

HIERARCHICAL BAYESIAN MODELING OF HEALTH INSURANCE CLAIMS

by

Guang Qu Yu

B.Com., University of Melbourne, 2011

A Project Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the

Department of Statistics and Actuarial Science
Faculty of Science

© Guang Qu Yu 2015

SIMON FRASER UNIVERSITY

Spring 2015

All rights reserved.

However, in accordance with the *Copyright Act of Canada*, this work may be reproduced without authorization under the conditions for "Fair Dealing". Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

APPROVAL

Name: Guang Qu Yu

Degree: Master of Science

Title: HIERARCHICAL BAYESIAN MODELING OF HEALTH INSURANCE CLAIMS

Examining Committee: Dr. Cary Chi-Liang Tsai
Associate Professor
Chair

Dr. Yi Lu
Associate Professor
Senior Supervisor

Dr. Tim Swartz
Professor
Co-Supervisor

Dr. Gary Parker
Associate Professor
External Examiner

Date Approved: March 6th, 2015

Partial Copyright Licence



The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the non-exclusive, royalty-free right to include a digital copy of this thesis, project or extended essay[s] and associated supplemental files ("Work") (title[s] below) in Summit, the Institutional Research Repository at SFU. SFU may also make copies of the Work for purposes of a scholarly or research nature; for users of the SFU Library; or in response to a request from another library, or educational institution, on SFU's own behalf or for one of its users. Distribution may be in any form.

The author has further agreed that SFU may keep more than one copy of the Work for purposes of back-up and security; and that SFU may, without changing the content, translate, if technically possible, the Work to any medium or format for the purpose of preserving the Work and facilitating the exercise of SFU's rights under this licence.

It is understood that copying, publication, or public performance of the Work for commercial purposes shall not be allowed without the author's written permission.

While granting the above uses to SFU, the author retains copyright ownership and moral rights in the Work, and may deal with the copyright in the Work in any way consistent with the terms of this licence, including the right to change the Work for subsequent purposes, including editing and publishing the Work in whole or in part, and licensing the content to other parties as the author may desire.

The author represents and warrants that he/she has the right to grant the rights contained in this licence and that the Work does not, to the best of the author's knowledge, infringe upon anyone's copyright. The author has obtained written copyright permission, where required, for the use of any third-party copyrighted material contained in the Work. The author represents and warrants that the Work is his/her own original work and that he/she has not previously assigned or relinquished the rights conferred in this licence.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2013

Abstract

The purpose of this project is to propose a statistical model for health insurance total claim amounts classified by age group, region of residence and time horizon of the insured population under Bayesian framework. This model can be used to predict future total claim amounts and thus to facilitate premium determination. The prediction is based on the past observed information and prior beliefs about the insured population, number of claims and amount of claims. The insured population growth is modelled by a generalized exponential growth model (GEGM), which takes into account the random effects in age, region and time classifications. The number of claims for each classified group is assumed Poisson distributed and independent of the size of the individual claims. A simulation study is conducted to test the effectiveness of modelling and estimation, and Markov chain Monte Carlo (MCMC) is used for parameter estimation. Based on the predicted values, the premiums are estimated using four premium principles and two risk measures.

Keywords: Collective risk model; Health Insurance; Hierarchical Bayesian model; Markov chain Monte Carlo (MCMC); Premium Principle; Risk Measures

To my beloved family!

"Essentially, all models are wrong, but some are useful"

--- *George E.P. Box, EMPIRICAL MODEL-BUILDING AND RESPONSE SURFACES,*
1987

Acknowledgements

There are so many people that inspired and supported me during the journey at Simon Fraser University. I would like to express the greatest gratitude to my senior supervisor Dr. Yi Lu. It is her professional guidance and enormous support that encouraged me throughout my graduate study. Without her enlightening instructions and persistent patience, this project would not be possible.

A huge thank goes to my co-supervisor Prof. Tim Swartz for his expertise in Bayesian inferences and always being there to help even beyond his duty. His course aroused my interest in Bayesian theories and built the foundation of this project. I would also like to thank Dr. Gary Parker for serving on my supervisory committee and providing valuable input on this project. I am very grateful that Dr. Cary Tsai has agreed to be the chair of the examining committee.

A special thank goes to Dr. Shuanming Li at the University of Melbourne, without whom I would not have chosen to pursue a master's degree at SFU. I will forever be grateful for his suggestion and reference.

I have owed a lot to my family for their everlasting love and support which I will use the rest of my life to repay. I am very fortunate to be friends with all lovely fellow graduate students and I truly wish them all the best for their study and career.

Contents

| | |
|---|-------------|
| Approval | ii |
| Partial Copyright License | iii |
| Abstract | iv |
| Dedication | v |
| Quotation | vi |
| Acknowledgements | vii |
| Contents | viii |
| List of Tables | x |
| List of Figures | xi |
| 1 Introduction | 1 |
| 1.1 Background and Motivation | 1 |
| 1.2 Literature Review | 2 |
| 1.3 Outline | 3 |
| 2 Models and Assumptions | 5 |
| 2.1 The Compound Collective Risk Model | 5 |
| 2.2 The Compound Poisson Process | 6 |
| 2.3 Bayesian Inference | 7 |
| 2.4 Generalized Exponential Growth Models | 9 |
| 2.5 Model Specification | 9 |

| | | |
|----------|--|-----------|
| 2.5.1 | The General Bayesian Model | 9 |
| 2.5.2 | Insured Population with Spatial Effect | 12 |
| 2.5.3 | Hierarchical Collective Risk Model Summary | 14 |
| 2.6 | Computation Using MCMC Algorithm | 16 |
| 3 | Simulation Studies and Model Fitting | 20 |
| 3.1 | Simulation Studies | 20 |
| 3.2 | Model Fitting | 22 |
| 3.2.1 | Prior Elicitation | 22 |
| 3.2.2 | Illustrative Results | 24 |
| 3.2.3 | Credible Intervals of the Posterior Means | 32 |
| 4 | Predictions and Premium Determination | 39 |
| 4.1 | Predictions in Bayesian Framework | 39 |
| 4.2 | Predictive Results | 40 |
| 4.3 | Premium Determination under Various Premium Principles | 49 |
| 4.3.1 | Net Premium Principle | 49 |
| 4.3.2 | Expected Value Premium Principle | 51 |
| 4.3.3 | Variance Premium Principle | 52 |
| 4.3.4 | Standard Deviation Premium Principle | 52 |
| 4.3.5 | Value-at-Risk | 53 |
| 4.3.6 | Tail Value-at-Risk | 53 |
| 4.3.7 | Numerical Premium Analysis | 54 |
| 5 | Conclusion | 58 |
| 5.1 | Concluding Summary | 58 |
| 5.2 | Further Remarks | 59 |
| | References | 62 |
| | Appendix A Derivation of Conditional Posterior Distributions | 64 |
| | Appendix B Simulated Data | 69 |
| | Appendix C Premiums for 21st Time Unit in Region 2 | 73 |

List of Tables

| | | |
|-----|---|----|
| 3.1 | True Values/Distributions of the Parameters | 22 |
| 3.2 | Statistic Summary of the Major Parameters | 38 |
| 4.1 | Predicted Insured Population, Claim Number and Total Claim Amount for 21 st Time Unit | 50 |
| 4.2 | Total Premiums for 21 st Time Unit in Region 1 Using Predicted Results . . . | 56 |
| 4.3 | Premium Per Policyholder for 21 st Time Unit in Region 1 | 56 |
| B.1 | Simulated Insured Population $M_{t,i,a}$ | 70 |
| B.2 | Simulated Total Claim Frequency $N_{t,i,a}$ | 71 |
| B.3 | Simulated Aggregate Loss $X_{t,i,a}$ | 72 |
| C.1 | Total Premiums for 21 st Time Unit in Region 2 Using Predicted Results . . . | 74 |
| C.2 | Premium Per Policyholder for 21 st Time Unit in Region 2 | 74 |

List of Figures

| | | |
|------|---|----|
| 2.1 | Region Map Example | 13 |
| 3.1 | Trace Plot and the Posterior Distribution of β_{a_0} , $a = 1, 2, \dots, 7$ | 27 |
| 3.2 | Trace Plot and the Posterior Distribution of β_{a_2} , $a = 1, 2, \dots, 7$ | 28 |
| 3.3 | Trace Plot and the Posterior Distribution of β_1 | 29 |
| 3.4 | Trace Plot and the Posterior Distribution of λ_a for $a = 1, 2, \dots, 7$ | 30 |
| 3.5 | Trace Plot and the Posterior Distribution of θ_a for $a = 1, 2, \dots, 7$ | 31 |
| 3.6 | Trace Plot and the Posterior Distribution of L_i for $i = 1, 2$ | 32 |
| 3.7 | Credible Interval of β_{a_0} for $a = 1, 2, \dots, 7$ | 33 |
| 3.8 | Credible Interval of β_{a_2} for $a = 1, 2, \dots, 7$ | 33 |
| 3.9 | Credible Interval of λ_a for $a = 1, 2, \dots, 7$ | 34 |
| 3.10 | Observed Number of Claims, Times 1 to 20, Age Groups 7 and 1. | 35 |
| 3.11 | Observed Number of Claims, Times 1 to 20, Age Groups 7 and 3. | 35 |
| 3.12 | Observed Number of Claims, Times 1 to 20, Age Groups 7 and 6. | 35 |
| 3.13 | Credible Interval of $1/\theta_a$ for $a = 1, 2, \dots, 7$ | 36 |
| 3.14 | Observed Average Claim Severity, Time 1 to 20, Age Groups 2 and 1. | 37 |
| 3.15 | Observed Average Claim Severity, Time 1 to 20, Age Groups 2 and 5. | 37 |
| 4.1 | Prediction of Insured Population for 21 st Time Unit -- Region 1 | 41 |
| 4.2 | Insured Population with Prediction for Age Groups 1 and 6 -- Region 1 | 42 |
| 4.3 | Insured Population with Prediction for Age Groups 1 and 5 -- Region 1 | 43 |
| 4.4 | Prediction of Number of Claims for 21 st Time Unit -- Region 1 | 44 |
| 4.5 | Claim Number with Prediction -- Age Groups 7 and 1 | 45 |
| 4.6 | Claim Number with Prediction -- Age Groups 7 and 3 | 45 |
| 4.7 | Claim Number with Prediction -- Age Groups 7 and 6 | 45 |
| 4.8 | Prediction of Total Claim Amount for 21 st Time Unit | 47 |

- 4.9 Average Amount Per Claim with Prediction in Region 1 -- Age Groups 2 and 1 48
- 4.10 Average Amount Per Claim with Prediction in Region 1 -- Age Groups 2 and 5 48

Chapter 1

Introduction

1.1 Background and Motivation

Among the prevailing topics for health insurance providers, determining a right level of premium to charge has been discussed widely in the industry. The expenditure in health-care has been increasing dramatically over the past decade. According to a report entitled "National Health Expenditure Trends, 1975 to 2014"¹ published by Canadian Institute for Health Information (CIHI) in October 2014, the total health-care costs in Canada doubled in the past decade, from \$98.6 billion in 2000 increased to \$205.4 billion in 2012. One of the causes of such phenomenon is the current fee schedule basis² (or fee-for-service model) such that the payments to doctors depend on the quantity of treatments. It gives the incentive for doctors to over-care the patients by increasing the number and length of the visits, creating pressure on future health care costs. Furthermore, there is evidence showing that the population in some countries like U.S.³ and Canada⁴ is increasing in certain areas over the past years. Consequently, the health insurance providers are facing challenges of increasing number of claims (frequency) and the amount of claims (severity)

¹This report can be downloaded free from the website of Canadian Institute for Health Information.

²Further discussion about the major cost drivers can be found in the report "Health Care Cost Drivers: The Facts".

³Source: United States Census Bureau. Access through: <http://www.census.gov/geography.html>.
Alternative Source: The World Bank. Access through: <http://data.worldbank.org/indicator/SP.POP.TOTL>

⁴Source: Statistics Canada, censuses of population, 1851 to 2011.
Access through: <http://www12.statcan.ca/census-recensement/2011/as-sa/98-310-x/2011001/fig/fig2-eng.cfm>

year after year. It is a paramount issue for them to be able to forecast future claim trends and hence to facilitate decision making in premium scheme.

The claim frequency and severity predictions involve the consideration of many uncertain factors, such as health-care regulation, demographic and geographic factors and so on. After years of experience, most likely the health insurance experts possess profound knowledge in some, if not all, of these areas. The expert opinions should be taken into consideration when forecasting the claims and premiums. The Bayesian framework is very good at dealing with such situations. Before conducting the analysis, a Bayesian framework requires prior information on the parameters, that is, the expert opinions and industrial information. The observed data over the past years of practice is a valuable input for the prediction process. Furthermore, with the assistance of the software OpenBUGS (or WinBUGS) the implementation of Bayesian calculations becomes handy with ease in interpretation. All the above reasons stimulate the incentive of considering health insurance problem under a Bayesian framework.

1.2 Literature Review

This project is inspired by the work of many authors. Here we briefly review some of the published work related to this project.

Migon and Moura (2005) propose a generalized collective risk model under Bayesian framework to determine the premium for health insurance. They recommend a full Bayesian model to take all uncertainty into account. The premium is determined based on the past information about the number and size of the claims and the population at risk, which is classified by time and age of the insured population. The proposed model assumes that the total claim amount is age dependent and priors are hierarchically distributed for each age class.

Migon et al. (2006) apply a similar methodology to two real data sets collected in Brazil. They discuss the implementation of the collective risk model under a Bayesian setting with stochastic simulation techniques. The premium is expressed by maximization of the insurer's expected utility under the Bayesian model. The insured population is assumed to follow a non-linear growth model. According to dynamic Bayesian forecasting techniques, generalized exponential growth models (GEGM) have been studied by Migon and

Gamerman (1993). This class of models is for processing data with non-negative and non-decreasing mean functions. They present and discuss the properties of the general modified exponential family that can be applied to model the population growth.

Souza et al. (2009) propose a method to predict the population for small areas given the census data. Since the population growth pattern for a municipal may be related to the development level of its surrounding neighbors, a spatial hierarchical model is proposed, which takes into account the extra uncertainty associated with the hyperparameters. A Markov chain Monte Carlo (MCMC) algorithm is applied to generate samples for numerical analysis.

Gschlöbl and Czado (2007) discuss the statistical models for the number of claims and average claim size in non-life insurance under Bayesian context. The premium is calculated based on the simulated total claim sizes. They analyze the claim frequency and claim size separately, which is initially discussed by Dimakos and Frigessi (2002), by assuming a spatial Poisson regression model for claim frequency and a Gamma model for the average claim size per policyholder. As summarized by Gschlöbl and Czado (2007), by considering the regression model for spatial data, it includes the correlated spatial random effects to describe the underlying spatial dependency pattern. Furthermore, the spatial dependencies are modeled by Gaussian conditional autoregressive (CAR) prior introduced by Pettitt et.al. (2002). The CAR models are based on the assumption that adjacent regions share similar features and hence have strong spatial dependencies.

Premium principles have been extensively discussed in the literature, e.g., Young (2004) and Goovaerts et al. (2010), among others. Young (2004) describes three methods that actuaries use to design premium principles. The author also lists common premium principles along with a discussion of the desirable properties of those premium principles. Hardy (2006) and Sarykalin et al. (2008) consider Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) as legitimate approaches to determine premium and give detailed discussion about their properties and application.

1.3 Outline

The purpose of this project is to make prediction of the future total claim amounts under the Bayesian framework to determine the premium rate. In light of this aim, the following

report consists of four chapters. The first half of Chapter 2 lists the preliminary topics that play major roles in our Bayesian modeling to be discussed. We briefly discuss the collective risk model and its mean and variance expression. Some important Bayesian concepts are also provided, such as the fundamental Bayes' Theorem and Gibbs sampling technique. The study of the generalized exponential growth model (GEGM) paves the way to the coming discussion about the population modeling. A hierarchical Bayesian model, along with the assumptions, are presented in the second half of Chapter 2. We model the insured population growth based on three covariates, namely the age class, location of residence and time of measurement. The insured population contributes to the claim numbers, which eventually contribute to the prediction of total claim amounts.

The aim of Chapter 3 is to test how effective the proposed model is in identifying the underlying parameters. We perform a simulation study with predetermined values (or distributions) of the parameters; then we fit the simulated data back to the model to see if it can capture the true values of the parameters. Once the predicted total claim amounts are obtained, the procedures of forecast premiums are discussed in Chapter 4. The first half of Chapter 4 presents a number of premium principles while the numerical premium analysis is conducted in the second half of that chapter. Chapter 5 wraps up this project by suggesting further possible developments for this topic.

Chapter 2

Models and Assumptions

This chapter is arranged as follows: firstly, classical compound collective risk models and generalized exponential growth models are briefly described, as these models serve as the building blocks for the key model studied in this project. Then the hierarchical Bayesian model and relevant derivations are presented.

2.1 The Compound Collective Risk Model

The collective risk model is well known and discussed intensively in the actuarial realm. Consider a portfolio of policies of a single type. Let N denote the total number of claims arising from a risk in a certain time period, and Z_j denote the amount of j^{th} claim. The aggregate total amount of claims is given by

$$X = \sum_{j=1}^N Z_j, \quad (2.1)$$

with $X = 0$ when $N = 0$. The main assumptions embedded in this model are:

- the individual claim amounts Z_j 's are positive independently and identically distributed random variables;
- the total number of claims N is a random variable and independent of the claim amounts Z_j 's.

The advantage of the collective risk model is obvious. That is, the claim frequency and claim severity can be separately modeled. For example, a general raise in cost of drugs

may affect the claim severity but with little influence on the claim frequency, whereas introducing another line of business would increase claim frequency without much alteration in claim severity. Furthermore, the measure of expected value (and variance) of the aggregate claim amount can be decomposed by the measure of mean and variance of the claim frequency and severity, namely,

$$\begin{aligned} E(X) &= E(E[X|N]) = E(N)E(Z), \\ Var(X) &= E(V[X|N]) + V(E[X|N]) = E(N)V(Z) + [E(Z)]^2V(N). \end{aligned} \quad (2.2)$$

When N has a Poisson distribution with parameter λ , we say that X in (2.1) has a compound Poisson distribution with parameters λ and F , where $F(x) = Pr(Z_1 \leq x)$ denotes the distribution function of individual claim amounts. It follows from (2.2) that in this case,

$$\begin{aligned} E(X) &= \lambda E(Z), \\ Var(X) &= \lambda V(Z) + [E(Z)]^2 \lambda = \lambda E(Z^2). \end{aligned}$$

It is worth mentioning that there are situations where the total number of claims is not independent of the claim amount. For example, patients with certain types of diseases need special treatments that require to have frequent doctor visits, and each visit may take more time than an ordinary visit. As a result, the number of claims for such patients increases as well as the claim amount of each visit. The original assumptions of the collective risk model may not be suitable under this circumstance.

2.2 The Compound Poisson Process

The model studied in this project has features in common with the compound Poisson process. We briefly discuss the compound Poisson process here for the completeness of the project.

The Poisson process is a special counting process describing the occurrence of events, which is the occurrence of claims in this project. Let $N(t)$ be the number of claims up to time t , $t > 0$. A Poisson process with parameter λ indicates that the sequence of inter-arrival times, that is the time between two consecutive claims, are independently exponentially distributed with mean $1/\lambda$. Hence the distribution of $N(t)$, for a fixed $t > 0$, is Poisson with parameter λt . In other words, if we have discrete time $t = 1, 2, \dots$ with unit length,

the number of claims between $t - 1$ and t follows an independent and identical Poisson distribution with parameter λ , that is,

$$N(t) - N(t - 1) \sim \text{Poisson}(\lambda), \quad t = 1, 2, 3, \dots$$

Let $Z_i, i \geq 1$, be a sequence of independently and identically distributed random variables with finite variance and independent of $N(t)$ for all $t > 0$. The compound Poisson process with Poisson parameter λ is denoted as

$$X(t) = \sum_{i=1}^{N(t)} Z_i$$

with $X(t) = 0$ when $N(t) = 0$.

$X(t)$ is a homogeneous Markov process that has stationary and independent increments. That is to say, for $0 < s < t$, the increment of the process $X(t) - X(s)$ depends only on $t - s$, not on the value of s and t . Furthermore, the increments over non-overlapping time intervals are independent, that is if $0 < s < t \leq u < v$, $X(t) - X(s)$ is independent of $X(v) - X(u)$. Refer to Dickson (2005) for detailed discussion related to this topic.

2.3 Bayesian Inference

In recent years, Bayesian methodology has aroused the attention of researchers in mathematics, statistics and actuarial sciences. One of the major merits of the Bayesian framework is that it allows the introduction of prior beliefs, which eventually leads to posterior beliefs. Therefore, the posterior beliefs of the random variable not only incorporate the prior beliefs but also the information contained the data.

In this section the fundamental Bayesian paradigm is presented. More advanced applications are discussed in the next section. See Klugman (1992) for further discussions on Bayesian statistics in actuarial science.

The prior beliefs about the values for d parameters of interest $\theta = (\theta_1, \theta_2, \dots, \theta_d)$, $d > 0$, can be expressed by the probability density function $\pi(\theta)$, representing our opinion on the possible values of θ and the relative chances of being true parameter. Suppose it is possible to obtain n observations, namely $\mathbf{X} = (x_1, x_2, \dots, x_n)$, whose joint density function is defined as $f(\mathbf{X})$. Denote $l(\mathbf{X}|\theta)$ as the likelihood function, $\pi(\theta, \mathbf{X})$ as the joint density function of θ and \mathbf{X} and $\pi(\theta|\mathbf{X})$ as the posterior distribution, which is the conditional probability

distribution of the parameters given the observed data. According to Bayes' Theorem the posterior distribution can be expressed as

$$\pi(\boldsymbol{\theta}|\mathbf{X}) = \frac{\pi(\boldsymbol{\theta}, \mathbf{X})}{f(\mathbf{X})} = \frac{\pi(\boldsymbol{\theta})l(\mathbf{X}|\boldsymbol{\theta})}{f(\mathbf{X})} \propto \pi(\boldsymbol{\theta})l(\mathbf{X}|\boldsymbol{\theta}).$$

Hence the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{X})$ is proportional to the prior density function times the likelihood function, that summarizes the modified beliefs of the parameter according to the observations.

If the posterior distribution follows a well-known distribution (say the Gamma distribution), it can be modeled or simulated with little difficulties. However, in reality it is quite common to come across parameters with unrecognizable posterior distributions, especially in models with high dimension. One of the prevalent methods is to utilize MCMC as an approximation algorithm. Generally speaking, an MCMC algorithm allows users to simulate samples from the posterior distribution when direct generation is complicated or impossible. The most commonly used MCMC algorithms are Metropolis-Hastings and the Gibbs sampler. The Gibbs sampling algorithm, which is a special case of Metropolis-Hastings, is a scheme based on successive generations from the full conditional distributions, denoted as $\pi(\theta_i|\theta_{-i}, \mathbf{X})$, $i = 1, 2, \dots, d$, where θ_{-i} represents all parameters in $\boldsymbol{\theta}$ but θ_i . The following is the procedure to generate draws for the parameter of interest. Further discussion can be found in Gamerman (1997), Gilks et al. (1995), Banerjee et al. (2003) and Lee (2012).

The Gibbs Sampling Algorithm:

1. Set initial values $\boldsymbol{\theta}^{(0)} = \{\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)}\}$,
2. With $\mathbf{X} = (x_1, x_2, \dots, x_n)$ being the observations, for $j = 1, \dots, J$, generate $\boldsymbol{\theta}^{(j)}$ by repeating the following:
 - Generate $\theta_1^{(j)}$ from $\pi(\theta_1 | \theta_2^{(j-1)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}, \mathbf{X})$;
 - Generate $\theta_2^{(j)}$ from $\pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}, \mathbf{X})$;
 - \vdots
 - Generate $\theta_d^{(j)}$ from $\pi(\theta_d | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{d-1}^{(j)}, \mathbf{X})$.

2.4 Generalized Exponential Growth Models

The insured population can be modeled using the generalized exponential growth model (GEGM) presented in Migon and Gamerman (1993). The importance of the growth trend has been well recognized and widely used in areas such as demography, biology and actuarial science. As it is stated by Migon and Gamerman (1993): "A major advantage of this approach is to keep the measurements in the original scale, making the interpretation easier". Assume that the size of the population at time t , π_t , characterized by the parameterization (a, b, γ, λ) , is modeled by a probability distribution in the exponential family with mean

$$\mu_t = [a + b \exp(\gamma t)]^{\frac{1}{\lambda}}, \quad t \geq 0.$$

The following are well-known cases shown in the literature, which ensure a non-negative non-decreasing function of the mean μ_t .

- (1) Logistic: when $\lambda = -1$, $\mu_t^{-1} = a + b \exp(\gamma t)$
- (2) Gompertz: when $\lambda = 0$, μ_t is defined such that $\ln(\mu_t) = a + b \exp(\gamma t)$
- (3) Modified exponential: when $\lambda = 1$, $\mu_t = a + b \exp(\gamma t)$

The constraints of the parameterization (a, b, γ, λ) depend on the characteristics of the population. Further discussions on this model are conducted in the following sections.

2.5 Model Specification

The aim of this section is to introduce the general model used in this project. Within the Bayesian structure, spatial random effects are incorporated in modeling the insured population. This is followed by an explanation of the computations in the corresponding MCMC algorithm.

2.5.1 The General Bayesian Model

One of the major issues facing the health insurance industry is to evaluate and determine the optimal premium. The premium is normally evaluated based on the past information in terms of claim severity, claim frequency and the information of the policyholders.

Migon and Moura (2005) propose a generalization of the collective risk model taking into account the evolution of the population at risk described by a hierarchical growth model. Their work was based on both age and time related parameters, arguing that population evolution is affected by both the age group and the time of measurement.

This project further elaborates their framework by introducing spatial related parameters into the model. Demographic characteristics can not be neglected when considering the population growth. For example, in some developing countries the rapid pace of urbanization stimulates the incentive for people to relocate from rural to urban regions. First-tier cities get most of the attention because of the high quality education, medical services as well as abundant career opportunities. Meanwhile, the regions surrounding first-tier cities would soon be urbanized and populated as the first-tier cities reaching the limit of their capacity. At the same time the population growth in the rural regions tend to be stable and diminishing. The demographic features significantly influence the health insurance industry and hence the trend in the insured population with respect to demography shall be considered when modeling the claim severity and frequency.

The basic collective risk model can be extended to incorporate age, time and region factors. Age is one of the important factors influencing the mortality and health condition of the policyholders. However, setting premiums for each age is redundant as people similar in age (for instance, ages from 26 to 29) show similar mortality and health conditions, given all else equal. It is convenient for the health insurance provider to classify the policyholders by age classes, denoted as $a = 1, 2, 3, \dots, A$. The insurer has the full freedom in terms of the age classification, such as the number of different ages in each age class. Furthermore, there is no restriction on the time unit, which could be yearly, quarterly, monthly or customized units. Generally, the frequency of the collection of input data can be a fair reflection of the time unit. It is suggested to be consistent with the unit once it has been decided.

Migon and Moura (2005) studied the compound collective risk model for a portfolio of policies classified by age class $a = 1, \dots, A$ and by time $t = 1, \dots, T$. In this project, we consider the region of residence of the policyholder as another classification category denoted as $i = 1, 2, \dots, I$. Let $N_{t,i,a}$ represent the total number of claims and $X_{t,i,a}$ represent the aggregate total claim amount occurred in the time interval $(t - 1, t)$ for age class a in region i . Let $Z_{t,i,a,j}$ denote the amount of the j^{th} claim occurring within this time interval for

age class a in region i . The compound collective risk model is given by

$$X_{t,i,a} = \sum_{j=1}^{N_{t,i,a}} Z_{t,i,a,j}, \quad i = 1, 2, \dots, I, \quad t = 1, \dots, T, \quad a = 1, \dots, A. \quad (2.3)$$

The assumptions for the collective risk model all hold, as stated in the previous section. For simplicity, from this point on we use the phrase "at time t " instead of "in the time interval $(t - 1, t)$ " to describe claims occurred between $t - 1$ to t , $t = 1, 2, \dots, T$.

Assume that the individual claim amount follows a Gamma distribution with parameter κ_a and θ_a for $a = 1, 2, \dots, A$, and the number of claims $N_{t,i,a}$ is Poisson distributed with mean $M_{t,i,a} \cdot \lambda_a$, that is,

$$Z_{t,i,a} | \kappa_a, \theta_a \sim \text{Gamma}(\kappa_a, \theta_a), \quad \kappa_a > 0, \quad \theta_a > 0, \quad (2.4)$$

$$N_{t,i,a} | \lambda_a, M_{t,i,a} \sim \text{Poisson}(M_{t,i,a} \lambda_a), \quad \lambda_a > 0, \quad (2.5)$$

where $M_{t,i,a}$ is the insured population at time t for age class a at region i , and λ_a is the average number of claim for each individual per unit of time. The notation $M_{t,i,a}$ implies a constant population over the time interval $(t - 1, t)$ as we do not model the population growth within this interval. For fixed t , i and a , knowing $N_{t,i,a} = n_{t,i,a}$, the claim sizes $Z_{t,i,a,j}$, $j = 1, 2, \dots, n_{t,i,a}$ are independent and identically distributed. The sum of these Gamma distributions forms another Gamma distribution, namely,

$$X_{t,i,a} | \theta_a, n_{t,i,a}, \kappa_a \sim \text{Gamma}(n_{t,i,a} \kappa_a, \theta_a), \quad \theta_a > 0, \quad \kappa_a > 0.$$

Refer to (2.4), the amount of individual claim Z is the only variable dependent on age in this model, and hence the notation in (2.3) and (2.4) can be simplified to $Z_{a,j}$ and Z_a , respectively. One can further investigate the feasibility of having κ and θ being time or region related. In (2.5), $M_{t,i,a} \lambda_a$ represents the average number of claims made by the insured population in age group a at time t in region i . Since for given age class a and region i , the insured population $M_{t,i,a}$ varies over time, the collection of the total claim amounts $\{X_{t,i,a}, t = 1, \dots, T, i = 1, \dots, I, a = 1, \dots, A\}$ are not identically distributed.

Note that in reality the health insurance companies normally keep the information of the policyholders, such as the number, amount and time of claims made as well as the age and region of residence of the claimants. Therefore, the total claim amount X , the total number of claims N and the insured population M are assumed observed, and hence being treated as inputs of the model.

2.5.2 Insured Population with Spatial Effect

According to the work by Migon and Gamerman (1993), the insured population can be modeled by a GEGM. For illustrative purposes it is assumed that the insured population follows a Normal distribution

$$M_{t,i,a} \sim N(\mu_{t,i,a}, \tau^{-1}), \quad \tau > 0,$$

with precision τ and mean

$$\mu_{t,i,a} = \beta_{a_0} + L_i + \beta_1 e^{t\beta_{a_2}}, \quad (2.6)$$

in which L_i denotes the spatial factor for region i and β_{a_0} and β_{a_2} are age related parameters. It should be mentioned that information about the age of the insured population is independent of the information about region. Knowledge about demographic and age features can be taken into account when selecting the prior distributions. Firstly the age related parameters are specified as follows:

$$\begin{aligned} \beta_{a_0} &= \beta_0 + \varepsilon_a^0, & \varepsilon_a^0 &\sim N(0, \tau_{\varepsilon_0}^{-1}), & \tau_{\varepsilon_0} &> 0, \\ \beta_{a_2} &= \beta_2 + \varepsilon_a^2, & \varepsilon_a^2 &\sim N(0, \tau_{\varepsilon_2}^{-1}), & \tau_{\varepsilon_2} &> 0, \end{aligned}$$

where τ_{ε_0} and τ_{ε_2} follow Gamma distributions with known parameters. The age related factors $\{\beta_{a_0}\}$ (same case for β_{a_2}) vary by ages but share the same mean. The hyperparameters are assumed to follow normal distributions with different parameters: $\beta_0 \sim N(\mu_0, \tau_0^{-1})$, $\beta_1 \sim N(\mu_1, \tau_1^{-1})$, $\beta_2 \sim N(\mu_2, \tau_2^{-1})$ where $\mu_0, \mu_1, \mu_2, \tau_0, \tau_1, \tau_2$ are known values.

Secondly, the spatial factor, based on the work by Gschl bl and Czado (2007), is assumed to follow a multivariate normal distribution, that is,

$$\mathbf{L} \sim MVN(\mathbf{0}, \sigma^{-1}\mathbf{Q}^{-1}), \quad (2.7)$$

where the $(g, h)^{\text{th}}$ element of the spatial precision matrix \mathbf{Q} is specified as

$$Q_{gh} = \begin{cases} 1 + |\eta| \cdot m_g, & g = h, \\ -\eta, & g \neq h, \quad g \sim h, \quad g, h = 1, 2, \dots, I, \\ 0, & \text{otherwise.} \end{cases} \quad (2.8)$$

This precision matrix describes three types of relative positions for a pair of regions g and h , namely the two regions coincide to be one (denoted as $g = h$); the two regions are

neighbors sharing a common border (denoted as $g \sim h$); or the two regions do not share any common area. The quantity m_g denotes the number of neighbors of region g .

The spatial effects are described by proper CAR priors based on the work by Pettitt et al, (2002). The η is called the degree of spatial dependence, with $\eta = 0$ indicating independent spatial effects. Large value of η means strong spatial dependency. A proper hyper prior is assumed for η . Since a non-negative correlation between two regions is expected, we have $\eta \geq 0$. The two-parameter Pareto distribution with parameters (1,1) and probability density function $1/(1 + \eta)^2$ is selected such that it takes high values for small η . See Gschlöbl and Czado (2007) for further discussion on this topic.

To give an example, consider the following figure as a "map" of an area. Regions 1 shares a same border with region 2 ($1 \sim 2$), but does not overlap with region 3, 4 nor 5. Hence region 1 has one neighbor ($m_1 = 1$). Likewise, region 3 shares borders with regions 2, 4 and 5 ($3 \sim 2$, $3 \sim 4$, and $3 \sim 5$) but does not have common area with region 1. Hence region 3 has three neighbors ($m_3 = 3$).

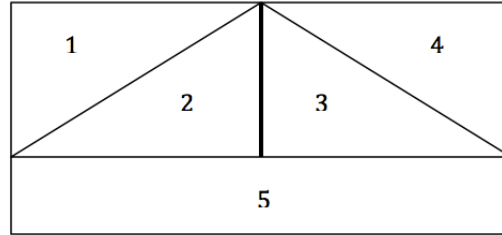


Figure 2.1: Region Map Example

Assuming the degree of spatial dependence is 0.9, the spatial precision matrix Q can be written as

$$Q = \begin{bmatrix} 1 + 0.9 * 1 & -0.9 & 0 & 0 & 0 \\ -0.9 & 1 + 0.9 * 3 & -0.9 & 0 & -0.9 \\ 0 & -0.9 & 1 + 0.9 * 3 & -0.9 & -0.9 \\ 0 & 0 & -0.9 & 1 + 0.9 * 1 & 0 \\ 0 & -0.9 & -0.9 & 0 & 1 + 0.9 * 2 \end{bmatrix},$$

and its inverse can be calculated as

$$Q^{-1} = \begin{bmatrix} 0.6157 & 0.1887 & 0.0752 & 0.0356 & 0.0848 \\ 0.1887 & 0.3983 & 0.1587 & 0.0752 & 0.1791 \\ 0.0752 & 0.1587 & 0.3983 & 0.1887 & 0.1791 \\ 0.0356 & 0.0752 & 0.1887 & 0.6157 & 0.0848 \\ 0.0848 & 0.1791 & 0.1791 & 0.0848 & 0.4723 \end{bmatrix}. \quad (2.9)$$

The quantity σ is chosen to be a non-informative prior following a Gamma distribution with known parameters. As expressed in (2.7), the variance-covariance matrix for the region factor L is the product of σ^{-1} and Q^{-1} . The diagonal elements of matrix (2.9) contribute to the variances of L for all regions. The correlations between adjacent regions are larger than that between non-adjacent regions. For instance, region 2 is a neighbor of region 1 (by sharing a common boundary) but regions 3, 4 and 5 are not. The covariance between regions 1 and 2 (expressed as $\sigma^{-1} \times 0.1887$) is higher than the covariance between region 1 and any other regions. Further discussion about modeling the spatial effects can be found in Gschlöbl and Czado (2007).

2.5.3 Hierarchical Collective Risk Model Summary

Priors are required in order to perform Bayesian inference. In general the use of improper priors can cause computational problems such as inability to obtain posterior distributions. In this project the priors are chosen such that they are proper but relatively less informative priors (referred to as reference priors). Some of the variables have been assigned with prior distributions in previous sections based on the practical knowledge or experience. The remaining variables are assigned with reference priors, due to the insufficiency of information. In this project, $\lambda_a, \theta_a, \kappa_a$ follow independent Gamma distributions, respectively, $G(\alpha_\lambda, \beta_\lambda), G(\alpha_\theta, \beta_\theta), G(\alpha_\kappa, \beta_\kappa)$. As $\alpha_\lambda, \beta_\lambda, \alpha_\theta, \beta_\theta, \alpha_\kappa, \beta_\kappa, \tau, \sigma$ are all non-negative quantities, Gamma distribution is assumed, generally denoted as $\psi \sim \text{Gamma}(\alpha_\psi, \beta_\psi)$.

A summary of the model is presented in the following. Note that $\text{Gamma}(\alpha, \beta)$ denotes a Gamma distribution with probability density function $\beta^\alpha x^{\alpha-1} e^{-\beta x} / \Gamma(\alpha)$ and mean α/β ; $\text{Poisson}(\lambda)$ denotes a Poisson distribution with mean λ ; $N(\mu, \sigma^2)$ denotes a Normal distribution with mean μ and variance σ^2 .

Distributions describing the value of claims, number of claims and insured population:

$$X_{t,i,a}|\theta_a, n_{t,i,a}, \kappa_a \sim \text{Gamma}(n_{t,i,a}\kappa_a, \theta_a), \quad \theta_a > 0, \kappa_a > 0,$$

$$N_{t,i,a}|\lambda_a, M_{t,i,a} \sim \text{Poisson}(M_{t,i,a}\lambda_a), \quad \lambda_a > 0,$$

$$M_{t,i,a} \sim N(\mu_{t,i,a}, \tau^{-1}),$$

where

$$\mu_{t,i,a} = \beta_{a_0} + L_i + \beta_1 e^{t\beta_{a_2}}.$$

Distributions describing age and region class hierarchy:

$$\theta_a|\alpha_\theta, \beta_\theta \sim \text{Gamma}(\alpha_\theta, \beta_\theta),$$

$$\lambda_a|\alpha_\lambda, \beta_\lambda \sim \text{Gamma}(\alpha_\lambda, \beta_\lambda),$$

$$\kappa_a|\alpha_\kappa, \beta_\kappa \sim \text{Gamma}(\alpha_\kappa, \beta_\kappa),$$

$$\beta_{a_0} = \beta_0 + \varepsilon_a^0, \quad \varepsilon_a^0 \sim N(0, \tau_{\varepsilon_0}^{-1}),$$

$$\beta_{a_2} = \beta_2 + \varepsilon_a^2, \quad \varepsilon_a^2 \sim N(0, \tau_{\varepsilon_2}^{-1}),$$

$$\mathbf{L} \sim \text{MVN}(\mathbf{0}, \sigma^{-1}\mathbf{Q}^{-1}),$$

where

$$Q_{gh} = \begin{cases} 1 + |\eta| \cdot m_g, & g = h, \\ -\eta, & g \neq h, g \sim h \forall g, h = 1, 2, \dots, I, \\ 0, & \text{otherwise.} \end{cases}$$

Distributions of the hyperparameter priors:

$$f(\eta) = \frac{1}{(1 + \eta)^2}, \quad \eta > 0,$$

$$\beta_0 \sim N(\mu_0, \tau_0^{-1}),$$

$$\beta_1 \sim N(\mu_1, \tau_1^{-1}),$$

$$\beta_2 \sim N(\mu_2, \tau_2^{-1}),$$

and $\psi \in \{\tau, \sigma, \tau_{\varepsilon_0}, \tau_{\varepsilon_2}, \alpha_\theta, \beta_\theta, \alpha_\lambda, \beta_\lambda, \alpha_\kappa, \beta_\kappa\}$ follows Gamma distributions with known non-negative parameters, that is, $\psi \sim \text{Gamma}(\alpha_\psi, \beta_\psi)$, where $\mu_0, \mu_1, \mu_2, \tau_0, \tau_1, \tau_2, \alpha_\psi, \beta_\psi$ are known values.

2.6 Computation Using MCMC Algorithm

In the following discussion and the numerical illustration, a simplified version of the model is adopted by assuming $\kappa_a = 1$, which means that the individual claim amount $Z_{t,i,a,j}$ in (2.3) is independent of time and region, written as $Z_{a,j}$. Two adjacent regions are to be studied (i.e., $I = 2$). The simplified model is sufficient to demonstrate the key features of the model via effective illustrations. The purpose of using the simplified model is to present the model implementation which, of course, could be expanded as needed.

Let Θ represent all parameters of the model, and $\mathbf{D}_T = \{(\mathbf{x}_t, \mathbf{n}_t, \mathbf{M}_t), t = 1, \dots, T\}$ represent all the data available. Assuming independence in time, age class and region, the likelihood function is given by

$$\begin{aligned} l(\Theta|\mathbf{D}_T) &\propto \prod_{t,i,a=1}^{T,I,A} f(x_{t,i,a}|\theta_a, n_{t,i,a}) f(n_{t,i,a}|\lambda_a, M_{t,i,a}) f(M_{t,i,a}) \\ &\propto \prod_{t,i,a=1}^{T,I,A} (\lambda_a \theta_a)^{n_{t,i,a}} \sqrt{\tau} \cdot e^{-(\theta_a x_{t,i,a} + \lambda_a M_{t,i,a} + \frac{\tau}{2}(M_{t,i,a} - \mu_{t,i,a})^2)}, \end{aligned}$$

where the product is over time $t = 1, 2, \dots, T$, region $i = 1, 2, \dots, I$ and age group $a = 1, 2, \dots, A$.

The full conditional posteriors of the model for Θ are required in order to implement the Gibbs sampler algorithm introduced in Section 2.3. The full conditional posterior distributions of λ_a and θ_a can be obtained in closed form as follows (the derivations are given in Appendix A):

$$\begin{aligned} \lambda_a | \Theta_{-\lambda_a}, \mathbf{D}_T &\sim \text{Gamma} \left(\alpha_\lambda + \sum_{t,i=1}^{T,I} n_{t,i,a}, \beta_\lambda + \sum_{t,i=1}^{T,I} M_{t,i,a} \right), \quad a = 1, 2, \dots, A, \\ \theta_a | \Theta_{-\theta_a}, \mathbf{D}_T &\sim \text{Gamma} \left(\alpha_\theta + \sum_{t,i=1}^{T,I} n_{t,i,a}, \beta_\theta + \sum_{t,i=1}^{T,I} x_{t,i,a} \right), \quad a = 1, 2, \dots, A. \end{aligned}$$

The full conditional distributions of β_0 , β_1 and ε_a^0 are also available in closed form, respectively, given by

$$\beta_0 | \Theta_{-\beta_0}, \mathbf{D}_T \sim N(\text{Mean}_{\beta_0}, (\text{Precision}_{\beta_0})^{-1}), \quad (2.10)$$

where

$$\begin{aligned} \text{Mean}_{\beta_0} &= \frac{\tau_0 \mu_0 + \tau \sum_{t,i,a=1}^{T,I,A} b_{t,i,a}}{\text{Precision}_{\beta_0}}, \\ \text{Precision}_{\beta_0} &= \tau_0 + \tau TIA, \\ b_{t,i,a} &= M_{t,i,a} - \varepsilon_a^0 - L_i - \beta_1 e^{t(\beta_2 + \varepsilon_a^2)}, \\ \beta_1 | \Theta_{-\beta_1}, \mathbf{D}_T &\sim N(\text{Mean}_{\beta_1}, (\text{Precision}_{\beta_1})^{-1}), \end{aligned}$$

where

$$\begin{aligned} \text{Mean}_{\beta_1} &= \frac{\tau_1 \mu_1 + \tau \sum_{t,i,a=1}^{T,I,A} d_{t,i,a} \cdot e^{t(\beta_2 + \varepsilon_a^2)}}{\text{Precision}_{\beta_1}}, \\ \text{Precision}_{\beta_1} &= \tau_1 + \tau \sum_{t,i,a=1}^{T,I,A} e^{2t(\beta_2 + \varepsilon_a^2)}, \\ d_{t,i,a} &= M_{t,i,a} - \beta_0 - \varepsilon_a^0 - L_i, \\ \varepsilon_a^0 | \Theta_{-\varepsilon_a^0}, \mathbf{D}_T &\sim N(\text{Mean}_{\varepsilon_a^0}, (\text{Precision}_{\varepsilon_a^0})^{-1}), \quad a = 1, 2, \dots, A, \end{aligned}$$

where

$$\begin{aligned} \text{Mean}_{\varepsilon_a^0} &= \frac{\tau \sum_{t,i=1}^{T,I} c_{t,i,a}}{\text{Precision}_{\varepsilon_a^0}}, \\ \text{Precision}_{\varepsilon_a^0} &= \tau_{\varepsilon_0} + \tau TI, \\ c_{t,i,a} &= M_{t,i,a} - \beta_0 - L_i - \beta_1 e^{t(\beta_2 + \varepsilon_a^2)}. \end{aligned}$$

However the full conditional posterior distributions of β_2 and ε_a^2 can not be obtained in closed form due to the nonlinear functions of the parameters; their relevant proportional expressions are given by

$$\begin{aligned} f(\beta_2 | \Theta_{-\beta_2}, \mathbf{D}_T) &\propto \exp \left(-\frac{\tau_2}{2} (\beta_2^2 - 2\beta_2 \mu_2) - \frac{\tau}{2} \sum_{t,i,a=1}^{T,I,A} (\beta_1 e^{t(\beta_2 + \varepsilon_a^2)} - d_{t,i,a})^2 \right), \\ f(\varepsilon_a^2 | \Theta_{-\varepsilon_a^2}, \mathbf{D}_T) &\propto \exp \left(-\frac{\tau_{\varepsilon_2}}{2} (\varepsilon_a^0)^2 - \frac{\tau}{2} \sum_{t,i=1}^{T,I} (\beta_1 e^{t(\beta_2 + \varepsilon_a^2)} - d_{t,i,a})^2 \right), \quad a = 1, 2, \dots, A. \end{aligned}$$

The full conditional posterior distributions of τ , τ_{ε_0} , τ_{ε_2} can be obtained in closed form and they follow Gamma distributions with parameters similar in structure; they are

$$\tau | \Theta_{-\tau}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_\tau + \frac{1}{2} TIA, \beta_\tau + \frac{1}{2} \sum_{t,i,a=1}^{T,I,A} (M_{t,i,a} - \mu_{t,i,a})^2 \right),$$

$$\tau_{\varepsilon_0} | \Theta_{-\tau_{\varepsilon_0}}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_{\tau_{\varepsilon_0}} + \frac{A}{2}, \beta_{\tau_{\varepsilon_0}} + \frac{1}{2} \sum_{a=1}^A (\varepsilon_a^0)^2 \right),$$

$$\tau_{\varepsilon_2} | \Theta_{-\tau_{\varepsilon_2}}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_{\tau_{\varepsilon_2}} + \frac{A}{2}, \beta_{\tau_{\varepsilon_2}} + \frac{1}{2} \sum_{a=1}^A (\varepsilon_a^2)^2 \right).$$

The hyperparameters α_θ and α_λ have the non-closed form conditional posteriors with similar structure whereas β_θ and β_λ have closed form conditional posteriors, given by

$$f(\alpha_\theta | \Theta_{-\alpha_\theta}, \mathbf{D}_T) \propto \alpha_\theta^{\alpha_\theta - 1} e^{-\beta_\theta \alpha_\theta} \cdot \beta_\theta^{A \cdot \alpha_\theta} [\Gamma(\alpha_\theta)]^{-A} \prod_{a=1}^A \theta_a^{\alpha_\theta - 1},$$

$$f(\alpha_\lambda | \Theta_{-\alpha_\lambda}, \mathbf{D}_T) \propto \alpha_\lambda^{\alpha_\lambda - 1} e^{-\beta_\lambda \alpha_\lambda} \cdot \beta_\lambda^{A \cdot \alpha_\lambda} [\Gamma(\alpha_\lambda)]^{-A} \prod_{a=1}^A \lambda_a^{\alpha_\lambda - 1},$$

$$\beta_\theta | \Theta_{-\beta_\theta}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_{\beta_\theta} + \alpha_\theta A, \beta_{\beta_\theta} + \sum_{a=1}^A \theta_a \right),$$

$$\beta_\lambda | \Theta_{-\beta_\lambda}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_{\beta_\lambda} + \alpha_\lambda A, \beta_{\beta_\lambda} + \sum_{a=1}^A \lambda_a \right).$$

We write the variance-covariance matrix of the multivariate normal distribution for spatial variables as

$$\sigma^{-1} \begin{pmatrix} 1 + |\eta| \cdot m_g & -\eta \\ -\eta & 1 + |\eta| \cdot m_g \end{pmatrix}^{-1} = \begin{pmatrix} P & S \\ S & P \end{pmatrix},$$

where

$$P = \frac{\sigma^{-1}(1 + |\eta|m_g)}{(1 + |\eta|m_g)^2 - \eta^2},$$

$$S = \frac{\sigma^{-1}\eta}{(1 + |\eta|m_g)^2 - \eta^2}.$$

Note that in our example, we assumed η to be positive. Since we only study two regions, the number of neighbors of each region is always 1, hence $m_g = 1$. Therefore, in the expression above, P and S can be simplified as

$$P = \frac{\sigma^{-1}(1 + \eta)}{1 + 2\eta},$$

$$S = \frac{\sigma^{-1}\eta}{1 + 2\eta}.$$

The correlation coefficient of this variance-covariance matrix can be expressed as ρ , such that

$$\rho = \frac{S}{P}.$$

Therefore, the conditional posterior distribution for the region related parameters can be obtained in non-closed form,

$$f(L_i|\Theta_{-L_i}, \mathbf{D}_T) \propto \exp\left(-\frac{L_1^2 + L_2^2 - 2\rho L_1 L_2}{2(1 - \rho^2)P^2}\right) \prod_{t,a=1}^{T,A} \exp\left(-\frac{\tau}{2}(M_{t,i,a} - \mu_{t,i,a})^2\right), \quad i = 1, 2,$$

where $\mu_{t,i,a}$ is expressed as (2.6), and

$$f(\eta|\Theta_{-\eta}, \mathbf{D}_T) \propto \frac{1}{(1 + \eta)^2 P \sqrt{1 - \rho^2}} \cdot \exp\left(-\frac{L_1^2 + L_2^2 - 2\rho L_1 L_2}{2P(1 - \rho^2)}\right),$$

$$f(\sigma|\Theta_{-\sigma}, \mathbf{D}_T) \propto \frac{\sigma^{\alpha_\sigma - 1} e^{-\sigma \beta_\sigma}}{P^2} \cdot \exp\left(-\frac{L_1^2 + L_2^2 - 2\rho L_1 L_2}{2P(1 - \rho^2)}\right).$$

At this point all the theoretical information of MCMC implementation is available. This model is capable of analyzing the implied parameters of the data set given the history of claim severity, amount and insured population. The model borrows strength from region and the age class enhancing the estimation of parameters. The parameters then can be applied to make predictions about future claims and insured population. WinBUGS (Lunn et al., 2000, Cowles 2004 and Ntzoufras 2011) or OpenBUGS (Spiegelhalter et al., 2007) can be used to implement the model with ease in practice. In the following chapters, a numerical example is presented with detailed illustration.

Chapter 3

Simulation Studies and Model Fitting

As presented in the previous chapter, the hierarchical Bayesian model requires data input of the historical number of claims, total claim amount and the corresponding insured population, which is commonly available for health insurance companies, to estimate parameters and hence make predictions for future claims. At this stage it would be of more interest to test the effectiveness of this model to ensure its capability of generating the correct parameters implied by the data. To serve this purpose, this chapter is organized in two main parts: simulating claim data with predetermined parameters, and fitting the model to see if it can capture the predetermined parameters.

3.1 Simulation Studies

The aim of this section is to present the process of simulating claim history data, that is the claim severity, claim frequency and the insured population. As long as the parameters are predetermined, the simulation can be carried out in R with "actuar" (Dutang et al., 2008) and other supplementary packages. The following discusses how the parameters are determined in order to carry out the simulation. The parameters are chosen mainly for the purpose of testing the effectiveness of this model. The priority for this section is to select some values for parameters that can pursue this purpose with simple values for execution, without losing too much generality. Meanwhile, hopefully, the selected value could contain as much practical meaning so that it helps to comprehend the elements of the model.

When considering claim frequency simulation, one can apply practical knowledge or experience in the assumptions for parameters. For example, the number of doctor's visits

for a type of medical treatment could be on average about 0.1 per time unit for age class 20-30 (say, age class 1), increasing to about 0.4 for age class 40-50 (say, age class 3). Hence the insurance company could anticipate on average 10% of the policyholders aged 20-30 as well as 40% of those aged 40-50 to report a claim. If we recall (2.5) in the previous chapter, the λ_a in that expression represents the claim made per person per unit time contributing to the average claim frequency. Consequently, one reasonable choice is to set the mean of λ_1 and λ_3 to be 0.1 and 0.4, respectively. For the sake of simplicity (but may be less likely in reality), it is assumed that the claim frequency is at the same level for all age groups in the simulation, namely $\lambda_a \sim G(40, 200)$ for any $a = 1, 2, \dots, A$. It means that 20% of the insured population would report a claim for each age class, with standard deviation of 3.16%. The variance is intentionally selected to be small under the assumption that the variation of the claim frequency is not excessively large. One has more freedom in this matter and can make other assumptions if a large variation in claim frequency is believed.

The selection of the claim severity parameter can be obtained in a similar fashion. Again consider the two age classes 20-30 (age class 1) and 40-50 (age class 3). Policyholders in older age groups tend to make claims with large amounts. Hence the average claim amount for the members in age class 1 can be assumed to be small, say \$15 per claim, while that of age class 3 may be higher, say \$30 per claim. Again in order to present a simple but effective model testing, in the simulation it is assumed that the average claim amount is about \$25, same among all age classes. According to (2.4), given that $\kappa_a = 1$ for $a = 1, 2, \dots, A$ and θ_a assumed known, the remaining question is the determination of θ_a . Again the variation of the claim severity is assumed small. Therefore in the simulation study, $\theta_a \sim G(400, 10000)$ for any $a = 1, 2, \dots, A$. There is no unique way to select the parameters. These values are chosen so that the model is easy to understand and the process is technically simple to implement.

The simulation of insured population relies on three predominant parameters estimation. In expression (2.6), the population mean is mainly determined by the value of β_j , where $j = 0, 1, 2$. In the simulation a quantitative criterion is assumed such that the mean of the insured population for a given age group and region is to be doubled in 20 time units. Again this criterion is not unique and subject to changes. The average populations at time 0 and 20 are measured by $E(\mu_{0,i,a}) \approx \beta_0 + \beta_1$ and $E(\mu_{20,i,a}) \approx \beta_0 + \beta_1 e^{20\beta_2}$, respectively. The approximate sign indicates that the measures represent estimates rather than the true

means of $\mu_{t,i,a}$. The expression involves exponential terms making the calculation of the true mean complicated. For the concern in simulation stage, such calculations are not required for effective simulations and hence are not necessary. One has freedom in determining the initial population level and only need to ensure the population doubles at time 20. The assumption adopted in this simulation is that the $E(\mu_{0,i,a}) \approx 70$ and $E(\mu_{20,i,a}) \approx 140$. A small increase in the value of β_2 could increase the mean value of population exponentially and hence a small value for β_2 is preferred. Set $\beta_0 = 50$, $\beta_1 = 20$ and $\beta_2 = 0.075$; it can be easily seen that these selected values satisfy the requirements stated above. One is not expected to have much knowledge in terms of the values for the region variables L . In expressions (2.7) and (2.8), the σ is randomly assumed to follow a Gamma distribution with large mean and variance and η follows a Pareto distribution with parameters 1 and 1, assigning high probability to small values.

Table 3.1 summarizes the assumptions, based on which the simulation study can be performed in R with auxiliary packages. The completed simulated data (regions 1 and 2) is give in Appendix B.

Table 3.1: True Values/Distributions of the Parameters

| Claim Frequency Parameters | Claim Severity Parameters | Population Parameters |
|--|--|--|
| $\lambda_a \sim \text{Gamma}(40, 200)$ for any $a = 1, 2, \dots, 7$ | $\kappa_a = 1$ $\theta_a \sim \text{Gamma}(400, 10000)$ for any $a = 1, 2, \dots, 7$ | $\beta_0 = 50$ $\beta_1 = 20$ $\beta_2 = 0.075$ $\eta \sim \text{Pareto}(1, 1)$ |

3.2 Model Fitting

3.2.1 Prior Elicitation

From this point on, the simulated data (Appendix B) is regarded as the observed data set and applied as input for the hierarchical model introduced in Chapter 2. Thus the information used in simulation process is not applied in the model fitting, that is, one is not supposed to know very much about the parameters of the data.

Once the data is available, there is only one step remaining before model application, that is the prior elicitation. As stated in Chapter 2, the Bayesian framework allows prior

belief and knowledge to play a role in the calculation. If little knowledge is available, reference priors carrying default information about the parameters can be used. In this case, the historical data with rich information can be regarded as a guideline that facilitate prior elicitation. Empirical Bayes is often selected as a method of using data to obtain priors.

The overall purpose of this chapter is to test if the model can catch the parameters incorporated in the data set. Selecting priors concentrated near the true value of parameters would be meaningless in conducting such effectiveness test. Therefore, it makes more sense, for the testing purpose, to choose as vague priors as possible. That is the general rule for the following prior discussion. Note that this rule is not in line with the insurers' intention, who would like to obtain the underlying parameters from the data with high accuracy. To serve the insurer's intention, the historical data can be examined and studied in great detail such that the prior¹ is a good estimate of the true value. For testing purposes, we do not present it in this project.

Bearing all of the above in mind, let us look at the insured population data. After 1 time unit, the average insured population across all age classes for regions 1 and 2 are about 68 and 75, respectively. A brave guess of the initial insured population could be some number in between, say 70. This implies that $\beta_0 + \beta_1 \approx 70$. Similarly the average population after 20 time units are 142 and 141 for regions 1 and 2. The guess could be about 140 and $\beta_0 + \beta_1 e^{20\beta_2} \approx 140$. Two equations are not enough to estimate 3 parameters. As high accuracy is not the priority at this stage, a wild guess can be made such that those conditions are satisfied, say $\beta_0 \approx 30$, $\beta_1 \approx 40$ and $\beta_2 \approx 0.05$. Furthermore, the variance of the distribution of β 's are assumed large, indicating less confidence of the true value of β 's and thus allowing the model to find the true value with great freedom. The priors are made as follows:

$$\begin{aligned}\beta_0 &\sim N(30, 10^6), \\ \beta_1 &\sim N(40, 10^6), \\ \beta_2 &\sim N(0.05, 10^2).\end{aligned}$$

¹Practical recommendation in setting priors in the GEGM class can be found in Migon and Gamerman (1993).

The priors for all other hyperparameters are reference priors that contain little information, and are given as follows:

$$\begin{aligned}\tau, \alpha_\lambda, \beta_\lambda, \alpha_\theta, \beta_\theta &\sim \text{Gamma}(0.001, 0.001), \\ \tau_{\varepsilon_0} &\sim \text{Gamma}(1, 10000), \\ \tau_{\varepsilon_2} &\sim \text{Gamma}(1, 100), \\ \sigma &\sim \text{Gamma}(1, 0.005), \\ \eta &\sim \text{Pareto}(1, 1).\end{aligned}$$

The priors for $\tau, \alpha_\lambda, \beta_\lambda, \alpha_\theta, \beta_\theta$ are set with large variance; τ_{ε_0} and τ_{ε_2} are set with small means such that the ε_a^0 and ε_a^2 have large variance and wide range of possible values to take; a two-parameter Pareto distribution allows η to take small values with high probability; σ with large mean contributes towards the region factor.

It is now time to implement the model using OpenBUGS (or WinBUGS) to obtain the posteriors of these parameters. The output is presented in the next subsection.

3.2.2 Illustrative Results

The OpenBUGS program can be called from R using the package "R2OpenBUGS" (Gelman et al., 2011 and Sturtz et al., 2005). We present two types of graphs, namely the history plot (or called trace plot) and the kernel density plot for the sampling parameters for each age class. Random initial values are given for the parameters to initiate the program. Three parallel chains are generated with 100,000 observations, each including burn-in of 65,000 iterations and the remaining 35,000 iterations for graphical presentation.

Burn-in refers to the operation of discarding an initial portion of a Markov chain sample so as to minimize the effect of initial values on the posterior inference. In theory, after infinite many runs of the Markov chain the effect of the initial values would vanish. However, it is practically inefficient and time consuming to reach infinitely many runs. Therefore, it is assumed that after a number of iterations (such number is the burn-in number), the chain would reach its target distribution. The early iterations are thrown away and the remaining samples are used for posterior inference.

The program allows each trace to start with different initial values so as to verify whether the sequences would eventually mix together. Every chain is superimposed on the same

plot. The overlapping of the chains represents a sign of convergence whereas an identifiable divergence in the trace indicates less confidence in convergence.

The analysis of the density plots allows us to identify potential departures from convergence such as the presence of multiple modes. The density plot format depends on whether the specified variable is discrete or continuous; if the variable is discrete then a histogram is produced whereas if it is continuous a kernel density estimate is produced instead. In statistics, kernel density estimation (KDE) is a non-parametric approach to estimate the probability density function of a random variable. Kernel density estimation provides a data smoothing solution based on a finite data sample. The extent of smoothness is determined by a free parameter called bandwidth. It is an index that strongly influences the estimated result. A large value of the bandwidth indicates strong force of smoothness being imposed, which could be exposed to the risk of oversmoothed curve obscuring the underlying structure. On the other hand, the undersmoothed curve with a small value of the bandwidth may contain too much redundant information and is not helpful towards decision making. The OpenBUGS program automatically selects the optimal bandwidth when graphing the density estimate. One can obtain useful information from the bandwidth chosen by the program under the optimal estimation. Further information about the kernel density estimation and relative topics can be found in literature such as Rosenblatt (1956) and Park and Marron (1990).

As stated in the previous section, priors are vaguely chosen with large possibility of taking a wide range of values. Consequently, the sampling values do not show signs of convergence until thousands of iterations. The traces of the three chains start close to the prior means for a long while but eventually intertwine and converge to the true values of parameters. The advantage is that the model can identify the correct values even with less informative priors, while the downside is the time consumption, implying the importance of good prior estimation. Priors with more accurate information would significantly reduce the calculation time and speed convergence. When the traces are not converging, it is possible that multiple modes would emerge in the posterior density functions. The convergence can be informally checked by the trace plot.

Figure 3.1 presents the trace and posterior distribution of β_{a_0} for $a = 1, 2, \dots, 7$. The square parentheses in the graph titles indicate the age groups. The vertical axis of the trace plot represents the value of each sample and the horizontal axis represents the number of

iterations. According to the output, the means of β_{a_0} , which are centered at about 50, do not differentiate by age groups. That is consistent with the assumption of simulation process. However, each age class is subject to the variation of the random error and hence the seven graphs are not exactly the same. Each trace cyclically alternates up and down and the average lines of the three traces superimpose, which is a strong indication of convergence. There are samples oscillating away from the average lines, for example, at about 80,000 and 100,000, dropping towards 20. However soon the trace reverts to its mean.

The density plots reveal further differences among age groups. "N=35000" represents the number of iteration recorded in the plot. The bandwidths for β_{a_0} 's are between 0.73 and 0.79, implying that the density plots are not heavily smoothed and yet display intuitive information. Age group 1 has the smallest mean and group 5 has the largest, but the overall means locate at around 50. Both age groups 1 and 5 show slightly double modes. The double modes are not far from each other and the curve is smooth with clear contour. It is deemed as a strong evidence of convergence. Note that the curve is left skewed, with a number of samples located at somewhere between 15 to 30. The skewness reflects a distinctive distance between the prior knowledge and the true underlying values.

Likewise, Figure 3.2 for β_{a_2} , for $a = 1, 2, \dots, 7$, shows a similar situation. On the trace plot the averages of the iterations locate at about 0.08, which is close enough to the true value of 0.075. The iterations are oscillating at about the average line except a visible spike reached 0.12 at around 83,000 iterations. The abnormal values soon vanish and the trace reverts to its mean. Differences can be further detected from the posterior distributions. The bandwidths are approximately 0.0012, even smaller than that of β_{a_0} 's. It implies that little smoothing has been done for the plot. Age group 1 displays large mean above 0.08 while the averages of other age groups seem between 0.074 and 0.08. The normal bell shape curve is more obvious for β_{a_2} than β_{a_0} .

Similar conclusions can be drawn on β_1 (Figure 3.3). The average of the three traces stabilizes at about 20 after 65,000 iterations. However, there are still a small number of iterations that jump above 40. The bandwidth shown as 0.5726, indicates a weak force towards smoothing. The density plot is bumpy with several humps around the mean value. These features are not unexpected. The exponential function can amplify the effects of even the minute change of parameters (i.e., time and β_{a_2}). Nevertheless, the overall shape of the posterior distribution and the aligned average of the three traces provide convincing

True Values of $\beta_{a_0} \forall a$
 $E(\beta_{a_0}) = \beta_0 = 50$

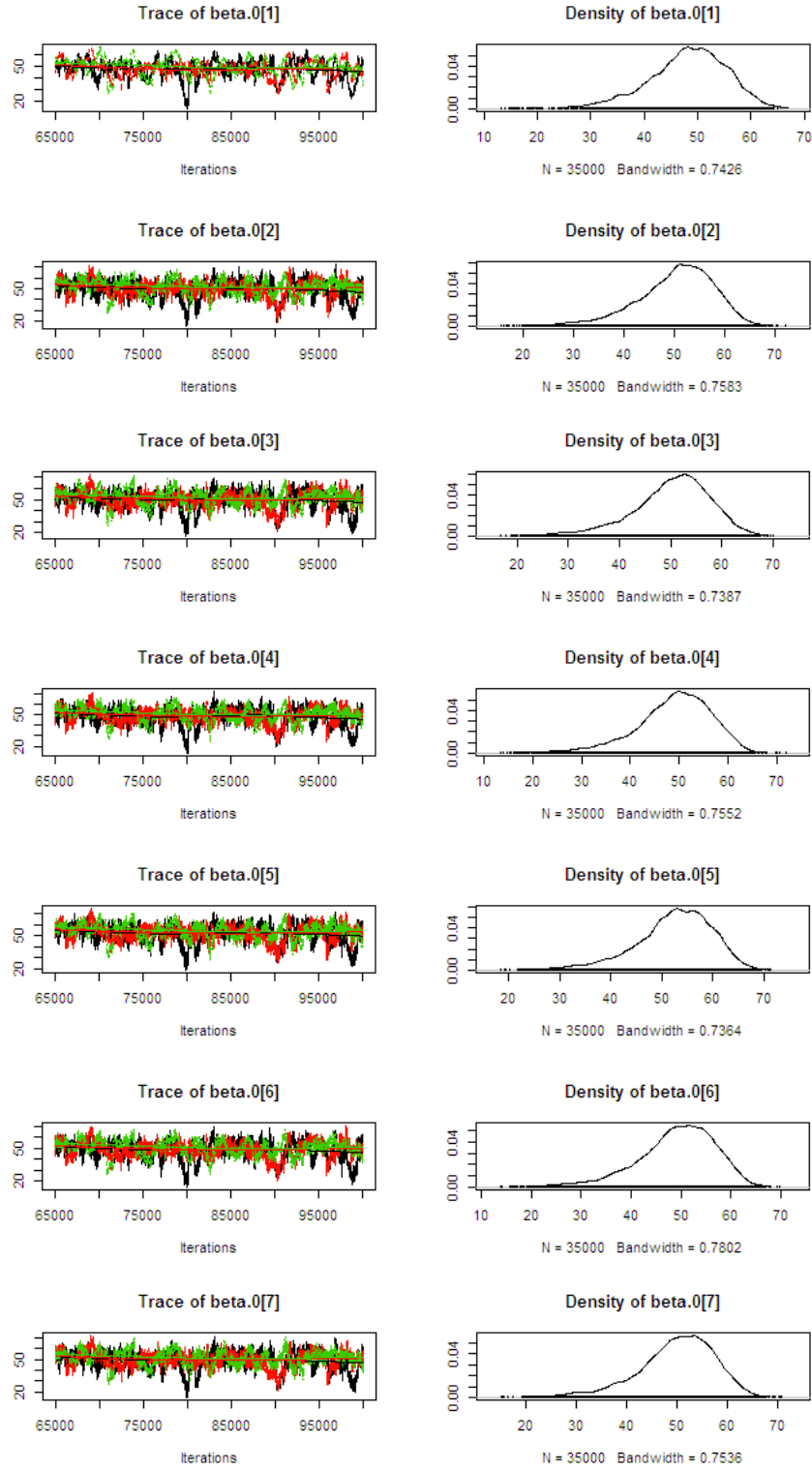


Figure 3.1: Trace Plot and the Posterior Distribution of β_{a_0} , $a = 1, 2, \dots, 7$.

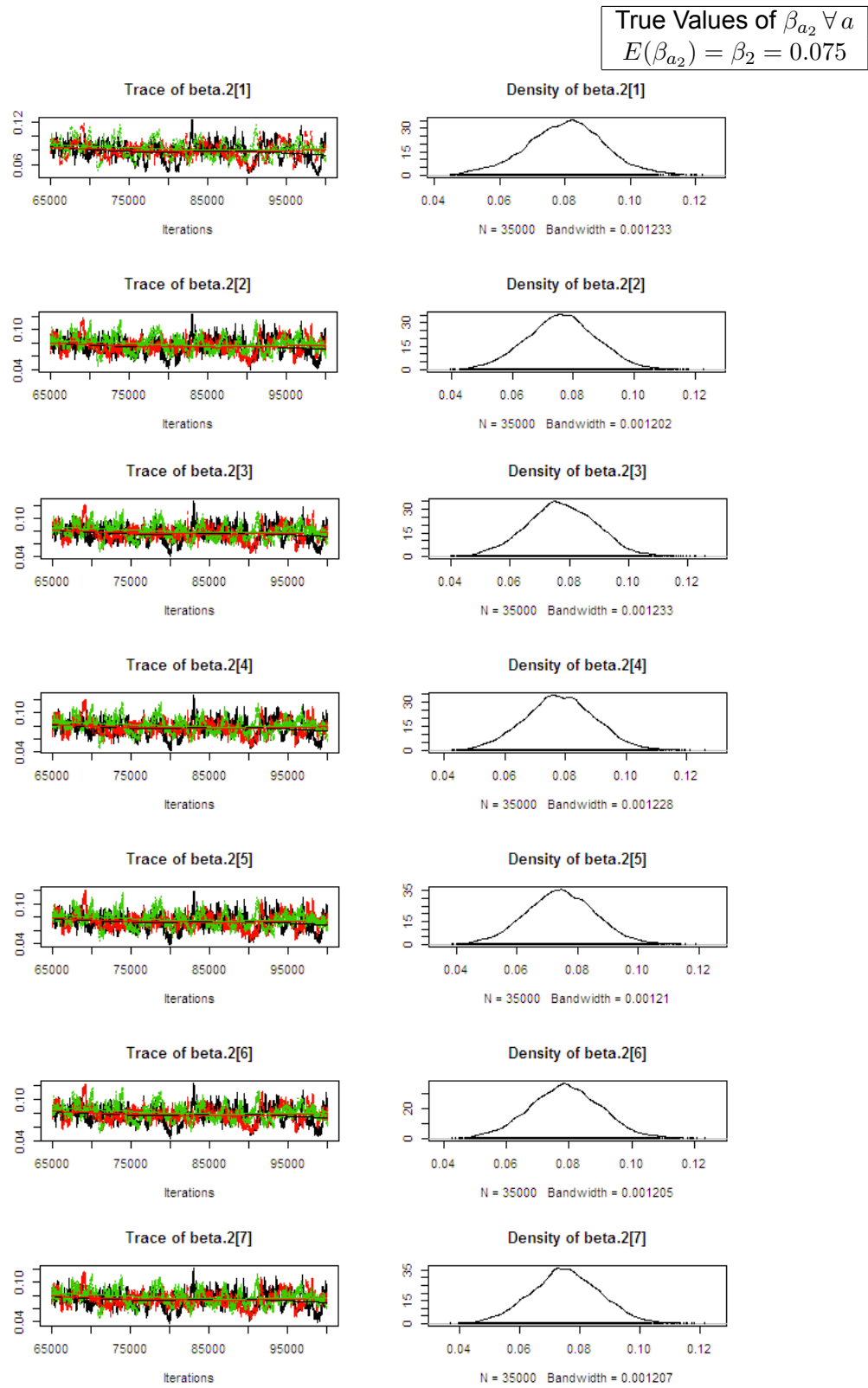


Figure 3.2: Trace Plot and the Posterior Distribution of β_{a_2} , $a = 1, 2, \dots, 7$.

evidence of a strong convergence. The posterior is right skewed with mass on the right tail, due to the distance between prior belief and the true underlying value.

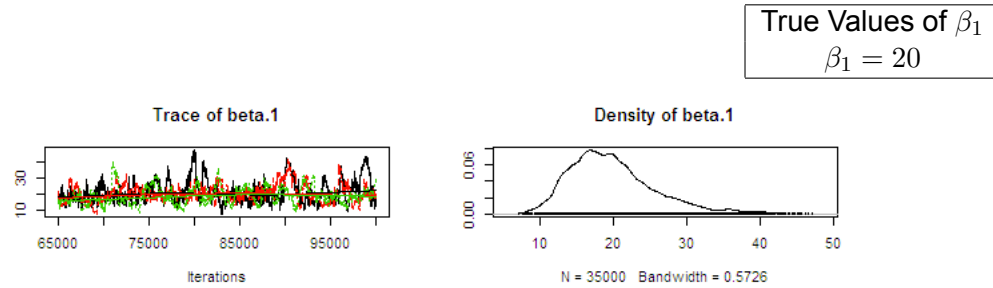


Figure 3.3: Trace Plot and the Posterior Distribution of β_1 .

The output for λ 's and θ 's shows no departure from convergence. The three parallel simulated chains are well superimposed. The small values of bandwidths imply little smooth effect of the plot. It is almost certain that the values of λ 's and θ 's would converge. Age groups 1 and 7 have larger mean value of λ than the others whereas age groups 2 and 5 have larger mean value of θ . The empirical meaning of the values for λ 's and θ 's are further discussed in the next subsection.

It is clear from the output of both region variables that the mean values are centered at zero. The data is obtained under the assumption that little information of region is imposed. The output is consistent with such assumptions. Nevertheless the program continues the attempt to seek for potential values of L 's implied by data. As a result there are occasional spikes in the trace plot and small tails in the density plots.

True Values of $\lambda_a \forall a$
 $E(\lambda_a) = 0.2$

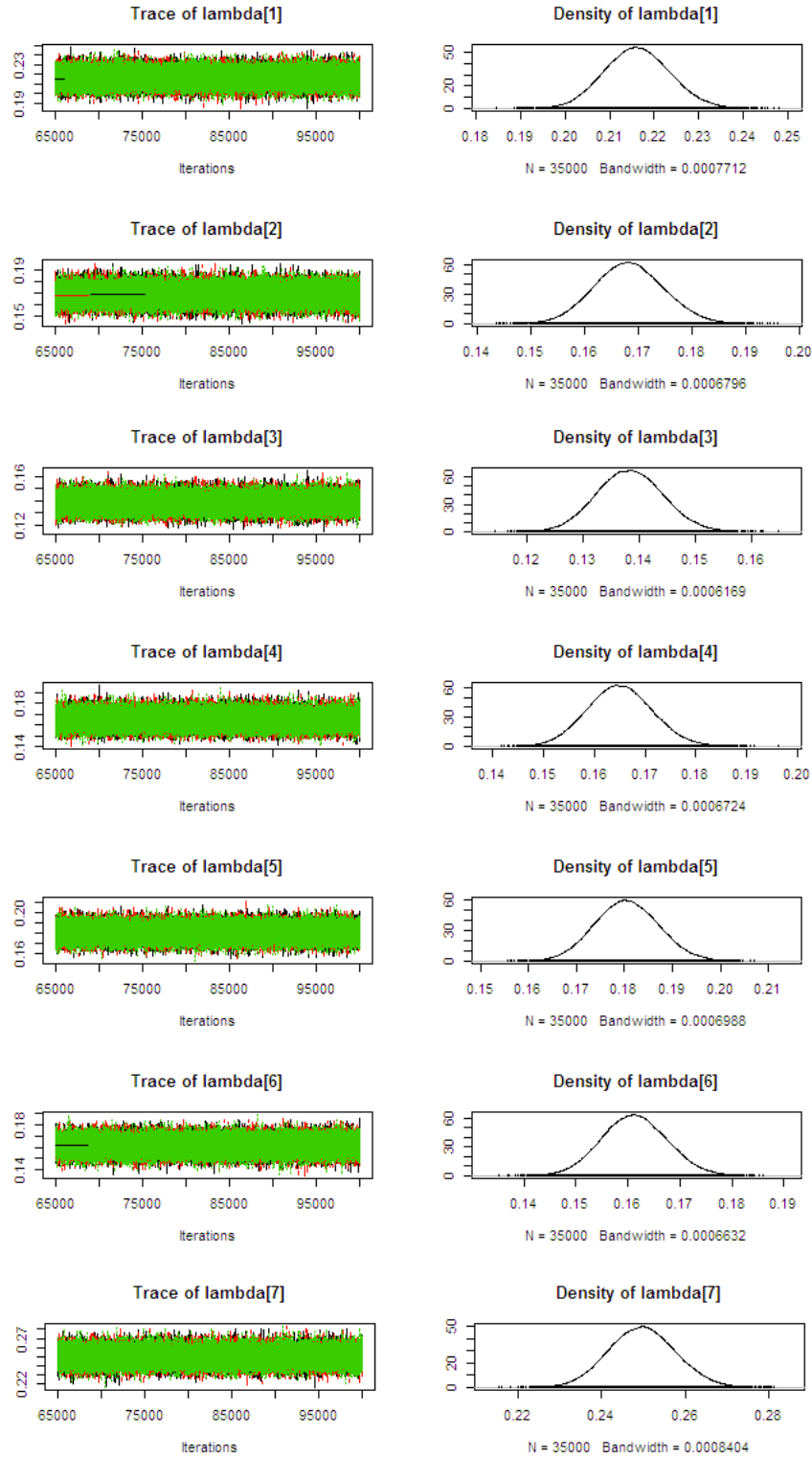


Figure 3.4: Trace Plot and the Posterior Distribution of λ_a for $a = 1, 2, \dots, 7$.
30

True Values of $\theta_a \forall a$
 $E(\theta_a) = 0.04$

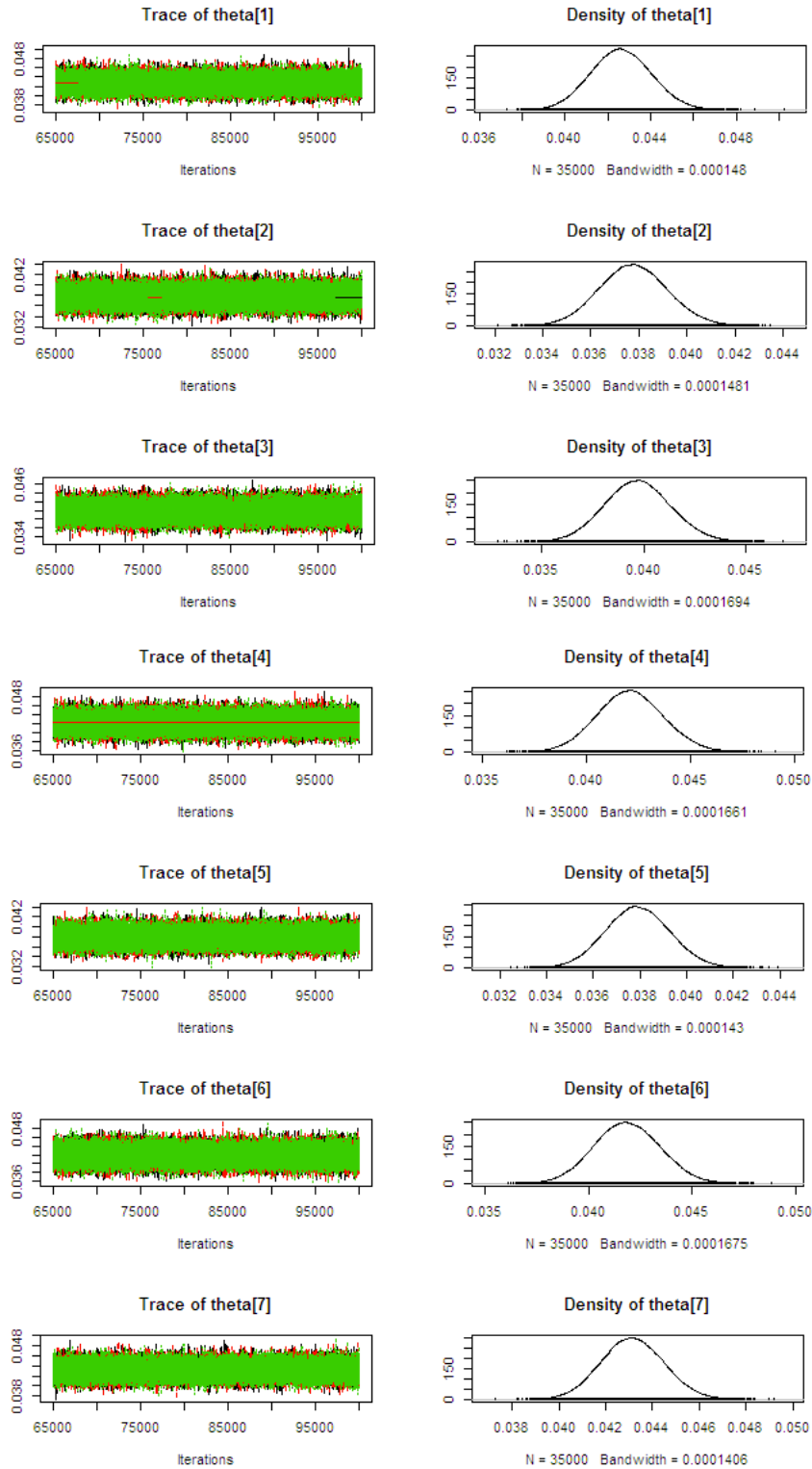


Figure 3.5: Trace Plot and the Posterior Distribution of θ_a for $a = 1, 2, \dots, 7$.

True Values of $L_i \forall i$
 $E(L_i) = 0$

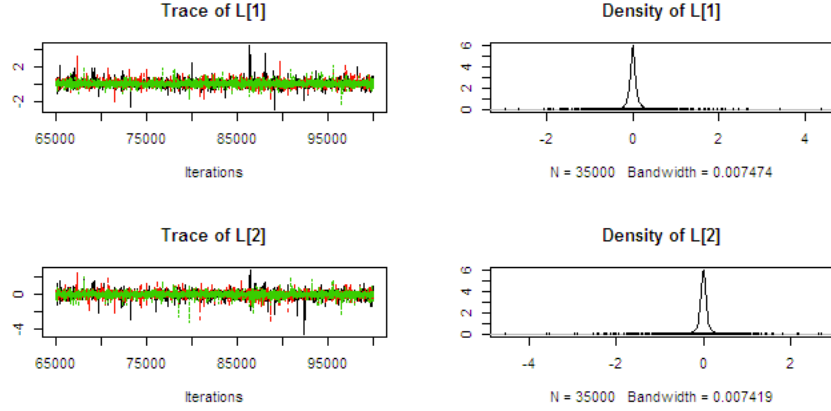


Figure 3.6: Trace Plot and the Posterior Distribution of L_i for $i = 1, 2$.

3.2.3 Credible Intervals of the Posterior Means

The aim of this subsection is to further discuss the posterior means of the age-related parameters along with the corresponding credible intervals.

Credible intervals in Bayesian theory are analogous to confidence intervals in frequentist inference. However, they are philosophically different. In frequentist inference the bounds of confidence intervals are treated as random variables and the parameters as fixed values. In contrast, in Bayesian theory the bounds are regarded as fixed while the parameters are random variables. Further discussion in terms of the difference between the credible intervals and confidence intervals can be found in the literature, such as Lindley (1965) and Jaynes and Kempthorne (1976), among others.

The blue solid line represents the posterior mean for each age group and the red dotted line represents the 2.5% and 97.5% credible intervals. The credible interval graphs provide clearer vision of the difference in parameter values over age groups. In practice this would be of more interest to the management level. In Figure 3.7 values of β_{a_0} are simulated under the same mean; therefore it is expected that the posterior means across age groups do not differ significantly. Similar expectation holds for β_{a_2} in Figure 3.8.

Not surprisingly the simulation of λ_a involves more volatility across age groups. Recall that λ represents the average claim frequency per person per unit of time. In Figure 3.9, age

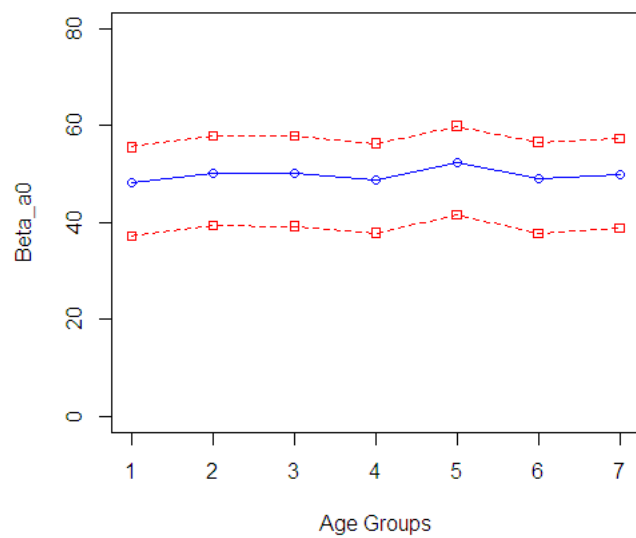


Figure 3.7: Credible Interval of β_{a_0} for $a = 1, 2, \dots, 7$.

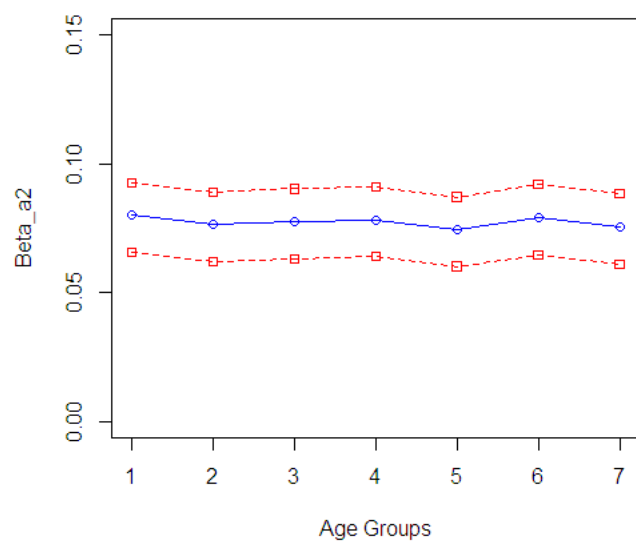


Figure 3.8: Credible Interval of β_{a_2} for $a = 1, 2, \dots, 7$.

groups 1 and 7 show higher average claim frequency among all age groups, with values of 0.22 and 0.25, respectively. Group 3 shows the lowest average claim frequency of 0.14, meaning that about 14% of the insured population in age group 3 would report a claim.

To verify if it is consistent with the data input, Figures 3.10, 3.11 and 3.12 display comparison in the number of reported claims between age groups 7 and 1, 7 and 3, 7 and 6, respectively, from times 1 to 20. The corresponding input can be found in Appendix B. According to Figure 3.10, age groups 1 and 7 show a similar pattern in number of claims in the 20 time units. By comparing the actual claim numbers for each time point, group 1 makes less claims than group 7 most of the time in the 20 time units. It is even clearer from Figure 3.11 and Figure 3.12 that group 7 makes the most number of claims on average over the 20 time units, whereas the difference between groups 3 and 6 is not that obvious. These conclusions are consistent with that drawn from Figure 3.9.

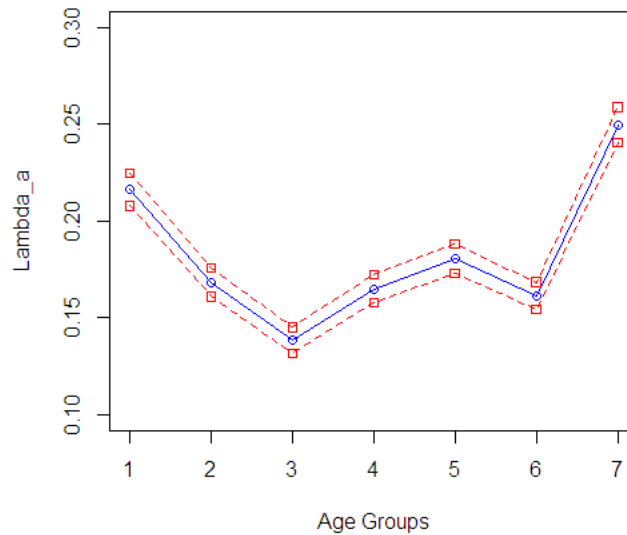


Figure 3.9: Credible Interval of λ_a for $a = 1, 2, \dots, 7$.

Similar analyses can be conducted for θ_a , the inverse of which represents the average claim severity. However, Figure 3.13 does not show too much difference in the claim severity across age groups. Age group 2 shows the average claim severity of $1/0.03777 \approx 26.5$,

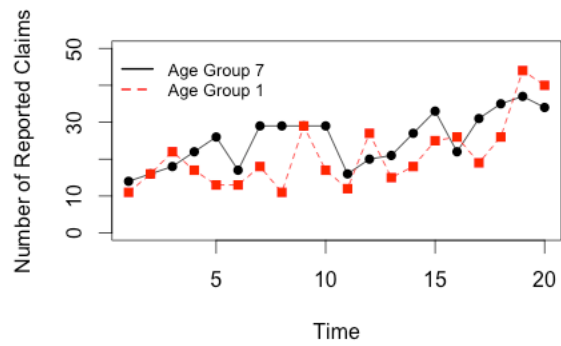


Figure 3.10: Observed Number of Claims, Times 1 to 20, Age Groups 7 and 1.

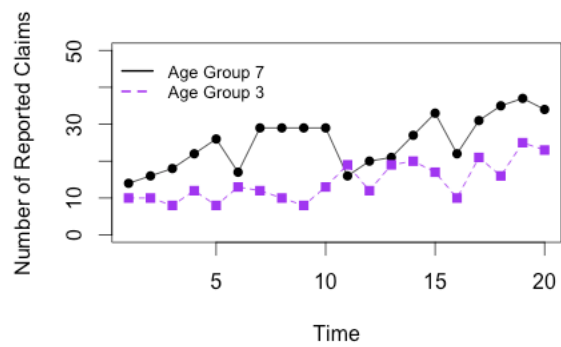


Figure 3.11: Observed Number of Claims, Times 1 to 20, Age Groups 7 and 3.

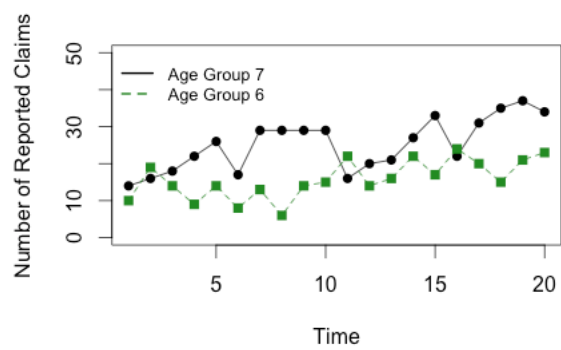


Figure 3.12: Observed Number of Claims, Times 1 to 20, Age Groups 7 and 6.

which is higher than group 5 of $1/0.03791 \approx 26.3$, group 1 of $1/0.04263 \approx 23.5$, but the differences are not significant, as expected.

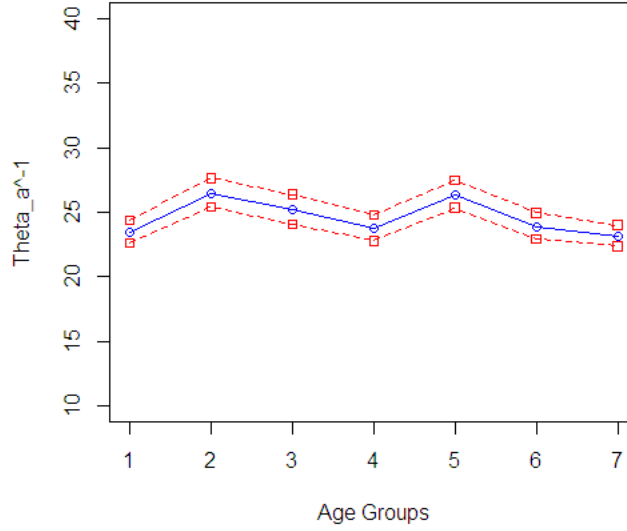


Figure 3.13: Credible Interval of $1/\theta_a$ for $a = 1, 2, \dots, 7$.

By plotting the input data, Figure 3.14 and Figure 3.15 show the comparison in the average of claim severity between age groups 2 and 1, 2 and 5 from time 1 to 20. The differences are too small to be identified easily, which is again consistent with the conclusion from Figure 3.13.

The statistics of posteriors are summarized in Table 3.2. It contains the mean, standard deviation, median and the credible interval bounds for the parameters. Credible interval figures are generated based on these statistics. The two region parameters centered at about 0, which is consistent with the fact of insufficient information in terms of the region effect in this example; β_{0_a} 's have means around 50 and standard deviations in between 7 and 8; β_1 centers at about 20; β_{2_a} 's have means between 0.074 and 0.081 with relatively small standard deviations; the claim frequency index, λ_a 's, reach the highest values at age groups 1 and 7, with mean values above 0.2; the claim severity index, θ_a 's, do not differ a lot across age classes.

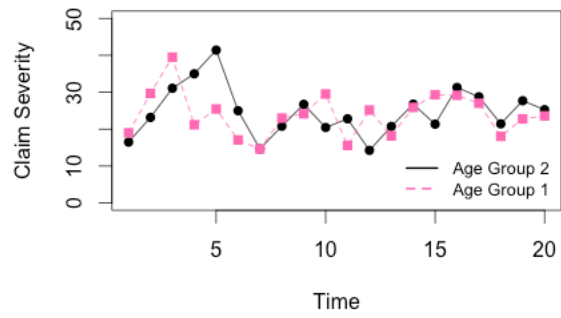


Figure 3.14: Observed Average Claim Severity, Time 1 to 20, Age Groups 2 and 1.

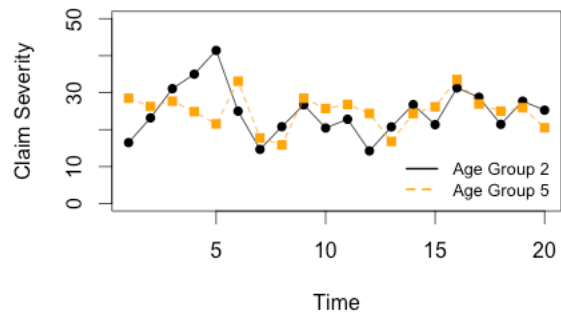


Figure 3.15: Observed Average Claim Severity, Time 1 to 20, Age Groups 2 and 5.

Table 3.2: Statistic Summary of the Major Parameters

| Parameter | Mean | Standard Deviation | 2.5% Percentile | Median | 97.5% Percentile |
|--------------|---------|--------------------|-----------------|---------|------------------|
| L_1 | 0.0074 | 0.142 | -0.2272 | 0.0028 | 0.2726 |
| L_2 | -0.0079 | 0.1474 | -0.2738 | -0.0026 | 0.2298 |
| β_{10} | 48.11 | 7.524 | 38.17 | 48.42 | 60.82 |
| β_{20} | 50.31 | 7.669 | 35.29 | 51.62 | 62.32 |
| β_{30} | 50.23 | 7.666 | 36.16 | 50.75 | 62.73 |
| β_{40} | 48.9 | 7.759 | 30.95 | 49.4 | 61.52 |
| β_{50} | 52.39 | 7.594 | 41.55 | 53.99 | 64.25 |
| β_{60} | 49.07 | 7.885 | 32.89 | 50.34 | 61.73 |
| β_{70} | 49.92 | 7.626 | 36.79 | 51.06 | 61.88 |
| β_1 | 20.03 | 6.086 | 11.73 | 18.93 | 32.08 |
| β_{12} | 0.0803 | 0.0120 | 0.0568 | 0.0803 | 0.1043 |
| β_{22} | 0.0763 | 0.0116 | 0.0549 | 0.0759 | 0.0997 |
| β_{32} | 0.0774 | 0.0118 | 0.0555 | 0.0770 | 0.1009 |
| β_{42} | 0.0783 | 0.0117 | 0.0566 | 0.0778 | 0.1020 |
| β_{52} | 0.0745 | 0.0116 | 0.0530 | 0.0741 | 0.0981 |
| β_{62} | 0.0789 | 0.0116 | 0.0576 | 0.0785 | 0.1023 |
| β_{72} | 0.0753 | 0.0116 | 0.0535 | 0.0749 | 0.0987 |
| λ_1 | 0.2162 | 0.0073 | 0.2018 | 0.2161 | 0.2313 |
| λ_2 | 0.1683 | 0.0065 | 0.1555 | 0.1682 | 0.1816 |
| λ_3 | 0.1384 | 0.0059 | 0.1267 | 0.1383 | 0.1505 |
| λ_4 | 0.1650 | 0.0064 | 0.1523 | 0.1650 | 0.1782 |
| λ_5 | 0.1805 | 0.0067 | 0.1673 | 0.1804 | 0.1942 |
| λ_6 | 0.1613 | 0.0063 | 0.1488 | 0.1612 | 0.1744 |
| λ_7 | 0.2497 | 0.0080 | 0.2340 | 0.2497 | 0.2659 |
| θ_1 | 0.0426 | 0.0014 | 0.0381 | 0.0426 | 0.0472 |
| θ_2 | 0.0378 | 0.0014 | 0.0332 | 0.0378 | 0.0424 |
| θ_3 | 0.0397 | 0.0016 | 0.0350 | 0.0397 | 0.0445 |
| θ_4 | 0.0421 | 0.0016 | 0.0374 | 0.0421 | 0.0468 |
| θ_5 | 0.0379 | 0.0014 | 0.0334 | 0.0379 | 0.0424 |
| θ_6 | 0.0419 | 0.0016 | 0.0371 | 0.0419 | 0.0466 |
| θ_7 | 0.0432 | 0.0013 | 0.0387 | 0.0431 | 0.0476 |

Chapter 4

Predictions and Premium Determination

Once the underlying parameter values of the data have been obtained, a natural question to ask is whether they can be applied for predicting the insured population, number of claims and most importantly the total claim amount for the next time unit. One of the major concerns to the health insurance providers is if there are sufficient premiums retained to cover the losses. Once we can predict the total claim amount, the premium can be determined using different premium principles.

This chapter focuses on the premium setting under different risk measures. Firstly the theoretical Bayesian theory of the prediction algorithm is presented; it is followed by the predictive results for the numerical example presented in Chapter 3; finally it also demonstrates a number of ways to determine the premium based on the predicted claim amounts under certain premium principles.

4.1 Predictions in Bayesian Framework

Making predictions in Bayesian theory has been discussed intensively in the literature. Most commonly the prediction can be made under two circumstances, with or without observations. If the historical observation is not available, the predictive distribution can be expressed as

$$f(X) = \int f(X|\theta) \cdot \pi(\theta) d\theta,$$

which is called the prior predictive density. This prediction is made based on the average of all possible parameters supported by the prior information.

It is also possible to predict the future observation X' based on the observed data X_{obs} , known as the posterior predictive density, given by

$$f(X'|X_{obs}) = \int f(X'|\theta) \cdot \pi(\theta|X_{obs})d\theta.$$

It averages the conditional density of the observations against the posterior knowledge of the predictions. This prediction algorithm is only applicable when observations are available.

As the historical data are most likely available for the insurers, the posterior prediction algorithm would be a better instrument to perform the prediction studies for the purpose of this project. Details are discussed in the next section. Refer to Gelfand (1996) for further knowledge about predictive distributions.

4.2 Predictive Results

Thanks to the OpenBUGS program, it makes prediction in Bayesian inferences straightforward to implement. The values to be predicted, marked as 'NA', along with the observed data (which is the simulated data in this project) are treated as input. Only the observed data determines the values of the parameters, based on which the prediction is made.

Our aim is to obtain the prediction of the total claim amount for the coming period. It is preferred to have a distribution for such predictions so that the insurers are able to determine the premium based on their risk tolerance index measured by, say, standard deviation or the Value at Risk (VaR). In reality, many insurers choose more handy ways to determine premiums, such as establishing manual rates or blended rates. One of the advantages of this model implemented under Bayesian framework is the convenience of obtaining the posterior predictive distributions, which provides valuable information for the experience rating process in practice.

Based on the data of the past 20 time units, the prediction of the 21st time unit in terms of the insured population, total number of claims and total claim amount are presented in Figures 4.1, 4.4 and 4.8, respectively. It is worth mentioning that the program actually allows to make predictions for more than one time unit. This project only presents the prediction and discussion for one time unit. The effectiveness of further predictions remains to be investigated.

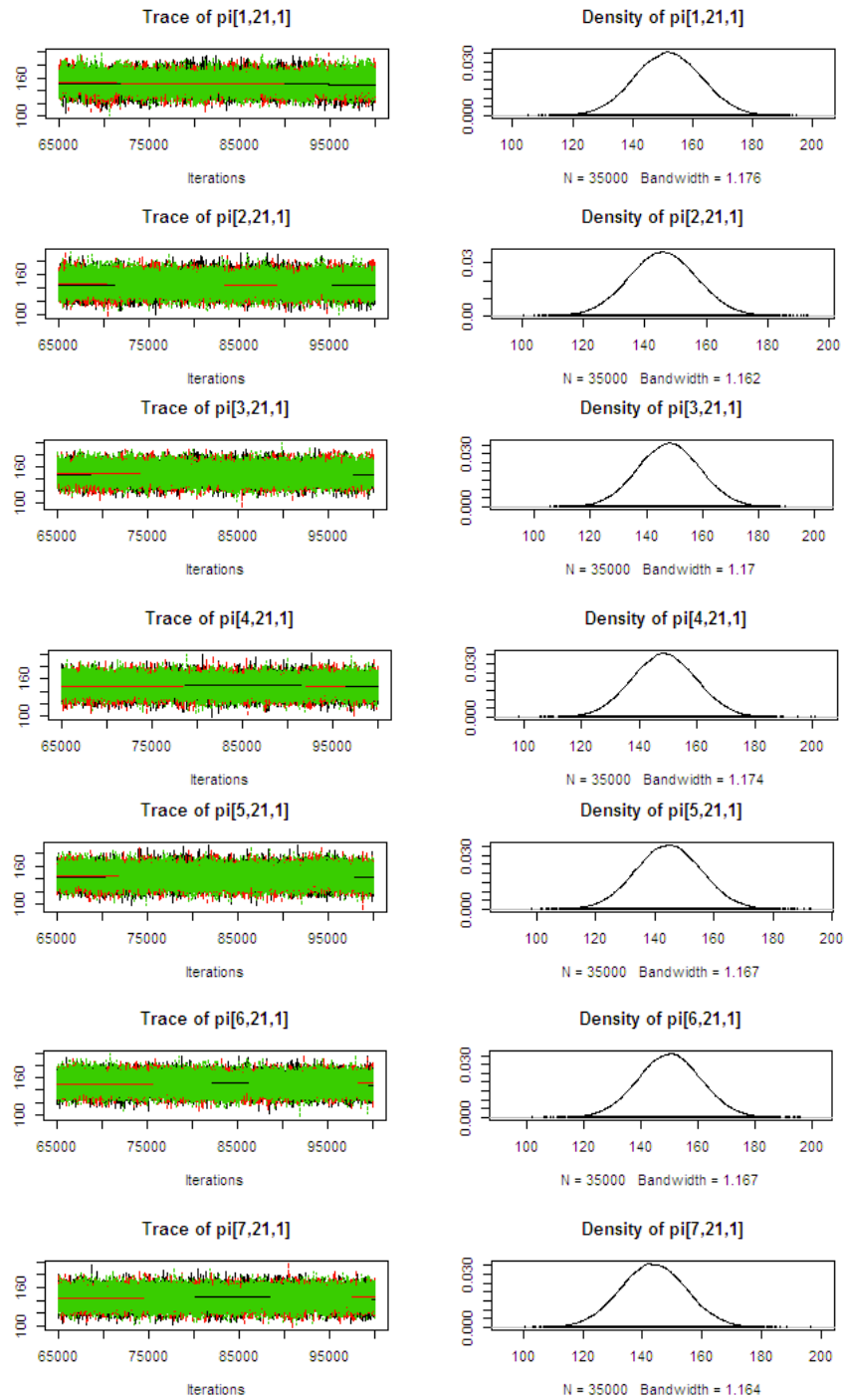


Figure 4.1: Prediction of Insured Population for 21st Time Unit -- Region 1

Figure 4.1 represents the predicted insured population for the 21st time unit by seven age groups in region 1. The predicted insured population in region 2 is not much different from region 1, which is in line with the input data as a result of lacking information, and hence is not presented here. Same as the previous chapter, the trace plot contains three parallel chains each with 35,000 iterations. For each age class, the three chains superimposed one another and the difference is hardly noticeable. As the data for insured population is discrete, it is expected to see a probability mass function of the predictions. In this example the domain of the predicted insured population could take any value varying from more than 100 to less than 200. Slicing the distribution to histogram-like would not be friendly to read. Therefore, it is presented in a continuous format with a probability density function. The superimposed traces are good indications of convergence.

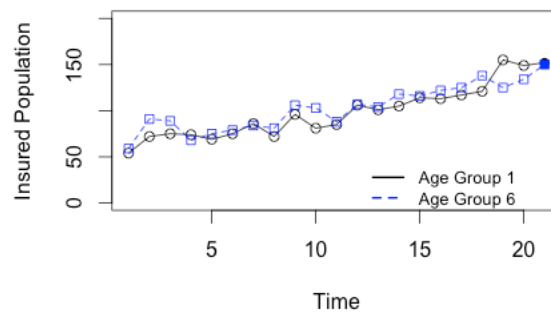


Figure 4.2: Insured Population with Prediction for Age Groups 1 and 6 -- Region 1

The statistical summary is presented in Table 4.1. The first column represents the data type with the subscripts in the order of age, time and region. The other columns list the mean, standard deviation, 2.5% quantile, the mode and the 97.5% quantile of the predicted distribution. According to the table, it is predicted that age groups 1 and 6 would have slightly higher insured population than other groups. Plotting the predicted values along with the historical insured population might give a better idea of whether the predictions are legitimate. Figures 4.2 and 4.3 present a comparison of the mean insured population between age groups 1 and 6, 1 and 5, respectively. The open dots are the historical data of the mean insured population while the filled dots represents the average predicted value for the 21st time unit. In both figures the mean predicted values lie in line with the historical data

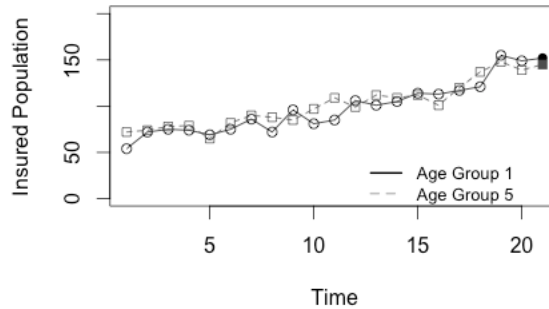


Figure 4.3: Insured Population with Prediction for Age Groups 1 and 5 -- Region 1

retaining the existing curvature. The past 20 time units of data reveal a steadily increasing trend that doubles the initial population. The average predicted values tend to increase if values in previous time units are relatively small; if the historical values are large, the average predicted values tend to decrease. In general, the predicted insured population reflects the features of the population growth and hence can be regarded as a reasonable estimation.

Figure 4.4 displays the posterior prediction for the number of claims for the 21st time unit in region 1. For each of the age groups the trace plots show strong evidence of convergence due to undistinguished superimposed iterations. The density plots are displayed in discrete manner, as the number of claims are discrete with a small domain. The distribution for each age class has a symmetrical bell shape but the average number of claims differ a lot. Age groups 1 and 7 have the average claims exceed 30 whereas age group 3 merely above 20. Detailed statistics are presented in Table 4.1. An interesting feature is that the age classes with higher predictive averages tend to have higher standard deviation than other age classes, indicating more volatility in the number of claims for the recorded time units.

The average predicted numbers of claims for the 21st time unit are placed along with the existing values for the past 20 time units. Figure 4.5 compares age group 7 with group 1, with the filled dots being the predicted mean. Note that the number of claims for age group 1 increases by a large amount from time unit 18 to 19 and decreases at time 20. Therefore, in time unit 21 it is expected that the average number of claims continue to fall, with a smaller predicted value than the previous time unit. Meanwhile group 7 predicts to grow steadily

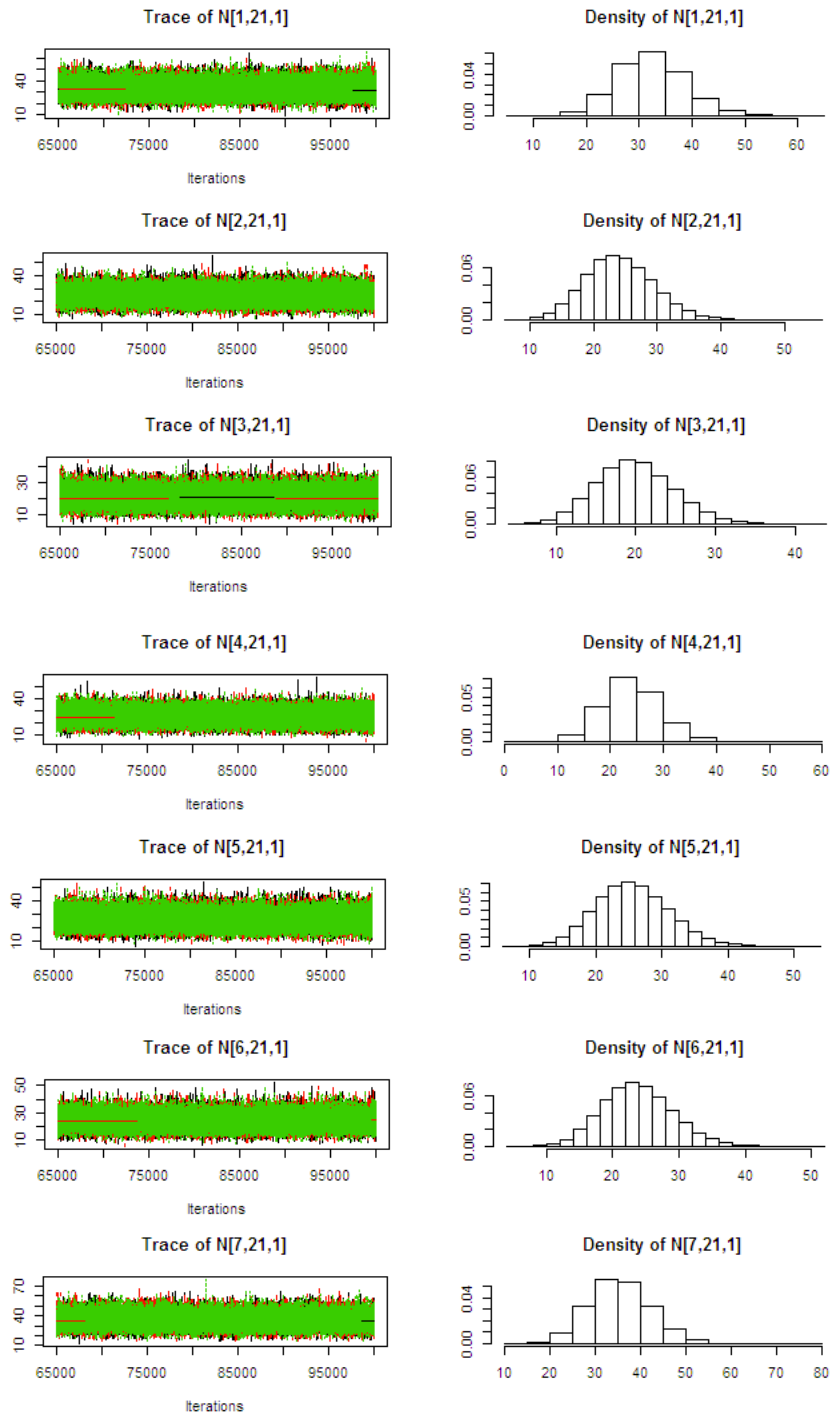


Figure 4.4: Prediction of Number of Claims for 21st Time Unit -- Region 1

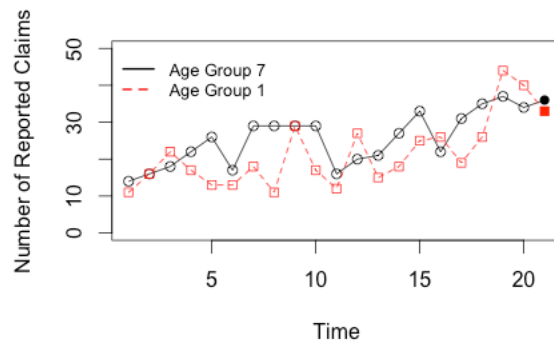


Figure 4.5: Claim Number with Prediction -- Age Groups 7 and 1

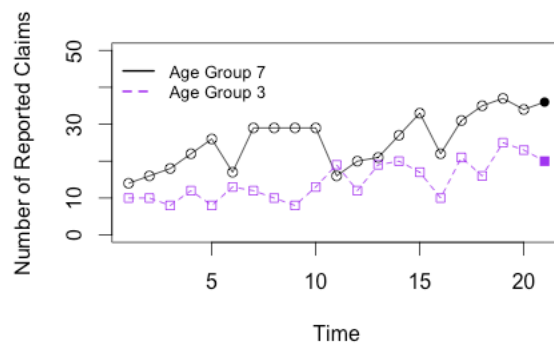


Figure 4.6: Claim Number with Prediction -- Age Groups 7 and 3

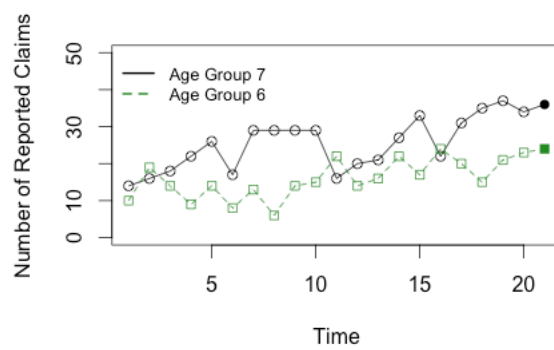


Figure 4.7: Claim Number with Prediction -- Age Groups 7 and 6

with a slight increase in claim number than that in the last time unit. In Figures 4.6 and 4.7, both age groups 3 and 6 present the average predicted values consistent with the existing trend of increase. For all age groups, the claim numbers continue the overall increasing pattern with minor corrections. Furthermore the average predicted values are consistent with the analysis of the parameter values of the λ 's. Based on Figure 3.9 in the previous chapter, the λ 's representing average claim frequency, have high values for age groups 1 and 7 and hence should in theory produce larger number of claims than those for other groups. Therefore, the average predicted claim numbers satisfy the gleaned knowledge about the claim frequency.

The predicted total claim amounts may be of the most interest from the insurers' point of view. Ultimately the claim amounts, without considering the deductible or coinsurance, indicate the actual liability to the claimants. It directly affects how the insurers determining the reserves for the corresponding time period. Figure 4.8 displays the predicted outcome of the claim amounts for all age groups in region 1 for the 21st time unit. As usual the traces of three chains overlap until they are almost indistinguishable, providing a good evidence of reaching the convergence state. Due to the wide range of values for a claim amount, the program automatically chooses to present continuous density functions with large bandwidth. It means that the smooth effect imposed on density curves is strong.

Note that in the statistics summary Table 4.1, the predicted total claim amounts X across all age groups show high standard deviations compared to the predicted insured population M and the number of claims N , resulting in large coefficients of variation in general. In fact this outcome is not surprising under the model design. One of the key features of a hierarchical Bayesian model is the feasibility to incorporate known information at multiple layers. At the same time introducing layers may also bring uncertainty to the parameters. The aggregate total claim amount is derived from the variables representing claim frequency and insured population. As defined previously, each of those variables bear uncertainty to some extent. The accumulation of such uncertainty may result in high volatility for the variables at the top of the hierarchy, in this case, the aggregate claim amount X . Therefore, it is suggested that in the premium calculation, the variance or standard deviation should be taken into account as a representation of the accumulated uncertainty; that is discussed in detail in the next section.

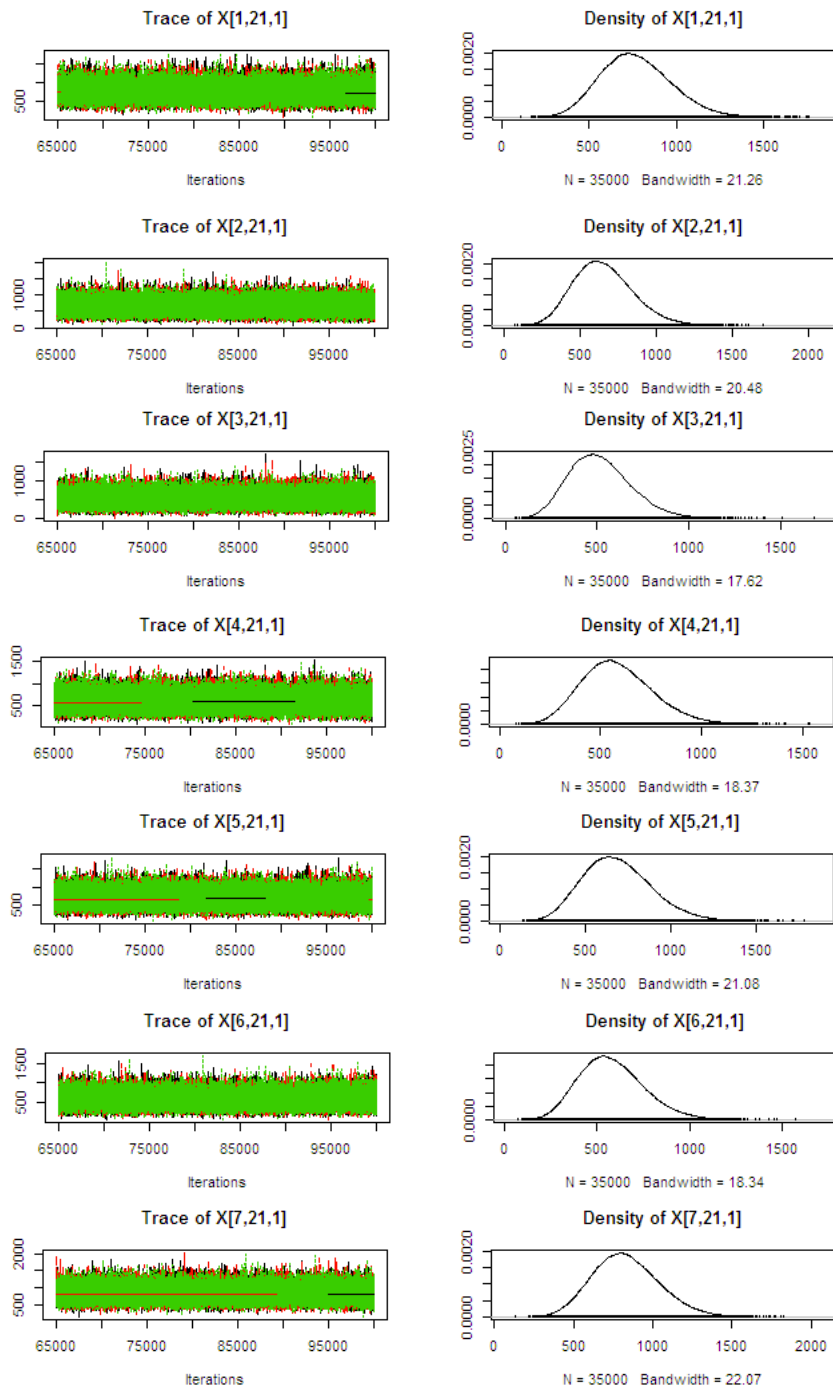


Figure 4.8: Prediction of Total Claim Amount for 21st Time Unit

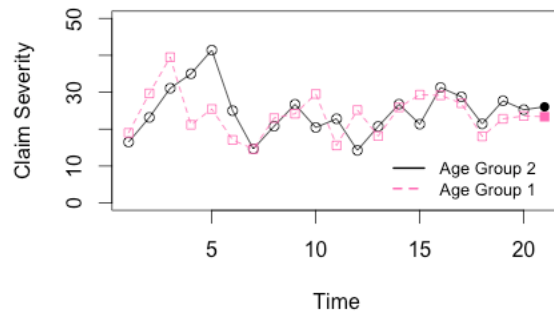


Figure 4.9: Average Amount Per Claim with Prediction in Region 1 -- Age Groups 2 and 1

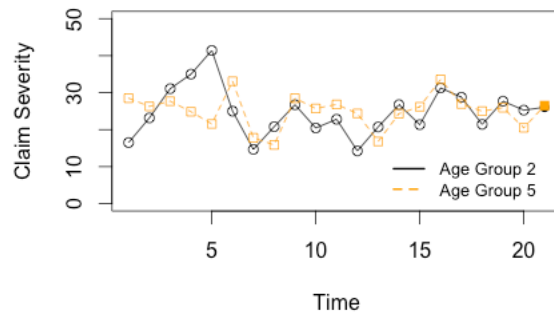


Figure 4.10: Average Amount Per Claim with Prediction in Region 1 -- Age Groups 2 and 5

Figures 4.9 and 4.10 present the predicted average claim severities for the 21st time unit for age groups 2 and 1, 2 and 5, respectively. The average claim severity is calculated by dividing the total claim amounts by the corresponding number of claims for each time unit. The figures are for region 1 only, which contain a lot of resemblance to region 2. The average claim severity for group 2 varies substantially for the first 7 time units and then stabilizes thereafter. The predicted average claim severity for age group 2 reaches about \$26. Likewise, age groups 1 and 5 present a similar pattern in the average claim severity with predicted values between \$23 to \$27. Overall the predictions are consistent with the assumption of \$25 per claim. Over the time units some age groups show a slight increase (or decrease) of the average claim severity, and the variations are within reasonable ranges.

4.3 Premium Determination under Various Premium Principles

In this section we present several approaches to calculate premiums based on the information at hand. Some approaches only require the mean and variance of the predicted variables while others require more details of the predictive distribution such as the percentiles. Each method has its own features and advantages. Some approaches are more conservative, with high premium scheme aiming to cover the extreme claims, while others are moderate making the products competitive in market. The insurers have the freedom to choose the one according to their level of risk tolerance. Further discussion can be found in Goovaerts et al. (2010), Hardy (2006), Laeven and Goovaerts (2008), Young (2004) and Embrechts et al. (1997), among others.

A premium principle, denoted as P , is a function assigning a real number to a random variable. In this project the random variable is the predicted total claim amounts (or the losses) for the coming time unit given the observations over the past 20 time units, denoted as $X|D_T$. To simplify the notation for premium principle illustrations we use X instead to represent the loss random variable.

4.3.1 Net Premium Principle

Among all the premium principles, the net premium principle is one of the commonly applied principles in the literature. It is feasible and simple in application and satisfies many nice properties. The fundamental theory under this principle is that the risk is eventually

Table 4.1: Predicted Insured Population, Claim Number and Total Claim Amount for 21st Time Unit

| Data _{a,t,i} | Mean | Standard Deviation | 2.5% Percentile | Median | 97.5% Percentile |
|-----------------------|-------|--------------------|-----------------|--------|------------------|
| $M_{1,21,1}$ | 151.9 | 11.2 | 130 | 151.9 | 173.9 |
| $M_{1,21,2}$ | 151.8 | 11.2 | 130.4 | 151.7 | 173.8 |
| $M_{2,21,1}$ | 145.8 | 11.07 | 124 | 145.8 | 167.8 |
| $M_{2,21,2}$ | 145.8 | 11.12 | 124 | 145.7 | 167.5 |
| $M_{3,21,1}$ | 148 | 11.15 | 126.2 | 148.1 | 169.8 |
| $M_{3,21,2}$ | 148.1 | 11.13 | 126.5 | 148.1 | 169.7 |
| $M_{4,21,1}$ | 148.5 | 11.18 | 126.6 | 148.5 | 170.9 |
| $M_{4,21,2}$ | 148.6 | 11.15 | 126.9 | 148.5 | 170.4 |
| $M_{5,21,1}$ | 144.6 | 11.12 | 122.8 | 144.6 | 166.4 |
| $M_{5,21,2}$ | 144.6 | 11.13 | 122.7 | 144.6 | 166.4 |
| $M_{6,21,1}$ | 150.1 | 11.12 | 128.3 | 150.1 | 171.9 |
| $M_{6,21,2}$ | 150 | 11.11 | 128.4 | 150 | 172 |
| $M_{7,21,1}$ | 143.6 | 11.09 | 121.9 | 143.6 | 165.3 |
| $M_{7,21,2}$ | 143.6 | 11.1 | 121.9 | 143.6 | 165.5 |
| $N_{1,21,1}$ | 32.83 | 6.326 | 21.88 | 33 | 46.51 |
| $N_{1,21,2}$ | 32.82 | 6.337 | 21 | 33 | 46 |
| $N_{2,21,1}$ | 24.55 | 5.371 | 15 | 24.86 | 36 |
| $N_{2,21,2}$ | 24.55 | 5.386 | 15 | 24.6 | 36 |
| $N_{3,21,1}$ | 20.47 | 4.87 | 12 | 20.83 | 31 |
| $N_{3,21,2}$ | 20.51 | 4.875 | 12 | 20 | 31 |
| $N_{4,21,1}$ | 24.52 | 5.381 | 15 | 24 | 36 |
| $N_{4,21,2}$ | 24.51 | 5.364 | 15 | 24 | 36 |
| $N_{5,21,1}$ | 26.09 | 5.578 | 16 | 26 | 37.09 |
| $N_{5,21,2}$ | 26.1 | 5.571 | 16 | 26 | 38 |
| $N_{6,21,1}$ | 24.2 | 5.313 | 14 | 24 | 35 |
| $N_{6,21,2}$ | 24.22 | 5.328 | 14.18 | 24 | 35 |
| $N_{7,21,1}$ | 35.89 | 6.707 | 23.77 | 36 | 50 |
| $N_{7,21,2}$ | 35.85 | 6.71 | 23 | 36 | 50 |
| $X_{1,21,1}$ | 771.1 | 202.6 | 412.6 | 756.5 | 1201 |
| $X_{1,21,2}$ | 771.8 | 202.3 | 413.4 | 759.6 | 1209 |
| $X_{2,21,1}$ | 651.1 | 195.1 | 312.7 | 634.4 | 1074 |
| $X_{2,21,2}$ | 651.2 | 196 | 313.5 | 635.6 | 1076 |
| $X_{3,21,1}$ | 516.1 | 168.6 | 226.2 | 500.7 | 881.5 |
| $X_{3,21,2}$ | 517.5 | 168.8 | 231.9 | 503.4 | 884.5 |
| $X_{4,21,1}$ | 583.6 | 175.2 | 282.4 | 569 | 964.2 |
| $X_{4,21,2}$ | 583.3 | 175.1 | 280 | 569.8 | 965.1 |
| $X_{5,21,1}$ | 689.2 | 200.8 | 340.3 | 672.9 | 1124 |
| $X_{5,21,2}$ | 690.4 | 202.3 | 338.8 | 673.9 | 1131 |
| $X_{6,21,1}$ | 579.3 | 175.4 | 278.7 | 564.4 | 963.1 |
| $X_{6,21,2}$ | 579.2 | 174.8 | 275.5 | 564.3 | 958.5 |
| $X_{7,21,1}$ | 832.6 | 210.3 | 457.9 | 816.7 | 1279 |
| $X_{7,21,2}$ | 831.3 | 210.4 | 459.5 | 819 | 1281 |

eliminated after selling a great many identical and independently distributed policies. Thus the premium would just to cover the claims only. It does not encompass any load for expenses or profit. This principle is defined as

$$P(X) = E(X).$$

In this project, the net premium for each age group in either region, the premium at time 21 is just the corresponding mean predicted total claim amounts that can be found in Table 4.1.

The advantage of the net premium principle is that it requires the least amount of information from the predicted posterior distribution with a handy calculation process. It is a crude method of providing estimation when there is no sophisticated analysis of the predicted variables. At the same time the disadvantages are too remarkable to be neglected. In reality it is almost impossible to sell infinitely many independent and identical policies. Bearing no risk loading makes the premiums exposed to extreme events and fluctuations such as very large claim amounts. Hence it is not recommended to apply the net premium principle in practice, but to treat it as an estimated measure.

4.3.2 Expected Value Premium Principle

The expected value premium principle, often regarded as the extension of the net premium principle, expresses as

$$P(X) = (1 + \xi)E(X), \quad \xi \geq 0,$$

where ξ is the loading factor. If $\xi = 0$, it is the same as the net premium principle. Clearly the premium under this principle is larger than the expected loss. The difference between the expected loss and the premium can be referred as the premium loading which provides protection against unexpected losses. Furthermore, according to the ruin theory if the loading is not applied, ruin would eventually occur with certainty. The loading factor can be determined based on the risk tolerance level of the insurers. A big value of ξ produces large protection margin while less attraction to the potential buyers. Therefore, it is suggested to pay attention to the loading factor and do constant testing to ensure that the factor is set at a right level.

4.3.3 Variance Premium Principle

The variance premium principle is another further development of the net premium principle. The premium depends not only on the expected value but also the variance of the loss. Unlike the other premium principles, the variance premium principle considers the the variability of the loss; the more variability the loss, the higher the premium. In contrast to the previous case that the risk loading is proportional to the expected loss, here it is proportional to the variance of the loss. The variance premium principle can be expressed as

$$P(X) = E(X) + \omega V(X), \quad \omega \geq 0.$$

Note that this is the same as the net premium principle if $\omega = 0$. Like the expected value premium principle, the insurers have the freedom to determine the risk load based on their risk tolerance. It is more reasonable to have the premium loading related to the variability of the loss. However, since the variance and the expectation have different units (the unit of the variance is the square of that of the expectation), the interpretation of the empirical indication may contain ambiguity.

4.3.4 Standard Deviation Premium Principle

The standard deviation premium principle has the same structure as the variance premium principle, with the variance replaced by the standard deviation of the loss. It is expressed as

$$P(X) = E(X) + \nu \sqrt{V(X)}, \quad \nu \geq 0.$$

Similar to the variance premium principle, it takes the variability of the loss into the premium determination. As the standard deviation and the expectation of the loss share the same unit, it is more convenient to interpret the underlying reasoning of the principle.

It is not the intention in this project to discuss the effectiveness of each premium principle, nor to determine the rational decision of the premium loadings. The aim is to illustrate how to apply the premium principles to the predicted results we have obtained so far. The net premium principle and the expected premium principle require only the expected loss to calculate the premium whereas the variance and standard deviation premium principles require expectation and variance of the loss. Each premium principle has its properties and features. Some would be better used as crude estimation while others can be seen as

legitimate decision for the premium. Further discussion among the premium principles can be found in Bühlmann (1970) and Gerber (1979).

4.3.5 Value-at-Risk

Recently the Value at Risk (VaR) has caught much attention in the insurance and banking industry, which is also known as the quantile risk measure or quantile premium principle. VaR calculation requires to specify a confidence level α . Suppose that the loss random variable X has the cumulative function $F_X(z) = P\{X \leq z\}$. The VaR with confidence level α is defined as

$$\text{VaR}_\alpha(X) = \min\{z | F_X(z) \geq \alpha\}, \quad \alpha \in [0, 1].$$

According to this definition, $\text{VaR}_\alpha(X)$ is a lower α -percentile of the random variable X . The choice of α depends on the risk tolerance level of the insurers, for example α , $0 \leq \alpha \leq 1$, can take the value of 90%, 95% or 99%. Generally the $\text{VaR}_\alpha(X)$ represents the threshold loss level such that with probability of α the loss does not exceed the threshold.

Unlike the premium principles presented previously which require the expectation and variance of the losses, the VaR is essentially measuring the percentile of the loss distribution function, providing a minimum value of the loss based on the confidence level. Setting premium based on the VaR tends to give the insurer more confidence in terms of the percentage losses capable to be covered. However, there are critics emphasizing the inadequacy to provide information for the losses beyond the VaR level. It may happen that the upper tail contains losses with severely large amount. Despite the fact that such kind of severe losses occur with small probability, once they do occur, the size of the losses can be large enough that it exposes the business to illiquidity or even insolvency. This arouses the necessity of considering the extension of VaR measure presented in the next subsection. Further discussion about VaR can be found in Peng (2009) and Hardy (2006), among others.

4.3.6 Tail Value-at-Risk

An alternative risk measure is the so-called tail value-at-risk (TVaR), which is also known as conditional VaR (CVaR), average VaR (AVaR), mean excess loss, mean shortfall, conditional tail expectancy (CTE), expected shortfall (ES), expected tail loss (ETL). TVaR is

chosen to address some of the shortages of VaR measure. Similar to the VaR measure, TVaR is defined based on a confidence level α , $0 \leq \alpha \leq 1$, whose typically values can be 90%, 95% or 99%. As suggested by the name, TVaR accounts for the less profitable outcomes in the tail exceeding VaR. It is sensitive to the shape of the distribution in the tail. It is the expected loss given that the loss belongs to the $(1 - \alpha)$ tail of the loss distribution, that is the conditional expectation of loss X subject to $X \geq \text{VaR}_\alpha(X)$.

Denote α -quantile risk measure as Q_α , the TVaR of the loss X with confidence level $\alpha \in (0, 1)$ can be expressed as (see Hardy 2006)

$$\text{TVaR}_\alpha = E(X|X > Q_\alpha),$$

or in other words,

$$\text{TVaR}_\alpha(X) = \frac{1}{1 - \alpha} \int_{Q_\alpha}^{\infty} z dF_X(z).$$

Note that this formula does not work if there is a probability mass at Q_α . Especially for a discrete loss distribution, there can be more than one outcomes equal to Q_α . In this project we use the following expression to calculate the TVaR

$$\widehat{\text{TVaR}}_\alpha = \frac{1}{N(1 - \alpha)} \sum_{j=N \cdot \alpha + 1}^N X_{(j)},$$

where N is the total sample size, $X_{(j)}$ is the j^{th} smallest values (or j^{th} order statistic) of X and $N(1 - \alpha)$ is assumed to be an integer.

The TVaR has become a very important risk measure in actuarial practice and financial risk management. It is easy to understand and to apply with simulation output. It is worth noting that, since the $\text{TVaR}_\alpha(X)$ is the average loss given that the loss is greater than VaR at confidence level α , the $\text{TVaR}_\alpha(X)$ is not less than $\text{VaR}_\alpha(X)$, providing a conservative risk measure. Many important properties of VaR and TVaR are beyond the scope of this project; refer to Rockafellar and Uryasev (2002), Hardy (2006), Sarykalin et al. (2008) and Peng (2009) for further reading.

4.3.7 Numerical Premium Analysis

As explained in early sections, the predicted distributions about the total claim amounts are available and thus it is not difficult to determine the premium under the premium principles listed above. Table 4.2 displays the premium for 21st time unit in region 1 using the

four premium principles, VaR and TVaR risk measures. The total premium under the net premium principle is just the average predicted total claim amount, same as the second column of Table 4.1. For the expected value premium principle, variance premium principle and standard deviation premium principle, risk loadings of $\xi = 0.5$, $\omega = 0.01$ and $\nu = 1.96$ are assumed, respectively. The risk loadings reflect very subjective opinion towards the coverage of the total claims. Higher loadings provide stronger protection against the claim uncertainty. Due to big value of the variance, the premium under variance premium principle is very sensitive to the value of ω and hence it should take extra caution in applying this premium principle.

For the standard deviation premium principle, the risk loading $\nu = 1.96$ can almost ensure that the company would pay all claims with probability of 97.5%, given that the predicted total claim distributions have approximately the shape of a normal distribution. In our numerical results shown in Figure 4.8, the predicted distributions are slightly right (or positively) skewed with a fat right tail. That means the premium with risk loading $\nu = 1.96$ can cover the total claims with slightly less than 97.5% of chance. This is indeed the case as under the VaR measure with confidence level of 97.5%, the premium is slightly higher than that under the standard deviation premium principle for all age classes. Conditioning on the premium exceeding $\text{VaR}_{97.5\%}$, the $\text{TVaR}_{97.5\%}$ gives even larger premiums.

The aim of Table 4.3 is to show different approaches of calculating premium per policyholder for the next time unit in region 1. Having the total premium and total insured population available, premium per policyholder can be obtained by averaging premium over population with ease. There is no such thing as the best premium design. The purpose of this section is to provide a perspective on determining the premium per policyholder, with no intention to justify which is the best fit. The results for region 2 can be found in Appendix C.

A crude estimation of the premium per policyholder is to use the premium under net premium principle over the mean predicted population for each age class. The premium per policyholder is higher for age classes 1 and 7 but generally between \$3 and \$6. This is consistent with the assumptions in Chapter 3. It is assumed that for every time unit 20% of the population would report claims, each worth about \$25, which indicates that the average cost per policyholder per time unit is about $20\% \times \$25 = \5 . As the simulation study generates higher claim frequency for age groups 1 and 7 (See Figure 3.9), it is reasonable to

Table 4.2: Total Premiums for 21st Time Unit in Region 1 Using Predicted Results

| Total Premium (\$) | Loading | Age Class | | | | | | |
|--------------------------------------|---------|-----------|----------|--------|----------|----------|----------|----------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Net Premium Principle | ----- | 771.10 | 651.10 | 516.10 | 583.60 | 689.20 | 579.30 | 832.60 |
| Expected Value Premium Principle | 0.5 | 1,156.65 | 976.65 | 774.15 | 875.40 | 1,033.80 | 868.95 | 1,248.90 |
| Variance Premium Principle | 0.01 | 1,181.57 | 1,031.74 | 800.36 | 890.55 | 1,092.41 | 886.95 | 1,274.86 |
| Standard Deviation Premium Principle | 1.96 | 1,168.20 | 1,033.50 | 846.56 | 926.99 | 1,082.77 | 923.08 | 1,244.79 |
| VaR _{97.5%} | ----- | 1,201.00 | 1,074.00 | 881.50 | 964.20 | 1,124.00 | 963.10 | 1,279.00 |
| TVaR _{97.5%} | ----- | 1,306.69 | 1,173.16 | 973.44 | 1,052.98 | 1,225.21 | 1,052.15 | 1,380.62 |

Table 4.3: Premium Per Policyholder for 21st Time Unit in Region 1

| Premium Per Person (\$) | | Age Class | | | | | | |
|-------------------------|--------------------------------------|-----------|------|------|------|------|------|-------|
| Premium Measure | Predicted Insured Population Measure | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Net Premium Principle | Average Insured Population | 5.08 | 4.47 | 3.49 | 3.93 | 4.77 | 3.86 | 5.80 |
| VaR _{97.5%} | Average Insured Population | 7.91 | 7.37 | 5.96 | 6.49 | 7.77 | 6.42 | 8.91 |
| VaR _{97.5%} | 2.5 Percentile Insured Population | 9.24 | 8.66 | 6.98 | 7.62 | 9.15 | 7.51 | 10.49 |
| TVaR _{97.5%} | Average Insured Population | 8.60 | 8.05 | 6.58 | 7.09 | 8.47 | 7.01 | 9.61 |
| TVaR _{97.5%} | 2.5 Percentile Insured Population | 10.05 | 9.46 | 7.71 | 8.32 | 9.98 | 8.20 | 11.33 |

anticipate the average cost being higher for these two groups. This premium determination does not involve any loadings and hence can only cover the average cost per policyholder.

Another approach is to average the $\text{VaR}_{97.5\%}$ over the mean insured population. This method provides more certainty that the company can fulfill its obligation. Since the insured population may be volatile, a more conservative measure is to use the 2.5 percentile of the insured population, to protect against overestimation of the population. This results in a higher estimation of premium per person. Under such estimation, age groups 1 and 7 have premiums higher than \$9 and \$10, respectively. Note that age groups 5 and 7 have low predicted insured populations that results in even higher premiums per policyholder.

A similar measure is to replace $\text{VaR}_{97.5\%}$ with $\text{TVaR}_{97.5\%}$ to perform the above calculation. Apparently the premium increases to almost double that under the net premium principle. Using $\text{TVaR}_{97.5\%}$ over 2.5 percentile of the insured population, the average premiums for policyholders in age groups 1 and 7 exceed \$10 and \$11, respectively, which are sufficient to cover most of the extreme claims.

Chapter 5

Conclusion

5.1 Concluding Summary

The main purpose of this project is the prediction of the total claim amount under a hierarchical Bayesian framework. Future claim prediction is a paramount issue that plays a significant role in risk measurement for health insurance providers. According to the work of Migon and Moura (2005), the claim total is related to the number of claims and the insured population for a given time period. Policyholders in different age bands present different patterns in the reported claim frequency and severity. It is hence reasonable to categorize the policyholders by age class when predicting the claim situation in each time unit. Inspired by their work, this project extends the model by introducing one more category to describe the regions of residence for the policyholders. The spatial factor can represent the combined random effect of the elements influencing the claim behavior, such as education, ability to access medical service, wealth level and even weather condition. Each of these elements can potentially influence the claim behavior but modeling each element separately is, if possible, redundant and unnecessary. Hence introducing a spatial factor independent of the existing age classification is practically achievable and comprehensible.

To achieve our goal, we firstly modify the model by Migon and Moura (2005) by adding a region factor to the insured population, arguing that the region factor, the age class and time of measurement together affect the average insured population. An MCMC algorithm enables us to perform posterior estimation given the prior knowledge of the parameters and historical information. Since the full conditional distributions of the parameters are not in closed forms, we utilize the Gibbs sampling algorithm performed by OpenBUGS (or

WinBUGS). In order to test whether the model can effectively detect the true value of the parameters, we do a simulation study on the insured population, number of claims and total claim amounts, for 20 time units, 7 age groups and 2 regions, with some of the parameter values predetermined while others randomly simulated using R and relevant packages. The model indeed can extract the correct value of the parameters. Due to the vague prior we proposed, the posterior distribution suggests that after about 60,000 iterations that the trace plots show signs of convergence. After generating the true parameter values, the predictions are made for the 21st time unit, including insured population, number of claims and claim amounts. Finally based on the predicted claim amounts, the premiums can be calculated under various premium principles. The premium to charge per policyholder can also be obtained easily by dividing the total premiums over the predicted insured population.

5.2 Further Remarks

One of the advantages of this model is that it allows the introduction of prior knowledge in the prediction process. Health insurance providers, after years of experience, most likely possess the professional acumen in the industry. Their opinion as prior information would be much more accurate than a vague prior for the parameters carrying no valuable indications. Better priors consume less time for the process to converge and stabilize. Another advantage is that the prediction can be made for more than one time unit, which may enable the insurers to set reserves for future and adjust some valuation assumptions. OpenBUGS (or WinBUGS) allows the user to generate many predictions with ease. However the accuracy and convergence of the prediction in the distant future can not be guaranteed. The less time unit the prediction made for, the more reliable the result may be assumed.

There are some remarks and suggestions for further development in this topic. In this project the insured population is modeled by a normal distribution with the expected mean following the modified exponential growth curve. In fact one can adjust the population distribution in light of the growth feature of a group. The population mean would then be modeled by other GEGMs with appropriate parameters. For example the Log-normal distribution could serve as an alternative, in which case the mean would follow a Gompertz growth curve. A numerical example of such a combination can be found in Migon and Moura (2005).

Instead of introducing the error terms (e.g., ε_{a_0}) being age related parameters, one can actually further modify the model by using an age related β_{a_0} containing the noise effect and thus the original noise term can be eliminated. The mean of this β_{a_0} can be age related so that it can capture the differences in insured population growth among age groups. For instance, some insurance products are designed targeting a specific age group. As a result, it is expected that the majority of the insured population is contributed by the targeted age group, whereas the insured population for other age groups are relatively small. Moreover, the variable β_1 can be modified to be age related as well, indicating that the multiplier of the exponential growth is also driven by the age discrepancy. This can be applied in the situation where the insured population of a particular age group grows a lot faster than others.

Apparently the population growth pattern varies substantially due to factors such as geography, economics, natural resources or even warfare. There is no unique model that can describe all the growth patterns simultaneously. One of the important questions to ask is that what is the feature of the insured population growth for the cities to be modeled. The population growth model should be in line with the key feature of empirical growth curve. One may notice that the insured population growth model utilized in this project has the potential to encounter over-sized population as time goes by. The mean population (2.6) increases exponentially given positive parameters β_0 , β_1 and β_2 . Refer to Figure 4.2 and Figure 4.3 in Chapter 4, the insured population curves slightly concave up and increasing faster in time. The slope of the curves are not unrealistically steep and the speed of increase is within acceptable limits. Some regions in countries like U.S. and Canada (See Chapter 1 for reference) which show rapid demographic growth may be suitable for this type of model rather than countries with stable or even decreasing demographic growth. The latter situation can be modeled using other set of parameters.

In this project, the simulation study facilitates the test of effectiveness of parameter estimation. It is important to keep in mind that the intention is to provide a procedure for statistical modeling of health insurance related data. The users have the freedom to modify the parameter structure according to the empirical data, which usually is subject to more scenarios than presented in this project. Then the users might be able to detect features from the empirical data, such that the prediction can be more reliable and closer to reality. For example, the age related parameters, such as λ_a for claim frequency and the inverse

of θ_a for claim severity, surely can have different means across age groups. One can customize the mean and variance for each age group based on the past performance. Despite that the age classification is assumed independent of the region factor in this project, one can alter this assumption by introducing correlations, arguing that some cities are more aged than the others.

Furthermore, if empirical data is available (say, for 20 time units), another way to assess the correctness of the model is to make prediction for the 20th time unit given that the data for previous 19 time units is available; then compare the real data for the 20th time unit with this prediction.

In this project, we adopt a simple aspect to demonstrate that the region factor can play a role in insured population prediction. We introduce a random location effect to capture possible small scale movements between the regions. We choose to give no information in the region factor L 's but to contribute variability to the total claim amounts. It turns out that this model can indeed reflect the fact of L 's bearing no information, as shown in Figure 3.6. Further research can be conducted, by imposing some information related to the region factor, to test how effective the model detects the true parameters.

It is also worth pointing out the possibility of considering policy deductible and policy limits for the primary insurers when calculating the premiums. It is common in practice that insurers do not cover small losses. Truncated distributions can be considered in that matter. One can further extend the algorithm to calculate the stop loss premium for reinsurance. Further information can be found in Pai (1997).

References

- Banerjee, S., Carlin, B., and Gelfand, A. (2003). *Hierarchical Modelling and Analysis for Spatial Data*. Chapman & Hall/CRC.
- Bühlmann, H. (1970). *Mathematical methods in risk theory*, volume 172. Springer Science & Business Media.
- Cowles, M. (2004). Review of WinBUGS 1.4. *The American Statistician*, 58(4):330--336.
- Dickson, D. C. (2005). *Insurance risk and ruin*. Cambridge University Press.
- Dimakos, X. and Frigessi, A. (2002). Bayesian premium rating with latent structure. *Scandinavian Actuarial Journal*, 2002(3):162--184.
- Dutang, C., Goulet, V., and Pigeon, M. (2008). Actuar: An R package for actuarial science. *Journal of Statistical Software*, 25(7). 38.
- Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*, volume 33. Springer Science & Business Media.
- Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press.
- Gelfand, A. E. (1996). Model determination using sampling-based methods. In *Markov chain Monte Carlo in practice*, pages 145--161. Chapman & Hall/CRC.
- Gelman, A., Sturtz, S., Ligges, U., Gorjanc, G., and Kerman, J. (2011). R2WinBUGS: Running WinBUGS and OpenBUGS from R/S-PLUS. *R package version*. 2-1.
- Gerber, H. U. (1979). An introduction to mathematical risk theory. 8. Philadelphia: SS Huebner Foundation for Insurance Education, Wharton School, University of Pennsylvania.
- Gilks, W., Richardson, S., and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo In Practice*, 1. 19. Chapman & Hall.
- Goovaerts, M., Dhaene, J., and Rachedi, O. (2010). Actuarial premium principles. In *Encyclopedia of Quantitative Finance*. John Wiley and Sons.
- Gschlöbl, S. and Czado, C. (2007). Spatial modeling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, 2007(3):202--225.
- Hardy, M. R. (2006). An introduction to risk measures for actuarial applications. SOA Syllabus Study Note.
- Jaynes, E. T. and Kempthorne, O. (1976). Confidence intervals vs Bayesian intervals. *Foundations of probability theory, statistical inference, and statistical theories of science*, pages 175--257.
- Klugman, S. A. (1992). *Bayesian statistics in actuarial science: with emphasis on credibility*, volume 15. Springer Science & Business Media.
- Laeven, R. J. and Goovaerts, M. J. (2008). Premium calculation and insurance pricing. In *Encyclopedia of Quantitative Risk Analysis and Assessment*. John Wiley and Sons.
- Lee, P. M. (2012). *Bayesian Statistics: An Introduction*. John Wiley & Sons.

- Lindley, D. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint (2 vols--Part I: Probability and Part II: Inference)*. Cambridge: Cambridge University Press.
- Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325--337.
- Migon, H. and Gamerman, D. (1993). Generalized exponential growth models-a Bayesian approach. *Journal of Forecasting*, 12(7):573--584.
- Migon, H. and Moura, F. (2005). Hierarchical Bayesian collective risk model: an application to health insurance. *Insurance: Mathematics and Economics*, 36(2):119--135.
- Migon, H. and Penna, E. (2006). Bayesian analysis of a health insurance model. *Journal of Actuarial Practice*, 13:61--80.
- Ntzoufras, I. (2011). *Bayesian Modeling Using WinBUGS*, volume 698. John Wiley & Sons.
- Pai, S. (1997). Bayesian analysis of compound loss distributions. *Journal of Econometrics*, 79(1):129--146.
- Park, B. and Marron, J. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85(409):66--72.
- Peng, J. (2009). Value at risk and tail value at risk in uncertain environment. In *Proceedings of the 8th International Conference on Information and Management Sciences*, pages 787--793.
- Pettitt, A. N., Weir, I. S., and Hart, A. G. (2002). A conditional autoregressive gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing*, 12(4):353--367.
- Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443--1471.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832--837.
- Sarykalin, S., Serraino, G., and Uryasev, S. (2008). Value-at-risk vs. conditional value-at-risk in risk management and optimization. *Tutorials in Operations Research*, pages 270--294. INFORMS Annual Meeting, Handover, MD.
- Souza, D., Moura, F., and Migon, H. (2009). Small area population prediction via hierarchical models. In *Survey Methodology*, volume 39 of *Catalogue No. 12-001-X*, pages 203--214. Statistics Canada.
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2007). OpenBUGS user manual, version 3.0.2. *MRC Biostatistics Unit, Cambridge*.
- Sturtz, S., Ligges, U., and Gelman, A. E. (2005). R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical software*, 12(3):1--16.
- Young, V. R. (2004). Premium principles. In *Encyclopedia of Actuarial Science*. John Wiley and Sons.

Appendix A

Derivation of Conditional Posterior Distributions

The following derivation is for the simplified model used as numerical illustration. However, the same logic can be applied to the full model. Let Θ represent all parameters of the model, and $\mathbf{D}_T = \{(\mathbf{x}_t, \mathbf{n}_t, \mathbf{M}_t), t = 1, \dots, T\}$ represent all the data available.

Assuming independence in time, age class and region, the likelihood function is given by

$$\begin{aligned}
 l(\Theta|\mathbf{D}_T) &\propto \prod_{t,i,a=1}^{T,I,A} f(N_{t,i,a}|\lambda_a, M_{t,i,a}) f(X_{t,i,a}|\theta_a, n_{t,i,a}) f(M_{t,i,a}) \\
 &= \prod_{t,i,a=1}^{T,I,A} \frac{(\lambda_a M_{t,i,a})^{n_{t,i,a}} e^{-\lambda_a M_{t,i,a}}}{n_{t,i,a}!} \cdot \frac{\theta_a^{n_{t,i,a}} x_{t,i,a}^{n_{t,i,a}-1} e^{-\theta_a x_{t,i,a}}}{\Gamma(n_{t,i,a})} \cdot \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(x_{t,i,a} - \mu_{t,i,a})^2} \\
 &\propto \prod_{t,i,a=1}^{T,I,A} (\lambda_a \theta_a)^{n_{t,i,a}} \sqrt{\tau} \cdot e^{-(\theta_a x_{t,i,a} + \lambda_a M_{t,i,a} + \frac{\tau}{2}(M_{t,i,a} - \mu_{t,i,a})^2)}.
 \end{aligned}$$

The full conditional posterior distributions are as follows.

$$\begin{aligned}
 \bullet \quad f(\lambda_a|\Theta_{-\lambda_a}, \mathbf{D}_T) &\propto f(\lambda_a) \prod_{t,i=1}^{T,I} f(N_{t,i,a}|\lambda_a, M_{t,i,a}) \\
 &= \frac{\beta_\lambda^{\alpha_\lambda}}{\Gamma(\alpha_\lambda)} \lambda_a^{\alpha_\lambda-1} e^{-\beta_\lambda \lambda_a} \prod_{t,i=1}^{T,I} \frac{(\lambda_a M_{t,i,a})^{n_{t,i,a}} e^{-\lambda_a M_{t,i,a}}}{n_{t,i,a}!} \\
 &\propto \lambda_a^{\alpha_\lambda + \sum_{t,i=1}^{T,I} n_{t,i,a} - 1} e^{-\lambda_a(\beta_\lambda + \sum_{t,i=1}^{T,I} M_{t,i,a})},
 \end{aligned}$$

and hence,

$$\lambda_a|\Theta_{-\lambda_a}, \mathbf{D}_T \sim \text{Gamma}\left(\alpha_\lambda + \sum_{t,i=1}^{T,I} n_{t,i,a}, \beta_\lambda + \sum_{t,i=1}^{T,I} M_{t,i,a}\right), \quad a = 1, 2, \dots, A.$$

$$\begin{aligned}
\bullet \quad f(\theta_a | \Theta_{-\theta_a}, \mathbf{D}_T) &\propto f(\theta_a) \prod_{t,i=1}^{T,I} f(X_{t,i,a} | \theta_a, n_{t,i,a}) \\
&= \frac{\beta_\theta^{\alpha_\theta}}{\Gamma(\alpha_\theta)} \theta_a^{\alpha_\theta-1} e^{-\beta_\theta \theta_a} \prod_{t,i=1}^{T,I} \frac{\theta_a^{n_{t,i,a}} x_{t,i,a}^{n_{t,i,a}-1} e^{-\theta_a x_{t,i,a}}}{\Gamma(n_{t,i,a})} \\
&\propto \theta_a^{\alpha_\theta + \sum_{t,i=1}^{T,I} n_{t,i,a} - 1} e^{-\theta_a (\beta_\theta + \sum_{t,i=1}^{T,I} x_{t,i,a})},
\end{aligned}$$

and hence,

$$\theta_a | \Theta_{-\theta_a}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_\theta + \sum_{t,i=1}^{T,I} n_{t,i,a}, \beta_\theta + \sum_{t,i=1}^{T,I} x_{t,i,a} \right), \quad a = 1, 2, \dots, A.$$

$$\begin{aligned}
\bullet \quad f(\beta_0 | \Theta_{-\beta_0}, \mathbf{D}_T) &\propto f(\beta_0) \prod_{t,i,a=1}^{T,I,A} f(M_{t,i,a}) \\
&= \sqrt{\frac{\tau_0}{2\pi}} \exp \left(\frac{-\tau_0 (\beta_0 - \mu_0)^2}{2} \right) \cdot \prod_{t,i,a=1}^{T,I,A} \sqrt{\frac{\tau}{2\pi}} \exp \left(\frac{-\tau (M_{t,i,a} - \mu_{t,i,a})^2}{2} \right) \\
&\propto \exp \left(-\frac{\tau_0}{2} (\beta_0 - \mu_0)^2 - \frac{\tau}{2} \sum_{t,i,a=1}^{T,I,A} (M_{t,i,a} - \mu_{t,i,a})^2 \right) \\
&\propto \exp \left(-\frac{1}{2} (\tau_0 + \tau TIA) \beta_0^2 + (\tau_0 \mu_0 + \tau \sum_{t,i,a=1}^{T,I,A} b_{t,i,a}) \beta_0 \right),
\end{aligned}$$

and hence,

$$\beta_0 | \Theta_{-\beta_0}, \mathbf{D}_T \sim N(\text{Mean}_{\beta_0}, (\text{Precision}_{\beta_0})^{-1}),$$

where

$$\text{Mean}_{\beta_0} = \frac{\tau_0 \mu_0 + \tau \sum_{t,i,a=1}^{T,I,A} b_{t,i,a}}{\text{Precision}_{\beta_0}},$$

$$\text{Precision}_{\beta_0} = \tau_0 + \tau TIA,$$

$$b_{t,i,a} = M_{t,i,a} - \varepsilon_a^0 - L_i - \beta_1 e^{t(\beta_2 + \varepsilon_a^2)}.$$

$$\begin{aligned}
\bullet \quad f(\varepsilon_a^0 | \Theta_{-\varepsilon_a^0}, \mathbf{D}_T) &\propto f(\varepsilon_a^0) \prod_{t,i=1}^{T,I} f(M_{t,i,a}) \\
&= \sqrt{\frac{\tau_{\varepsilon_0}}{2\pi}} \exp \left(\frac{-\tau_{\varepsilon_0} (\varepsilon_a^0)^2}{2} \right) \cdot \prod_{t,i=1}^{T,I} \sqrt{\frac{\tau}{2\pi}} \exp \left(\frac{-\tau (M_{t,i,a} - \mu_{t,i,a})^2}{2} \right) \\
&\propto \exp \left(-\frac{\tau_{\varepsilon_0} (\varepsilon_a^0)^2}{2} - \frac{\tau}{2} \sum_{t,i=1}^{T,I} (M_{t,i,a} - \mu_{t,i,a})^2 \right) \\
&\propto \exp \left(-\frac{1}{2} (\tau_{\varepsilon_0} + \tau TI) (\varepsilon_a^0)^2 + \tau \sum_{t,i=1}^{T,I} c_{t,i,a} \varepsilon_a^0 \right),
\end{aligned}$$

and hence,

$$\varepsilon_a^0 | \Theta_{-\varepsilon_a^0}, \mathbf{D}_T \sim N(\text{Mean}_{\varepsilon_a^0}, (\text{Precision}_{\varepsilon_a^0})^{-1}), \quad a = 1, 2, \dots, A,$$

where

$$\begin{aligned} \text{Mean}_{\varepsilon_a^0} &= \frac{\tau \sum_{t,i=1}^{T,I} c_{t,i,a}}{\text{Precision}_{\varepsilon_a^0}}, \\ \text{Precision}_{\varepsilon_a^0} &= \tau_{\varepsilon_0} + \tau T I, \\ c_{t,i,a} &= M_{t,i,a} - \beta_0 - L_i - \beta_1 e^{t(\beta_2 + \varepsilon_a^2)}. \end{aligned}$$

$$\begin{aligned} \bullet \quad f(\beta_1 | \Theta_{-\beta_1}, \mathbf{D}_T) &\propto f(\beta_1) \prod_{t,i,a=1}^{T,I,A} f(M_{t,i,a}) \\ &= \sqrt{\frac{\tau_1}{2\pi}} \exp\left(\frac{-\tau_1(\beta_1 - \mu_1)^2}{2}\right) \cdot \prod_{t,i,a=1}^{T,I,A} \sqrt{\frac{\tau}{2\pi}} \exp\left(\frac{-\tau(M_{t,i,a} - \mu_{t,i,a})^2}{2}\right) \\ &\propto \exp\left(-\frac{\tau_1}{2}(\beta_1 - \mu_1)^2 - \frac{\tau}{2} \sum_{t,i,a=1}^{T,I,A} (M_{t,i,a} - \mu_{t,i,a})^2\right) \\ &\propto \exp\left(-\frac{1}{2} \left(\tau_1 + \tau \sum_{t,i,a=1}^{T,I,A} e^{2t(\beta_2 + \varepsilon_a^2)} \right) \beta_1^2 + \left(\tau_1 \mu_1 + \tau \sum_{t,i,a=1}^{T,I,A} d_{t,i,a} \cdot e^{t(\beta_2 + \varepsilon_a^2)} \right) \beta_1 \right), \end{aligned}$$

and hence,

$$\beta_1 | \Theta_{-\beta_1}, \mathbf{D}_T \sim N(\text{Mean}_{\beta_1}, (\text{Precision}_{\beta_1})^{-1}),$$

where

$$\begin{aligned} \text{Mean}_{\beta_1} &= \frac{\tau_1 \mu_1 + \tau \sum_{t,i,a=1}^{T,I,A} d_{t,i,a} \cdot e^{t(\beta_2 + \varepsilon_a^2)}}{\text{Precision}_{\beta_1}}, \\ \text{Precision}_{\beta_1} &= \tau_1 + \tau \sum_{t,i,a=1}^{T,I,A} e^{2t(\beta_2 + \varepsilon_a^2)}, \\ d_{t,i,a} &= M_{t,i,a} - \beta_0 - \varepsilon_a^0 - L_i. \end{aligned}$$

$$\begin{aligned} \bullet \quad f(\beta_2 | \Theta_{-\beta_2}, \mathbf{D}_T) &\propto f(\beta_2) \prod_{t,i,a=1}^{T,I,A} f(M_{t,i,a}) \\ &= \sqrt{\frac{\tau_2}{2\pi}} \exp\left(\frac{-\tau_2(\beta_2 - \mu_2)^2}{2}\right) \cdot \prod_{t,i,a=1}^{T,I,A} \sqrt{\frac{\tau}{2\pi}} \exp\left(\frac{-\tau(M_{t,i,a} - \mu_{t,i,a})^2}{2}\right) \\ &\propto \exp\left(-\frac{\tau_2}{2}(\beta_2^2 - 2\beta_2 \mu_2) - \frac{\tau}{2} \sum_{t,i,a=1}^{T,I,A} (\beta_1 e^{t(\beta_2 + \varepsilon_a^2)} - d_{t,i,a})^2\right). \end{aligned}$$

$$\begin{aligned}
\bullet \quad f(\varepsilon_a^2 | \Theta_{-\varepsilon_a^2}, \mathbf{D}_T) &\propto f(\varepsilon_a^2) \prod_{t,i=1}^{T,I} f(M_{t,i,a}) \\
&= \sqrt{\frac{\tau \varepsilon_2}{2\pi}} \exp\left(\frac{-\tau \varepsilon_2 (\varepsilon_a^2)^2}{2}\right) \cdot \prod_{t,i=1}^{T,I} \sqrt{\frac{\tau}{2\pi}} \exp\left(\frac{-\tau (M_{t,i,a} - \mu_{t,i,a})^2}{2}\right) \\
&\propto \exp\left(-\frac{\tau \varepsilon_2}{2} (\varepsilon_a^2)^2 - \frac{\tau}{2} \sum_{t,i=1}^{T,I} (\beta_1 e^{t(\beta_2 + \varepsilon_a^2)} - d_{t,i,a})^2\right), \quad a = 1, 2, \dots, A.
\end{aligned}$$

$$\begin{aligned}
\bullet \quad f(\tau | \Theta_{-\tau}, \mathbf{D}_T) &\propto f(\tau) \prod_{t,i,a=1}^{T,I,A} f(M_{t,i,a}) \\
&\propto \tau^{\alpha_\tau - 1} e^{-\beta_\tau \tau} \cdot \prod_{t,i,a=1}^{T,I,A} \sqrt{\frac{\tau}{2\pi}} \exp\left(\frac{-\tau (M_{t,i,a} - \mu_{t,i,a})^2}{2}\right) \\
&\propto \tau^{\alpha_\tau + \frac{1}{2}TIA - 1} \cdot \exp\left(-\tau(\beta_\tau + \frac{1}{2} \sum_{t,i,a=1}^{T,I,A} (M_{t,i,a} - \mu_{t,i,a})^2)\right),
\end{aligned}$$

and hence,

$$\tau | \Theta_{-\tau}, \mathbf{D}_T \sim \text{Gamma}\left(\alpha_\tau + \frac{1}{2}TIA, \beta_\tau + \frac{1}{2} \sum_{t,i,a=1}^{T,I,A} (M_{t,i,a} - \mu_{t,i,a})^2\right).$$

$$\begin{aligned}
\bullet \quad f(\alpha_\theta | \Theta_{-\alpha_\theta}, \mathbf{D}_T) &\propto f(\alpha_\theta) \prod_{a=1}^A f(\theta_a) \\
&\propto \alpha_\theta^{\alpha_\theta - 1} e^{-\beta_{\alpha_\theta} \alpha_\theta} \prod_{a=1}^A \frac{\beta_\theta^{\alpha_\theta}}{\Gamma(\alpha_\theta)} \theta_a^{\alpha_\theta - 1} e^{-\beta_\theta \theta_a} \\
&\propto \alpha_\theta^{\alpha_\theta - 1} e^{-\beta_{\alpha_\theta} \alpha_\theta} \cdot \beta_\theta^{A \cdot \alpha_\theta} [\Gamma(\alpha_\theta)]^{-A} \prod_{a=1}^A \theta_a^{\alpha_\theta - 1}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
f(\alpha_\lambda | \Theta_{-\alpha_\lambda}, \mathbf{D}_T) &\propto f(\alpha_\lambda) \prod_{a=1}^A f(\lambda_a) \\
&\propto \alpha_\lambda^{\alpha_\lambda - 1} e^{-\beta_{\alpha_\lambda} \alpha_\lambda} \cdot \beta_\lambda^{A \cdot \alpha_\lambda} [\Gamma(\alpha_\lambda)]^{-A} \prod_{a=1}^A \lambda_a^{\alpha_\lambda - 1}.
\end{aligned}$$

$$\begin{aligned}
\bullet \quad f(\beta_\theta | \Theta_{-\beta_\theta}, \mathbf{D}_T) &\propto f(\beta_\theta) \prod_{a=1}^A f(\theta_a) \\
&\propto \beta_\theta^{\alpha_{\beta_\theta} - 1} e^{-\beta_{\beta_\theta} \beta_\theta} \prod_{a=1}^A \frac{\beta_\theta^{\alpha_\theta}}{\Gamma(\alpha_\theta)} \theta_a^{\alpha_\theta - 1} e^{-\beta_\theta \theta_a} \\
&\propto \beta_\theta^{\alpha_{\beta_\theta} + \alpha_\theta A - 1} \cdot e^{-\beta_\theta (\beta_{\beta_\theta} + \sum_{a=1}^A \theta_a)},
\end{aligned}$$

and hence,

$$\beta_\theta | \Theta_{-\beta_\theta}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_{\beta_\theta} + \alpha_\theta A, \beta_{\beta_\theta} + \sum_{a=1}^A \theta_a \right).$$

Similarly,

$$\beta_\lambda | \Theta_{-\beta_\lambda}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_{\beta_\lambda} + \alpha_\lambda A, \beta_{\beta_\lambda} + \sum_{a=1}^A \lambda_a \right).$$

$$\bullet \quad \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^{-1} \begin{pmatrix} 1 + |\eta| \cdot m_g & -\eta \\ -\eta & 1 + |\eta| \cdot m_g \end{pmatrix}^{-1} \right],$$

which can be denoted as

$$\text{MVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P & S \\ S & P \end{pmatrix} \right],$$

with the correlation coefficient

$$\rho = \frac{S}{P}.$$

$$\begin{aligned} \bullet \quad f(L_i | \Theta_{-L_i}, \mathbf{D}_T) &\propto \exp \left(-\frac{L_1^2 + L_2^2 - 2\rho L_1 L_2}{2(1 - \rho^2)P^2} \right) \prod_{t,a=1}^{T,A} \exp \left(-\frac{\tau}{2} (M_{t,i,a} - \mu_{t,i,a})^2 \right), \quad i = 1, 2. \\ \bullet \quad f(\eta | \Theta_{-\eta}, \mathbf{D}_T) &\propto \frac{1}{(1 + \eta)^2 P \sqrt{1 - \rho^2}} \cdot \exp \left(-\frac{L_1^2 + L_2^2 - 2\rho L_1 L_2}{2P(1 - \rho^2)} \right). \\ \bullet \quad f(\sigma | \Theta_{-\sigma}, \mathbf{D}_T) &\propto \frac{\sigma^{\alpha_\sigma - 1} e^{-\sigma \beta_\sigma}}{P^2} \cdot \exp \left(-\frac{L_1^2 + L_2^2 - 2\rho L_1 L_2}{2P(1 - \rho^2)} \right). \\ \bullet \quad f(\tau_{\varepsilon_0} | \Theta_{-\tau_{\varepsilon_0}}, \mathbf{D}_T) &\propto \tau_{\varepsilon_0}^{\alpha_{\tau_{\varepsilon_0}} - 1} e^{-\beta_{\tau_{\varepsilon_0}} \tau_{\varepsilon_0}} \cdot \prod_{a=1}^A \sqrt{\frac{\tau_{\varepsilon_0}}{2\pi}} e^{-\frac{\tau_{\varepsilon_0} (\varepsilon_a^0)^2}{2}} \\ &\propto \tau_{\varepsilon_0}^{\alpha_{\tau_{\varepsilon_0}} + \frac{A}{2} - 1} \cdot e^{-\tau_{\varepsilon_0} (\beta_{\tau_{\varepsilon_0}} + \frac{1}{2} \sum_{a=1}^A (\varepsilon_a^0)^2)}, \end{aligned}$$

and hence,

$$\tau_{\varepsilon_0} | \Theta_{-\tau_{\varepsilon_0}}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_{\tau_{\varepsilon_0}} + \frac{A}{2}, \beta_{\tau_{\varepsilon_0}} + \frac{1}{2} \sum_{a=1}^A (\varepsilon_a^0)^2 \right).$$

Similarly,

$$\tau_{\varepsilon_2} | \Theta_{-\tau_{\varepsilon_2}}, \mathbf{D}_T \sim \text{Gamma} \left(\alpha_{\tau_{\varepsilon_2}} + \frac{A}{2}, \beta_{\tau_{\varepsilon_2}} + \frac{1}{2} \sum_{a=1}^A (\varepsilon_a^2)^2 \right).$$

Appendix B

Simulated Data

Table B.1: Simulated Insured Population $M_{t,i,a}$

| Insured Population | | Age Class | | | | | | |
|--------------------|------|-----------|-----|-----|-----|-----|-----|-----|
| | Time | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Region 1 | 1 | 54 | 63 | 73 | 83 | 72 | 59 | 70 |
| | 2 | 72 | 77 | 75 | 63 | 74 | 91 | 70 |
| | 3 | 75 | 98 | 66 | 77 | 78 | 89 | 75 |
| | 4 | 74 | 89 | 78 | 97 | 79 | 68 | 86 |
| | 5 | 69 | 82 | 83 | 92 | 65 | 75 | 72 |
| | 6 | 75 | 80 | 83 | 74 | 82 | 79 | 66 |
| | 7 | 86 | 95 | 92 | 70 | 90 | 84 | 91 |
| | 8 | 72 | 75 | 71 | 85 | 88 | 81 | 89 |
| | 9 | 96 | 93 | 81 | 79 | 85 | 106 | 100 |
| | 10 | 81 | 79 | 93 | 84 | 97 | 103 | 92 |
| | 11 | 85 | 110 | 130 | 89 | 109 | 88 | 84 |
| | 12 | 106 | 98 | 98 | 123 | 99 | 107 | 84 |
| | 13 | 101 | 124 | 109 | 94 | 112 | 104 | 86 |
| | 14 | 105 | 96 | 110 | 92 | 109 | 118 | 105 |
| | 15 | 114 | 119 | 119 | 137 | 112 | 116 | 126 |
| | 16 | 113 | 102 | 104 | 111 | 101 | 122 | 106 |
| | 17 | 117 | 125 | 133 | 121 | 120 | 125 | 117 |
| | 18 | 121 | 140 | 129 | 125 | 137 | 138 | 120 |
| | 19 | 155 | 115 | 133 | 137 | 148 | 125 | 133 |
| | 20 | 149 | 146 | 147 | 138 | 139 | 134 | 137 |
| Region 2 | 1 | 80 | 75 | 85 | 65 | 79 | 53 | 84 |
| | 2 | 87 | 70 | 61 | 80 | 86 | 57 | 70 |
| | 3 | 81 | 44 | 77 | 74 | 93 | 44 | 68 |
| | 4 | 70 | 78 | 61 | 75 | 73 | 76 | 76 |
| | 5 | 71 | 77 | 91 | 72 | 81 | 94 | 88 |
| | 6 | 81 | 72 | 89 | 62 | 70 | 80 | 83 |
| | 7 | 76 | 77 | 81 | 87 | 86 | 69 | 85 |
| | 8 | 92 | 87 | 96 | 97 | 81 | 105 | 89 |
| | 9 | 94 | 96 | 74 | 66 | 100 | 94 | 90 |
| | 10 | 105 | 102 | 99 | 91 | 88 | 96 | 87 |
| | 11 | 96 | 89 | 88 | 89 | 92 | 103 | 88 |
| | 12 | 98 | 107 | 100 | 109 | 93 | 101 | 109 |
| | 13 | 117 | 93 | 104 | 94 | 111 | 100 | 107 |
| | 14 | 104 | 103 | 101 | 105 | 105 | 103 | 104 |
| | 15 | 109 | 110 | 94 | 130 | 120 | 113 | 94 |
| | 16 | 118 | 109 | 118 | 125 | 100 | 109 | 125 |
| | 17 | 127 | 125 | 128 | 118 | 107 | 118 | 123 |
| | 18 | 117 | 126 | 110 | 136 | 139 | 129 | 114 |
| | 19 | 147 | 146 | 128 | 119 | 126 | 131 | 144 |
| | 20 | 136 | 128 | 152 | 145 | 132 | 148 | 141 |

Table B.2: Simulated Total Claim Frequency $N_{t,i,a}$

| Total Claim Number | | Age Class | | | | | | |
|--------------------|------|-----------|----|----|----|----|----|----|
| | Time | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Region 1 | 1 | 11 | 10 | 10 | 16 | 16 | 10 | 14 |
| | 2 | 16 | 17 | 10 | 11 | 14 | 19 | 16 |
| | 3 | 22 | 13 | 8 | 14 | 16 | 14 | 18 |
| | 4 | 17 | 19 | 12 | 22 | 7 | 9 | 22 |
| | 5 | 13 | 12 | 8 | 18 | 11 | 14 | 26 |
| | 6 | 13 | 20 | 13 | 6 | 11 | 8 | 17 |
| | 7 | 18 | 12 | 12 | 18 | 14 | 13 | 29 |
| | 8 | 11 | 5 | 10 | 14 | 16 | 6 | 29 |
| | 9 | 29 | 11 | 8 | 8 | 14 | 14 | 29 |
| | 10 | 17 | 18 | 13 | 14 | 18 | 15 | 29 |
| | 11 | 12 | 16 | 19 | 18 | 21 | 22 | 16 |
| | 12 | 27 | 8 | 12 | 17 | 21 | 14 | 20 |
| | 13 | 15 | 23 | 19 | 18 | 20 | 16 | 21 |
| | 14 | 18 | 21 | 20 | 15 | 24 | 22 | 27 |
| | 15 | 25 | 20 | 17 | 29 | 22 | 17 | 33 |
| | 16 | 26 | 21 | 10 | 11 | 26 | 24 | 22 |
| | 17 | 19 | 17 | 21 | 9 | 24 | 20 | 31 |
| | 18 | 26 | 30 | 16 | 22 | 23 | 15 | 35 |
| | 19 | 44 | 23 | 25 | 17 | 27 | 21 | 37 |
| | 20 | 40 | 26 | 23 | 15 | 27 | 23 | 34 |
| Region 2 | 1 | 21 | 18 | 13 | 9 | 19 | 10 | 19 |
| | 2 | 15 | 13 | 12 | 17 | 15 | 10 | 13 |
| | 3 | 18 | 7 | 5 | 12 | 21 | 6 | 21 |
| | 4 | 17 | 15 | 7 | 12 | 12 | 14 | 14 |
| | 5 | 12 | 10 | 12 | 8 | 18 | 15 | 23 |
| | 6 | 20 | 12 | 10 | 13 | 10 | 11 | 22 |
| | 7 | 20 | 14 | 17 | 11 | 22 | 13 | 22 |
| | 8 | 17 | 17 | 13 | 12 | 16 | 21 | 25 |
| | 9 | 14 | 12 | 7 | 10 | 15 | 13 | 24 |
| | 10 | 22 | 9 | 15 | 18 | 14 | 18 | 22 |
| | 11 | 21 | 16 | 13 | 17 | 11 | 17 | 19 |
| | 12 | 19 | 14 | 8 | 18 | 21 | 25 | 26 |
| | 13 | 23 | 16 | 11 | 23 | 20 | 13 | 21 |
| | 14 | 20 | 15 | 12 | 21 | 13 | 17 | 26 |
| | 15 | 27 | 11 | 11 | 23 | 26 | 12 | 16 |
| | 16 | 21 | 27 | 18 | 18 | 10 | 15 | 30 |
| | 17 | 23 | 18 | 19 | 25 | 15 | 15 | 35 |
| | 18 | 26 | 19 | 12 | 26 | 20 | 25 | 25 |
| | 19 | 39 | 25 | 23 | 24 | 16 | 25 | 31 |
| | 20 | 39 | 28 | 16 | 14 | 28 | 21 | 38 |

Table B.3: Simulated Aggregate Loss $X_{t,i,a}$

| Total Claim Amount | | Age Class | | | | | | |
|--------------------|------|-----------|-----|-----|-----|-----|-----|-----|
| | Time | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Region 1 | 1 | 209 | 165 | 363 | 641 | 456 | 244 | 477 |
| | 2 | 475 | 394 | 483 | 300 | 368 | 361 | 483 |
| | 3 | 869 | 404 | 199 | 245 | 443 | 436 | 429 |
| | 4 | 360 | 665 | 334 | 472 | 174 | 228 | 555 |
| | 5 | 331 | 497 | 294 | 378 | 237 | 202 | 623 |
| | 6 | 222 | 500 | 361 | 124 | 364 | 323 | 336 |
| | 7 | 263 | 176 | 177 | 493 | 249 | 315 | 518 |
| | 8 | 253 | 104 | 179 | 182 | 254 | 175 | 635 |
| | 9 | 701 | 294 | 85 | 172 | 399 | 325 | 642 |
| | 10 | 502 | 368 | 292 | 297 | 463 | 575 | 524 |
| | 11 | 187 | 365 | 621 | 335 | 563 | 424 | 453 |
| | 12 | 680 | 114 | 225 | 488 | 512 | 329 | 667 |
| | 13 | 272 | 477 | 385 | 408 | 336 | 348 | 571 |
| | 14 | 466 | 562 | 488 | 343 | 584 | 569 | 493 |
| | 15 | 733 | 427 | 372 | 707 | 576 | 282 | 730 |
| | 16 | 758 | 657 | 246 | 219 | 873 | 474 | 581 |
| | 17 | 513 | 489 | 392 | 185 | 646 | 481 | 503 |
| | 18 | 469 | 643 | 549 | 534 | 575 | 183 | 984 |
| | 19 | 1003 | 637 | 572 | 258 | 700 | 709 | 951 |
| | 20 | 943 | 657 | 459 | 290 | 554 | 556 | 694 |
| Region 2 | 1 | 344 | 437 | 334 | 297 | 502 | 265 | 425 |
| | 2 | 221 | 305 | 255 | 397 | 549 | 253 | 317 |
| | 3 | 354 | 201 | 132 | 313 | 704 | 227 | 636 |
| | 4 | 507 | 534 | 114 | 299 | 346 | 300 | 282 |
| | 5 | 350 | 437 | 589 | 287 | 578 | 369 | 713 |
| | 6 | 445 | 212 | 461 | 359 | 173 | 185 | 405 |
| | 7 | 359 | 459 | 532 | 206 | 709 | 426 | 593 |
| | 8 | 244 | 705 | 225 | 279 | 529 | 385 | 381 |
| | 9 | 280 | 402 | 249 | 298 | 443 | 206 | 653 |
| | 10 | 453 | 199 | 285 | 580 | 620 | 592 | 671 |
| | 11 | 619 | 347 | 245 | 484 | 218 | 313 | 656 |
| | 12 | 487 | 303 | 141 | 412 | 730 | 531 | 508 |
| | 13 | 344 | 448 | 259 | 412 | 588 | 363 | 451 |
| | 14 | 589 | 240 | 277 | 767 | 220 | 308 | 430 |
| | 15 | 409 | 427 | 347 | 522 | 576 | 348 | 322 |
| | 16 | 521 | 765 | 411 | 294 | 152 | 367 | 542 |
| | 17 | 592 | 556 | 451 | 539 | 371 | 394 | 707 |
| | 18 | 610 | 408 | 283 | 615 | 407 | 381 | 589 |
| | 19 | 996 | 736 | 512 | 532 | 436 | 755 | 716 |
| | 20 | 995 | 863 | 478 | 263 | 809 | 547 | 693 |

Appendix C

Premiums for 21st Time Unit in Region 2

Table C.1: Total Premiums for 21st Time Unit in Region 2 Using Predicted Results

| Total Premium (\$) | Loading | Age Class | | | | | | |
|--------------------------------------|---------|-----------|----------|--------|----------|----------|----------|----------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Net Premium Principle | ----- | 771.80 | 651.20 | 517.50 | 583.30 | 690.40 | 579.20 | 831.30 |
| Expected Value Premium Principle | 0.5 | 1,157.70 | 976.80 | 776.25 | 874.95 | 1,035.60 | 868.80 | 1,246.95 |
| Variance Premium Principle | 0.01 | 1,181.05 | 1,035.36 | 802.43 | 889.90 | 1,099.65 | 884.75 | 1,273.98 |
| Standard Deviation Premium Principle | 1.96 | 1,168.31 | 1,035.36 | 848.35 | 926.50 | 1,086.91 | 921.81 | 1,243.68 |
| VaR _{97.5%} | ----- | 1,209.00 | 1,076.00 | 884.50 | 965.10 | 1,131.00 | 958.50 | 1,281.00 |
| TVaR _{97.5%} | ----- | 1,305.89 | 1,177.64 | 975.19 | 1,054.53 | 1,235.32 | 1,046.18 | 1,385.43 |

Table C.2: Premium Per Policyholder for 21st Time Unit in Region 2

| Premium Per Person (\$) | | Age Class | | | | | | |
|-------------------------|--------------------------------------|-----------|------|------|------|-------|------|-------|
| Premium Measure | Predicted Insured Population Measure | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Net Premium Principle | Average Insured Population | 5.08 | 4.47 | 3.49 | 3.93 | 4.77 | 3.86 | 5.79 |
| VaR _{97.5%} | Average Insured Population | 7.96 | 7.38 | 5.97 | 6.49 | 7.82 | 6.39 | 8.92 |
| VaR _{97.5%} | 2.5 Percentile Insured Population | 9.27 | 8.68 | 6.99 | 7.61 | 9.22 | 7.46 | 10.51 |
| TVaR _{97.5%} | Average Insured Population | 8.60 | 8.08 | 6.58 | 7.10 | 8.54 | 6.97 | 9.65 |
| TVaR _{97.5%} | 2.5 Percentile Insured Population | 10.01 | 9.50 | 7.71 | 8.31 | 10.07 | 8.15 | 11.37 |