064	0.746	0.987	0.900	0.065	0.738	0.987
Λ_{11}	2.683	0.643	1.560	4.057	2.699	0.657	1.581	4.149	2.700	0.642	1.584	4.124	2.668	0.651	1.555	4.102
Λ_{22}	2.556	0.638	1.473	3.970	2.567	0.638	1.500	3.974	2.556	0.623	1.505	3.912	2.575	0.642	1.495	3.964
Λ_{12}	0.022	0.470	-0.747	1.085	0.038	0.475	-0.744	1.103	0.043	0.471	-0.737	1.081	0.037	0.470	-0.735	1.089
Σ_{11}	0.410	0.113	0.263	0.689	0.409	0.113	0.259	0.695	0.407	0.109	0.258	0.683	0.413	0.114	0.262	0.698
Σ_{22}	0.433	0.123	0.256	0.732	0.432	0.125	0.256	0.713	0.432	0.121	0.260	0.724	0.431	0.131	0.257	0.717
$\frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$	0.006	0.182	-0.369	0.346	-0.001	0.182	-0.365	0.334	-0.003	0.181	-0.371	0.337	0.000	0.182	-0.376	0.344

where Λ_{ij} is the element of precision matrix $\boldsymbol{\Lambda}$ in position (i, j), and Σ_{ij} is the element of matrix $\boldsymbol{\Sigma} = \Lambda^{-1}$ in position (i, j), this Σ_{11} is the spatial variance component of $\boldsymbol{U}$ and Σ_{22} is the spatial variance component of $\boldsymbol{V}$, $\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$ denote their correlation.

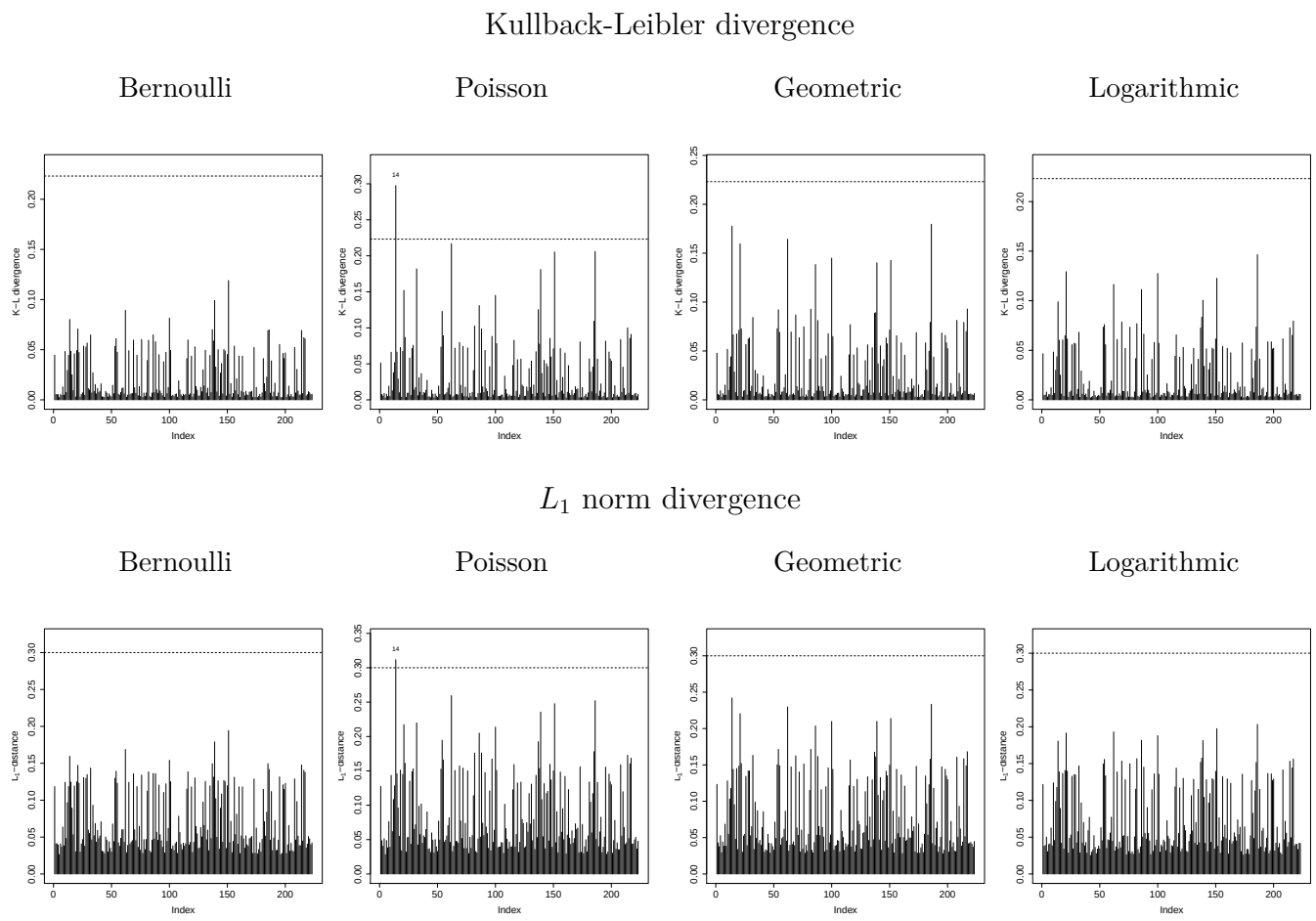

Figure 4.1: Index plots of K-L and L_1 divergence measures from the fitted cure models.

To reveal the impact of these possible influential observations on the parameter estimates and inference, we removed these observations and refitted the model. We also calculated the relative variations (RV) for the posterior mean of the parameters, defined by $RV = (\hat{\vartheta}_{d,-\{14\}} - \hat{\vartheta}_d)/\hat{\vartheta}_d \times 100$, for all d, where d is the index of the parameters, $\hat{\vartheta}_{d,-\{72,151\}}$ denotes the posterior mean of $\vartheta_{d,-\{14\}}$, after the set of observations $\{14\}$ has been removed. The posterior summaries of the parameters for the refitted model and RV for the posterior mean of the parameters are presented in Table 4.3. We note that RV of the the parameters $b_{\text{consumption}}$, b_{duration}, $\beta_{\text{consumption}}$, β_{duration} and $\frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$ have higher values. However, these parameters are not sensitive since their posterior means are still near zero. The estimated standard deviation $\Sigma_{11}^{1/2}$ of random spatial effects in the survival model is 0.640, and the estimated standard deviation $\Sigma_{22}^{1/2}$ of random spatial effects in cure rate is 0.658 which indicate there is considerable heterogeneity among the clusters. Moreover, there are no correlations between the spatial effects $\boldsymbol{U}$ and $\boldsymbol{V}$.

Figure 4.2 maps the posterior means and standard deviations of frailties $\boldsymbol{U}$ and $\boldsymbol{V}$ in the complementary promotion time cure model. For the frailties $\boldsymbol{U}$ for which the high value indicates a high relapse rate, we can note that the north regions and some cities of the south region have higher values, that is, the individuals in these regions have higher relapse rates than the others.

Table 4.3: Posterior summaries of the parameter of the complementary promotion time cure model without detected individual 14.

	Mean	RV	SD	2.50%	97.50%
$b_{\text{intercept}}$	0.136	-47	0.747	-1.346	1.616
b_{sex}	0.082	-1	0.458	-0.836	0.982
$b_{\text{treatment}}$	-0.397	-21	0.466	-1.370	0.476
$b_{\text{consumption}}$	0.044	45	0.032	-0.041	0.089
b_{duration}	-0.038	61	0.042	-0.100	0.068
β_{sex}	0.435	17	0.304	-0.133	1.078
$\beta_{\text{treatment}}$	-0.261	-2	0.321	-0.840	0.417
$\beta_{\text{consumption}}$	-0.006	-381	0.021	-0.047	0.030
β_{duration}	0.012	156	0.023	-0.027	0.061
μ	0.699	3	0.233	0.396	1.244
σ^2	0.797	1	0.419	0.409	1.705
ς	1.454	-8	0.809	0.317	3.444
a_1	0.900	0	0.065	0.740	0.987
a_2	0.899	0	0.066	0.736	0.987
Λ_{11}	2.698	0	0.637	1.611	4.058
Λ_{22}	2.573	0	0.630	1.516	3.971
Λ_{12}	0.038	0	0.466	-0.742	1.042
Σ_{11}	0.407	0	0.109	0.262	0.672
Σ_{22}	0.429	0	0.121	0.257	0.707
$\frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$	-0.001	136	0.179	-0.360	0.344

In contrast, the center region (Rochester city) and some northeast cities show slightly better than avenge cessation behavior, which also can be observed by the frailties $\boldsymbol{V}$, for which the high value indicates lower cure probability. In general, all center regions have close cure probabilities. The posterior standard deviations of frailties $\boldsymbol{U}$ and $\boldsymbol{V}$ also have close values. Both maps show that the cities round the center regions have lower values and Waseca has the highest value. The DIC value for the fitted models is 382.24, which is lower than in the model for the data without removing the detected observation.

Figure 4.3 presented the survival functions under the Complementary Promotion time cure model stratified by treatments and sex for patients who residence in Rochester city with duration of smoking habit equal to 20, 25 and 33 years, and cigarette consumption equal to 25, 31 and 35 cigarettes per day, which correspond to the first, second and third quantiles of duration of smoking habit and cigarette consumption. The surviving probability decreases more rapidly for patients received the usual care than special intervention; the surviving probability of the female patients is lower than male patients under the same condition, but we note that the surviving probability of the female patients have special care is close the male patients only received usual care; the surviving probability of the female patients, who have high level of duration of smoking habit and cigarette consumption (Duration = 33 and Cigarette = 35), just a bit lower than the patients who have low

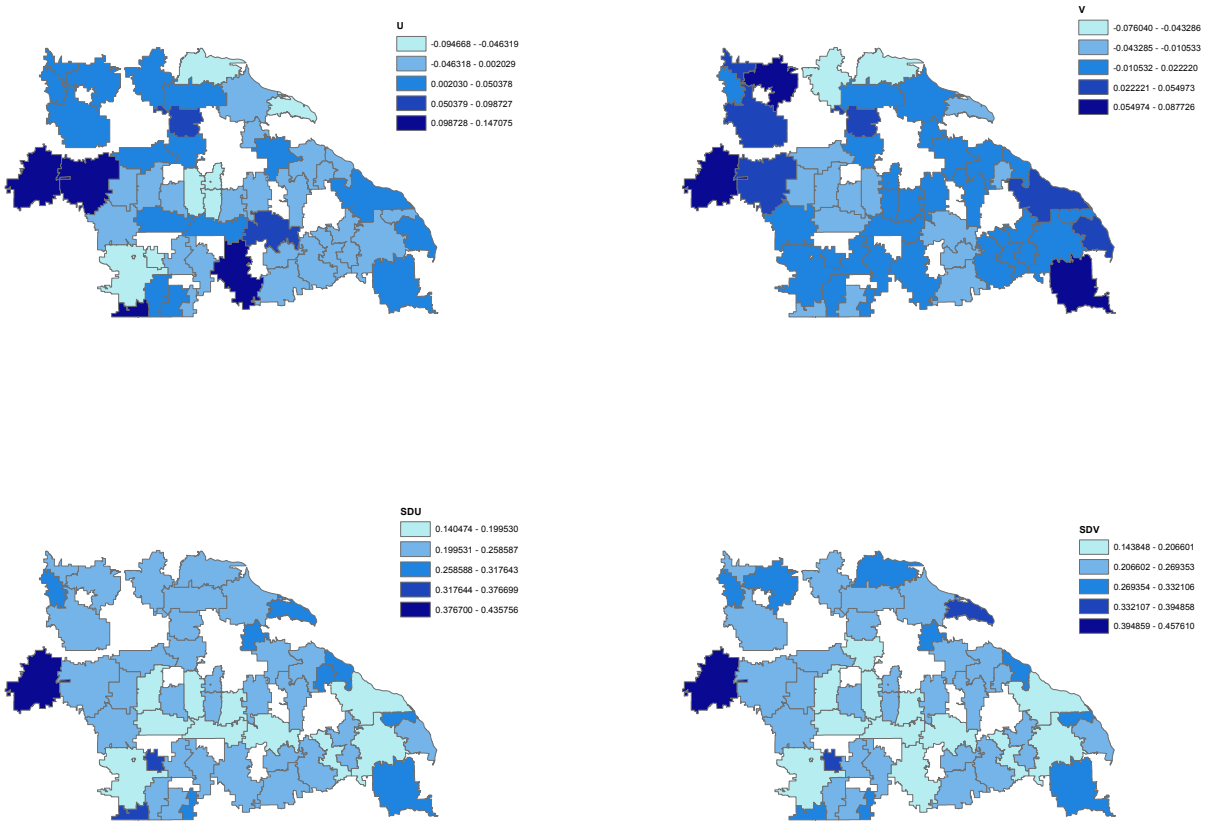

Figure 4.2: Maps of posterior means for frailties U (upper-left panel) and V (upper-right panel) and posterior standard derivations for frailties U (lower-left panel) and V (lower-right panel).

and median levels of duration of smoking habit and cigarette consumption; the surviving probability of the male patients does not influence by duration of smoking habit and cigarette consumption.

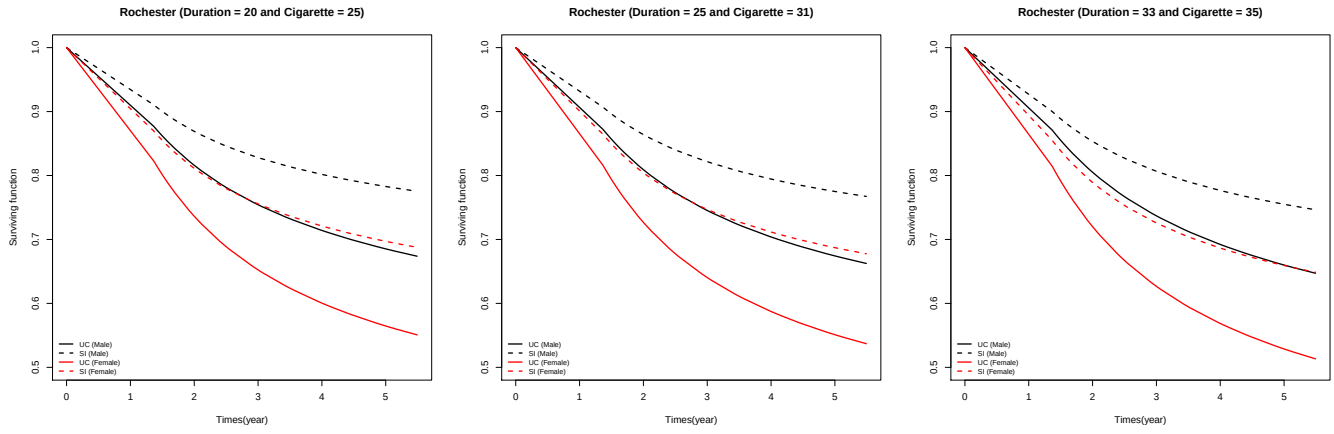

Figure 4.3: Surviving function under the complementary promotion time cure model stratified by treatments and sex for patients who residence in Rochester city with three levels of duration of smoking habit and cigarette consumption.

Figure 4.3 presents the survival functions under the complementary promotion time cure model stratified by treatments and sex for patients residing in Rochester with durations of smoking

habit equal to 20, 25 and 33 years, and daily cigarette consumption equal to 25, 31 and 35, which correspond to the first, second and third quantiles of duration of smoking habit and cigarette consumption. The survival probability decreases more rapidly for patients that received the usual care than those receiving special intervention. Also, the survival probability of the female patients is lower than males under the same condition, but we note that the survival probability of the female patients submitted to special care is close to the male patients who only received usual care; the survival probability of the female patients who had long smoking habit and high cigarette consumption (Duration = 33 and Cigarettes = 35) is just a bit lower than the patients who had low and medium levels of duration and cigarette consumption; and the survival probability of male patients was not influenced by duration of smoking habit and cigarette consumption.

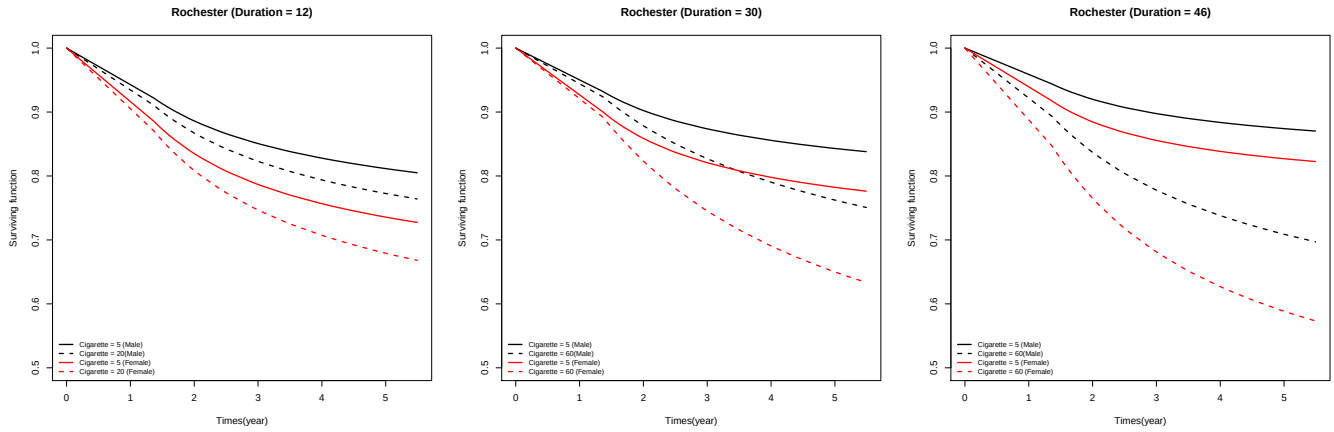

Figure 4.4: Surviving function under the complementary promotion time cure model stratified by cigarette consumption and sex for patients who residence in Rochester city with duration of smoking habit equal 12, 30 and 46 years.

4.6 Conclusions

In this work, we propose a PS cure rate model for spatially correlated interval-censored data based on the generalized extreme value distribution. The proposed model is very flexible and generalizes the Bernoulli, geometric, Poisson, and logarithm models. Furthermore, it can be tested for best fit in a straightforward way. We use MCMC methods with Bayesian inference for our models and the Bayesian comparison criterion for model comparison. The results of the application show that the proposed model has better fit than the WNBCR and CWNBCR models. Our model also performs better than the PHNBCR and CPHNBCR models, and also has the advantage of fewer parameters. Moreover, we also conducted a Bayesian case deletion influence diagnosis to examine outlying and influential observations in the application, observing that the proposed model is not sensitive to influential observations.

Chapter 5

D-Measure: A Bayesian model selection criterion for survival data

5.1 Introduction

Model assessment and comparison is of extreme importance in statistical analysis. Focusing on Bayesian model assessment and comparison, various methods have been proposed in recent decades, particularly relying on Bayes factors, boosted by significant advances in computer technology. However, it is well known that proper prior distributions are needed for using these methods. Later, Spiegelhalter *et al.* (2002) proposed the deviance information criterion (DIC), which is one of most used Bayesian criteria for the model comparison. Several recent papers use DIC for comparing models, including the survival models which consider random effects or frailties (Banerjee & Carlin, 2004; Carvalho Lopes & Bolfarine, 2012; Pan *et al.*, 2014; Li Dan & Dey, 2015),Carvalho Lopes & Bolfarine (2012), Pan *et al.* (2014), Li Dan & Dey (2015), just to name a few). It is well known that, in order to ensure the consistency of parameters, it is needed first integrated out the random effects or frailties, and then compute the criterion (Ando, 2007). However, a numerical integration always gives unstable results, i.e, different values of a criterion may be obtained by considering different numerical iteration methods.

Another alternative is to use criteria which are constructed from the posterior predictive distribution. Let $y = (y_1, \ldots, y_n)$ denotes the observed data from an experiment, with the joint sampling density of the y_i's denoted by $p(y|\mathbf{\Omega})$, where $\mathbf{\Omega}$ is a vector of indexing parameters. Let $z = (z_1, \ldots, z_n)$ denotes the future values of an imagined replicated experiment, i.e, z is a future response vector with the same sampling density of $y|\mathbf{\Omega}$. A good model should make predictions close to what has been observed for an identical experiment. With this idea, Ibrahim & Laud (1994) proposed

L measure as the expected squared Euclidean distance between y and z, $L = E\left[(z-y)'(z-y)\right]$, where the expectation is taken with respect to the posterior predictive distribution of $z|y$, $p(z|y) = \int p(z|\Omega)p(\Omega|y)d_\Omega$. Afterwards, Ibrahim *et al.* (2001b) extended this measure to a general Bayesian criterion for modelling assessment. Recently, Gu *et al.* (2011) proposed the M-measure for a right censoring data, by using counting process of number of deaths over time to be compared with the number of observed deaths over time, in order to define the measure of model adequacy.

In this work we propose a new model assessment and comparison criteria based on the well known survival function, which plays a key role in the survival analysis. Such criterion is a distance based Bayesian model selection criterion for survival data, hereafter the D-measure. The D-measure is constructed from the posterior predictive distribution of the data, it can be viewed as a Bayesian goodness-of-fit statistic, which measures the performance of a model by a combination of how close its predictions are from the observed data. It can be used for all kind of survival data sets, such as uncensored data, right censored data and interval censored data. Moreover, it is an alternative criterion which can be used to compare cure rate models in presence of random effects or frailties.

5.2 D-measure

Let $t_1, \ldots, t_n$ be random samples from the density function $f(t|\boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is a vector of indexing parameters and assuming that $\boldsymbol{\Omega}$ has a prior distribution $\pi(\Omega)$. In the context of survival analysis, t represents the lifetime or time to failure of an individual, usually, it is assumed to be continuous. In this thesis, we denote the observed data by $\boldsymbol{D}$ and we allow the t_i's to be fully observed, right censored, or interval censored. In the right censored case, t_i may be a failure time or a censored time. In the interval censored case, we only observe the interval $(L_i, R_i]$ in which t_i occurred. Now, let $t_{1p}, \ldots, t_{np}$ be random samples from the posterior predictive distribution

$$\pi(t_{ip}|\boldsymbol{D}) = \int f(t_{ip}|\boldsymbol{\Omega})\pi(\boldsymbol{\Omega}|\boldsymbol{D})d\boldsymbol{\Omega}, \tag{5.1}$$

where $f(t_{ip}|\boldsymbol{\Omega})$ is the sampling density for ith subject conditional upon $\boldsymbol{\Omega}$ being known and $\pi(\boldsymbol{\Omega}|\boldsymbol{D})$ is the posterior distribution of $\boldsymbol{\Omega}$. The D-measure for model m, is defined as

$$D_m = E_{T_{ip}|\boldsymbol{D}}\left[\int_0^\tau \parallel \hat{S}_0(t) - \hat{S}_p(t) \parallel dt\right] = E\left[\int_0^\tau \parallel \hat{S}_0(t) - \hat{S}_p(t) \parallel dt \mid \boldsymbol{D}\right], \tag{5.2}$$

where $\tau > 0$, $\hat{S}_0(t)$ and $\hat{S}_p(t)$ denote the (smoothing) non-parametric survival function from observed data $\boldsymbol{D}$ and from the posterior predicted samples, respectively. $\parallel\ \parallel$ denotes the norm, thus, different norms can be used in the formulation, such as the absolute value, the square and the maximum

value. Smaller values of D-measure indicate better fit to the observed data as well as more precise predictive fit for the model.

The proposed D-measure can be computed with a two-step procedure. First, we sample the posterior samples $\{\Omega_k\}_{k=1}^{G}$ from the posterior density $\pi(\boldsymbol{\Omega}|\boldsymbol{D})$ for the model m. Second, we simulate the samples from the predictive distribution. Note that, given the posterior samples $\{\Omega_{(j)}\}_{j=1}^{G}$ from $\pi(\boldsymbol{\Omega}|\boldsymbol{D})$, $t_{ip(j)}$ can be sampled from $f(t_{ip}|\boldsymbol{\Omega} = \boldsymbol{\Omega}_{(j)})$ for $i = 1, \ldots, n$ and $j = 1, \ldots, G$, that is, we obtained the samples $\{t_{ip(j)}\}_{j=1}^{G}$ from the posterior predictive distribution of the ith subject, via $\pi(t_{ip}|\boldsymbol{D})$. The procedures to compute the proposed measure for each kind of data (uncensored data, right censoring data and interval censoring data) are presented as follow.

5.2.1 Uncensored data

For uncensored data, the proposed measure given in (5.2) is obtained fixing $\tau = \infty$, $\hat{S}_0(t)$ and $\hat{S}_p(t)$ are the empiric survival function from observed data $\boldsymbol{D}$ and from the posterior predicted samples, respectively. Given the posterior samples $\boldsymbol{\Omega}_{(j)}$, the sampling of $t_{ip(j)}$ can be easily done. One of the most used methods is using the inverse cumulative distribution function (quantile function). The numerical approximation of D-measure given in (5.2)is given by

$$D_m \approx \frac{1}{G}\sum_{j=1}^{G}\sum_{q=1}^{Q} \mid \hat{S}_0(a_q) - \hat{S}_{pj}(a_q) \mid \Delta_q,$$

where $0 = a_0 < a_1 < \ldots < a_Q = \infty$ are all distinct points and $\Delta_q = a_{q+1} - a_q$.

5.2.2 Right censored data

For a right censored data, let $D_i = (y_i, \delta_i, \boldsymbol{x}_i)$ denotes the observed data for the ith individual, where $\boldsymbol{x}_i$ is a covariate vector, $y_i = \min(T_i, C_i)$ is the observed failure time, $\delta_i = I[T_i \leq C_i]$ is the censoring indicator, in which T_i is the non-informative random failure time and C_i is the non-informative random censoring time. In this case, because the censoring indicator δ_i is also part of the observed data, the closeness between the observed $Y_i = \min(T_i, C_i)$ and the predicted Y_{ip} is not appropriated. Thus, the proposed measure for a right censoring data for model m, which we denoted by D_m^r, is defined as

$$D_m^r = E\left[\int_0^{\tau} \parallel \hat{S}_0(t) - \hat{S}_p(t) \parallel dt|\boldsymbol{D}, F_C = \hat{F}_{CKM}\right], \tag{5.3}$$

where $\tau = \max_i\{y_i\}$, $\hat{S}_0(t)$ and $\hat{S}_p(t)$ denote the the Kaplan-Meier estimator of the survival function from observed data $\boldsymbol{D}$ and from the posterior predicted samples, respectively. The distribution of censoring variable C_i is assumed to be known as $\hat{F}_{CKM}$, the Kaplan-Meier estimator of the cumulative distribution function of C from $\boldsymbol{D}$.

Given the posterior samples $\{\Omega_{(j)}\}_{j=1}^{G}$ from the posterior density $\pi(\boldsymbol{\Omega}|\boldsymbol{D})$ for the model m, the $\{y_{ipj}, \delta_{ipj}\}$ need to be simulated from the predictive distribution (Y_{ip}, δ_{ip}) assuming the distribution of C to be $\hat{F}_{CKM}$. The algorithm of generating the samples $\{y_{ipj}, \delta_{ipj}\}$ for $j = 1 \ldots, G$ is presented as follow:

Step (i): Sample each C_{ipj} from the Kaplan-Meier cumulative density function $\hat{F}_{CKM}$;

Step (ii): Sample $\delta_{ipj} \sim$ Bernulli$(F(C_{ipj}))$ where $F(C_{ipj}) = P(T_{ip} \leq C_{ipj})$ is the c.d.f. of traditional survival model or cure rate model and set $y_{ipj} = C_{ipj}$ if $\delta_{ipj} = 0$;

Step (iii): If $\delta_{ipj} = 1$, sample $U_{ij} \sim U(0, 1)$ and set $y_{ipj} = F_c^{-1}(U_{ij})$, where $F_c(y) = F(T_{ip} \leq y|\delta_{ip} = 1) = F(y)/F(C_{ipj})$ for $0 < y < C_{ipj}$, if $F(\cdot)$ is c.d.f. of cure rate model; else let $F_c(y) = F(y)$.

Once the posterior predictive samples are obtained, the numerical approximation of D-measure given in (5.3) are given by

$$D_m^r \approx \frac{1}{G}\sum_{j=1}^{G}\sum_{q=1}^{\Upsilon_j} \mid \hat{S}_0(a_q) - \hat{S}_{pj}(a_q) \mid \Delta_q,$$

where $0 = a_0 < a_1 < \ldots < a_{\Upsilon_j} = \tau$ are distinct points where $S_0(a_q)$ and $S_{pj}(a_q)$ have jumps and $\Delta_q = a_{q+1} - a_q$ and a_{Υ_j} is defined as the maximum observed failure time of the observed data and the posterior predictive sample for each j.

5.2.3 Interval censored data

For interval censored data, unlike right censored data, the exact failure time can not be observed. In the literature, there are two types of interval censored data: Case 1 - interval censored data (or current status data), and Case 2 - interval censored data.

For the Case 1 of interval censored data, the only knowledge about the exact failure time is whether it has occurred before observed time or not. Let T_i's denote the unobservable failure times and assume that the observed time have the form $\{(L_i, \delta_i), i = 1, \cdots, n\}$, where L_i denotes the observation time for subject i independent of T_i and $\delta_i = I(T_i \leq L_i)$. Thus, $\delta_i = 1$ indicates that the event of interest occurred before the observed time L_i and $\delta_i = 0$ indicates the event of interest

maybe occur after observed time L_i (right censored) or never occur the event (cured). For the Case 2 of interval censored data, the time intervals $(L_i, R_i]$ are observed, where $0 \leq L_i < R_i \leq \infty$. $R_i < \infty$ indicates that the event of interest can occur in any time of the observed time interval $(L_i, R_i]$ and $R_i = \infty$ indicates a right censoring or cure. The interval censoring indicator is defined as $\delta_i = I(R_i < \infty)$. For the special case in which the survival time is right-(left-) censored, $R_i = \infty (L_i = 0)$, whereas for exact observations, $L_i = R_i$.

Let C denotes the censoring variables, and $D_i = (l_i, , \delta_i, \boldsymbol{x}_i)$ and $D_i = ((l_i, r_i], \delta_i, \boldsymbol{x}_i)$ denote the observed data for the ith individual for "case 1" interval censoring data and the "case 2" interval censoring data, respectively. The D measure for the interval censoring data for model m is defined as

$$D_m^I = E\left[\int_0^\tau \parallel \hat{S}_0(t) - \hat{S}_p(t) \parallel dt | \boldsymbol{D}, F_C = \hat{F}_{KM^*}\right], \tag{5.4}$$

where $\tau = \max_i\{l_i\}$ for the Case 1 (interval censoring data) and $\tau = \max_i\{l_i, r_i; r_i < \infty\}$ for the Case 2 (interval censoring data); $\hat{S}_0(t)$ and $\hat{S}_p(t)$ are the nonparametric maximum likelihood estimator (NPMLE) of a survival function for the observed data and the generated sample, respectively. For Case 1, the NPMLE of a survival function has a close form, which can be obtained by using the max-min formula for isotonic regression (more detail see Barlow *et al.* (1972); Härdle (1989)). For Case 2, there are some algorithms which were proposed in the literature. The first one is the self-consistency algorithm that was developed by Turnbull (1976). It can be regarded as an application of the EM algorithm (Dempster *et al.*, 1977). Lately, Groeneboom (1995) introduced the ICM algorithms, which was then modified by Jongbloed (1998). At the same period, Wellner & Zhan (1997) proposed a hybrid algorithm, which is known as the EM-ICM algorithm. It basically combines the self-consistency algorithm and the ICM algorithm. The distribution of censoring variable is assumed to be known as $\hat{F}_{KM^*}$, the Kaplan-Meier estimator of the cumulative distribution function of the censoring variable from the data $\boldsymbol{D}^*$, where $\boldsymbol{D}^*$ is obtained by transforming the interval censoring data $\boldsymbol{D}$ to the right censoring data assuming that the midpoint of intervals are the exact failure times, i.e., $\{t_i \ ; \ t_i = 0.5l_i, \delta_i = 1, i = 1 \cdots, n\}$ and $\{t_i \ ; \ t_i = 0.5(r_i - l_i) + l_i, \delta_i = 1, i = 1 \cdots, n\}$ for Cases 1 and 2, respectively.

First we present the algorithm for generating the samples (l_{ipj}, δ_{ipj}), $j = 1 \ldots, G$, for the Case 1 of interval censoring data as follows:

Step (i): Sample each C_{ipj} from the Kaplan-Meier cumulative density function $\hat{F}_{KM^*}$;

Step (ii): Sample $\delta_{ipj} \sim$ Bernulli$(F(C_{ipj}))$ where $F(C_{ipj}) = P(T_{ip} \leq C_{ipj})$ is the (population) c.d.f. and set $l_{ipj} = C_{ipj}$ if $\delta_{ipj} = 0$;

Step (iii): If $\delta_{ipj} = 1$, sample $U_{ij} \sim U(0, 1)$ and set $t_{ipj} = F_c^{-1}(U_{ij})$, where $F_c(y) = F(T_{ip} \leq y | \delta_{ip} =$

1) = $F(y)/F(C_{ipj})$ for $0 < y < C_{ipj}$, if $F(\cdot)$ is c.d.f. of cure rate model; else let $F_c(y) = F(y)$;

Step (vi): Now, generate l_{ij} from distribution $U(0, l_{(\mathrm{m})})$, where $l_{(m)} = \max\{l_i; \delta_i = 1, i = 1, \ldots, n\}$. Then, l_{ipj} is chosen as that satisfying $t_{ipj} \leq l_{ij}$.

Given the posterior predictive samples, the the numerical approximation of M-measure given in (5.4) is obtained as

$$D_m^I \approx \frac{1}{G} \sum_{j=1}^{G} \sum_{q=1}^{\Upsilon_j} \mid \hat{S}_0(a_q) - \hat{S}_{pj}(a_q) \mid \Delta_q,$$

where $0 = a_0 < a_1 < \ldots < a_{\Upsilon_j} = \tau$ are the unique ordered elements of $\{0, l_i, l_{ipj}, ; \delta_i, \delta_{ip} = 1$ and $i, ip = 1, \ldots, N\}$. They are distinct points with the length $\Delta_q = a_q - a_{q-1}$, for $q = 1, \ldots, \Upsilon_q$, and a_{Υ_j} is defined as the maximum finite observed time of the observed data and the posterior predictive sample for each j.

For the Case 2 of interval censoring data, the algorithm for generating the samples $(l_{ipj}, r_{ipj}, \delta_{ipj})$, for $j = 1 \ldots, G$, is presented as follows:

Step (i): Sample each C_{ipj} from the Kaplan-Meier cumulative density function $\hat{F}_{KM^*}$;

Step (ii): Sample $\delta_{ipj} \sim$ Bernulli$(F(C_{ipj}))$ where $F(C_{ipj}) = P(T_{ip} \leq C_{ipj})$ is the (population) c.d.f. and set $l_{ipj} = C_{ipj}$ and $r_{ipj} = \infty$ if $\delta_{ipj} = 0$;

Step (iii): If $\delta_{ipj} = 1$, sample $U_{ij} \sim U(0, 1)$ and set $y_{ipj} = F_c^{-1}(U_{ij})$, where $F_c(y) = F(T_{ip} \leq y | \delta_{ip} = 1) = F(y)/F(C_{ipj})$ for $0 < y < C_{ipj}$, if $F(\cdot)$ is c.d.f. of cure rate model; else let $F_c(y) = F(y)$;

Step (vi): Now, create len_{ij} from distribution $U(d_{\min}, d_{\max})$ and s_{ij} from $U(0, 0.5\, l_{(1)})$. Then, from $(0, l_{ij}], (l_{ij}, s_{ij} + len_{ij}], \ldots, (s_{ij} + d \times len_{ij}, \infty]$, $d = 1, 2, \ldots,$ (l_{ipj}, r_{ipj}) is chosen as that satisfying $l_{ipj} < t_{ipj} \leq r_{ipj}$, where $l_{(1)}$ is the positive minimum observed value, $d_{\min}$ is the minimum finite length and $d_{\max}$ is the maximum finite length of observed data. (i.e, $d_{\min} = \min\{d_i, i = 1 \ldots, n\}$, $d_{\max} = \max\{d_i, i = 1 \ldots, n; d_i < \infty\}$, where $d_i = r_i - l_i$ denotes the length of the time interval for the ith subject of the observed data.)

Given the posterior predictive samples, the numerical approximation of M-measure given in (5.4) is obtained by

$$D_m^I \approx \frac{1}{G} \sum_{j=1}^{G} \sum_{q=1}^{\Upsilon_j} \mid \hat{S}_0(a_q) - \hat{S}_{pj}(a_q) \mid \Delta_q,$$

where $0 = a_0 < a_1 < \ldots < a_{\Upsilon_j} = \tau$ are the unique ordered elements of $\{0, l_i, r_i, l_{ipj}, r_{ipj}\, ;\, r_i, r_{ip} < \infty$ and $i, ip = 1, \ldots, N\}$. They are distinct points with the length $\Delta_q = a_q - a_{q-1}$, for $q = 1, \ldots, \Upsilon_q$, and a_{Υ_j} is defined as the maximum finite observed time of the observed data and the posterior predictive sample for each j.

5.3 Simulation

In this section, we consider some simulation studies to compare models using the proposed D-measure for three most used parametric models in survival analysis: survival models with right censoring, survival models with with right censoring and cure fraction, and cure rate models with interval censoring.

5.3.1 Survival model with right censored data

In order to generate survival data with right censoring, we consider the Weibull and Gamma distributions, which are two of the most used models in survival analysis. Then, the DIC and proposed D-measure are calculated for comparison. Therefore, we first generate the current data y_i from the Weibull model,

$$p(y_i|\alpha,\lambda_i) = \alpha^{y_i-1}\exp\{\lambda_i - y_i^{\alpha}\exp(\lambda_i)\}, \tag{5.5}$$

where $\lambda_i = \beta_0 + \beta_1 x_i$, $i = 1,\cdots,n$, n denotes the sample size, $\alpha = 2$, $\beta_0 = 1$ and $\beta_1 = -1$, the covariates x_i are i.i.d. Bernulli(0.5) variates and the observations are randomly right censored with censoring times $t_i = 0.75 \times y_i$. The prior distributions for the parameters are $\alpha \sim N(0, 10^2)I_{(0,\infty)}$, $\beta_0 \sim N(0, 10^2)$ and $\beta_1 \sim N(0, 10^2)$. Here $N(\mu, \sigma^2)I_{(a,b)}$ denotes the truncated normal distribution, which is the probability distribution of a normally distributed random variable whose value is bounded in (a, b).

We also generate the data y_i from the Gamma model,

$$p(y_i|k,\theta_i) = \frac{y_i^{k-1}}{\Gamma(k)\theta_i^k}\exp\left(\frac{y_i}{\theta_i}\right), \tag{5.6}$$

where $\theta_i = \beta_0^* + \beta_1^* x_i$, $i = 1,\cdots,n$, $k = 3.8$, $\beta_0^* = -1.8$ and $\beta_1^* = 0.5$, and the observations are randomly right censored with censoring times $t_i = 0.75 \times y_i$. The prior distributions for the parameters are $k^* \sim N(0, 10^2)I_{(0,\infty)}$, $\beta_0^* \sim N(0, 10^2)$ and $\beta_1^* \sim N(0, 10^2)$.

After generating the data sets, both models are used to be fitted by the generated date sets and the DIC and the D-measure are computed. In order to verify the performance of the D-measure, we calculated the percentage of samples in which the adjusted model was indicated as the best model according to the DIC and D-measure. Hereafter we call this percentage as correct rate. The different censoring levels and sample sizes are considered in the study and 1,000 replicates are conducted in each configuration. The results of simulations are presented in the Table 5.1. We observe that the correct rates obtained by proposed the D-measure are higher than DIC in the most parts of cases.

Although there are some cases that the correct rates of D-measure are lower than DIC, they have approximately the same values. On the other hand, considering the censoring level of 60% and the sample size of $n = 50$, the correct rates of the D-measure are much higher than the DIC ones.

Table 5.1: Percentage of samples in which the adjusted survival model was indicated as the best model according to the criteria for the right censoring data sets

True Model	Censoring level	Sample Size	Criteria			
				D-measure		
		n	DIC	Abs.	Square	Max
Weibull	30%	50	76.2%	**78.9%**	71.3%	70.6%
		100	85.8%	**88.2%**	75.2%	77.1%
		200	94.9%	**97.6%**	84.0%	85.7%
	60%	50	73.8%	**97.0%**	**95.0%**	**83.8%**
		100	87.9%	**98.9%**	**97.5%**	**91.0%**
		200	95.7%	**99.9%**	**99.4%**	**96.1%**
Gamma	30%	50	64.9%	**72.4%**	**80.6%**	**94.7%**
		100	70.8%	**78.5%**	**91.2%**	**98.8%**
		200	85.8%	**87.0%**	**97.8%**	**99.9%**
	60%	50	55.5%	44.7%	49.1%	**70.2%**
		100	59.5%	52.1%	**61.7%**	**83.4%**
		200	69.1%	60.9%	**76.5%**	**91.5%**

5.3.2 Cure rate model with right censored data

For right censored data in presence of cure rate, we consider the cure rate proportional odds model (CRPO model) studied by Gu *et al.* (2011). Note that this model can be characterized by the latent factors model of Cooner *et al.* (2007) with a geometric distribution for the number of latent factors and it may be derived in a context in which relapse occurs in patients with cancer.

Let M_i denotes the number of carcinogenic cells in the beginning of a treatment for the ith individual, and assume that M_i has a Geometric distribution with the mean θ. Let Y_j for $j = 1, \ldots, M_i$ denotes the failure time due to the jth latent cause, that is, the time until jth carcinogenic cell produces a detectable cancer. Supposing that given M_i, the random variables Y_j are assumed to be independent and identically distributed (i.i.d.) with c.d.f. $F(\cdot) = 1 - S(\cdot)$ and the presence of any of latent risk (i.e., $M_i \geq 1$) will ultimately lead to the occurrence of the event. Thus, the time to event of interest (time to detect cancer) is defined by the random variable $T = \min\{Y_j, j = 0, \cdots, M_i\}$, where $P(Y_0 = \infty) = 1$. The survival function for the population is given by $S_{pop}(t) = [1 + \theta F(t)]^{-1}$ and its cure fraction can be obtained by $p_0 = \lim_{t\to\infty} S_{pop}(t) = (1 + \theta)^{-1}$. As it is well known, the cure fraction plays a key role in the survival models with a cure fraction. Thus, we consider the parametrization of the model in terns of cure fraction. Therefore, the survival

function of the CRPO model can be written as

$$S_{pop}(t) = \left[1 + (p_0^{-1} - 1)F(t)\right]^{-1}.$$

The non-negative random variables Y_j's can have various different distributions, here we assume that they are Weibull and Gamma distributed, denoting the cure rate models as CRPO-W and CRPO-G, respectively. Similarly to the simulation study for the data set with right censoring, we generate the current data from the CRPO-W model,

$$S_{pop}(t_i) = \left\{1 + (p_{0i}^{-1} - 1)\left[1 - \exp(-t_i^{\alpha}\exp(\lambda))\right]\right\}^{-1}, \tag{5.7}$$

where $p_{0i} = \exp\{\beta_0 + \beta_1 x_i\}/(1 + \exp\{\beta_0 + \beta_1 x_i\})$, $i = 1, \cdots, n$, n, $\alpha = 2$, $\lambda = -2$, $\beta_0 = -0.5$ and $\beta_1 = -2$ and the covariates x_i are i.i.d. Bernoulli(0.5) variates. Here, the right censoring observations are generated from a uniform distribution $U(0, c)$, where $c = 30$ for the low censoring level (data with 30% censoring observations) and $c = 7$ for the high censoring level (data with 60% censoring observations). The prior distributions for the parameters are assumed to be $\alpha \sim N(0, 10^2)I_{(0,\infty)}$, $\lambda \sim N(0, 10^2)$, $\beta_0 \sim N(0, 10^2)$ and $\beta_1 \sim N(0, 10^2)$.

For the comparison purpose, we generate the data sets from the CRPO-G model,

$$S_{pop}(t_i) = \left\{1 + (p_{0i}^{-1} - 1)\left[\frac{1}{\Gamma(k)}\gamma\left(k, \frac{t_i}{\theta}\right)\right]\right\}^{-1}, \tag{5.8}$$

where $\gamma(k, t/\theta)$ is the lower incomplete gamma function, $p_{0i} = \exp\{\beta_0 + \beta_1 x_i\}/(1 + \exp\{\beta_0 + \beta_1 x_i\})$, $i = 1, \cdots, n$, n, $k = 2$, $\theta = 0.5$, $\beta_0 = -1$ and $\beta_1 = 1$. Here, the right censoring observations are generated from a uniform distribution $U(0, c)$, where $c = 30$ for the low censoring level (data with 30% censoring observations) and $c = 7$ for the high censoring level (data with 60% censoring observations). The prior distributions for the parameters are assumed to be $k \sim N(0, 10^2)I_{(0,\infty)}$, $\theta \sim N(0, 10^2)I_{(0,\infty)}$, $\beta_0 \sim N(0, 10^2)$ and $\beta_1 \sim N(0, 10^2)$. The Percentage of samples in which the adjusted cure rate model was indicated as the best model according to the criteria for the right censoring data sets in presence of a cure fraction are presented in the Table 5.2. We note that the correct rates obtained for the D-measure are higher than the DIC ones in most of the cases. In some cases the correct rates of D-measure are lower than the DIC, but they are almost the same values. The correct rates of the D-measure are much higher than DIC for the data sets which were generated from the CRPO-W model with 60% censoring observations. Moreover, we observe that the D-measure outperforms the DIC for small sample sizes.

Table 5.2: Percentage of samples in which the adjusted cure rate model was indicated as the best model according to the criteria for the right censoring data sets

True Model	Censoring level	Sample Size	Criteria			
				D-measure		
		n	DIC	abs	square	max
CRPO-W	30%	50	56.0%	**82.0%**	**80.2%**	54.1%
		100	78.3%	**80.3%**	**78.7%**	75.1%
		200	89.9%	**91.9%**	88.5%	81.4%
	60%	50	17.5%	**56.6%**	**48.3%**	**31.5%**
		100	19.0%	**66.6%**	**60.6%**	**45.2%**
		200	19.3%	**76.8%**	**75.2%**	**69.5%**
CRPO-G	30%	50	35.5%	**54.6%**	**49.7%**	**60.4%**
		100	41.5%	**56.3%**	**53.5%**	**64.7%**
		200	61.3%	61.3%	60.2%	**66.2%**
	60%	50	62.6%	**66.4%**	**69.3%**	**72.7%**
		100	66.8%	60.0%	**61.2%**	**70.0%**
		200	71.4%	63.5%	64.3%	**71.8%**

5.3.3 Cure rate model with interval censored data

In this study, we consider a flexible cure rate model proposed by Cancho *et al.* (2011), which encompasses as a special case three of most used cure rate models: the Mixture model (Berkson & Gage, 1952), the promotion time cure model (Chen *et al.*, 1999) and CRPO model.

Let M_i be the latent risk, which denote the initial number of competing causes related to the occurrence of an event and assume that M_i has a Negative Binomial (NB) distribution with parameters θ and η (Piegorsch, 1990), with the probability mass function

$$P(M_i = m) = \frac{\Gamma(\eta^{-1} + m)}{\Gamma(\eta^{-1})m!}\left(\frac{\eta\theta}{1+\eta\theta}\right)^m (1+\eta\theta)^{-1/\eta}, \quad m = 0, 1, 2\cdots. \tag{5.9}$$

where $\theta > 0$, $\eta > -1/\theta$, so that $E(M) = \theta$ and $Var(M) = \theta\,(1+\eta\theta)$. Here, η is a dispersion parameter (Saha & Paul, 2005), values of $\eta > 0$ ($\eta < 0$) corresponds to over (under) dispersion relative to the Poisson distribution. Particularly, when $\eta \to 0$, the NB approaches to the Poisson distribution and the geometric distribution with parameter $1/(1+\theta)$ can be obtained when $\eta = 1$.

Let Y_j for $j = 1, \ldots, M_i$, denotes the failure time due to the jth latent cause, and assume that, given M_i, the random variables Y_j are i.i.d. with c.d.f. $F(\cdot) = 1 - S(\cdot)$ and the presence of any of latent risk (i.e., $M_i \geq 1$) will ultimately lead to the occurrence of the event. Thus, the time to event of interest (time to detect cancer) is defined by the random variable $T = \min\{Y_j, j = 0, \cdots, M_i\}$, where $P(Y_0 = \infty) = 1$. The survival function for the population is given by $S_{pop}(t) = [1 + \eta\theta F(t)]^{-1/\eta}$. The cure fraction has expression $p_0 = (1+\theta)^{-1/\eta}$. Similarly, considering the parametrization of the

model in terms of cure fraction, the survival function of this cure rate model can be written as

$$S_{pop}(t_i) = \left[1 + (p_{0i}^{-\eta} - 1)F(t_{ij})\right]^{-1/\eta}. \tag{5.10}$$

Here, we also assume that Y_j's are Weibull and Gamma distributed. Because this cure rate model can be characterized by the latent factors model of Cooner *et al.* (2007) with a Negative Binomial distribution for the number of latent factors, we called it as Negative Binomial cure rate model, denoting it as NBCR-W and NBCR-G. Similarly to the simulation studies above, we generate the current data from the NBCR-W model,

$$S_{pop}(t_i) = \left[1 + (p_{0i}^{-\eta} - 1)\left[1 - \exp(-t_i^{\alpha}\exp(\lambda_i))\right]\right]^{-1/\eta}. \tag{5.11}$$

where $p_{0i} = \exp\{b_0 + b_1 x_i\}/(1 + \exp\{b_0 + b_1 x_i\})$ and $\lambda_i = \beta x_i$ for$i = 1, \cdots, n$. We fixed parameters $b_0 = -1$, $b_1 = -0.5$, $\beta = -2$, $\alpha = 2$ and $\eta = 0.4$. The covariates x_i are assumed to be i.i.d. Bernulli(0.5) variates. Recently, in many clinical researches, the collected data set with more than 65% censoring observations appear frequently more. Therefore, we consider 70% censoring level (data with 70% censoring observations) and they were generated from a uniform distribution $U(0, 1)$. The prior distributions for the parameters are assumed to be $b_0 \sim N(-1, 0.25^2)$ and $b_1 \sim N(0.5, 0.25^2)$, $\beta \sim N(0, 5^2)$, $\alpha \sim N(0, 5^2)I_{(0,\infty)}$ and $\eta \sim N(0.4, 0.1^2)I_{(0,\infty)}$.

The data sets were also generated from the NBCR-G model,

$$S_{pop}(t_i) = \left[1 + (p_{0i}^{-\eta} - 1)\left[\frac{1}{\Gamma(k)}\gamma\left(k, \frac{t_i}{\theta_i}\right)\right]\right]^{-1/\eta}. \tag{5.12}$$

where $p_{0i} = \exp\{b_0 + b_1 x_i\}/(1 + \exp\{b_0 + b_1 x_i\})$ and $\theta_i = \exp\{\beta x_i\}$ for $i = 1, \cdots, n.$, In this case, we fixed $b_0 = -3$, $b_1 = 0.4$, $\beta = 1$, $\alpha = 3$, $\eta = 2$, and the 70% censoring observations are generated from a uniform distribution $U(0, 1)$. The prior distributions for the parameters are assumed to be $b_0 \sim N(-3, 0.1^2)$ and $b_1 \sim N(0.4, 0.1^2)$, $\beta \sim N(0, 1^2)$, $\alpha \sim N(0, 1^2)I_{(0,\infty)}$ and $\eta \sim N(2, 0.1^2)I_{(0,\infty)}$.

The results of this simulation study are presented in the Table 5.3, where the percentage of samples in which the adjusted cure rate model was indicated as the best model according to the criteria for the interval censoring data sets are presented. For the cure rate model with right censoring data, it shows that the D-measure outperforms the DIC for small sample sizes, though we only consider the sample size $n = 50$ and $n = 100$. We can observe that the correct rate obtained from the D-measure are higher than the DIC in some cases, but they have approximate values generally.

Table 5.3: Percentage of samples in which the adjusted cure rate model was indicated as the best model according to the criteria for the interval censoring data sets

True Model	Censoring level	Sample Size	Criteria			
				D-measure		
		n	DIC	Abs.	Square	Max
NBCR-W	70%	50	95.2%	**98.4%**	**99.2%**	**95.6%**
		100	97.6%	91.2%	97.2%	92.8%
NBCR-G	70%	50	80.4%	76.4%	80.4%	**92.8%**
		100	83.2%	81.2%	81.6%	**97.2%**

5.4 Application

In this section, we illustrate the applicability of the D-measure in two real data sets. The first data set is the melanoma data which evaluates the effectiveness of the implementation of a high dose of interferon alfa-2b in order to prevent the recurrence of cancer. The second data set is the smoking cessation data which evaluates the effectiveness of a special anti-smoking intervention scheme.

5.4.1 Melanoma data

Data were collected between 1991 and 1995, but there was monitoring of patients until 1998. The response variable is the time to death of the patient or the censor time. Further details of this data set can be found in Ibrahim *et al.* (2001c), with a total of $n = 417$ patients, with 56% of censored observations. The variables considered in this study include t: observed time (years, mean = 3.179, SD = 1.692), x_{1i}: type of treatment (0: observation, $n = 204$, 1: interferon, $n = 213$); x_{2i}: age (years, mean = 48.000, SD = 13.121), x_{3i}: presence of positive nodes at lymphadenectomy (0: no, $n = 111$, 1: yes, $n = 306$), $i = 1, \cdots, 417$ and x_{4i}: patient sex (0: male $n = 263$, 1 : female $n = 154$); x_{5i}: functional capacity (0: active, $n = 363$, 1: other, $n = 54$) and x_{6i} thickness of the tumor (in mm, mean = 3.941 and 3.204 standard deviation).

We fitted the Weibull Negative Binomial regression model with cure rate under the first, last and random activation mechanisms (denoted by WNBcr-FA, WNBcr-LA, and WNBcr-RA) and its particular sub-models: Weibull geometric regression model and the Weibull Poisson regression model under the first, last activation (denoted by WGcr-FA, WGcr-LA, WPcr-FA and WPcr-LA), and considering

$$\log\left(\frac{p_{0i}}{1-p_{0i}}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6}.$$

The Bayesian criteria DIC and D-measure were calculated and presented in Table 5.4. We note that both criteria indicate that the cure rate models under first activation are more adequate than the models under last activation, and the cure rate model under random activation has the largest criteria values. Moreover, both criteria show that the Weibull Negative Binomial regression model with cure rate is equivalent its sub models under the same activation, although DIC of WNBcr-FA has smallest value.

Table 5.4: Bayesian criteria for the fitted models for the melanoma data.

	Criteria			
	D-measure			
Model	Abs.	Square	Max	DIC
WNBcr-FA	99.6	5.9	34.2	1039
WGcr-FA	98.4	5.9	33.7	1047
WPcr-FA	98.9	5.9	33.5	1051
WNBcr-LA	103.2	6.3	34.5	1064
WGcr-LA	106.2	6.6	35.2	1069
WPcr-LA	103.6	6.4	34.6	1063
WNBcr-RA	236.5	53.8	68.9	1089

5.4.2 Smoking cessation data

We now apply the propose D-measure to the interval-censored smoking cessation data presented in Section 1.1. We fitted

We fitted the some flexile cure models considering the different spatial frailties in the models to the data set, there are Weibull negative binomial cure rate model, complementary Weibull negative binomial cure rate model, proportional hazard negative binomial cure rate model, complementary proportional hazard negative binomial cure rate and their sub-models. These cure models with spatial frailties are presented in Section 3.2.

Prior distributions for the parameters $\boldsymbol{b}$, $\boldsymbol{\beta}$ and η are $b_j \sim N(0, 100)$, $j = 0, \ldots, 4$, $\beta_j \sim N(0, 100)$, $j = 1, \ldots, 4$, and $\eta \sim N(0, 100)I_{(0,\infty)}$ and a prior distribution for the shape parameter of WNBCR and CWNBCR model is $\alpha \sim N(0, 100)I_{(0,\infty)}$. As we know, the Piecewise Exponential distribution has better approximation to any unknown function when the length of each interval becomes smaller. Therefore, we partition the time axis so that they denoted the ordered distinct time points of all observed interval end points. Thus, we have 178 risk parameters need to estimate. Prior distributions for the risk parameters are $\alpha_i \sim N(0, 100)I_{(0,\infty)}$, $i = 1, \ldots, 178$. For the sub-models of the PHNBCR and CPHNBCR, we used the informative prior distributions for the parameters $\boldsymbol{b}$ and $\boldsymbol{\beta}$, where the priors are based on the posterior distributions of these parameters of PHNBCR

and CPHNBCR models, i.e., $b_j \sim N(0,1)$, $j = 0, \ldots, 4$, and $\beta_j \sim N(0, 0.6)$, $j = 1, \ldots, 4$. The prior distributions for spatial parameters are $a \sim \text{Beta}(18, 2)$ and $\mathbf{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$, where $\text{Diag}(0.1, 0.1) = 0.1\boldsymbol{I}_{2\times 2}$.

The Bayesian model selection criteria DIC and D-measure for the fitted cure rate models are presented in Table 5.5. We observe that the PHNBCR and CPHNBCR models have the two smallest values of DIC and D-measure among all fitted cure models, considering the norm as absolute or square value. According to the criteria, we also note that the both cure models are almost equivalent with their sub-models and the cure models with piecewise exponential distribution are more appropriate than the cure models with Weibull distribution.

Table 5.5: Bayesian criteria for the fitted models for the smoking cessation data.

	Criteria			
	D measure			
Model	Abs.	Square	Max	DIC
Weibull negative binomial cure rate model	160.3	53.7	110.1	402
Weibull geometric cure rate model	156.9	52.3	110.7	417
Weibull promotion time cure model	162.4	55.0	112.5	417
Complementary Weibull negative binomial cure rate model	156.5	52.2	109.7	417
Complementary Weibull geometric cure rate model	163.2	55.8	113.1	418
Complementary Weibull promotion time cure model	157.4	52.3	109.6	419
PH negative binomial cure rate model	144.0	45.7	102.8	388
PH geometric cure rate model	150.2	47.5	103.2	395
PH promotion time cure model	148.3	46.6	102.3	404
Complementary PH Negative-Binomial cure rate model	144.8	46.2	102.5	382
Complementary PH geometric cure rate model	146.4	46.3	102.3	395
Complementary PH promotion time cure model	150.0	48.1	103.6	405

5.5 Conclusion

In this paper, we propose the D-measure, which measures the goodness of a model by comparing how close its predictions are from the observed data. The propose D-measure can be viewed as a Bayesian goodness-of-fit statistic which measures the performance of a model by a combination of how close its predictions are from the observed data based on the survival functions. It can also be used for all kind of survival data sets and it is an alternative criterion which can be used to compare cure rate models, even in presence of random effects or frailties. The D-measure was compared to the DIC via simulation, where we noted that the D-measure outperforms the DIC in presence of small sample size and high censoring level.

Chapter 6

Concluding Remarks

In this work, we firstly described approaches to extend geometric cure rate model and negative binomial cure rate model their complementary models to allow for spatial correlations by including spatial frailty for the interval-censored data set. The negative binomial cure rate models are more flexible because they encompass the geometric cure rate model and several well known cure rate models as its particular cases. The MCMC method was used in Bayesian inference approach for the proposed models and the DIC was used for the model comparison. The results of the applications show that

- The cure rate models with Weibull distribution (WGCR and WNBCR models) have better fit than their complementary models (CWGCR and CWNBCR models) for all prior distributions considered for the parameters.

- The cure rate models and their complementary models with Piecewise exponential distribution (PHGCR, PHNBCR, CPHGCR and CPHNBCR) have are more adequate than cure rate models and their complementary models with Weibull distribution (WGCR, WNBCR, CWGCR and CWNBCR) for all prior distributions considered for the parameters.

- According to the DIC, the CPHGCR model and PHGCR model are equivalent and it is also can be observed for CPHNBCR and PHNBCR model.

- The Negative Binomial cure rate models have smaller DIC values than the Geometric cure rate models.

- Comparing the proposed cure rate models with models introduced Carlin & Banerjee (2003) and Pan *et al.* (2014), it is showed that the PHGCR, CPHGCR, PHNBCR and CPHNBCR models are more adequate.

Moreover, the proposed models are not sensitive with influential observations, which can be observed through the influence diagnostic in the simulation studies as well as in the application. The interpretation of the covariates is easy due to the parameterization of the models considered in the cure rate and the MCAR prior can be used even if frailties effects are low or they are not correlated.

Then, we propose power series cure rate model for spatially correlated interval-censored data based on generalized extreme value distribution. The proposed model is very flexible and generalizes the Bernoulli, geometric, Poisson, and logarithm models, whose may be tested for the best fitting in a straightforward way. The MCMC method was also used in Bayesian inference approach for the proposed models and the DIC was used for the model comparison. The results of the applications show that the proposed model has better fittings than the WNBCR and CWNBCR models. Comparing the proposed models with PHNBCR and CPHNBCR models, we conclude that the proposed models are more adequate, indeed it has much less parameters. From the results of Bayesian case deletion influence diagnostics, we also observed that the proposed cure rate models are not sensible with influence observations.

In the penultimate chapter, we propose D-measure which measures the performance of a model by a combination of how close its predictions are to the observed data based on survival function. The measure can be used for all kind of survival data in presence of censoring. It can also be used to compare cure rate models, even in presence of random effects or frailties. The D-measure was compared to the DIC via simulation, where we noted that the D-measure outperforms the DIC in presence of small sample size and high censoring level. Finally, it is applied in two real data sets.

For the future works, we present some suggestions as follow:

- Propose the spatial fragilities in the destructive weighted Poisson cure rate models (Rodrigues *et al.*, 2010b);
- Propose spatial temporal fragilities in cure rate models. Considering both space and time are discrete, the fragilities can be modeled based on a Markov random field (MRF) structure in the form of the CAR specifications (Martínez-Beneito *et al.*, 2008);
- Propose a non-parametric estimation for estimating the baseline functions of the proposed models;
- Introduce the classic approach to estimation for the spatial fragilities cure rate models.

Bibliography

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models. *Biometrika*, **94**(2), 443–458.

Banerjee, S. & Carlin, B. P. (2004). Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics*, **60**(1), 268–275.

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.

Banerjee, S., Gelfand, A. E. & Carlin, B. P. (2004). *Hierarchical modeling and analysis for spatial data*. Crc Press.

Barlow, R. E., Bartholomew, D. J., Bremner, J. & Brunk, H. D. (1972). *Statistical inference under order restrictions: The theory and application of isotonic regression*. Wiley New York.

Berkson, J. & Gage, R. P. (1952). Survival cure for cancer patients following treatment. *Journal of the 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)*, pages 192–236.

Boag, J. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B*, **11**, 15–53.

Cancho, V., Ortega, E. & Bolfarine, H. (2009). The Log-exponentiated-Weibull Regression Models with Cure Rate: Local Influence and Residual Analysis. *Journal of Data Science*, **7**, 433–458.

Cancho, V., Ortega, E. M. M. & Paula, G. (2010). On estimation and influence diagnostics for log-birnbaum-saunders student-t regression models: Full bayesian analysis. *Journal of Statistical Planning and Inference*, **140**, 2486–2496.

Cancho, V. G., Rodrigues, J. & Castro, M. (2011). A flexive model for survival with a cure rate: A bayesian approach. *Journal of Applied Statistics*, **38**, 57–70.

Carlin, B. & Banerjee, S. (2003). *Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion)*. Oxford University Press.

Carlin, B. P. & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*. CRC Press.

Carvalho Lopes, C. M. & Bolfarine, H. (2012). Random effects in promotion time cure rate models. *Computational Statistics & Data Analysis*, **56**(1), 75–87.

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.

Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall/CRC, Boca Raton, FL.

Cooner, F., Banerjee, S., Carlin, B. & Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, **102**, 560–572.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B*, **34**, 187–220.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Dey, D. & Birmiwal, L. (1994). Robust bayesian analysis using divergence measures. *Statistics & Probability Letters*, **20**(4), 287–294.

Duchateau, L. & Janssen, P. (2007). *The frailty model*. Springer.

Ewell, M. & Ibrahim, J. G. (1997). The large sample distribution of the weighted log rank statistic under general local alternatives. *Lifetime Data Analysis*, **3**(1), 5–12.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, pages 1041–1046.

Feigl, P. & Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, pages 826–838.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. Wiley, 3rd edition, New York.

Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics*, pages 101–113.

Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, Boca Raton, second edition.

Gelfand, A. E. & Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika*, **85**(1), 1–11.

Gelfand, A. E. & Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, **4**(1), 11–15.

Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*. Clarendon Press, Oxford, UK.

Groeneboom, P. (1995). Nonparametric estimators for interval censoring problems. *Lecture Notes-Monograph Series*, pages 105–128.

Gu, Y., Sinha, D. & Banerjee, S. (2011). Analysis of cure rate survival data under proportional odds model. *Lifetime data analysis*, **17**(1), 123–134.

Härdle, W. (1989). Robertson, t., wright, ft and rl dykstra: Order restricted statistical inference. *Statistical Papers*, **30**(1), 316–316.

Hougaard, P. & Hougaard, P. (2000). *Analysis of multivariate survival data*, volume 564. Springer New York.

Ibrahim, J., Chen, M. & Sinha, D. (2001a). *Bayesian survival analysis*. Springer Verlag.

Ibrahim, J. G. & Laud, P. W. (1994). A predictive approach to the analysis of designed experiments. *Journal of the American Statistical Association*, **89**(425), 309–319.

Ibrahim, J. G., Chen, M.-H. & Sinha, D. (2001b). Criterion-based methods for bayesian model assessment. *Statistica Sinica*, **11**(2), 419–444.

Ibrahim, J. G., Chen, M. H. & Sinha, D. (2001c). *Bayesian Survival Analysis*. Springer, New York, NY.

Jenkinson, A. (1969). Estimation of maximum floods. *World Meteorological Organization, Technical Note*, (98), 183–257.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, **81**(348), 158–171.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics*, **7**(3), 310–321.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica: Journal of the Econometric Society*, pages 939–956.

Li Dan, W. X. & Dey, D. K. (2015). A flexible cure rate model for spatially correlated survival data based on generalized extreme value distribution and gaussian process priors. *Biometrical Journal*.

Lindsey, J. C. & Ryan, L. M. (1998). Methods for interval-censored data. *Statistics in medicine*, **17**(2), 219–238.

Lopes, C., Mendes, C. & Bolfarine, H. (2012). Random effects in promotion time cure rate models. *Computational Statistics and Data Analysis*, **56**, 75–87.

Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure fraction. *Statistica Sinica*, **20**, 661.

Ma, X. & Xiang, L. (2013). Testing for the presence of a cure fraction in clustered interval-censored survival data. *Australian & New Zealand Journal of Statistics*, **55**(2), 173–190.

Maller, R. A. & Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Wiley, New York.

Mardia, K. (1988). Multi-dimensional multivariate gaussian markov random fields with application to image processing. *Journal of Multivariate Analysis*, **24**(2), 265–284.

Martínez-Beneito, M., López-Quilez, A. & Botella-Rocamora, P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in medicine*, **27**(15), 2874–2889.

Murray, R. P., Anthonisen, N. R., Connett, J. E., Wise, R. A., Lindgren, P. G., Greene, P. G., Nides, M. A. *et al.* (1998). Effects of multiple attempts to quit smoking and relapses to smoking on pulmonary function. *Journal of clinical epidemiology*, **51**(12), 1317–1326.

Noack, A. (1950). A class of random variables with discrete distributions. *The Annals of Mathematical Statistics*, pages 127–132.

Pan, C., Cai, B., Wang, L. & Lin, X. (2014). Bayesian semiparametric model for spatially correlated interval-censored survival data. *Computational Statistics & Data Analysis*, **74**, 198–208.

Pearson, K. (1895). Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A*, pages 343–414.

Peng, F. & Dey, D. (1995). Bayesian analysis of outlier problems using divergence measures. *Canadian Journal of Statistics*, **23**(2), 199–213.

Peng, Y. (2003). Estimating baseline distribution in proportional hazards cure models. *Computational statistics & data analysis*, **42**(1), 187–201.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Applied Statistics*, pages 86–91.

Piegorsch, W. W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics*, **46**, 863–867.

Plummer, M., Best, N., Cowles, K. & Vines, K. (2005). Output analysis and diagnostics for mcmc. *R package version 0.10-3, URL http://cran. rproject. org*.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rodrigues, J., Cancho, V., Castro, M. & Louzada-Neto, F. (2009a). On the unification of long-term survival models. *Statistics and Probability Letters*, **79**, 753–759.

Rodrigues, J., Cancho, V. G., de Castro, M. & Louzada-Neto, F. (2009b). On the unification of long-term survival models. *Statistics & Probability Letters*, **79**(6), 753–759.

Rodrigues, J., Cancho, V. G., Castro, M. & Balakrishnan, N. (2010a). A Bayesian destructive weighted Poisson cure rate model and an application to a cutaneous melanoma data. *Statistical Methods in Medical Research*.

Rodrigues, J., Castro, M., Balakrishnan, N. & Cancho, V. (2010b). Destructive weighted Poisson cure rate models. *Lifetime Data Analysis*.

Roy, V. & Dey, D. K. (2014). Propriety of posterior distributions arising in categorical and survival models under generalized extreme value distribution. *Statistica Sinica*, **24**, 699–722.

Rücker, G. & Messerer, D. (1988). Remission duration: An example of interval-censored observations. *Statistics in Medicine*, **7**(11), 1139–1145.

Saha, K. & Paul, S. (2005). Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics*, **61**(1), 179–185.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, **64**, 583–639.

Stangl, D. K. & Greenhouse, J. B. (1998). Assessing placebo response using bayesian hierarchical survival models. *Lifetime Data Analysis*, **4**(1), 5–28.

Sun, X. & Chen, C. (2010). Comparison of finkelstein's method with the conventional approach for interval-censored data analysis. *Statistics in Biopharmaceutical Research*, **2**(1), 97–108.

Therneau, T. M. (2000). *Modeling survival data: extending the Cox model*. Springer.

Tournoud, M. & Ecochard, R. (2007). Application of the promotion time cure model with time-changing exposure to the study of HIV/AIDS and other infectious diseases. *Statistics in medicine*, **26**, 1008–1021.

Tsodikov, A. D., Ibrahim, J. G. & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, **98**, 1063–1078.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 290–295.

Vaupel, J. W., Manton, K. G. & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**(3), 439–454.

Weibull, W. (1939). A statistical theory of strength of materials. *Ingeniors Vetenskaps Akademien*, **151**(5), 45–55.

Weibull, W. (1951). Regression models and life-tables. *J. Appl. Mech.*, **18**, 293–297.

Weiss, R. (1996). An approach to bayesian sensitivity analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 739–750.

Wellner, J. A. (1995). Interval censoring, case 2: alternative hypotheses. *Lecture Notes-Monograph Series*, pages 271–291.

Wellner, J. A. & Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association*, **92**(439), 945–959.

Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, pages 32–52.

Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, New Jersey.

Yau, K. K. & Ng, A. S. (2001). Long-term survivor mixture model with random effects: application to a multi-centre clinical trial of carcinoma. *Statistics in medicine*, **20**(11), 1591–1607.

Appendix

Algorithm

In this work, for the MCMC updates, the Metropolis-Hastings algorithm is used to generate a posteriori samples for the parameters $\boldsymbol{\varphi}^* = (\boldsymbol{b}, \boldsymbol{\beta}, \kappa, \boldsymbol{\zeta}, \boldsymbol{U}, \boldsymbol{V})$ of the independents priors assumption and for the parameters $\boldsymbol{\varphi}^* = (\boldsymbol{b}, \boldsymbol{\beta}, \eta, \boldsymbol{\zeta}, \boldsymbol{\psi}, \boldsymbol{\rho})$ of the dependents priors assumption. Although the block Metropolis-Hastings algorithm is computationally efficient, it is difficult to obtain a good approximation of variance covariance proposal which directly affects the convergence of chains. Therefore, we used one-dimensional random walk Metropolis algorithm where at each iteration we generate new values from a univariate normal candidate distributions whose variances were calibrated to obtain good acceptance rates. Let $\varphi^*_{i,(t)}$ denotes the state of ith parameter of $\boldsymbol{\varphi}^*$ ($\varphi^*_{i,}$) at the end of iteration t, the Metropolis-Hastings algorithm implemented for ith parameter of $\boldsymbol{\varphi}^*$ are given by:

(1) start with any point $\varphi^*_{i,(0)}$ and stage indicator $t = 0$;

(2) generate a point φ'_i from the transitional kernel distribution $q\left(\varphi'_i, \varphi^*_{i,(t)}\right) = N\left(\varphi_{i,(t)}, \sigma\right)$, where σ is variance of $\varphi^*_{i,(t)}$ is same in any stage;

(3) update $\varphi^*_{i,(t)}$ to $\varphi^*_{i,(t+1)} = \varphi'_i$ with probability $p_j = \min\left\{1, \pi(\varphi'_i|\boldsymbol{\mathcal{D}}) \,/\, \pi(\varphi^*_{i,(t)}|\boldsymbol{\mathcal{D}})\right\}$;

(4) repeat steps (2) and (3) by increasing the stage indicator until the process reaches a stationary distribution.

Finally, we repeat this algorithm for all parameters of $\boldsymbol{\varphi}^*$.

Prior sensitive Analysis

Considering the dependence of the random frailties, and we assume the random frailties take the traditional MCAR distribution $MCAR(a, \boldsymbol{\Lambda})$, where a prior distribution for parameter a taken $a \sim \text{Uniform}(0, 1)$ or $a \sim \text{Beta}(18, 2)$ and $\Lambda \sim \text{Wishart}(2, \Lambda_0)$ following Carlin & Banerjee (2003), Gelfand & Vounatsou (2003) and Banerjee & Carlin (2004). However, Gelfand & Vounatsou (2003) and Banerjee & Carlin (2004) considered Λ_0 equals $\boldsymbol{I}$ and $0.01\boldsymbol{I}$ in their papers, respectively, where $\boldsymbol{I}$ denote a identity matrix. Both authors also commented that they had no prior knowledge regarding the nature or extent of dependence for the parameter $\boldsymbol{\Lambda}$. Note that $\boldsymbol{\Lambda}^{-1}$ describe the relative variability and covariance relationship between the different diseases given the neighboring site. Thus, if Λ_0 has small values, we assumed high relative variability between neighborhood and we assumed low relative variability between neighborhood if Λ_0 have big values. Thus, it is necessary to

conduct a prior study for the parameter Λ_0 to verify the influence of Λ_0 in the estimation, in order to have a value for appropriate Λ_0. Next, we will conduct sensitive analysis for the parameter $\boldsymbol{\Lambda}$ of WNBCR and CWNBCR models, in which we fixe $\boldsymbol{\Lambda}_0$ equals to $\boldsymbol{I}_{2\times2}$, $0.1\boldsymbol{I}_{2\times2}$, $0.01\boldsymbol{I}_{2\times2}$ and $0.001\boldsymbol{I}_{2\times2}$, and the prior distributions for others parameters $\boldsymbol{b}$, $\boldsymbol{\beta}$ and η are $b_j \sim N(0, 100)$, $j = 0, \ldots, 4$, $\beta_j \sim N(0, 100)$, $j = 1, \ldots, 4$, and $\eta \sim N(0, 100)I_{(0,\infty)}$ and a prior distribution for the shape parameter is $\alpha \sim N(0, 100)I_{(0,\infty)}$.

Posterior summaries of the parameters of the WNBCR and CWNBCR models considering the prior distribution $a \sim Uniform(0, 1)$ and $a \sim beta(18, 2)$ for different values for $\boldsymbol{\Lambda}_0$ are presented in Table 1, 2, 3 and 4, respectively. Note that the smaller values are taken by Λ_0, the higher relative variability between neighborhood are assumed, which can be observed in Table 1, 2, 3 and 4. We note that the posterior estimative of elements of the covariance matrix $\boldsymbol{\Sigma}$ are deceasing with the diagonal elements of Λ_0, i.e, when we fixed $\Lambda_0 = 0.001\boldsymbol{I}$, the posterior estimative of diagonal elements of the covariance matrix $\boldsymbol{\Lambda}^{-1} = \boldsymbol{\Sigma}$ have very high value (For WNBCR model, $\Sigma_{11} = 34.348$ and $\Sigma_{22} = 37.164$). On the other hand, the posterior estimative of diagonal elements of the covariance matrix $\boldsymbol{\Sigma}$ have very low values when we fixed $\Lambda_0 = \boldsymbol{I}$. (For WNBCR model, $\Sigma_{11} = 0.041$ and $\Sigma_{22} = 0.043$.) Since $\boldsymbol{\Lambda}_0$ restrict the posterior estimate of $\boldsymbol{\Lambda}$ and the $\boldsymbol{\Sigma}$ interprets the variability and covariance relationship between the different diseases given the neighboring site, thus the too higher or too lower values are not adequate.

Tables 1 and 2 show that the posterior estimative of the parameters b_0, b_2, β_2 and β_3 decreasing with Λ_0 increasing, and the parameters b_1, α and η increasing with Λ_0 increasing, however, the posterior estimative of the parameters $\boldsymbol{\beta}$, $\boldsymbol{b}$, α, a and η have close values when $\Lambda_0 = 0.01$ and $\Lambda_0 = 0.1$. Tables 3 and 4 show that the posterior estimative of the parameters b_0, β_2 and α decreasing with Λ_0 increasing, and the parameters b_1 and η increasing with Λ_0 increasing. Posterior summaries of the parameters of the CPHNBCR model considering the prior distribution $a \sim beta(18, 2)$ for different values for $\boldsymbol{\Lambda}_0$ is presented in Table 5. We note that the parameter β_2 decreasing with Λ_0 increasing, and the parameters b_1, β_1, $\boldsymbol{\alpha}$ and η increasing with Λ_0 increasing. However, the posterior estimative of the parameters $\boldsymbol{\beta}$, $\boldsymbol{b}$, $\boldsymbol{\alpha}$ and η have close values for the CWNBCR model, which can be observed in Figure 1 and Table 5. Moreover, we observe that the posterior estimate of the parameters $\boldsymbol{\beta}$, $\boldsymbol{b}$, α, a and η have close values when $\Lambda_0 = 0.01$ and $\Lambda_0 = 0.1$ in all five tables. Therefore, in this work, we fixed $\Lambda_0 = 0.1\boldsymbol{I}$, which lead to the posterior estimative of diagonal elements of the covariance matrix $\boldsymbol{\Sigma}$ between zero and one.

Table 1: Posterior summaries of the parameters of the WNBCR model considering the prior distribution $a \sim Uniform(0,1)$ and different values for $\boldsymbol{\Lambda}_0$.

	$\Lambda_0 = 0.001\boldsymbol{I}$				$\Lambda_0 = 0.01\boldsymbol{I}$				$\Lambda_0 = 0.1\boldsymbol{I}$				$\Lambda_0 = \boldsymbol{I}$			
Parameters	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%
b_0	3.865	2.639	0.270	10.829	1.820	0.753	0.485	3.474	1.493	0.570	0.390	2.635	1.393	0.567	0.420	2.475
b_1	-4.119	4.489	-17.219	-0.049	-0.423	0.495	-1.371	0.433	-0.277	0.379	-1.067	0.418	-0.205	0.355	-0.927	0.497
b_2	2.835	2.279	-0.109	8.390	0.852	0.641	-0.121	2.501	0.558	0.475	-0.219	1.687	0.125	0.649	-1.234	1.332
b_3	-0.040	0.142	-0.268	0.353	-0.077	0.037	-0.171	-0.015	-0.064	0.031	-0.134	-0.006	-0.046	0.025	-0.101	-0.002
b_4	-0.212	0.286	-0.932	0.098	0.005	0.030	-0.058	0.059	0.010	0.026	-0.048	0.061	-0.006	0.024	-0.052	0.044
β_1	-2.207	2.160	-8.517	0.386	-0.234	0.817	-1.832	1.356	-0.252	1.089	-2.548	1.826	-0.597	2.356	-6.310	3.157
β_2	1.910	1.419	-0.511	4.927	1.616	1.107	-0.164	4.248	1.491	1.230	-0.631	4.205	-0.779	4.462	-12.590	3.974
β_3	-0.055	0.123	-0.281	0.138	-0.172	0.075	-0.313	-0.029	-0.211	0.090	-0.378	-0.004	-0.228	0.108	-0.440	0.009
β_4	-0.159	0.145	-0.416	0.074	-0.043	0.073	-0.219	0.062	-0.041	0.088	-0.292	0.074	-0.128	0.157	-0.493	0.068
α	2.354	0.666	1.435	3.873	3.105	0.554	2.019	4.019	3.347	0.586	2.047	4.087	3.615	0.506	2.308	4.128
a	0.501	0.286	0.028	0.970	0.500	0.289	0.025	0.976	0.503	0.289	0.025	0.975	0.506	0.287	0.026	0.975
η	0.642	0.666	0.013	2.480	3.425	2.189	0.380	8.463	6.825	3.577	0.602	13.737	10.564	4.217	1.668	17.976
Λ_{11}	0.032	0.007	0.020	0.048	0.273	0.065	0.163	0.416	2.669	0.657	1.562	4.130	26.822	6.460	15.707	41.258
Λ_{22}	0.030	0.007	0.018	0.045	0.283	0.066	0.169	0.429	2.581	0.639	1.498	3.971	25.567	6.365	14.935	39.868
Λ_{12}	-0.004	0.005	-0.014	0.006	-0.008	0.046	-0.099	0.084	-0.012	0.462	-0.916	0.916	-0.095	4.595	-9.189	8.933
Σ_{11}	34.348	8.375	21.495	53.751	4.010	1.015	2.466	6.363	0.413	0.110	0.249	0.679	0.041	0.011	0.025	0.066
Σ_{22}	37.164	10.323	22.721	62.668	3.869	0.998	2.378	6.283	0.427	0.122	0.254	0.711	0.043	0.012	0.026	0.071
ρ_Σ	0.124	0.164	-0.199	0.456	0.030	0.170	-0.300	0.368	0.004	0.180	-0.353	0.358	0.004	0.179	-0.343	0.359

where Λ_{ij} is the element of precision matrix $\boldsymbol{\Lambda}$ in position (i,j), and Σ_{ij} is the element of matrix $\boldsymbol{\Sigma} = \Lambda^{-1}$ in position (i,j), this Σ_{11} is the spatial variance component of $\boldsymbol{U}$ and Σ_{22} is the spatial variance component of $\boldsymbol{V}$, $\rho_\Sigma = \Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$ denote their correlation.

Table 2: Posterior summaries of the parameters of the WNBCR model considering the prior distribution $a \sim \text{Beta}(18, 2)$ and different values for $\boldsymbol{\Lambda}_0$.

	$\Lambda_0 = 0.001\boldsymbol{I}$				$\Lambda_0 = 0.01\boldsymbol{I}$				$\Lambda_0 = 0.1\boldsymbol{I}$				$\Lambda_0 = \boldsymbol{I}$			
Parameters	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%
b_0	7.371	4.279	1.284	19.003	1.760	0.798	0.172	3.316	1.425	0.584	0.200	2.516	1.407	0.537	0.463	2.528
b_1	-8.427	6.910	-24.109	1.435	-0.973	2.179	-9.249	0.390	-0.298	0.517	-1.607	0.535	-0.153	0.432	-1.003	0.615
b_2	4.935	3.842	-1.357	15.730	1.128	0.980	-0.133	3.976	0.570	0.577	-0.401	2.081	0.107	0.663	-1.589	1.100
b_3	0.199	0.201	-0.135	0.560	-0.073	0.044	-0.185	-0.006	-0.051	0.034	-0.120	0.024	-0.042	0.020	-0.084	-0.005
b_4	-0.869	0.481	-1.762	-0.054	-0.008	0.052	-0.151	0.067	-0.005	0.042	-0.111	0.061	-0.009	0.027	-0.066	0.042
β_1	-3.439	2.953	-9.644	1.370	-0.641	1.876	-6.986	1.449	-0.199	1.642	-4.481	3.067	0.010	2.469	-5.794	5.593
β_2	2.118	2.078	-1.194	7.947	1.854	1.184	-0.210	4.331	1.693	1.900	-1.733	6.053	-0.554	5.029	-16.629	4.518
β_3	0.109	0.094	-0.070	0.263	-0.154	0.076	-0.288	0.016	-0.171	0.122	-0.365	0.133	-0.217	0.106	-0.424	0.000
β_4	-0.446	0.184	-0.709	-0.111	-0.068	0.072	-0.213	0.055	-0.110	0.151	-0.471	0.070	-0.133	0.157	-0.574	0.090
α	1.796	0.289	1.361	2.517	3.031	0.689	1.692	4.068	3.318	0.599	2.069	4.075	3.649	0.439	2.474	4.103
a	0.899	0.066	0.735	0.987	0.899	0.066	0.734	0.987	0.899	0.067	0.735	0.987	0.901	0.065	0.743	0.987
η	0.419	0.183	0.089	0.818	3.480	2.746	0.246	10.002	6.967	3.624	0.815	13.816	10.734	3.433	3.173	16.820
Λ_{11}	0.034	0.007	0.022	0.050	0.272	0.065	0.163	0.417	2.663	0.661	1.534	4.098	26.857	6.463	15.902	40.765
Λ_{22}	0.028	0.007	0.017	0.043	0.281	0.066	0.170	0.424	2.624	0.650	1.534	4.038	25.591	6.322	14.809	39.455
Λ_{12}	-0.004	0.005	-0.014	0.006	-0.008	0.046	-0.101	0.081	-0.025	0.464	-0.948	0.881	0.043	4.567	-8.885	8.951
Σ_{11}	32.373	7.752	20.761	50.093	4.019	1.010	2.459	6.339	0.414	0.110	0.250	0.678	0.041	0.011	0.025	0.066
Σ_{22}	39.745	11.251	23.754	65.726	3.887	1.003	2.394	6.301	0.421	0.119	0.252	0.697	0.043	0.012	0.026	0.072
ρ_Σ	0.127	0.167	-0.199	0.465	0.031	0.169	-0.298	0.373	0.010	0.179	-0.339	0.364	-0.001	0.178	-0.357	0.353

Table 3: Posterior summaries of the parameters of the CWNBCR model considering the prior distribution $a \sim \text{Uniform}(0, 1)$ and different values for $\boldsymbol{\Lambda}_0$.

	$\Lambda_0 = 0.001\boldsymbol{I}$				$\Lambda_0 = 0.01\boldsymbol{I}$				$\Lambda_0 = 0.1\boldsymbol{I}$				$\Lambda_0 = \boldsymbol{I}$			
Parameters	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%
b_0	0.814	1.240	-1.584	3.194	0.407	1.027	-1.599	2.503	0.055	0.937	-1.722	1.926	-0.009	0.866	-1.674	1.734
b_1	-0.902	0.658	-2.297	0.308	-0.594	0.505	-1.616	0.341	-0.430	0.388	-1.203	0.308	-0.383	0.376	-1.122	0.355
b_2	1.067	0.584	0.004	2.311	0.936	0.511	-0.008	1.970	0.811	0.435	-0.012	1.700	0.818	0.439	-0.027	1.718
b_3	-0.070	0.033	-0.135	-0.004	-0.053	0.026	-0.104	-0.004	-0.044	0.022	-0.089	-0.003	-0.045	0.024	-0.097	-0.004
b_4	0.047	0.044	-0.046	0.129	0.042	0.033	-0.026	0.105	0.047	0.029	-0.010	0.105	0.049	0.030	-0.008	0.110
β_1	-0.083	0.544	-1.151	0.969	0.202	0.403	-0.618	0.952	0.361	0.361	-0.372	1.039	0.377	0.354	-0.350	1.046
β_2	0.371	0.663	-0.852	1.679	0.232	0.473	-0.652	1.186	0.070	0.418	-0.709	0.932	0.030	0.425	-0.731	0.917
β_3	-0.050	0.037	-0.120	0.022	-0.042	0.025	-0.090	0.007	-0.041	0.023	-0.089	0.002	-0.038	0.024	-0.086	0.010
β_4	-0.038	0.033	-0.106	0.022	-0.031	0.024	-0.083	0.013	-0.021	0.022	-0.069	0.017	-0.021	0.022	-0.073	0.016
α	2.630	0.433	1.801	3.506	2.177	0.330	1.530	2.820	1.971	0.290	1.422	2.543	1.894	0.291	1.373	2.494
a	0.499	0.289	0.024	0.975	0.496	0.291	0.023	0.973	0.495	0.291	0.027	0.975	0.506	0.288	0.024	0.976
η	1.858	3.030	0.020	11.411	3.129	3.527	0.071	12.797	4.506	4.548	0.085	16.466	4.906	4.834	0.074	16.699
Λ_{11}	0.033	0.007	0.021	0.049	0.273	0.063	0.164	0.413	2.651	0.653	1.552	4.077	26.570	6.424	15.731	40.611
Λ_{22}	0.032	0.007	0.020	0.047	0.275	0.065	0.164	0.417	2.572	0.641	1.489	3.974	25.593	6.394	14.807	39.715
Λ_{12}	-0.001	0.005	-0.011	0.009	-0.001	0.046	-0.091	0.089	0.001	0.466	-0.928	0.908	-0.054	4.579	-8.992	9.009
Σ_{11}	32.559	7.460	20.917	49.886	3.990	0.992	2.482	6.366	0.416	0.113	0.251	0.682	0.041	0.011	0.025	0.067
Σ_{22}	34.312	8.351	21.746	53.659	3.978	1.034	2.447	6.440	0.432	0.128	0.255	0.728	0.043	0.012	0.026	0.073
ρ_Σ	0.033	0.158	-0.279	0.342	0.002	0.170	-0.333	0.336	0.001	0.182	-0.358	0.362	0.002	0.180	-0.355	0.352

Table 4: Posterior summaries of the parameters of the CWNBCR model considering the prior distribution $a \sim \text{Beta}(18, 2)$ and different values for $\boldsymbol{\Lambda}_0$.

	$\Lambda_0 = 0.001\boldsymbol{I}$				$\Lambda_0 = 0.01\boldsymbol{I}$				$\Lambda_0 = 0.1\boldsymbol{I}$				$\Lambda_0 = \boldsymbol{I}$			
Parameters	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%
b_0	0.430	1.521	-2.685	3.373	0.171	1.023	-1.851	2.072	0.022	0.986	-1.954	1.922	-0.169	0.950	-2.087	1.628
b_1	-0.738	0.745	-2.265	0.847	-0.559	0.465	-1.527	0.322	-0.439	0.386	-1.198	0.315	-0.369	0.379	-1.142	0.381
b_2	1.085	0.570	-0.002	2.271	0.952	0.504	0.028	2.059	0.837	0.464	0.003	1.800	0.812	0.420	0.015	1.687
b_3	-0.069	0.037	-0.140	0.004	-0.051	0.025	-0.101	-0.004	-0.044	0.025	-0.095	-0.002	-0.042	0.022	-0.088	0.000
b_4	0.053	0.050	-0.061	0.140	0.047	0.036	-0.026	0.119	0.048	0.030	-0.014	0.102	0.052	0.030	-0.007	0.115
β_1	0.047	0.582	-1.099	1.190	0.205	0.402	-0.614	0.965	0.348	0.349	-0.370	0.988	0.384	0.349	-0.327	1.029
β_2	0.214	0.700	-1.105	1.628	0.209	0.520	-0.787	1.236	0.091	0.420	-0.695	0.933	0.032	0.401	-0.719	0.839
β_3	-0.040	0.040	-0.116	0.032	-0.038	0.028	-0.094	0.017	-0.041	0.024	-0.089	0.008	-0.041	0.024	-0.089	0.006
β_4	-0.037	0.035	-0.115	0.025	-0.032	0.027	-0.089	0.016	-0.022	0.022	-0.070	0.016	-0.019	0.022	-0.070	0.017
α	2.547	0.452	1.774	3.480	2.178	0.344	1.533	2.865	1.989	0.309	1.406	2.603	1.938	0.279	1.432	2.516
a	0.899	0.066	0.740	0.987	0.898	0.066	0.734	0.987	0.899	0.066	0.742	0.986	0.901	0.064	0.749	0.988
η	3.773	5.670	0.021	20.242	3.603	4.095	0.046	15.593	4.311	4.490	0.050	16.316	5.356	5.596	0.075	20.012
Λ_{11}	0.033	0.007	0.021	0.049	0.274	0.064	0.164	0.418	2.652	0.647	1.540	4.091	26.805	6.509	15.741	41.149
Λ_{22}	0.031	0.007	0.019	0.046	0.273	0.064	0.165	0.412	2.593	0.642	1.495	4.016	25.375	6.210	14.739	39.012
Λ_{12}	-0.001	0.005	-0.011	0.009	0.001	0.046	-0.088	0.094	0.007	0.466	-0.935	0.914	-0.066	4.665	-9.381	9.272
Σ_{11}	32.484	7.840	20.921	50.224	3.986	1.003	2.441	6.358	0.416	0.113	0.253	0.683	0.041	0.011	0.025	0.066
Σ_{22}	34.895	9.271	22.096	55.956	4.002	1.037	2.457	6.395	0.428	0.140	0.254	0.726	0.044	0.012	0.026	0.073
ρ_Σ	0.024	0.161	-0.293	0.338	-0.004	0.171	-0.345	0.330	-0.002	0.182	-0.357	0.363	0.003	0.183	-0.358	0.361

Table 5: Posterior summaries of the parameters of the CPHNBCR model considering the prior distribution $a \sim \text{Beta}(18, 2)$ and different values for $\boldsymbol{\Lambda}_0$.

	$\Lambda_0 = 0.001\boldsymbol{I}$				$\Lambda_0 = 0.01\boldsymbol{I}$				$\Lambda_0 = 0.1\boldsymbol{I}$				$\Lambda_0 = \boldsymbol{I}$			
Parameters	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%
b_0	-0.411	0.870	-2.032	1.301	-0.319	0.762	-1.852	1.120	-1.735	1.800	-6.064	0.889	-0.713	0.699	-2.174	0.590
b_1	-0.313	0.645	-1.580	1.010	-0.255	0.497	-1.197	0.754	-0.038	0.767	-1.313	1.782	-0.093	0.428	-0.894	0.827
b_2	0.860	0.556	-0.269	1.938	0.847	0.506	-0.145	1.826	1.318	1.092	-0.366	3.979	0.570	0.456	-0.362	1.428
b_3	-0.071	0.054	-0.156	0.046	-0.044	0.047	-0.115	0.061	-0.031	0.050	-0.117	0.066	-0.069	0.037	-0.148	-0.003
b_4	0.065	0.058	-0.060	0.161	0.040	0.055	-0.079	0.122	0.051	0.082	-0.118	0.210	0.083	0.036	0.013	0.152
β_1	0.185	0.306	-0.418	0.783	0.166	0.271	-0.376	0.704	0.172	0.347	-0.602	0.807	0.191	0.241	-0.273	0.676
β_2	-0.273	0.342	-0.898	0.465	-0.278	0.318	-0.866	0.396	-0.689	0.442	-1.686	0.081	-0.388	0.253	-0.879	0.125
β_3	0.004	0.022	-0.051	0.039	-0.002	0.020	-0.047	0.032	0.004	0.016	-0.028	0.037	0.009	0.021	-0.033	0.051
β_4	-0.046	0.025	-0.101	0.004	-0.046	0.020	-0.088	-0.009	-0.039	0.023	-0.080	0.010	-0.044	0.015	-0.073	-0.014
a	0.900	0.065	0.748	0.988	0.900	0.067	0.735	0.987	0.900	0.066	0.737	0.986	0.901	0.065	0.742	0.987
η	11.741	7.511	0.195	27.162	12.529	6.854	1.057	27.957	13.991	6.089	4.242	26.668	18.442	6.438	6.713	31.891
Λ_{11}	0.035	0.007	0.022	0.051	0.287	0.067	0.174	0.436	2.712	0.656	1.579	4.157	26.807	6.519	15.669	41.211
Λ_{22}	0.031	0.007	0.019	0.046	0.271	0.065	0.160	0.414	2.563	0.645	1.466	3.987	25.721	6.283	14.843	39.269
Λ_{12}	0.000	0.005	-0.010	0.011	0.004	0.047	-0.089	0.095	0.015	0.464	-0.911	0.947	0.025	4.686	-9.293	9.174
Σ_{11}	30.662	6.929	19.868	47.016	3.807	0.937	2.358	5.974	0.406	0.106	0.247	0.663	0.041	0.011	0.025	0.067
Σ_{22}	35.261	8.839	21.925	55.973	4.054	1.090	2.453	6.623	0.433	0.124	0.257	0.728	0.043	0.012	0.026	0.072
ρ_Σ	-0.009	0.158	-0.315	0.301	-0.014	0.171	-0.347	0.321	-0.006	0.180	-0.374	0.345	-0.001	0.182	-0.358	0.359

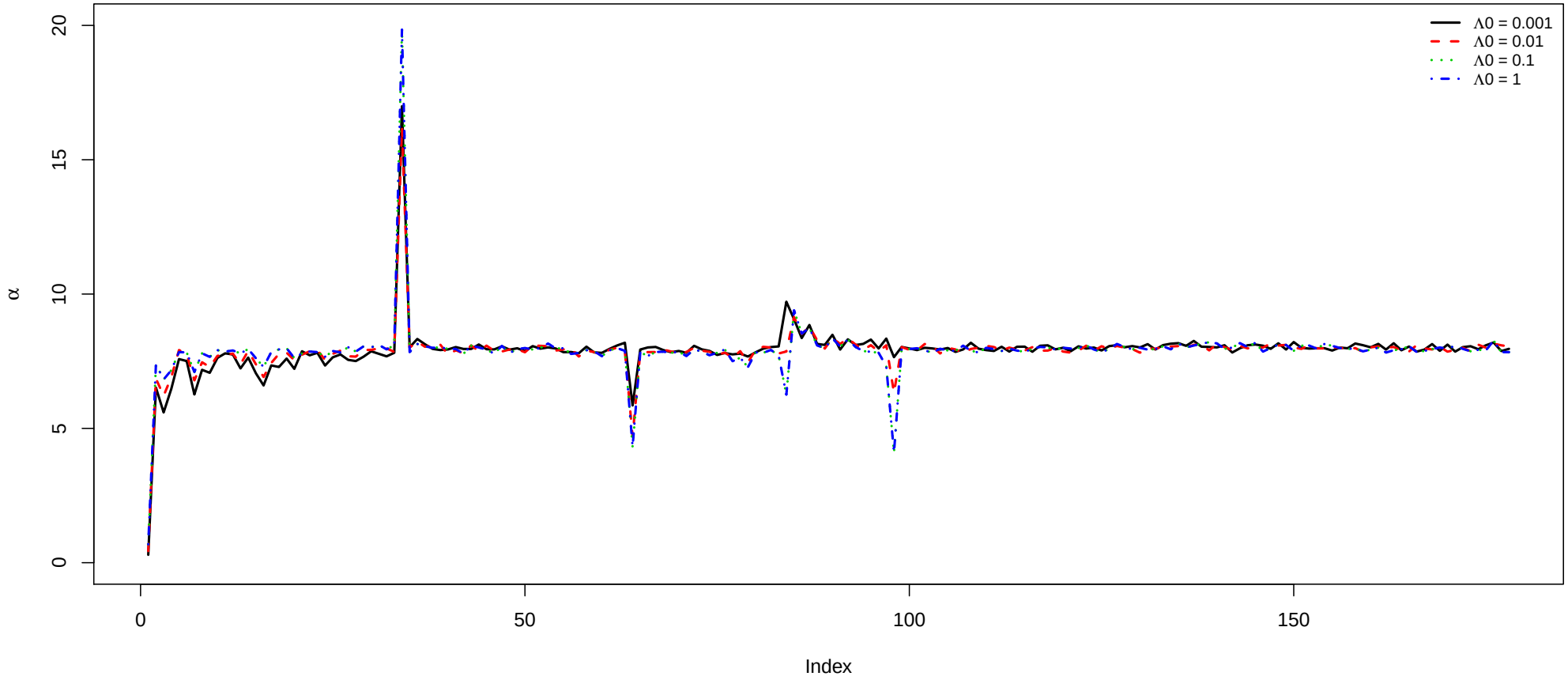

Figure 1: Posterior means of α_i's with $\Lambda_0 = 0.001$, 0.01, 0.1 and 1.