UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Matemáticas y Computacionales

Hierarchical Bayesian Model for Insurance Claims

Trabajo de integración curricular presentado como requisito para la obtención del título de Matemática.

Autor:

Rivadeneira Pérez Erika Lizbeth

Tutor:

PhD. Infante Quirpa Saba Rafael

Urcuquí, julio 2020

Urcuquí, 27 de febrero de 2020

SECRETARÍA GENERAL
(Vicerrectorado Académico/Cancillería)
ESCUELA DE CIENCIAS MATEMÁTICAS Y COMPUTACIONALES
CARRERA DE MATEMÁTICA
ACTA DE DEFENSA No. UITEY-ITE-2020-00003-AD

En la ciudad de San Miguel de Urcuquí, Provincia de Imbabura, a los 27 días del mes de febrero de 2020, a las 12:00 horas, en el Aula CHA-02 de la Universidad de Investigación de Tecnología Experimental Yachay y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Dr. FONSECA DELGADO, RIGOBERTO SALOMON , Ph.D.
Miembro No Tutor	Dr. CAMACHO , FRANKLIN , Ph.D.
Tutor	Dr. INFANTE QUIRPA, SABA RAFAEL , Ph.D.

Se presenta el(la) señor(ita) estudiante **RIVADENEIRA PEREZ, ERIKA LIZBETH**, con cédula de identidad No. **1723987952**, de la **ESCUELA DE CIENCIAS MATEMÁTICAS Y COMPUTACIONALES**, de la Carrera de **MATEMÁTICA**, aprobada por el Consejo de Educación Superior (CES), mediante Resolución **RPC-SO-15-No.174-2015**, con el objeto de rendir la sustentación de su trabajo de titulación denominado: **HIERARCHICAL BAYESIAN MODEL FOR INSURANCE CLAIMS.**, previa a la obtención del título de **MATEMÁTICO/A.**

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dr. INFANTE QUIRPA, SABA RAFAEL , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Tutor	Dr. INFANTE QUIRPA, SABA RAFAEL , Ph.D.	10,0
Presidente Tribunal De Defensa	Dr. FONSECA DELGADO, RIGOBERTO SALOMON , Ph.D.	10,0
Miembro Tribunal De Defensa	Dr. CAMACHO , FRANKLIN , Ph.D.	10,0

Lo que da un promedio de: **10 (Diez punto Cero)**, sobre 10 (diez), equivalente a: **APROBADO**

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

RIVADENEIRA PEREZ, ERIKA LIZBETH
Estudiante

Dr. FONSECA DELGADO, RIGOBERTO SALOMON , Ph.D.
Presidente Tribunal de Defensa

CERTIFICACIÓN ELECTRÓNICA
BANCO CENTRAL DEL ECUADOR
Firmado Digitalmente por: RIGOBERTO SALOMON FONSECA DELGADO
Hora oficial Ecuador: 13/07/2020 09:54

Dr. INFANTE QUIRPA, SABA RAFAEL , Ph.D.
Tutor

SABA RAFAEL INFANTE QUIRPA
Firmado digitalmente por SABA RAFAEL INFANTE QUIRPA
Fecha: 2020.07.13 10:08:48 -05'00'

UNIVERSIDAD YACHAY TECH

Dr. CAMACHO , FRANKLIN , Ph.D.
Miembro No Tutor

FRANKLIN JOSE CAMACHO
Firmado digitalmente por FRANKLIN JOSE CAMACHO
Fecha: 2020.07.13 15:42:10 -05'00'

TORRES MONTALVÁN, TATIANA BEATRIZ
Secretario Ad-hoc

Autoría

Yo, **Erika Lizbeth Rivadeneira Pérez**, con cédula de identidad 1723987952, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autor(a) del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, Julio 2020.

Erika Lizbeth Rivadeneira Pérez
CI: 1723987952

Autorización de publicación

Yo, **Erika Lizbeth Rivadeneira Pérez**, con cédula de identidad 1723987952, cedo a la Universidad de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Urcuquí, Julio 2020.

Erika Lizbeth Rivadeneira Pérez
CI: 1723987952

Dedication

I dedicate this work to my beloved family, friends and those who contributed to my training both professionally and as a human being.

Acknowledgements

During this journey at Yachay Tech University, many people accompanied me and supported me. I would like to express the greatest gratitude, especially to my professor Aracelis Hernández Ph.D. due to her professional support and guidance. That was essential for the realization of this project, without her this work would not have been possible.

Besides, I thank my professors at the School of Mathematical and Computational Sciences which endowed me with knowledge and made me fall in love with mathematics. I also thank professor Saba Infante Ph.D. who taught the Bayesian statistics course and aroused my interest in Bayesian theories, the basis of this project.

A special thanks to my dear friend Pedro Cajilima due to his contribution to the translation of the work and his enormous support during this trajectory.

I owe a lot to my family and especially my mother for her love and eternal support. I feel greatly grateful and very fortunate to have this wonderful family and friends, who accompanied me during my entire career.

Resumen

Las compañías de seguro estiman los modelos de riesgo para predecir la magnitud de los siniestros y así poder determinar el valor de las primas que deben cobrar al asegurado con el objetivo de evitar pérdidas en el futuro. La frecuencia de los siniestros y la severidad exigida a las predicciones, involucra la consideración de muchos factores, tales como factores de regulación, factores demográficos, factores geográficos, entre muchos otros. Además la experiencia y opinión de los expertos en el área de seguros también deben de tomarse en cuenta. Bajo el paradigma Bayesiano se tiene la ventaja de poder tomar en consideración todos estos factores. En este sentido, el objetivo de este trabajo es proponer un modelo estadístico jerárquico de riesgo, bajo el contexto Bayesiano, para el número de siniestros en seguros clasificados por grupos de edad, región de residencia y horizonte temporal del seguro. La predicción estará basada en la información observada en el pasado, en las suposiciones a priori acerca de la población asegurada, y en el número y frecuencia de los siniestros. El crecimiento de la población asegurada se basará en un modelo de crecimiento exponencial generalizado (GEGM) que toma en cuenta los efectos aleatorios de la edad, la región de residencia, y el horizonte temporal del seguro. Se asumirá que la frecuencia de los siniestros, para cada grupo clasificado, sigue una distribución Gamma mientras que el número de siniestros sigue una distribución Poisson compuesta. La estimación de los parámetros del modelo se hará usando métodos de Monte Carlo por Cadenas de Markov (MCMC), y se probará la efectividad del modelo ajustado. Posteriormente, se estimarán el valor de las primas en base a las predicciones del modelo ajustado y al uso de dos medidas de riesgo en conjuntos con diversos principios de primas.

Palabras clave: Compañías de seguro; Modelos de Riesgo; Modelo Jerárquico Bayesiano; Métodos de Monte Carlo por Cadenas de Markov (MCMC); Principios de primas

Abstract

Insurance companies estimate the risk models to predict the magnitude of the claims and thus be able to determine the premiums that must be paid to the insured in order to avoid future losses. The frequency of the claims and the severity required for the predictions involves the consideration of many factors, such as regulatory factors, demographic factors, and geographical factors, among many others. In addition, the experience and opinion of experts in the insurance area should also be taken into account. Under the Bayesian paradigm there is the advantage of being able to take into account all of these factors. In this sense, the objective of this project is to propose a hierarchical statistical risk model, under the Bayesian context, for the number of insurance claims classified by age groups, residence regions and temporary insurance horizon. The prediction will be based on the information observed in the past, on the a priori assumptions about the insured population, and on the number and frequency of claims. The growth of the insured population will be based on a generalized exponential growth model (GEGM) that takes into account the random effects of age, the region of residence, and the temporary insurance horizon ([1],[2]). It will be assumed that the frequency of claims, for each classified group, follows a Gamma distribution while the number of claims follows a composite Poisson distribution. The estimation of the parameters of the model will be done using Markov Chain Monte Carlo (MCMC) methods. The effectiveness of the adjusted model will then be tested. Subsequently, the value of the premiums will be estimated based on the predictions of the adjusted model and the use of two risk measures in conjunction with different premium principles.

Keywords: Insurance companies; Risk Models; Hierarchical Bayesian Model; Markov Chain Monte Carlo (MCMC) methods; Premium Principles

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Justification

The goal of insurance companies is to produce profits and to generate premiums that allow them to cover the losses due to expenses caused by various reasons depending on the category of the insured item.

Insurance providers are facing challenges of increasing number of claims (frequency) and the amount of each claim (severity) year after year. For insurance companies, it is fundamental to be able to foresee the evolution of claims and, consequently, to be able to facilitate decision-making regarding the value that premiums should have.

The premium charged for an insurance contract for vehicles, health, housing, etc., is based on, among other factors, the age of the person who is getting insured and his medical history, the amount of the deductible, and the insurance plan chosen [3]. Therefore, pricing actuaries must use past information to develop probabilistic models that allow them to model the most important uncertainties involved in the process of losses. For example, in the development of a health insurance model there are several areas that will generate uncertainty: care must be taken in the proper selection of people who do not have a medical history of diseases, as well as solving the conflicting interests between the needs of the doctor, the owner of the insurance policy and the insurance company, for example whether in a service payment plan a doctor can order unnecessary diagnostic tests to increase their income and at the same time be protected against lawsuits for malpractice.

Therefore, insurers must establish a statistical control model that allows for the reduction of unnecessary expenses [4]. Nevertheless, many actuarial models do not adequately address uncertainties such as those that arise in the estimation of parameters [5].

Predictions of the frequency and severity of claims imply the consideration of many uncertain factors such as demographic and geographic factors, regulation of healthcare, etc. After years of experience, it is most likely that insurance experts have a deep understanding in some, if not all, of these areas. Consequently, for these problems associated with uncertainty and the setting of premium values, the opinion of the experts must be considered and, in this sense, Bayesian statistics allows us to include this experience of the experts through the preliminary information that is reflected in the choice of the a priori distributions that each of the parameters will have, in addition to being able to include all the information of the sector.

In this sense, the purpose of this paper is to develop a methodology, under the Bayesian paradigm, allowing predictions to be made of the total future amounts of claims in order to determine the rate of premiums using a Bayesian hierarchical structure.

The prediction of future claims is a key point in risk measurement for insurance providers. According to Migon and Moura (2005), the total of claims is related to the insured population and the number of claims of that population during a given period of time. The people insured in different age groups have different patterns in the frequency and severity of the reported claims and therefore it is reasonable to classify the insured by age groups with the idea of being able to predict the situation of claims in each unit of time, such as by year. This research work proposes an extension of the model introduced by Migon and Moura (2005) by introducing an additional category with the idea of being able to describe the regions of residence of the insured. This spatial factor comes to represent the combined random effect of many elements on the severity and frequency of claims, such as the level of education, the ability to access medical services, the economic level, and even weather conditions. Each of these elements can potentially influence the behavior of the claims, so modeling each of these elements separately is redundant and unnecessary. In this way, introducing a spatial factor that is independent of age classification is practically feasible and understandable.

In Migon and Moura (2005) a generalized collective risk model was proposed under the Bayesian paradigm to determine a health insurance premium. This premium was determined based on historical information about the number and volume of claims, and the population at risk was classified according to age. The proposed model assumes that the total amount of claims depends on the age and it also assumes that the a priori distributions are distributed hierarchically according to age.

Migon and Penna (2006) applied a similar methodology to two sets of real data and discussed the implementation of a collective risk model under Bayesian methodology with stochastic simulation techniques. The value of the premium is given by the maximization of the utility expected by the insurance company while assuming that the insured population follows a non-linear growth model called the generalized exponential growth model (GEGM) (Migon and Gamerman, 1993). This class of models allows for the processing of data with non-negative and non-decreasing means.

Souza et al. (2009) proposed a method to predict population growth in small areas from population census data. Given that the growth pattern of the population of a region may be related to the level of development in the surrounding neighboring regions, these researchers proposed a hierarchical spatial model associated with hyperparameters. On the other hand, [6] discussed the statistical methods under the Bayesian context for the average number and size of claims. The value of the premium was calculated based on the total size of claims, analyzing the frequency and size of the claims separately as it is proposed in [7], assuming a spatial model of Poisson regression for the frequency of the claims and a Gamma model for the average size of claims per insured. The regression model for spatial data includes the random correlated spatial effects that describe the underlying spatial dependence pattern. These spatial dependencies were modeled using a Gaussian conditional autoregressive model (CAR) introduced by [8]. These CAR models are based on the assumption that adjacent regions share similar characteristics and therefore have strong spatial dependencies.

As for the principles of premiums, these have been widely discussed in literature such as Young (2004) and Goovaerts et al. (2010), among others. [9] describes three methods that actuaries use to design premium principles and it lists the common principles of premiums along with an analysis of the desired properties of these principles. On the other hand, [10] and [11] consider the Value at Risk (VaR) and the Tail Value at Risk (TVaR) as the approaches that allow to determine the premium, in addition they establish a detailed discussion of their properties and applications.

1.2 Contribution

Given the importance of insurance companies having a risk measure, this paper proposes to estimate the total number of claims in a specific category under a hierarchical Bayesian structure in a given unit of time that takes into account the insured population, classified according to age, adding a spatial factor that represents the region of residence. This spatial factor comes to represent the combined random effect of different elements that together characterize the frequency and severity of accidents, such as levels of education, the ability to access medical services, socio-economic levels, and even meteorological conditions.

Under the Bayesian paradigm, the Markov Chain Monte Carlo (MCMC) methods will be used to estimate the parameters of the model from the distributions posterior to each one of them, taking into account the a priori distributions which involve the experience and knowledge of the experts, plus historical information on insurance in different areas.

Finally, once the prediction of the number of claims for a given unit of time is made, it is proposed to calculate the value of the premiums using different premium principles so that the value of the premium assigned to each insured can be obtained by diving the total value of the premiums by the insured population.

1.3 Thesis overview

This work is divided into seven main Chapters named as follows: Introduction, Objectives, Theoretical Framework, Methodology, Simulation Studies and Model Fitting, Predictions and Premium Determination, and Conclusions.
In section 1.1, the problem statement, the justification of this work are presented. Also, the scientific contributions of this research is presented in section 1.2.
Chapter 2 states the general and specifics objectives of the project.
In chapter 3 a bibliographic review was carried out in order to obtain a solid theoretical framework with the aim of proposing the desired model.
Chapter 4 establishes the collective risk model together with the spatial effect. After this, the a priori distributions of each parameter involved in the model are presented. From here, the likelihood of the data is calculated and consequently the posterior distributions of the parameters are achieved.
On the other hand, chapter 5 presents the results of the model fitting and the simulation studies.
In chapter 6 are presented: the Bayesian theory of the prediction algorithm, followed by the predictive results for the numerical example presented in Chapter 5. Finally, it demonstrates ways to determine the premium based on the claim amounts provided under certain premium principles.
In Chapter 7, the conclusions obtained from this work are presented. Also, future works that can improve the proposed methodology and help to establish open issues are mentioned.

Chapter 2

Objectives

2.1 General Objective

To predict the total number of claims using a hierarchical Bayesian model as a risk measure for insurance companies and to calculate the value of premiums under different principles.

2.2 Specific Objectives

The next specific objectives will be followed in order to achieve the main goal.

- Categorize the insured population by age classes in a specific unit of time. Add a spatial factor corresponding to the region of residence that represents the combined random effects of elements that influence the characterization of the claims.
- Implement a Markov Chain Monte Carlo (MCMC) algorithm that allows for the estimation of each of the parameters of the proposed model using a priori knowledge, given by the experts in the matter, and for the information provided by the historical data of insurance companies in different items.
- Based on the number of estimated claims, calculate the value of the premiums according to the different premium principles.

Chapter 3

Theoretical Framework

3.1 The Collective Compound Risk Model

The collective risk model is well known and discussed intensively in the actuarial field. Consider a portfolio of single-type policies. Let N be the total number of claims that arise from a risk in a given period of time and Z_j denote the amount of the $j-$th claim. The total amount of claims is given by

$$X = \sum_{j=1}^{N} Z_j, \tag{3.1}$$

with $X = 0$ when $N = 0$. The main assumptions of this model are

- The amounts of individual claims Z_j are identically distributed and random independent positive variables.
- The total number of claims N is a random variable independent from the amounts of claims Z_j.

The advantage of the collective risk model is that the frequency and severity of the claim can be modeled separately. For example, a general increase in the cost of medications may affect the severity of the claim but have little influence on the frequency of the claim, while the introduction of another line of business would increase the frequency of the claim without altering much the severity of the claim. In addition, the measure of the expected value (and variance) of the amount of the added claim can be decomposed by measuring the average and the variance of the frequency and severity of the claim, that is,

$$E(X) = E(E[X|N]) = E(N)E(Z), \tag{3.2}$$

$$Var(X) = E(V[X|N]) + V(E[X|N]) = E(N)V(Z) + [E(Z)]^2V(N). \tag{3.3}$$

When N follows a Poisson distribution with parameter λ, it is said that in (3.1) X follows a Poisson distribution composed with parameters λ and F, where $F(x) = P(Z_1 \leq z)$

denotes the function for distributing individual amounts of claims. From (3.3) it follows that in this case,

$$E(X) = \lambda E(Z),$$
$$Var(X) = \lambda V(Z) + [E(Z)]^2\lambda = \lambda E(Z^2)$$

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

3.2 The Compound Poisson Process

Let $X_1, X_2, ..., X_n$ be random i.i.d. variables and let N be a random variable independent of the X_n and whose possible values are all integers. Then, the variable

$$S_N = \sum_{k=1}^{N} X_k$$

is a compound random variable.

To calculate the expected value and variance of S_N, remember that the conditional expected value of X given $Y = y$ is defined as

$$E(X|Y = y) = \sum_{j=1}^{\infty} x_j P_{X|Y}(x_j|y)$$

for the discrete case and

$$E(X|Y = y) = \int_{j=1}^{\infty} x_j P_{X|Y}(x_j|y)dx$$

for the continuous case.

Then, $E(g(x))$ is a constant, but $E(g(x)|Y = y)$ is a function of y, where y is a particular value of the random variable Y. Furthermore, $E(g(x)|Y)$ is a random variable whose average can be calculated. By the properties of the expected value and the probability distributions of a random variable established in [12] we have the following propositions:

Proposition 1.

$$E(g(x)) = E(E(g(x)|Y)) \tag{3.4}$$

From this proposition it can be deduced that

$$E(X) = E(E(X|Y)) = \begin{cases} \sum_{k=1}^{\infty} E(X|Y = y_k)P_Y(y_k) \\ \int_{k=1}^{\infty} E(X|Y = y_k)f_Y(y)dy \end{cases}$$

and also,

$$Var(X) = E(E(X^2|Y)) + [E(E(X|Y))]^2.$$

Now, if $X_1, X_2, \ldots$ are random i.i.d. variables, then it follows that

$$\begin{aligned} E(X_k) &= E(X) \quad \forall k = 1, 2, ... \\ Var(X_k) &= Var(X) \end{aligned}$$

and let N be a random variable independent of the X_k with values $1, 2, \ldots,$ then by (3.4) it follows that

$$\begin{aligned} E(S_N) &= E\left(\sum_{k=1}^{N} X_k\right) \\ &= E\left(E\left(\sum_{k=1}^{N} X_k|N\right)\right) \\ &= E(N)E(X) \end{aligned}$$

$$\begin{aligned} Var(S_N) &= Var\left(\sum_{k=1}^{N} X_k\right) \\ &= E(N)Var(X) + Var(N)(E(X))^2 \end{aligned}$$

Let's see the demonstration of the formulas

Proposition 2.

$$E(S_N) = E\left(\sum_{k=1}^{N} X_k\right) = E(N)E(X_1) \tag{3.5}$$

$$Var(S_N) = Var\left(\sum_{k=1}^{N} X_k\right) = E(N)Var(X_1) + Var(N)(E(X))^2 \tag{3.6}$$

Demonstration.

$$E(S_N) = E\left(\sum_{k=1}^{N} X_k\right) = \sum_{k=1}^{N} E(X_k) = nE(X_1)$$

Now with N independent of the X_k's, we can write

$$E\left(\sum_{k=1}^{N} X_k | N = n\right) = E\left(\sum_{k=1}^{n} X_k\right) = nE(X_1)$$

or

$$E\left(\sum_{k=1}^{N} X_k | N\right) = NE(X_1)$$

Then, by proposition 1.

$$\begin{aligned}
E(S_N) &= E\left(\sum_{k=1}^{N} X_k\right) \\
&= E\left(E\left(\sum_{k=1}^{N} X_k | N\right)\right) \\
&= E(NE(X_1)) \\
&= E(N)E(X_1)
\end{aligned}$$

although to reach this result it is not necessary for the X_k's to be independent among themselves.

Moreover, with X_k being i.i.d.

$$\begin{aligned}
Var\left(\sum_{k=1}^{N} X_k | N = n\right) &= nVar(X_1) \\
\Rightarrow Var(S_N|N) = Var\left(\sum_{k=1}^{N} X_k | N\right) &= NVar(X_1)
\end{aligned}$$

Now with the help of the conditional variance

$$\begin{aligned}
Var(X|Y) &= E[(X - E(X|Y))^2|Y] \\
&= E[(X^2 - 2XE(X|Y) + (E(X|Y))^2)|Y] \\
&= E(X^2|Y) - 2(E(X|Y))^2 + (E(X|Y))^2 \\
&= E(X^2|Y) - (E(X|Y))^2
\end{aligned}$$

Taking the expected value from both sides we get:

$$\begin{aligned}
&\Rightarrow E(Var(X|Y)) = E(E(X^2|Y)) - E((E(X|Y))^2)\\
&\Rightarrow E(Var(X|Y)) = E(X^2) - E((E(X|Y))^2)\\
&\Rightarrow E(Var(X|Y)) = Var(X) + (E(X))^2 - E((E(X|Y))^2)\\
&\Rightarrow E(Var(X|Y)) = Var(X) + (E(E(X|Y)))^2 - E((E(X|Y))^2)\\
&\Rightarrow E(Var(X|Y)) = Var(X) - Var(E(X|Y))
\end{aligned}$$

Then, we get that

$$\Rightarrow Var(X) = E(Var(X|Y)) + Var(E(X|Y)) \tag{3.7}$$

Now, using (3.7) in (3.6) we obtain:

$$\begin{aligned}
Var(S_N) &= E(Var(S_N|N)) + Var(E(S_N|N))\\
Var\left(\sum_{k=1}^{N} X_k\right) &= E(NVar(X_1)) + Var(NE(X_1))\\
&= E(N)Var(X_1) + (E(X_1))^2 Var(N)
\end{aligned}$$

Now, consider that $\{N(t), t \geq 0\}$ is a Poisson process with a rate of λ and let $X_1, X_2, ...$ be a random i.i.d. variable independent of $\{N(t), t \geq 0\}$, then the stochastic process $\{Y(t), t \geq 0\}$ defined as

$$Y(t) = \sum_{k=1}^{N(t)} X_k \quad \forall t \geq 0$$

and

$$Y(t) = 0 \quad si \quad N(t) = 0$$

is called the compound Poisson process.

A Poisson process $\{N(t), t \geq 0\}$ only counts the number of events that occur in an interval $[0, t]$, while the process $\{Y(t), t \geq 0\}$, for example, gives the length of phone calls that occur in $[0, t]$, or the number of people who are involved in a traffic accident in that interval $[0, t]$, etc. For this compound Poisson process it must be assumed that the X_k are i.i.d. and so the two-dimensional stochastic process

$$(N(t), Y(t)), t \geq 0$$

can be considered to retain all the information of interest.

Now, using **proposition 2**:

$$\begin{aligned}
E(Y(t)) &= E(N(t))E(X_1) = \lambda t E(x_1) \\
Var(Y(t)) &= E(N(t))Var(X_1) + Var(N(t))(E(X_1))^2 \\
&= \lambda t(Var(X_1)) + \lambda t(E(X_1))^2 \\
&= \lambda t(Var(X_1) + (E(X_1))^2) \\
&= \lambda t E(X_1^2))
\end{aligned}$$

We can also calculate the moment generating function (m.g.f.) of $Y(t)$:
If $M_1(s) = M_{X_1}(s) = E(e^{sX_1})$ then

$$M_{Y(t)}(s) = E(e^{sY(t)}) = E(e^{s(X_1+\ldots+X_{N(t)})})$$

by **proposition 1**

$$= E(E(e^{s(X_1+\ldots+X_{N(t)})}|N(t)))$$

by the definition of an expected value we have that

$$\begin{aligned}
&= \sum_{n=0}^{\infty} E(e^{s(X_1+\ldots+X_{N(t)})}) \cdot P(N(t) = n) \\
&= \sum_{n=0}^{\infty} (M_{X_1}(s))^n \cdot \frac{e^{-\lambda t}(\lambda t)^n}{n!} \\
&= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(M_{X_1}(s)\lambda t)^n}{n!} \\
&= e^{-\lambda t} e^{M_{X_1}(s)\lambda t} \\
&= e^{\lambda t(M_{X_1}(s)-1)}
\end{aligned}$$

The formula of the m.g.f. allows for the deduction or demonstration of the formula of the expected value for $Y(t)$:

$$\begin{aligned}
E(Y(t)) &= \frac{\partial(M_{Y(t)}(s))}{\partial s} \\
&= \frac{\partial(e^{\lambda t(M_{X_1}(s)-1)})}{\partial s} \\
&= \lambda t M'_{X_1}(s) e^{\lambda t(M_{X_1}(s)-1)}|_{s=0} \\
&= M_{Y(t)}(0) \lambda t M'_{X_1}(0)
\end{aligned}$$

On the other hand, we have that

$$\begin{aligned}
M_{Y(t)}(0) &= e^{\lambda t(M_{X_1}(0)-1)} \\
&= e^0 = 1
\end{aligned}$$

and

$$M_{X_1}(0) = E(e^{0X_1}) = 1$$

Therefore, we get that

$$E(Y(t)) = \lambda t M'_{X_1}(0) = \lambda t E(X_1)$$

When X_1 is a discrete random variable, with possible values $1, 2, ..., j$, we can write

$$Y(t) = \sum_{i=1}^{j} i N_i(t)$$

where $N_i(t)$ is the number of the random variable X_k (associated with some random event) that takes the values of i in the interval $[0, t]$.

Proposition 3.
The stochastic process $\{N_i(t), t \geq 0\}$ is a Poisson process with rates $\lambda_i = \lambda p$; for $i = 1, 2, ..., j$ so then $\{N_i(t), t \geq 0\}$ is an independent Poisson process with rates $\lambda p_{X_1}(i)$ for $i = 1, 2, \ldots, j$.

This representation of the process $\{Y(t), t \geq 0\}$ can be generalized in the case where X_1 is an arbitrary discrete random variable.

In the discrete case, the m.g.f. of the random variable $Y(t)$ is

$$\begin{aligned} M_{Y(t)}(s) &= E(e^{sY(t)}) \\ &= E(e^{s(1N_1(t)+...+jN_j(t))}) \\ &= e^{\lambda t \sum_{i=1}^{j}[(e^{si}-1)p_{X_1}(i)]} \end{aligned}$$

Since $\lim_{t\to\infty} N(t) = \infty$, then by the central limit theorem we have the following proposition.

Proposition 4.
For t large enough, we can write

$$Y(t) \approx N(\lambda t E(X_1), \lambda t E(X_1^2))$$

Recall that a good approximation occurs when number of variables in the sum is ≥ 30, or even a little less depending on the degree of asymmetry in the distribution of the random variable X_1 with respect to its average.

Finally, let $\{Y_1(t), t \geq 0\}$ and $\{Y_2(t), t \geq 0\}$ be two compound Poisson processes, defined by

$$Y(t) = \sum_{k=1}^{N_i(t)} X_{i,k} \qquad \forall t \geq 0 \quad with \quad Y_i(t) = 0 \quad if \quad N_i(t) = 0$$

where $\{N_i(t), t \geq 0\}$ is a Poisson process with rate $\lambda_i, i = 1, 2$.

We know that $N(t) = N_1(t) + N_2(t), \forall t \geq 0$ is a Poisson process with rate

$$\lambda = \lambda_1 + \lambda_2$$

because $N_1(t)$ and $N_2(t)$ are independent Poisson processes.
Let X_k be the random variable associated to the $k-$th event of the process $\{N(t), t \geq 0\}$, then we can write

$$X_k = \begin{cases} X_{1,k}, & p = \frac{\lambda_1}{\lambda_1+\lambda_2} \\ X_{2,k}, & 1-p \end{cases}$$

That is, X_k has the same distribution of $X_{1,k}$ with probability p or of $X_{2,k}$ with probability $1-p$; therefore, we have that

$$P(X_k \leq x) = P(X_{1,k} \leq x) \cdot p + P(X_{2,k} \leq x)(1-p)$$

Since the random variables X_1, X_2 are i.i.d. and independent of the Poisson process $\{N(t), t \geq 0\}$, then the process $\{Y(t), t \geq 0\}$ defined by

$$Y(t) = Y_1(t) + Y_2(t) \qquad t \geq 0$$

is also a Poisson process.

3.3 Bayesian Inference

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

In this section, the fundamental Bayesian paradigm is presented. The most advanced applications are discussed in the next section. See Klugman (1992) [13] for further discussion on Bayesian statistics in actuarial sciences.

Previous beliefs about the values for d parameters of interest $\theta = (\theta_1, \theta_2, ..., \theta_d), d > 0$ can be expressed by the probability density function $\pi(\theta)$, representing our opinion about the possible values of θ and the relative possibilities of being a true parameter. Suppose that it is possible to obtain n observations, in other words, $X = (x_1, x_2, \ldots, x_n)$ whose joint density function is defined as $f(X)$. Denote $l(X|\theta)$ as the likelihood function, $\pi(\theta, X)$ as the joint density function of θ and X, and $\pi(\theta|X)$ 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

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

Where the distribution $\pi(\theta|X)$ is proportional to the a priori density function times the likelihood function, which summarizes the modified beliefs of the parameter according to the observations. If the posterior distribution follows a known distribution (for example, the Gamma distribution), it can be modeled or simulated with few difficulties. However, it is actually quite common to find parameters with unrecognizable posterior distributions, especially in models with a high dimension. One of the predominant methods is to use Markov chain Monte Carlo (MCMC) as an approximation algorithm. In general terms, a MCMC algorithm allows users to simulate samples of the posterior distribution when direct generation is complicated or impossible. The most 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 of complete conditional distributions, denoted as $\pi(\theta_i|\theta_{-i}, X), \quad i = 1, 2, ..., d,$ where θ_{-i} represents every parameter in θ but θ_i. Further discussion can be found in Gamerman (1997) [14].

3.4 Generalized Exponential Growth Model (GEGM)

Assume that the observations Y_t are modeled through a probability distribution in the exponential family with an average response function

$$\mu_t = E[Y_t|\theta_t]$$

where θ_t is a vector of parameters. A broad class of exponential growth models is characterized by the parameterization (a, b, γ, λ), and is defined as:

$$\mu_t = [a + be^{\gamma t}]^{1/\lambda}$$

An important advantage of this approach is maintaining the measurements at the original scale, making interpretation simpler. Some special and known cases in literature are:

1. *Logistic*: when $\lambda = -1$, $\mu_t^{-1} = a + be^{\gamma t}$
2. *Gompertz*: when $\lambda = 0$, (1) is defined as $ln(\mu_t) = a + be^{\gamma t}$
3. *Modified Exponential*: when $\lambda = 1$, $\mu_t = a + be^{\gamma t}$

The process of interest always has a non-negative, non-declining μ_t function.
Assume that π_t, the size of the population in a time interval t, characterized by the parameterization (a, b, γ, λ), is modeled by a probability distribution in the exponential family with average μ_t, that is

$$\pi_t \sim Exp(\mu_t)$$

with

$$\mu_t = [a + be^{\gamma t}]^{1/\lambda}, \qquad t \geq 0$$

3.5 Diagnosis

There are many useful diagnoses to analyze the results of a chain, and since none of them can guarantee that it works, Sinharay recommends that several of the many available techniques be used. It is also necessary to guarantee the convergence of all the parameters involved. In this project it is considered the following tests:

3.5.1 MCMC plots

For the basic monitoring of simulated chains, graphs reflecting their sequential behavior are used. The following plots are generated for each parameter considered:

- **Plot of the string values in the form of a time series:** Better known as trace plots, they are used to make sure that the a priori distributions are well calibrated which is indicated by the parameters having sufficient state changes as the MCMC algorithm runs.
- **Plot of the estimated density from these values**
- **Autocorrelation plots:** The autocorrelation plots obtained from the simulated samples show that the samples can be effectively independent when observing the behavior of autocorrelation of each of the simulated chains for each parameter. If a slow zero decay is shown, it may be indicative of poor mixing, which may suggest reparametrization or some other approximation.

3.5.2 Heidelberger and Welch convergence test

Once the simulated traces are obtained from the subsequent distribution of a parameter of interest; It is necessary to make a long-term convergence diagnosis to test the null hypothesis that the simulated samples come from a stationary distribution. One of these tests is the convergence diagnosis by Heidelberger and Welch (1983) which uses the Cramer-Von-Mises statistic.

The test is applied successively, first to the entire chain, then, after discarding the first 10%, 20%. . . , of the chain until the null hypothesis is accepted, or 50% of the chain is discarded. If the test fails, the seasonality is not met and indicates that a longer simulation is necessary. If on the contrary the test passes, the number of burned tensions and the number of iterations that a trace with stationary distribution maintains.

Chapter 4

Methodology

4.1 Hierarchical Collective Risk Model

One of the main problems faced by the insurance industry is evaluating and determining the optimal premium. The premium is normally evaluated based on past information in terms of the severity of the claim, the frequency of the claim, and the information of the insured.

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

Later, spatial parameters were introduced into this structure. The demographic characteristics can not be ignored when the population grows due to the fact that they significantly influence the health insurance industry and, therefore, the tendency in the insured population with respect to the demography should be considered when the severity and frequency of the claims are modeled.

The basic collective risk model can be extended to incorporate factors of age, time and region. Age is one of the important factors that influence the mortality and health status of the insured. However, establishing premiums for each age is redundant since people of similar ages (for example, the ages of 26 and 29 years) show similar mortality and health conditions, given that everything else be the same. It is convenient for the health insurance provider to classify the insured by age classes, denoted by

$$a = 1, 2, \dots, A$$

The insurer has full freedom in terms of age classification, as well as the number of ages incorporated into each class. On the other hand, there is no restriction regarding the unit of time, this could be annual, quarterly, monthly or personalized. Generally the frequency of data collection can be a fair reflection of the unit of time.

Migon and Moura (2005) work the collective risk model for a portfolio of policies classified by age classes $a = 1, ..., A$ an by time $t = 1, ..., T$. This model is extended considering the region of residence of the insured as another category of classification, indicated as $i = 1, 2, ..., I$. Then, $(N_{t,i,a}, X_{t,i,a})$ denote, respectively, the total number of claims and the added quantity of claims produced by a policy portfolio in a given period of time $t = 1, 2, ..., T$ for an age class a in a region i.

The collective compound risk model, by Cramer and Lundberg, is given by

$$X_{t,i,a} = \sum_{j=1}^{N_{t,i,a}} Z_{t,i,a,j}$$

with $i = 1, .., , I$, $t = 1, ..., T$ and $a = 1, ..., A$, in the time interval $(t-1, t)$. Where $Z_{t,i,a,j}$ is the amount of the $j-$th claim that occurred within the time interval $(t, t-1)$ for an age class a in a region i.

The main assumptions in the Cramer and Lundberg process are:

1. The number of claims in the interval $(t-1, t)$ is a random variable denoted as $N_{t,a}$
2. For a fixed t, i, a with $N_{t,i,a} = n_{t,i,a}$, the amount of the $j-$th claim $Z_{t,i,a,j}, j = 1, 2, ..., n_{t,i,a}$ are identically distributed, random independent positive variables with a finite average $\mu_a = E[Z_{t,i,a,j}]$ and variance $\sigma_a^2 = var(Z_{t,i,a,j}) < \infty$
3. The time of claim occurs at random instances $t_{1,a} \leq t_{2,a} \leq ...$ and the time between arrivals $T_{k,a} = t_{k,i,a} - t_{k-1,a}$ are assumed to be identically and exponentially distributed, random independent variables with a finite average $E[T_{k,a}] = \lambda_a^{-1}$.

Assuming that the sequences T_k and Z_k are independent of each other and that the conditions mentioned above are met [15], it follows that

$$Z_{t,i,a} \sim Gamma(\kappa_a, \theta_a) \quad with \quad \kappa_a > 0, \theta_a > 0 \tag{4.1}$$

and

$$N_{t,i,a} \sim Poiss(M_{t,i,a}\lambda_a) \quad \lambda_a > 0 \tag{4.2}$$

where $Z_{t,i,a}$ is the individual amount of the claim and $M_{t,i,a}$ is the number of people insured in a time t for an age class a in the region i, λ_a is the average number of claims per individual per unit of time. $M_{t,i,a}$ implies a constant population in the time interval $(t-1, t)$ since the growth of the population is not modeled in this model.

Then, the sum of these Gamma is also a Gamma, that is

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

In (4.1) we see that the individual claim $Z_{t,i,a} \sim Gamma(\kappa_a, \theta_a)$ is the only variable that depends on age in this model. An extension of the model is to see the feasibility of having κ and θ also depend on time t and region i.

On the other hand, in (4.2), $M_{t,i,a}\lambda_a$ represents the average number of claims made by the insured population in age class a and the region i in time t. Given that for a certain age class a and region i the insured population $M_{t,i,a}$ varies over time, then the total number of claims

$$\left\{X_{t,i,a}, \quad t = 1, ..., T, \quad i = 1, ..., I, \quad a = 1, ..., A\right\}$$

are not identically distributed.

It should be taken into account that insurance companies normally keep information of the insured, such as the number, amount, and time of the claims made, as well as the age and region of residence of the claimants. Therefore, the total amount of the claims X, the total number of claims N, and the insured population M are assumed to be observed and, therefore, are considered model entries.

4.2 Spatial Effect

According to the work of 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

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

with precision τ and average

$$\mu_{t,i,a} = \beta_{a_0} + L_i + \beta_1 e^{t\beta_{a_2}} \tag{4.3}$$

where L_i represents the spatial factor for the region i and β_{a_0}, β_{a_2} are parameters related to age [14]. It's important to mention that the age of the insured population is independent of the information related to the region. The knowledge about age and demographic characteristics should be taken into account in the a priori distributions. The parameters related to age are specified as:

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

where τ_{ε_0} and τ_{ε_2} follow the Gamma distributions with known parameters. The age-related factors β_{a_0} and β_{a_0} vary with age but show the same average. It is assumed that hyperparameters follow normal distributions with different parameters, that is

$$\beta_0 \sim N(\mu_0, \tau_0^{-1})$$
$$\beta_1 \sim N(\mu_1, \tau_1^{-1})$$
$$\beta_2 \sim N(\mu_2, \tau_2^{-1})$$

where $\mu_0, \mu_1, \mu_2, \tau_0, \tau_1, \tau_2$ are unknown values.

The spatial factor, based on the work of Gschlöbl and Czado (2007), [16], is assumed to follow a normal multivariate distribution, i.e.

$$L \sim NMV(0, \sigma^{-1}Q^{-1}) \tag{4.4}$$

where the $(g, h)-$th element of the spatial precision matrix is given by

$$Q_{gh} = \begin{cases} 1 + |\eta| \cdot m_g, & g = h \\ -\eta, & g \neq h, g, h = 1, 2, ..., I \\ 0, & \text{otherwise} \end{cases} \tag{4.5}$$

The precision matrix describes three types of positions for the pair of regions g and h:

- If region g coincides with region h

$$g = h$$

- The two regions are neighboring and share a common border

$$g \sim h$$

- The two regions do not share any common border

The amount m_g denotes the number of regions neighboring the region g. Spatial effects are adequately described by CAR prioris based on the work of Pettitt et al, (2002), [8]. The η is called the degree of spatial dependence. If $\eta = 0$, this indicates the independence of the spatial effects and if η has a large value this signifies a strong spatial dependence.

To assign η its own priori it is known that a non-negative correlation is expected between two regions

$$\Rightarrow \eta \geq 0$$

A Pareto distribution with parameters $(1, 1)$ and density function

$$\frac{1}{(1+\eta)^2}$$

is selected in such a way that it takes large values for small values of η. (See Gschlöbl and Czado (2007), [6])

4.3 A Priori Distributions

Using improper prioris can cause computational problems such as the inability to obtain posterior distributions. In this work the prioris were chosen in such a way that they are their own but are relatively less informative, they are called reference prioris.

Some of the variables have already been assigned prioris based on practical knowledge or experience. The remaining variables are assigned reference prioris since there is not enough information.

$$\lambda_a \sim Gamm(\alpha_\lambda, \beta_\lambda)$$
$$\theta_a \sim Gamm(\alpha_\theta, \beta_\theta)$$
$$\kappa_a \sim Gamm(\alpha_\kappa, \beta_\kappa)$$

These distributions are independent with $\alpha_\lambda, \beta_\lambda, \alpha_\theta, \beta_\theta, \alpha_\kappa, \beta_\kappa$ non-negative amounts.

1. Distributions that describe the value of the claims, the number of claims and the insured population:

$$\begin{aligned} X_{t,i,a} &\sim Gamm(n_{t,i,a}\kappa_a, \theta_a), \quad \theta_a > 0, \kappa_a > 0, \\ N_{t,i,a} &\sim Poiss(M_{t,i,a}\lambda_a), \quad \lambda_a > 0, \\ M_{t,i,a} &\sim N(\mu_{t,i,a}, \tau^{-1}), \end{aligned}$$

   where

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

2. Distributions that describe the age and the regions

$$\begin{aligned} \theta_a &\sim Gamm(\alpha_\theta, \beta_\theta), \\ \kappa_a &\sim Gamm(\alpha_\kappa, \beta_\kappa), \\ \lambda_a &\sim Gamm(\alpha_\lambda, \beta_\lambda), \\ \beta_{a_0} &= \beta_0 + \varepsilon_a^0, \text{ with } \varepsilon_a^0 \sim N(0, \tau_{\varepsilon_0}^{-1}), \\ \beta_{a_2} &= \beta_2 + \varepsilon_a^2, \text{ with } \varepsilon_a^2 \sim N(0, \tau_{\varepsilon_2}^{-1}), \\ L &\sim NMV(0, \tau^{-1}Q^{-1}), \end{aligned}$$

   where

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

3. Distributions of the a priori hyperparameters

$$\begin{aligned} \beta_0 &\sim N(\mu_0, \tau_0^{-1}) \\ \beta_1 &\sim N(\mu_1, \tau_1^{-1}) \\ \beta_2 &\sim N(\mu_2, \tau_2^{-1}) \\ f(\eta) &= \frac{1}{(1+\eta)^2}, \quad \eta > 0 \end{aligned}$$

and $\psi = (\tau, \sigma, \tau_{\varepsilon_0}, \tau_{\varepsilon_2}, \alpha_\theta, \beta_\theta, \alpha_\lambda, \beta_\lambda, \alpha_\kappa, \beta_\kappa)$ follow Gamma distributions with known parameters

$$\Rightarrow \psi \sim Gamm(\alpha_\psi, \beta_\psi)$$

where $\mu_0, \mu_1, \mu_2, \tau_0, \tau_1, \tau_2, \alpha_\psi, \beta_\psi$ are known values.

4.4 Posterior Distributions

It is assumed that $\kappa_a = 1$, where the amount of individual claim $Z_{t,i,a,j}$ is independent of the time and the region of residence of the insured, i.e.

$$Z_{a,j}.$$

In addition, two adjacent regions, $I = 2$, will be considered. The simplified model explains the implementation of the methodology and then, as an extension, it could be expanded considering more than two adjacent regions.

Let Θ be the parameter vector and $D_t = (\tilde{X}_t, \tilde{n}_t, \tilde{M}_t), t = 1, ..., T$ the available data. Assuming independence in the time, age classes and geographical region, we have the following *likelihood function*:

$$\begin{aligned}
P(\Theta|D_t) &\propto \prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} f(X_{t,i,a}, n_{t,i,a}, M_{t,i,a}|\Theta) \\
&\propto \prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} f(X_{t,i,a}|\theta_a, n_{t,i,a}) \cdot f(n_{t,i,a}|\lambda_a, M_{t,i,a}) \cdot f(M_{t,i,a}) \\
&\propto \prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} \left[\frac{\theta_a^{n_{t,i,a}}}{\Gamma(n_{t,i,a})} X_{t,i,a}^{n_{t,i,a}^{-1}} e^{-X_{t,i,a}\theta_a} \cdot (\lambda_a M_{t,i,a})^{n_{t,i,a}} \cdot \frac{e^{-\lambda_a M_{t,i,a}}}{n_{t,i,a}!} \cdot \sqrt{\tau} e^{-\frac{\tau}{2}(M_{t,i,a}-\mu_{t,i,a})^2} \right] \\
&\propto \prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} \left[(\theta_a\lambda_a)^{n_{t,i,a}} \frac{X_{t,i,a}^{n_{t,i,a}^{-1}} M_{t,i,a}^{n_{t,i,a}}}{\Gamma(n_{t,i,a}) n_{t,i,a}!} \sqrt{\tau} \cdot e^{-\left(\theta_a X_{t,i,a} + \lambda_a M_{t,i,a} + \frac{\tau}{2}(M_{t,i,a}-\mu_{t,i,a})^2\right)} \right]
\end{aligned}$$

Posterior Distribution for θ_a

For a single a we have that

$$\begin{aligned}
P(\theta_a|\Theta_{-\theta_a}, D_t) &\propto P(D_t|\Theta)P(\theta_a) \\
&\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I} \left[\theta_a^{n_{t,i,a}} e^{-\theta_a X_{t,i,a}} \right] \right) \cdot \frac{\beta_\theta^{\alpha_\theta}}{\Gamma(\alpha_\theta)} \theta_a^{\alpha_\theta - 1} e^{-\theta_a \beta_\theta} \\
&\propto \prod_{t=1}^{T}\prod_{i=1}^{I} \theta_a^{n_{t,i,a}+\alpha_\theta-1} \cdot e^{-\theta_a(X_{t,i,a}+\beta_\theta)} \\
&\propto \theta_a^{\alpha_\theta + \sum_{t=1}^{T}\sum_{i=1}^{I} n_{t,i,a} - 1} \cdot e^{-\theta_a(\sum_{t=1}^{T}\sum_{i=1}^{I} X_{t,i,a} + \beta_\theta)} \quad \text{for } a = 1, 2, ..., A
\end{aligned}$$

then

$$P(\theta_a|\Theta_{-\theta_a}, D_t) \propto Gamm(\alpha_\theta + \sum_{t=1}^{T}\sum_{i=1}^{I} n_{t,i,a}, \sum_{t=1}^{T}\sum_{i=1}^{I} X_{t,i,a} + \beta_\theta) \quad \text{for } a = 1, ..., A$$

Posterior Distribution for λ_a

$$\begin{aligned}
P(\lambda_a|\Theta_{-\lambda_a}, D_t) &\propto P(D_t|\Theta)P(\lambda_a) \\
&\propto \prod_{t=1}^{T}\prod_{i=1}^{I} \left(\lambda_a^{n_{t,i,a}} e^{-\lambda_a M_{t,i,a}}\right) \cdot \frac{\beta_\lambda^{\alpha_\lambda}}{\Gamma(\alpha_\lambda)} \lambda_a^{\alpha_\lambda - 1} e^{-\lambda_a \beta_\lambda} \\
&\propto \prod_{t=1}^{T}\prod_{i=1}^{I} \lambda_a^{n_{t,i,a}+\alpha_\lambda - 1} \cdot e^{-\lambda_a (M_{t,i,a} + \beta_\lambda)} \\
&\propto \lambda_a^{\alpha_\lambda + \sum_{t=1}^{T}\sum_{i=1}^{I} n_{t,i,a} - 1} \cdot e^{-\lambda_a (\sum_{t=1}^{T}\sum_{i=1}^{I} M_{t,i,a} + \beta_\lambda)} \quad \text{for } a = 1, 2, ..., A
\end{aligned}$$

then

$$P(\lambda_a|\Theta_{-\lambda_a}, D_t) \propto Gamm(\alpha_\lambda + \sum_{t=1}^{T}\sum_{i=1}^{I} n_{t,i,a}, \sum_{t=1}^{T}\sum_{i=1}^{I} M_{t,i,a} + \beta_\lambda) \quad \text{for } a = 1, ..., A$$

Posterior Distribution for β_0

$$\begin{aligned}
P(\beta_0|\Theta_{-\beta_0}, D_t) &\propto P(D_t|\Theta)P(\beta_0) \\
&\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} e^{-\frac{\tau}{2}(M_{t,i,a} - \mu_{t,i,a})^2}\right) \cdot \frac{\sqrt{\tau_0}}{\sqrt{2\pi}} e^{-\frac{\tau_0}{2}(\beta_0 - \mu_0)^2} \\
&\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} e^{-\frac{\tau}{2}(\mu_{t,i,a}^2 - 2\mu_{t,i,a} M_{t,i,a})}\right) \cdot \frac{\sqrt{\tau_0}}{\sqrt{2\pi}} e^{-\frac{\tau_0}{2}(\beta_0 - \mu_0)^2}
\end{aligned}$$

but we have that

$$\begin{aligned}
\mu_{t,i,a}^2 &= \left(\beta_{a_0} + L_i + \beta_1 e^{t\beta_{a_2}}\right)^2 \\
&= \left(\beta_0 + \varepsilon_a^0 + L_i + \beta_1 e^{t\beta_{a_2}}\right)^2 \\
&= \beta_0^2 + 2\left(\varepsilon_a^0 + L_i + \beta_1 e^{t\beta_{a_2}}\right) + \left(\varepsilon_a^0 + L_i + \beta_1 e^{t\beta_{a_2}}\right)^2
\end{aligned}$$

furthermore,

$$\begin{aligned}
2\mu_{t,i,a} M_{t,i,a} &= 2\left(\beta_{a_0} + L_i + \beta_1 e^{t\beta_{a_2}}\right) \cdot M_{t,i,a} \\
&= 2\left(\beta_0 + \varepsilon_a^0 + L_i + \beta_1 e^{t\beta_{a_2}}\right) \cdot M_{t,i,a} \\
&= 2\beta_0 M_{t,i,a} + 2\left(\varepsilon_a^0 + L_i + \beta_1 e^{t\beta_{a_2}}\right) M_{t,i,a}
\end{aligned}$$

So, taking only what depends on β_0, we obtain

$$\begin{aligned}
P(\beta_0|\Theta_{-\beta_0}, D_t) &\propto \prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A}\left[e^{-\frac{\tau}{2}\left(\beta_0^2+2\beta_0\left(\varepsilon_a^0+L_i+\beta_1 e^{t\beta_{a_2}}\right)-2\beta_0 M_{t,i,a}\right)}\right]\cdot e^{-\frac{\tau_0}{2}(\beta_0^2-2\mu_0\beta_0)} \\
&\propto \prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A}\left[e^{-\frac{\tau}{2}\left(\beta_0^2-2\beta_0\left(M_{t,i,a}-\varepsilon_a^0-L_i-\beta_1 e^{t\beta_{a_2}}\right)\right)}\right]\cdot e^{-\frac{\tau_0}{2}(\beta_0^2-2\mu_0\beta_0)} \\
&\propto e^{-\frac{\tau}{2}TIA\beta_0^2+\tau\beta_0\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}\left(M_{t,i,a}-\varepsilon_a^0-L_i-\beta_1 e^{t\beta_{a_2}}\right)}\cdot e^{-\frac{\tau_0}{2}\beta_0^2-\tau_0\mu_0\beta_0} \\
&\propto e^{-\frac{1}{2}(\tau TIA+\tau_0)\beta_0^2+\beta_0\left(\tau_0\mu_0+\tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}\left(M_{t,i,a}-\varepsilon_a^0-L_i-\beta_1 e^{t\beta_{a_2}}\right)\right)} \\
&\propto e^{-\frac{1}{2}(\tau TIA+\tau_0)\left(\beta_0^2-2\beta_0\left(\tau_0\mu_0+\tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}\left(M_{t,i,a}-\varepsilon_a^0-L_i-\beta_1 e^{t\beta_{a_2}}\right)\right)/(\tau TIA+\tau_0)\right)}
\end{aligned}$$

Completing the square in the exponent in β_0 we obtain

$$\propto e^{-\frac{1}{2}(\tau TIA+\tau_0)\left(\beta_0-\frac{\tau_0\mu_0+\tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}\left(M_{t,i,a}-\varepsilon_a^0-L_i-\beta_1 e^{t\beta_{a_2}}\right)}{\tau TIA+\tau_0}\right)^2}$$

then

$$P(\beta_0|\Theta_{-\beta_0}, D_t) \propto N\left(\frac{\tau_0\mu_0+\tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}\left(M_{t,i,a}-\varepsilon_a^0-L_i-\beta_1 e^{t\beta_{a_2}}\right)}{\tau TIA+\tau_0}, (\tau TIA+\tau_0)^{-1}\right)$$

Posterior Distribution for β_1

$$\begin{aligned}
P(\beta_1|\Theta_{-\beta_1}, D_t) &\propto P(D_t|\Theta)P(\beta_1) \\
&\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} e^{-\frac{\tau}{2}(M_{t,i,a}-\mu_{t,i,a})^2}\right)\cdot\frac{\sqrt{\tau_1}}{\sqrt{2\pi}}e^{-\frac{\tau_1}{2}(\beta_1-\mu_1)^2} \\
&\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} e^{-\frac{\tau}{2}(\mu_{t,i,a}^2-2\mu_{t,i,a}M_{t,i,a})}\right)\cdot\frac{\sqrt{\tau_1}}{\sqrt{2\pi}}e^{-\frac{\tau_1}{2}(\beta_1-\mu_1)^2}
\end{aligned}$$

but we have that

$$\begin{aligned}
\mu_{t,i,a}^2 &= \left(\beta_{a_0}+L_i+\beta_1 e^{t\beta_{a_2}}\right)^2 \\
&= (\beta_{a_0}+L_i)^2+2(\beta_{a_0}+L_i)\beta_1 e^{t\beta_{a_2}}+\beta_1^2 e^{2t\beta_{a_2}}
\end{aligned}$$

furthermore,

$$\begin{aligned}
2\mu_{t,i,a}M_{t,i,a} &= 2\left(\beta_{a_0}+L_i+\beta_1 e^{t\beta_{a_2}}\right)\cdot M_{t,i,a} \\
&= 2(\beta_{a_0}+L_i)M_{t,i,a}+2\beta_1 e^{t\beta_{a_2}}M_{t,i,a}
\end{aligned}$$

Then, taking only what depends on β_1, we obtain

$$
\begin{aligned}
P(\beta_1|\Theta_{-\beta_1}, D_t) &\propto \left(\prod_{t=1}^{T} \prod_{i=1}^{I} \prod_{a=1}^{A} e^{-\frac{\tau}{2}\left(\beta_1^2 e^{2t\beta_{a_2}} + 2(\beta_{a_0}+L_i)\beta_1 e^{t\beta_{a_2}} - 2\beta_1 e^{t\beta_{a_2}} M_{t,i,a}\right)} \right) \cdot e^{-\frac{\tau_1}{2}(\beta_1^2 - 2\mu_1\beta_1)} \\
&\propto e^{-\frac{\tau I A}{2}\left(\sum_{t=1}^{T} e^{2t\beta_{a_2}}\right)\beta_1^2 \left(\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} e^{-\frac{\tau}{2}\beta_1\left(2(\beta_{a_0}+L_i)e^{t\beta_{a_2}} - 2e^{t\beta_{a_2}} M_{t,i,a}\right)}\right)} \cdot e^{-\frac{\tau_1}{2}(\beta_1^2 - 2\mu_1\beta_1)} \\
&\propto e^{-\frac{\tau}{2}\left[\beta_1^2\left(\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{et\beta_{a_2}}\right) - 2\beta_1\left(\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{t\beta_{a_2}} M_{t,i,a} - \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(\beta_{a_0}+L_i)e^{t\beta_{a_2}}\right)\right] - \frac{\tau_1}{2}\beta} \\
&\propto e^{\beta_1^2\left(-\frac{\tau}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t\beta_{a_2}} - \frac{\tau_1}{2}\right) + 2\beta_1\left(\frac{\tau}{2}\left(\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(M_{t,i,a} - \beta_{a_0} - L_i)e^{t\beta_{a_2}}\right) + \frac{\mu_1\tau_1}{2}\right)} \\
&\propto e^{-\frac{1}{2}\left[\beta_1^2\left(\tau_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t\beta_{a_2}}\right) - 2\beta_1\left(\tau_1\mu_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(M_{t,i,a} - \beta_{a_0} - L_i)e^{t\beta_{a_2}}\right)\right]} \\
&\propto e^{-\frac{1}{2}\left(\tau_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t\beta_{a_2}}\right)\left[\beta_1^2 - 2\beta_1 \frac{\tau_1\mu_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(M_{t,i,a} - \beta_{a_0} - L_i)e^{t\beta_{a_2}}}{\tau_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t\beta_{a_2}}}\right]}
\end{aligned}
$$

Completing the square in the exponent in β_1 we obtain

$$
\propto e^{-\frac{1}{2}\left(\tau_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t\beta_{a_2}}\right)\left(\beta_1 - \frac{\tau_1\mu_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(M_{t,i,a} - \beta_{a_0} - L_i)e^{t\beta_{a_2}}}{\tau_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t\beta_{a_2}}}\right)^2}
$$

in consequence

$$
\begin{aligned}
P(\beta_1|\Theta_{-\beta_1}, D_t) &\propto N\left(\frac{\tau_1\mu_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(M_{t,i,a} - \beta_{a_0} - L_i)e^{t\beta_{a_2}}}{\tau_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t\beta_{a_2}}}, \right. \\
&\qquad \left. \left(\tau_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t\beta_{a_2}}\right)^{-1} \right) \\
\Rightarrow P(\beta_1|\Theta_{-\beta_1}, D_t) &\propto N\left(\frac{\tau_1\mu_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(M_{t,i,a} - \beta_0 - \varepsilon_a^0 - L_i)e^{t(\beta_2 + \varepsilon_a^2)}}{\tau_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t\beta_{a_2}}}, \right. \\
&\qquad \left. \left(\tau_1 + \tau\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A} e^{2t(\beta_2 + \varepsilon_a^2)}\right)^{-1} \right)
\end{aligned}
$$

Posterior Distribution for ε_a^0

$$
\begin{aligned}
P(\varepsilon_a^0|\Theta_{-\varepsilon_a^0}, D_t) &\propto P(D_t|\Theta)P(\varepsilon_a^0) \\
&\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I} e^{-\frac{\tau}{2}(M_{t,i,a} - \mu_{t,i,a})^2} \right) \cdot \frac{\sqrt{\tau_{\varepsilon_0}}}{\sqrt{2\pi}} e^{-\frac{\tau_{\varepsilon_0}}{2}(\varepsilon_a^0)^2} \\
&\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I} e^{-\frac{\tau}{2}(\mu_{t,i,a}^2 - 2\mu_{t,i,a}M_{t,i,a})} \right) \cdot \frac{\sqrt{\tau_{\varepsilon_0}}}{\sqrt{2\pi}} e^{-\frac{\tau_{\varepsilon_0}}{2}(\varepsilon_a^0)^2} \quad \text{for } a = 1, ..., A
\end{aligned}
$$

but we have that

$$\begin{aligned}\mu_{t,i,a}^2 &= \left(\beta_{a_0} + L_i + \beta_1 e^{t\beta_{a_2}}\right)^2 \\ &= (\beta_0 + \varepsilon_a^0 + L_i + \beta_1 e^{t\beta_{a_2}}) \\ &= (\varepsilon_a^0)^2 + 2\varepsilon_a^0(\beta_0 + L_i + \beta_1 e^{t\beta_{a_2}}) + (\beta_0 + L_i + \beta_1 e^{t\beta_{a_2}})^2\end{aligned}$$

furthermore,

$$\begin{aligned}2\mu_{t,i,a}M_{t,i,a} &= 2M_{t,i,a}\cdot\left(\beta_{a_0} + L_i + \beta_1 e^{t\beta_{a_2}}\right) \\ &= 2M_{t,i,a}\cdot\left(\beta_0 + \varepsilon_a^0 + L_i + \beta_1 e^{t\beta_{a_2}}\right) \\ &= 2M_{t,i,a}\varepsilon_a^0 + 2M_{t,i,a}\left(\beta_0 + L_i + \beta_1 e^{t\beta_{a_2}}\right)\end{aligned}$$

Then, taking only what depends on ε_a^0, we obtain

$$\begin{aligned}P(\varepsilon_a^0|\Theta_{-\varepsilon_a^0}, D_t) &\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I} e^{-\frac{\tau}{2}}\left[(\varepsilon_a^0)^2 + 2\varepsilon_a^0\left(\beta_0 + L_i + \beta_1 e^{t\beta_{a_2}}\right) - 2M_{t,i,a}\varepsilon_a^0\right]\right)\cdot e^{-\frac{\tau_{\varepsilon_0}}{2}(\varepsilon_a^0)^2} \\ &\propto e^{-\frac{\tau}{2}\left(TI(\varepsilon_a^0)^2 + 2\varepsilon_a^0\sum_{t=1}^{T}\sum_{i=1}^{I}\left(\beta_0 + L_i + \beta_1 e^{t\beta_{a_2}}\right) - 2\varepsilon_a^0\sum_{t=1}^{T}\sum_{i=1}^{I}M_{t,i,a}\right) - \frac{\tau_{\varepsilon_0}}{2}(\varepsilon_a^0)^2} \\ &\propto e^{-\frac{1}{2}(\tau TI + \tau_{\varepsilon_0})(\varepsilon_a^0)^2 - 2\varepsilon_a^0\tau\sum_{t=1}^{T}\sum_{i=1}^{I}\left(M_{t,i,a} - \beta_0 - L_i - \beta_1 e^{t\beta_{a_2}}\right)} \\ &\propto e^{-\frac{1}{2}(\tau TI + \tau_{\varepsilon_0})\left[(\varepsilon_a^0)^2 - 2\varepsilon_a^0\frac{\tau\sum_{t=1}^{T}\sum_{i=1}^{I}\left(M_{t,i,a} - \beta_0 - L_i - \beta_1 e^{t\beta_{a_2}}\right)}{(\tau TI + \tau_{\varepsilon_0})}\right]}\end{aligned}$$

Completing the square in the exponent in ε_a^0 we obtain

$$\propto e^{-\frac{1}{2}(\tau TI + \tau_{\varepsilon_0})\left(\varepsilon_a^0 - \frac{\tau\sum_{t=1}^{T}\sum_{i=1}^{I}\left(M_{t,i,a} - \beta_0 - L_i - \beta_1 e^{t(\beta_2 + \varepsilon_a^2)}\right)}{(\tau TI + \tau_{\varepsilon_0})}\right)^2}$$

in consequence

$$P(\varepsilon_a^0|\Theta_{-\varepsilon_a^0}, D_t) \propto N\left(\frac{\tau\sum_{t=1}^{T}\sum_{i=1}^{I}\left(M_{t,i,a} - \beta_0 - L_i - \beta_1 e^{t(\beta_2 + \varepsilon_a^2)}\right)}{(\tau TI + \tau_{\varepsilon_0})}, (\tau TI + \tau_{\varepsilon_0})^{-1}\right)$$

Posterior Distribution for β_2

$$\begin{aligned}P(\beta_2|\Theta_{-\beta_2}, D_t) &\propto P(D_t|\Theta)P(\beta_2) \\ &\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} e^{-\frac{\tau}{2}(M_{t,i,a} - \mu_{t,i,a})^2}\right)\cdot\frac{\sqrt{\tau_2}}{\sqrt{2\pi}}e^{-\frac{\tau_2}{2}(\beta_2^2)^2} \\ &\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} e^{-\frac{\tau}{2}(\mu_{t,i,a}^2 - 2\mu_{t,i,a}M_{t,i,a})}\right)\cdot e^{-\frac{\tau_2}{2}(\beta_2 - \mu_2)^2}\end{aligned}$$

but we have that

$$\mu_{t,i,a}^2 = \left(\beta_{a_0} + L_i + \beta_1 e^{t\beta_{a_2}}\right)^2$$
$$= (\beta_0 + L_i)^2 + 2(\beta_0 + L_i)\beta_1 e^{t\beta_{a_2}} + \beta_1^2 e^{2t(\beta_2+\varepsilon_a^2)}$$

furthermore,

$$2\mu_{t,i,a}M_{t,i,a} = 2M_{t,i,a} \cdot \left(\beta_{a_0} + L_i + \beta_1 e^{t(\beta_2+\varepsilon_a^2)}\right)$$
$$= 2M_{t,i,a} \cdot (\beta_{a_0} + L_i) + 2\beta_1 M_{t,i,a} e^{t(\beta_2+\varepsilon_a^2)}$$

Then, taking only what depends on β_2, we obtain

$$P(\beta_2|\Theta_{-\beta_2}, D_t) \propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A} e^{-\frac{\tau}{2}\left(\beta_1^2 e^{2\tau(\beta_2+\varepsilon_a^2)} + 2(\beta_{a_0}+L_i)\beta_1 e^{t\beta_2+\varepsilon_a^2} - 2\beta_1 M_{t,i,a} e^{t(\beta_2+\varepsilon_a^2)}\right)}\right) \cdot e^{-\frac{\tau_2}{2}(\beta_2^2 - 2\beta_2\mu_2)}$$
$$\propto e^{-\frac{\tau}{2}\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}\left(\beta_1^2\left(e^{t(\beta_2+\varepsilon_a^2)}\right)^2 - 2\beta_1 e^{\beta_2+\varepsilon_a^2}(M_{t,i,a} - \beta_{a_0} - L_i)\right)\right] - \frac{\tau_2}{2}(\beta_2^2 - 2\beta_2\mu_2)}$$

completing the square in the first term, we obtain

$$\propto e^{-\frac{\tau_2}{2}(\beta_2^2 - 2\beta_2\mu_2) - \frac{\tau}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}\left(\beta_1 e^{t(\beta_2+\varepsilon_a^2)} - (M_{t,i,a} - \beta_{a_0} - L_i)\right)^2}$$

therefore,

$$P(\beta_2|\Theta_{-\beta_2}, D_t) \propto e^{-\frac{\tau_2}{2}(\beta_2^2 - 2\beta_2\mu_2) - \frac{\tau}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}\left(\beta_1 e^{t(\beta_2+\varepsilon_a^2)} - (M_{t,i,a} - \beta_{a_0} - L_i)\right)^2}$$

which is not closed form.

Posterior Distribution for ε_a^2

$$P(\varepsilon_a^2|\Theta_{-\varepsilon_a^2}, D_t) \propto P(D_t|\Theta)P(\varepsilon_a^2)$$
$$\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I} e^{-\frac{\tau}{2}(M_{t,i,a} - \mu_{t,i,a})^2}\right) \cdot \frac{\sqrt{\tau_{\varepsilon_2}}}{\sqrt{2\pi}} e^{-\frac{\tau_{\varepsilon_2}}{2}(\varepsilon_a^2)^2}$$
$$\propto \left(\prod_{t=1}^{T}\prod_{i=1}^{I} e^{-\frac{\tau}{2}(\mu_{t,i,a}^2 - 2\mu_{t,i,a}M_{t,i,a})}\right) \cdot \frac{\sqrt{\tau_{\varepsilon_2}}}{\sqrt{2\pi}} e^{-\frac{\tau_{\varepsilon_2}}{2}(\varepsilon_a^2)^2} \quad \text{for } a = 1, ..., A$$

but we have that

$$\mu_{t,i,a}^2 = \left(\beta_{a_0} + L_i + \beta_1 e^{t\beta_{a_2}}\right)^2$$
$$= (\beta_0 + L_i)^2(\beta_0 + L_i)\beta_1 e^{t(\beta_2+\varepsilon_a^2)} + \beta_1^2 e^{2t(\beta_2+\varepsilon_a^2)}$$

furthermore,

$$\begin{aligned}2\mu_{t,i,a}M_{t,i,a} &= 2M_{t,i,a}\cdot\left(\beta_{a_0}+L_i+\beta_1 e^{t\beta_{a_2}}\right)\\ &= 2M_{t,i,a}\cdot(\beta_{a_0}+L_i)+2\beta_1 M_{t,i,a}e^{t(\beta_2+\varepsilon_a^2)}\end{aligned}$$

Then, taking only what depends on ε_a^2, we obtain

$$\begin{aligned}P(\varepsilon_a^2|\Theta_{-\varepsilon_a^2},D_t) &\propto \left[\prod_{t=1}^{T}\prod_{i=1}^{I} e^{-}\frac{\tau}{2}\left(\beta_1^2 e^{2t(\beta_2+\varepsilon_a^2)}+2(\beta_{a_0}+L_i)\beta_1 e^{t(\beta_2+\varepsilon_a^2)}-2\beta_1 M_{t,i,a}e^{t(\beta_2+\varepsilon_a^2)}\right)\right]\\ &e^{-\frac{\tau_{\varepsilon_2}}{2}(\varepsilon_a^2)^2}\\ &\propto e^{-\frac{\tau}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\left[\beta_1^2 e^{2t(\beta_2+\varepsilon_a^2)}-2\beta_1 e^{t(\beta_2+\varepsilon_a^2)}(M_{t,i,a}-\beta_{a_0}-L_i)\right]-\frac{\tau_{\varepsilon_2}}{2}(\varepsilon_a^2)^2}\end{aligned}$$

completing the square in the term of the summation, we obtain

$$\propto e^{-\frac{\tau}{2}\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\left(\beta_1 e^{t(\beta_2+\varepsilon_a^2)}-(M_{t,i,a}-\beta_{a_0}-L_i)\right)^2\right]-\frac{\tau_{\varepsilon_2}}{2}(\varepsilon_a^2)^2}$$

therefore

$$P(\varepsilon_a^2|\Theta_{-\varepsilon_a^2},D_t)\propto e^{-\frac{\tau_{\varepsilon_2}}{2}(\varepsilon_a^2)^2-\frac{\tau}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\left(\beta_1 e^{t(\beta_2+\varepsilon_a^2)}-(M_{t,i,a}-\beta_{a_0}-L_i)\right)^2}\quad \text{for } a=1,...,A$$

which is not closed form.

Posterior Distribution for τ

$$\begin{aligned}P(\tau|\Theta_{-\tau},D_t) &\propto P(D_t|\Theta)P(\tau)\\ &\propto \left[\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A}\left(\sqrt{\tau}e^{-\frac{\tau}{2}(M_{t,i,a}-\mu_{t,i,a})^2}\right)\right]\tau^{\alpha_\tau-1}e^{-\beta_\tau\tau}\\ &\propto \tau^{\frac{1}{2}TIA}e^{-\frac{\tau}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(M_{t,i,a}-\mu_{t,i,a})^2}\cdot\tau^{\alpha_\tau-1}e^{-\beta_\tau\tau}\\ &\propto \tau^{(\frac{1}{2}TIA+\alpha_\tau)-1}e^{-\tau\left[\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(M_{t,i,a}-\mu_{t,i,a})^2+\beta_\tau\right]}\end{aligned}$$

therefore

$$P(\tau|\Theta_{-\tau},D_t)\propto Gamm\left(\frac{1}{2}TIA+\alpha_\tau,\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{a=1}^{A}(M_{t,i,a}-\mu_{t,i,a})^2+\beta_\tau\right)$$

Posterior Distribution for τ_{ε_0}

$$
\begin{aligned}
P(\tau_{\varepsilon_0}u|\Theta_{-\tau_{\varepsilon_0}}, D_t) &\propto P(D_t|\Theta)P(\varepsilon_0|\tau_{\varepsilon_0})P(\tau_{\varepsilon_0}) \\
&\propto \left[\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A}\left(e^{-\frac{\tau}{2}\left(M_{t,i,a}-\mu_{t,i,a}\right)^2}\right)\right]\prod_{a=1}^{A}\left[\tau_{\varepsilon_0}^{1/2}e^{-\frac{\tau_{\varepsilon_0}}{2}(\varepsilon_0^0)^2}\right]\frac{\alpha_{\tau_{\varepsilon_0}}^{-1}}{\tau_{\varepsilon_0}}e^{-\beta_{\tau_{\varepsilon_0}}\tau_{\varepsilon_0}} \\
&\propto \tau_{\varepsilon_0}^{\frac{A}{2}}e^{-\frac{\tau_{\varepsilon_0}}{2}\sum_{a=1}^{A}(\varepsilon_a^0)^2}\tau_{\varepsilon_0}^{\alpha_{\tau_{\varepsilon_0}}-1}e^{-\beta_{\tau_{\varepsilon_0}}\tau_{\varepsilon_0}} \\
&\propto \tau_{\varepsilon_0}^{\left(\frac{A}{2}+\alpha_{\tau_{\varepsilon_0}}\right)-1}e^{-\left(\beta_{\tau_{\varepsilon_0}}+\frac{1}{2}\sum_{a=1}^{A}(\varepsilon_a^0)^2\right)}
\end{aligned}
$$

therefore

$$
P(\tau_{\varepsilon_0}|\Theta_{-\tau_{\varepsilon_0}}, D_t) \propto Gamm\left(\frac{A}{2}+\alpha_{\tau_{\varepsilon_0}}, \beta_{\tau_{\varepsilon_0}}+\frac{1}{2}\sum_{a=1}^{A}(\varepsilon_a^0)^2\right)
$$

Posterior Distribution for τ_{ε_2}

$$
\begin{aligned}
P(\tau_{\varepsilon_2}u|\Theta_{-\tau_{\varepsilon_2}}, D_t) &\propto P(D_t|\Theta)P(\varepsilon_a^2|\tau_{\varepsilon_2})P(\tau_{\varepsilon_2}) \\
&\propto \left[\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{a=1}^{A}\left(e^{-\frac{\tau}{2}\left(M_{t,i,a}-\mu_{t,i,a}\right)^2}\right)\right]\cdot\left[\prod_{a=1}^{A}\tau_{\varepsilon_2}^{1/2}e^{-\frac{\tau_{\varepsilon_2}}{2}(\varepsilon_a^2)^2}\right]\cdot\tau_{\varepsilon_2}^{\alpha_{\tau_{\varepsilon_2}}-1}e^{\beta_{\tau_{\varepsilon_2}}\tau_{\varepsilon_2}} \\
&\propto \tau_{\varepsilon_2}^{\frac{A}{2}}e^{-\frac{\tau_{\varepsilon_2}}{2}\sum_{a=1}^{A}(\varepsilon_a^2)^2}\tau_{\varepsilon_2}^{\alpha_{\tau_{\varepsilon_2}}-1}e^{-\beta_{\tau_{\varepsilon_2}}\tau_{\varepsilon_2}} \\
&\propto \tau_{\varepsilon_2}^{\left(\frac{A}{2}+\alpha_{\tau_{\varepsilon_2}}\right)-1}e^{-\left(\beta_{\tau_{\varepsilon_2}}+\frac{1}{2}\sum_{a=1}^{A}(\varepsilon_a^2)^2\right)\tau_{\varepsilon_2}}
\end{aligned}
$$

therefore

$$
P(\tau_{\varepsilon_2}u|\Theta_{-\tau_{\varepsilon_2}}, D_t) \propto Gamm\left(\frac{A}{2}+\alpha_{\tau_{\varepsilon_2}}, \beta_{\tau_{\varepsilon_2}}+\frac{1}{2}\sum_{a=1}^{A}(\varepsilon_a^2)^2\right)
$$

Posterior Distribution for α_θ

$$
\begin{aligned}
P(\alpha_\theta|\Theta_{-\alpha_\theta}, D_t) &\propto P(D_t|\Theta)P(\alpha_\theta) \\
&\propto \left[\prod_{a=1}^{A}P(\alpha_\theta|\alpha_\theta)\right]\cdot P(\alpha_\theta) \\
&= \left(\prod_{a=1}^{A}\left[\frac{\theta_a^{\alpha_\theta-1}e^{-\beta_\theta\theta_a}\beta_\theta^{\alpha_\theta}}{\Gamma(\alpha_\theta)}\right]\right)\alpha_\theta^{\alpha_{\alpha_\theta}-1}e^{-\beta_{\alpha_\theta}\alpha_\theta} \\
&\propto \Gamma(\alpha_\theta)^{-A}\beta_\theta^{A\alpha_\theta}\alpha_\theta^{\alpha_{\alpha_\theta}-1}e^{-\beta_{\alpha_\theta}\alpha_\theta}\left(\prod_{a=1}^{A}\theta_a^{\alpha_\theta-1}\right)
\end{aligned}
$$

which is not closed form.

Posterior Distribution for α_λ

$$
\begin{aligned}
P(\alpha_\lambda|\Theta_{-\alpha_\lambda}, D_t) &\propto P(D_t|\Theta)P(\alpha_\lambda) \\
&\propto \left[\prod_{a=1}^{A} P(\alpha_\lambda|\alpha_\lambda)\right] \cdot P(\alpha_\lambda) \\
&= \left(\prod_{a=1}^{A} \left[\frac{\lambda_a^{\alpha_\lambda - 1} e^{-\beta_\lambda \lambda_a} \beta_\lambda^{\alpha_\lambda}}{\Gamma(\alpha_\lambda)}\right]\right) \alpha_\lambda^{\alpha_{\alpha_\lambda}-1} e^{-\beta_{\alpha_\lambda}\alpha_\lambda} \\
&\propto \Gamma(\alpha_\lambda)^{-A} \beta_\lambda^{A\alpha_\lambda} \alpha_\lambda^{\alpha_{\alpha_\lambda}-1} e^{-\beta_{\alpha_\lambda}\alpha_\lambda} \left(\prod_{a=1}^{A} \lambda_a^{\alpha_\lambda - 1}\right)
\end{aligned}
$$

which is not closed form.

Posterior Distribution for β_θ

$$
\begin{aligned}
P(\beta_\theta|\Theta_{-\beta_\theta}, D_t) &\propto P(D_t|\Theta)P(\beta_\theta) \\
&\propto \left(\prod_{a=1}^{A} P(\theta_a|\beta_\theta)\right) \cdot P(\beta_\theta) \\
&\propto \left[\prod_{a=1}^{A} \beta_\theta^{\alpha_\theta} e^{-\beta_\theta \theta_a}\right] \beta_\theta^{\alpha_{\beta_\theta}-1} e^{-\beta_{\beta_\theta}\beta_\theta} \\
&\propto \beta_\theta^{A\alpha_\theta} e^{-\beta_\theta \sum_{a=1}^{A} \theta_a} \beta_\theta^{\alpha_{\beta_\theta}-1} e^{-\beta_\theta \beta_{\beta_\theta}} \\
&\propto \beta_\theta^{(A\alpha_\theta + \alpha_{\beta_\theta})-1} e^{-\beta_\theta\left(\beta_{\beta_\theta} + \sum_{a=1}^{A} \theta_a\right)}
\end{aligned}
$$

therefore

$$
P(\beta_\theta|\Theta_{-\beta_\theta}, D_t) \propto Gamma\left(A\alpha_\theta + \alpha_{\beta_\theta}, \beta_{\beta_\theta} + \sum_{a=1}^{A} \theta_a\right)
$$

Posterior Distribution for β_λ

$$\begin{aligned}
P(\beta_\lambda|\Theta_{-\beta_\lambda}, D_t) &\propto P(D_t|\Theta)P(\beta_\lambda) \\
&\propto \left(\prod_{a=1}^{A} P(\lambda_a|\beta_\lambda)\right) \cdot P(\beta_\lambda) \\
&\propto \left[\prod_{a=1}^{A} \beta_\lambda^{\alpha_\lambda} e^{-\beta_\lambda \lambda_a}\right] \beta_\lambda^{\alpha_{\beta_\lambda}-1} e^{-\beta_{\beta_\lambda}\beta_\lambda} \\
&\propto \beta_\lambda^{A\alpha_\lambda} e^{-\beta_\lambda \sum_{a=1}^{A}\lambda_a} \beta_\lambda^{\alpha_{\beta_\lambda}-1} e^{-\beta_\lambda \beta_{\beta_\lambda}} \\
&\propto \beta_\lambda^{(A\alpha_\lambda+\alpha_{\beta_\lambda})-1} e^{-\beta_\lambda\left(\beta_{\beta_\lambda}+\sum_{a=1}^{A}\lambda_a\right)}
\end{aligned}$$

therefore

$$P(\beta_\lambda|\Theta_{-\beta_\lambda}, D_t) \propto Gamma\left(A\alpha_\lambda + \alpha_{\beta_\lambda}, \beta_{\beta_\lambda} + \sum_{a=1}^{A}\lambda_a\right)$$

For the spatial variable, we have to study mg neighboring regions, then the variance-covariance matrix of the multivariate normal distribution is given by

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

Then,

$$\begin{pmatrix} P & S \\ S & P \end{pmatrix} = \frac{\sigma^{-1}}{(1+|\eta|\cdot mg)^2-\eta^2} \cdot \begin{pmatrix} 1+|\eta|\cdot mg & \eta \\ \eta & 1+|\eta|\cdot mg \end{pmatrix}$$

Therefore, for P and S we have that

$$\begin{cases} P = \frac{\sigma^{-1}(1+|\eta|\cdot mg)}{(1+|\eta|\cdot mg)^2-\eta^2} \\ S = \frac{\sigma^{-1}\eta}{(1+|\eta|\cdot mg)^2-\eta^2} \end{cases}$$

Let's assume that $\eta > 0$ and two neighboring regions, i.e. $mg = 1$. Then we have that

$$\begin{cases} P = \frac{\sigma^{-1}(1+|\eta|\cdot mg)}{(1+|\eta|)^2-\eta^2} = \frac{\sigma^{-1}(1+\eta)}{1+2\eta+\eta^2-\eta^2} = \frac{\sigma^{-1}(1+\eta)}{1+2\eta} \\ S = \frac{\sigma^{-1}\eta}{(1+|\eta|)^2-\eta^2} = \frac{\sigma^{-1}\eta}{1+2\eta} \end{cases}$$

Therefore, the correlation coefficient can be written as

$$\begin{aligned}\rho &= \frac{cov(x,y)}{\sqrt{var(x)var(y)}}\\ &= \frac{S}{P}\\ &= \frac{\frac{\sigma^{-1}\eta}{1+2\eta}}{\frac{\sigma^{-1}(1+\eta)}{1+2\eta}}\\ &= \frac{\eta}{1+\eta}\end{aligned}$$

Let's see then the a posterior distribution for the parameters related to the regions (spatial section)

Posterior distribution for L_i

$$\begin{aligned}P(L_i|\Theta_{-L_i}, D_t) &\propto P(D_t|\Theta)P(L_i)\\ &\propto \left[\prod_{t=1}^{T}\prod_{a=1}^{A} e^{-\frac{\tau}{2}(M_{t,i,a}-\mu_{t,i,a})^2}\right]\cdot e^{-\frac{1}{2(1-\rho^2)}\left(\frac{L_1^2}{P^2}+\frac{L_2^2}{P^2}-\frac{2\rho L_i L_2}{P^2}\right)}\end{aligned}$$

for $i = 1, 2$

$$\propto \left[\prod_{t=1}^{T}\prod_{a=1}^{A} e^{-\frac{\tau}{2}(M_{t,i,a}-\mu_{t,i,a})^2}\right]\cdot e^{-\frac{1}{2P^2(1-\rho^2)\left(L_1^2+L_2^2-2\rho L_1 L_2\right)}}$$

and it has a not closed form.

Posterior distribution for η

$$\begin{aligned}P(\eta|\Theta_{\eta}, D_t) &\propto P(D_t|\Theta)P(L|\eta)P(\eta)\\ &\propto \frac{1}{P^2\sqrt{1-\rho^2}e^{-\frac{1}{2(1-\rho^2)}\left(\frac{L_1^2+L_2^2-2\rho L_1 L_2}{P^2}\right)}}\\ &= \frac{1}{(1+\eta)^2 P\sqrt{1-\rho^2}}e^{-\frac{1}{2P^2(1-\rho^2)}(L_1^2+L_2^2-2\rho L_1 L_2)}\end{aligned}$$

which is a not closed form.

Posterior distribution for σ

$$\begin{aligned}P(\sigma|\Theta_{-\sigma}, D_t) &\propto P(D_t|\Theta)P(L|\sigma)P(\sigma)\\ &\propto \frac{1}{P^2}\sigma^{\alpha_\sigma-1}e^{-\sigma\beta_\sigma}e^{-\frac{1}{2P^2(1-\rho^2)}(L_1^2+L_2^2-2\rho L_1 L_2)}\end{aligned}$$

which is a not closed form.

4.5 Monte Carlo Inference

Up to this point the theoretical information necessary for the implementation of MCMC has already been obtained. Given the history of severity and quantity of the claim and the insured population, it is possible to analyze the implicit parameters from the data set [17]. To improve the estimation of the parameters, the model considers the region and the age class of the insured. It is used MCMC algorithms in order to simulate the desired samples, specifically Metropolis-Hastings and Gibbs sampler are used. The Gibbs sampler algorithm, which is a special case of Metropolis-Hastings, generate posterior samples by sweeping through each variable (or block of variables) to sample from its conditional distribution with the remaining variables fixed to their current values. The conditional distributions are denoted as $\pi(\theta_i|\theta_{-i}, X), i = 1 : d$, where θ_{-i} represents all parameters in Θ but θ_i. The Gibbs sampler algorithm is the following:

Step 1
Initialize $\theta^{(0)} = \{\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_d^{(0)}\}$
Step 2
For iteration $j = 1, 2, ...$ **do**

- Generate $\theta_1^{(j)}$ from $\pi(\theta_1|\theta_2^{(j-1)}, \theta_2^{(j-1)}, ..., \theta_d^{(j-1)}, X)$
- Generate $\theta_2^{(j)}$ from $\pi(\theta_2|\theta_1^{(j-1)}, \theta_3^{(j-1)}, ..., \theta_d^{(j-1)}, X)$
  ⋮
- Generate $\theta_d^{(j)}$ from $\pi(\theta_d|\theta_1^{(j-1)}, \theta_2^{(j-1)}, ..., \theta_{d-1}^{(j-1)}, X)$

end for

The theory of MCMC guarantees that the stationary distribution of the samples genera ted under the Gibbs sampler algorithm is the target joint posterior that we are interested in [18].

On the other hand, the Metropolis–Hastings algorithm is a Markov chain Monte Carlo (MCMC) method for obtaining a sequence of random samples from a probability distribution from which direct sampling is difficult. The Metropolis Algorithm for sampling from a target distribution π, using transition kernel Q, consists of the following steps [19]:

Step 1:
Initialize $\theta^{(0)} = \{\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_d^{(0)}\}$

Step 2:
For iteration $i = 1, 2...$ **do**

- Sample γ from $Q(\gamma|\theta_i)$. (Think of γ as a "proposed" value for θ_{i+1})
- Compute
$$A = min\left(1, \frac{\pi(\gamma)Q(\theta_i|\gamma)}{\pi(\theta_i)Q(\gamma|\theta_i)}\right).$$
A is often called the "acceptance probabilty".
- With probability A "accept" the proposed value, and set
$$\theta_{i+1} = \gamma$$
. Otherwise set
$$\theta_{i+1} = \theta_i$$

end for
To implement these algorithms of the Gibbs and Metropolis-Hastings sampler in the adjustment of the proposed model, free software environment R for statistical computing and graphics will be used (www.r-project.org/).

Chapter 5

Simulation Studies and Model Fitting

As mentioned in the section of technical aspects of the project to carry out the simulation of the Bayesian hierarchical model, it is necessary to enter data on the number, total amount of insurance and region and age class of the insured. Insurance companies usually have this information to modify parameters and predictions for future claims.

This chapter contains the predetermination certain parameters for the simulation of claims data, followed by the model adjustment to see if it can capture the predetermined parameters.

5.1 Simulation Studies

This section presents the simulation of the history of insurance claims data, that is, the severity and frequency of the claim, and finally the insured population data. The simulation is carried out using the R packages *actuar*, *random*, *coda*, *MASS*, *mvtnorm*, *corpcor*, among other complementary packages. This as long as the parameters are already predetermined (Dutang et al., 2008). In oder to execute the simulation, the determination of the parameters is previously analyzed. The parameters are chosen primarily to test the complexity of the proposed model. Without losing of generality, the main objective of this section is to select values for the parameters in order to execute the model. In addition, if you properly select the parameter value, it could hopefully contain important practical meaning to help you understand the elements of the proposed model. To execute the simulation, the determination of the parameters is previously analyzed. The parameters are chosen primarily to test the complexity of the proposed model. The main objective of this section is to select parameters' values which have to be simple for the execution of the model, without losing of generality. In addition, if the parameter values are properly selected, it could hopefully contain important practical meaning to understand easily the elements of the proposed model.

For the parameters one can apply practical knowledge or experience in the assumptions when claim frequency is considered. For example, in health insurance instance, 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 2), increasing to about 0.4 for age class 40-50 (say, age class 4). 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 [20]. Let's consider (4.2), λ_a in that expression contribute to the average claim frequency and represents the claim made per person per unit time. Therefore, one possible selection of the value of this parameter is to set $\lambda_1 = 0.1$ and $\lambda_3 = 0.4$. For simplicity, but less likely in reality, it is assumed that for all age group the claim frequency in the simulation is at the same level, i.e., $\lambda_a \sim G(40, 200)$ for any $a = 1, 2, ..., A$. It means that 20% of the insured population would report a claim for each age class, with standard deviation of 3.16% [20]. Under the assumption that the variation of the claim frequency is not excessively large it has been selected intentionally a small value of the variance. If it is believed that there is a large variation in the claim frequency other assumptions can be made.

Now, likewise let's obtain the selection of the claim severity parameter. Let's consider again the two age classes 20-30 (age class 1) and 40-50 (age class 3). It can be assumed that the average claim amount for members in age class 1 is small, say \$15 per claim, while that in age class 3 may be higher, say \$30 per claim since policyholders in older age groups tend to make larger claims than policyholders in younger age groups. Once more, for simplicity and effectiveness of the model testing it is assumed that the average claim amount is about \$25 in all age classes. Let's consider (4.1), given that $\kappa_a = 1$ for $a = 1, 2, ..., A$ and θ_a are assumed known. It remains to select the value of the parameter θ_a. In the simulation, it is been considered that $\theta_a \sim G(400, 10000)$ for any $a = 1, 2, ..., A$. Again, this value has been selected to ensure a small variance. The selected values were chosen in order to the model be easy to understand and the process is technically simple to implement. As previously stated, these values have been chosen in such way that the model be easy to understand and so that the simulation not difficult to implement.

On the other hand, three parameters must be estimated in order to simulate insured population. In equation (4.3), mainly the value of β_j, with $j = 0, 1, 2$, determine the population mean. In the simulation it is assumed a quantitative criterion such that the mean of insured population for a given age group and region is to be doubled in 20 time units ([20]). At time 0 and 20 the average population 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 measure represent estimates rather than the true mean of $\mu_{t,i,a}$ which is indicated by the approximate sign. Also, the calculation of the true mean is complicated due to the exponential terms and those calculations are not required for effective simulations therefore 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 $E(\mu_{20,i,a}) \approx 140$ since it has been selected the population level in such way that at time 20 the population doubles. It can be seen that a small increase in the value β_2 increment the mean value of population exponentially then a small value for β_2 is selected. By the conditions previously stated let's consider the values $\beta_0 = 59, \beta_1 = 20$ and $\beta_2 = 0.075$. It is expected not to have much knowledge in terms of the values for the region parameters L. For the spatial effects let's consider equations (4.4) and (4.5), σ is randomly assumed to follow a Gamma distribution with large mean and variance and ind order to assign high probability to small values η follows a Pareto distribution with parameters 1 and 1.

So far, summarizing the assumptions mentioned above we have that:

- **Claim Frecuency Parameters**
  - $\lambda_a \sim Gamma(40, 200)$ for any $a = 1, 2, ..., 7$
- **Claim Severity Parameters**
  - $\kappa_a = 1$ for any $a = 1, 2, ..., 7$
  - $\theta_a \sim Gamma(400, 10000)$ for any $a = 1, 2, ..., 7$
- **Population Parameters**
  - $\beta_0 = 50$
  - $\beta_1 = 20$
  - $\beta_2 = 0.075$
  - $\eta \sim Pareto(1, 1)$

5.2 Model Fitting

5.2.1 Prior Elicitation

It is important to specify that the information used for the simulation process of data shown in Appendix A, set of observed data, is not applied in the adjustment of the model since in theory there is not much knowledge about the parameters of the data. It means that, prior elicitation is the remaining step for the application of the model.

The general purpose of this chapter is to test whether the model can capture the parameters incorporated in the data set. In addition, selecting concentrated priori near the true value of the parameters would not make sense when performing such an effectiveness test, therefore, it makes more sense, for the purpose of the test, to choose as vague background as possible.
Taking into account the above, consider the data of the insured population:

- For all age classes
  - Average after 1 unit of time: 43 for region 1
  - Average after 1 unit of time: 35 for region 2

So, an assumption for the initial insured population may be $\beta_0 + \beta_1 \approx 39$.

Similarly, after 20 units of time the mean increase to 62 and 55 for region 1 and 2 respectively. Therefore, the guess could be about 59, i.e., $\beta_0 + \beta_1 e^{20\beta_2} \approx 59$. Since two equations are not enough to estimate 3 parameters, a no accuracy estimating about the three parameters can be made. Let's say that the values could be $\beta_0 \approx 20$, $\beta_1 \approx 30$ and $\beta_2 \approx 0.015$. In order to indicate less confidence of the true value of β's and allowing the model to find the true value with great freedom it is assumed that the variance of the distribution of β's are large. Then, the priors are made as follows:

$$\beta_0 \sim N(20, 10^4)$$
$$\beta_1 \sim N(30, 10^4)$$
$$\beta_2 \sim N(0.015, 10^2)$$

For all other hyperparameters, the priors which contain little information are presented as follows:

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

The prior selected for $\tau, \alpha_\lambda, \beta_\lambda, \alpha_\theta, \beta_\theta$ are set with large variance; on the other hand τ_{ε_0} and τ_{ε_2} are set with small means in order to ε_a^0 and ε_a^2 have large variance. The prior of η allows to take small values with high probability. Finally, prior of σ is set with large mean and contributes towards region factor.

Now, let's implement the model using R tool in order to obtain the posterior of the parameters mentioned above. The results are shown in the next chapter.

5.3 Results

This section presents the adjustment of the proposed model using the simulated data classified by age group of the insured, region of residence of the insured and the time horizon of the insurance, in order to study its trend and seasonality. In addition, to determining the scale and shape behavior of the parameters involved in the model.

Using the Monte Carlo sequence methods, in particular, the Metropolis-Hastings algorithm and the Gibbs sampler described in section 4.5, 50000 random samples were generated for each parameter from their respective posterior distributions, as explained in chapter 4, involved in the collective compound risk model from their respective subsequent distribution.

The algorithm used to implement the Gibbs sample and the simulation from the predictive model was coded in the statistical programming language *R*, version 1.1.453 (Free Software); and the convergence analysis of the simulated traces was performed using the Software (CODA) with *R* routine.

The early iterations are thrown away and the remaining samples are used for posterior inference since 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 several iterations, the chain would reach its target distribution. Several simulations with different number of iterations were performed to compare the compilation time of the Gibbs sampler algorithm as can be seen in the following table:

Processor: 1.6 GHz Intel Core i5 Memory: 4 GB 1600 MHz DDR3			
Iterations	**Time in seconds**		
	User	System	Elapsed
100	0.786	0.185	1.544
1000	40.806	3.131	45.936
10000	4196.177	339.822	4861.823
Processor: Intel(R) Core(TM) i7-6700 CPU Memory: 16GB			
Iterations	**Time in seconds**		
	User	System	Elapsed
100	0.30	0.02	0.29
3300	340.6	0.23	341.2
50000	47849.56	88.72	47981.91

Table 5.1: Compilation time of the Gibbs sampler algorithm

In table 5.1, the *user time* is the CPU time charged for the execution of user instructions of the calling process. The *system time* is the CPU time charged for execution by the system on behalf of the calling process and the *elapsed time* is the time charged to the CPU(s) for the expression. As shown in the table 5.1, the execution time of the algorithm on a PC with an i7 processor and 16 GB of memory for 50,000 iterations was approximately 13 hours and 29 minutes.

The trace plot shows the sampled values of a parameter over time. This plot helps to judge how quickly the MCMC procedure converges in distribution, that is, how quickly it forgets its starting values. Also, the empirical density of each trace, in statistics known as kernel density estimation (KDE), which 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 [21]. The following figures show the trace and density plots of the major parameters used in the model where the plots on the left of the figures represent the traces of the parameter while those on the right the density or the parameter posterior distribution,

Figures 5.1 and 5.2 present the trace and posterior distribution of θ for $a = 1, 2, 3, 4$ and $a = 5, 6, 7$ respectively. 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 graph, the means of θ_a are centered near to 0.04 this do not differentiate by age groups. That is consistent with the assumption of simulation process. However, each age class is subject to random error variation and, therefore, the graphs for each age class are not exactly the same.

In figures 5.1 and 5.2, convergence is highlighted for each age class since each trace cyclically alternates up and down and the average lines of the three traces overlap, which is a strong indication of convergence. In the empirical density graphs of the trace, convergence can also be appreciated, besides, the density of each age class is shown to be stabilized almost symmetrically around the means after which they converge.

On the other hand, in figure 5.3 can be seen the θ_a's autocorrelations, for $a = 1, 2, ..., 7$, observed for each of the samples indicating that these generated traces effectively are an independent sample.

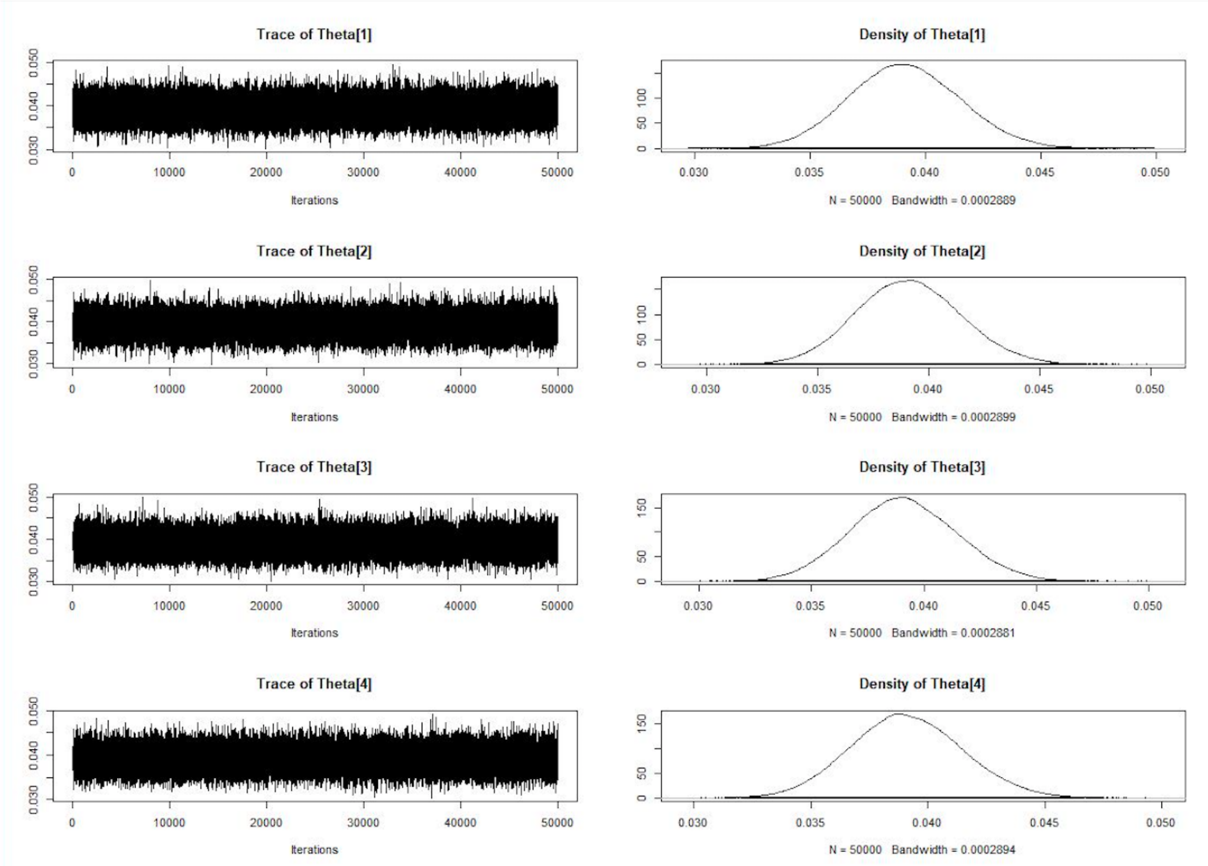

Figure 5.1: Trace plot and the posterior distribution of θ for $a = 1, 2, 3, 4$

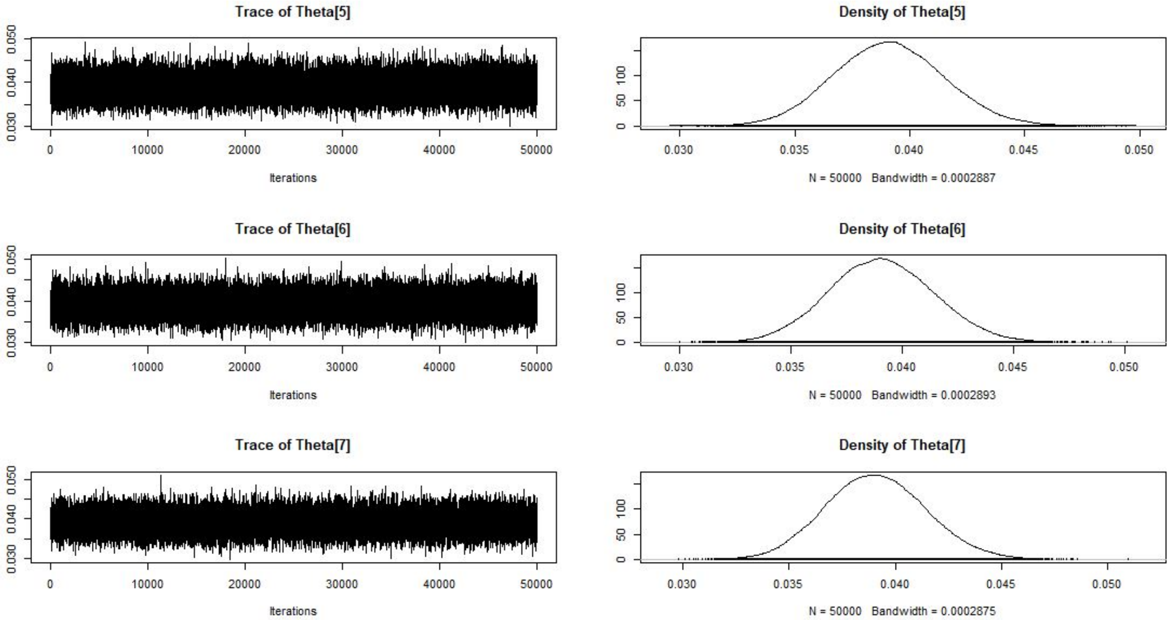

Figure 5.2: Trace plot and the posterior distribution of θ for $a = 5, 6, 7$

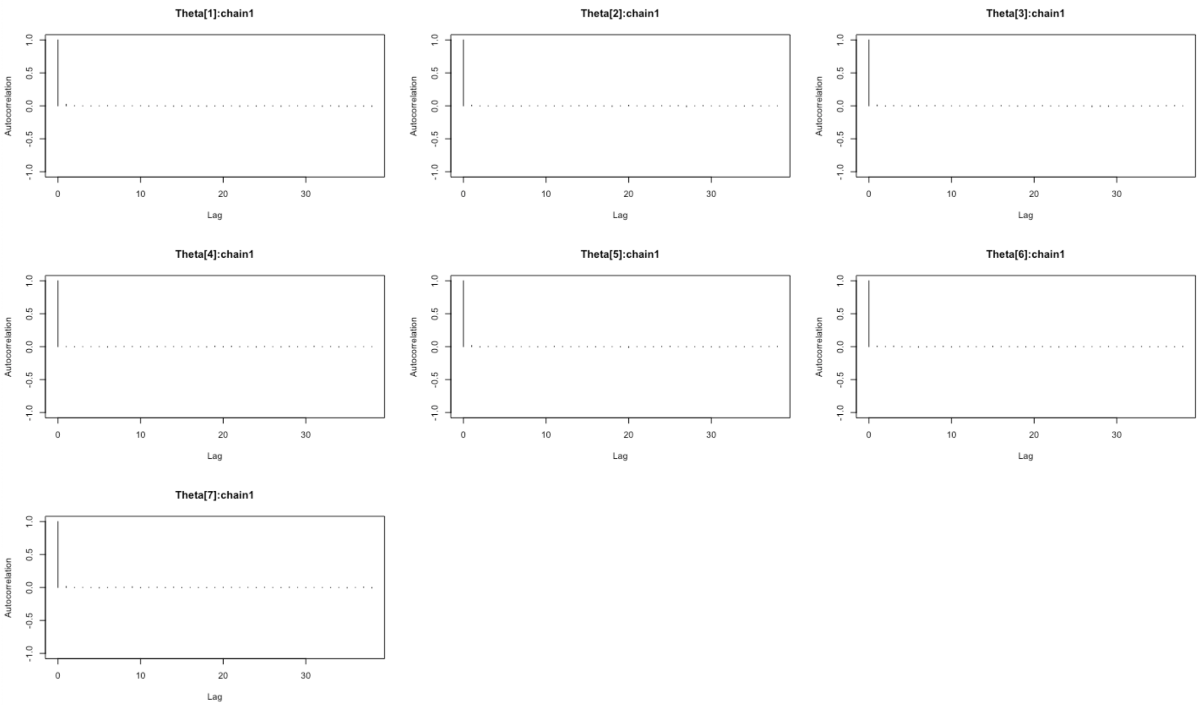

Figure 5.3: Autocorrelation of θ, for $a = 1, .., 7$

Trace, density and autocorrelation plots were generated for the parameters λ_a, β_{a_0}, β_1, β_{a_2}, ε_{a_0}, ε_{a_2}, α_θ, β_θ, α_λ, β_λ and L_i, with $a = 1, 2, ..., 7$ and $i = 1, 2$, whose respective figures are shown below (from figure 5.4 to figure 5.27).

Like with θ, in each parameter, it is observed that each of them converges around a value that is the posterior mean, as is also observed for the empirical density functions of each trace, where the approximate symmetry of each of them around a central value represents the posterior mean, confirming again that each chain has obtained convergence. On the other hand, autocorrelation plots have shown that the considered parameters are independent samples.

Also, it is clear from the output of both region parameters, L_1 and L_2, in figure 5.26, that the mean values are centered at 2. The data is obtained under the assumption that little information about the region is imposed. The output is consistent with such assumptions. The density plot of these parameters is bumpy with several humps around the mean value. This may be because the program continues the attempt to search for potential values of L's implied in the data. Again, for spatial factor L_i, $i = 1, 2$, in figure 5.27 the independence of the parameter can be observed.

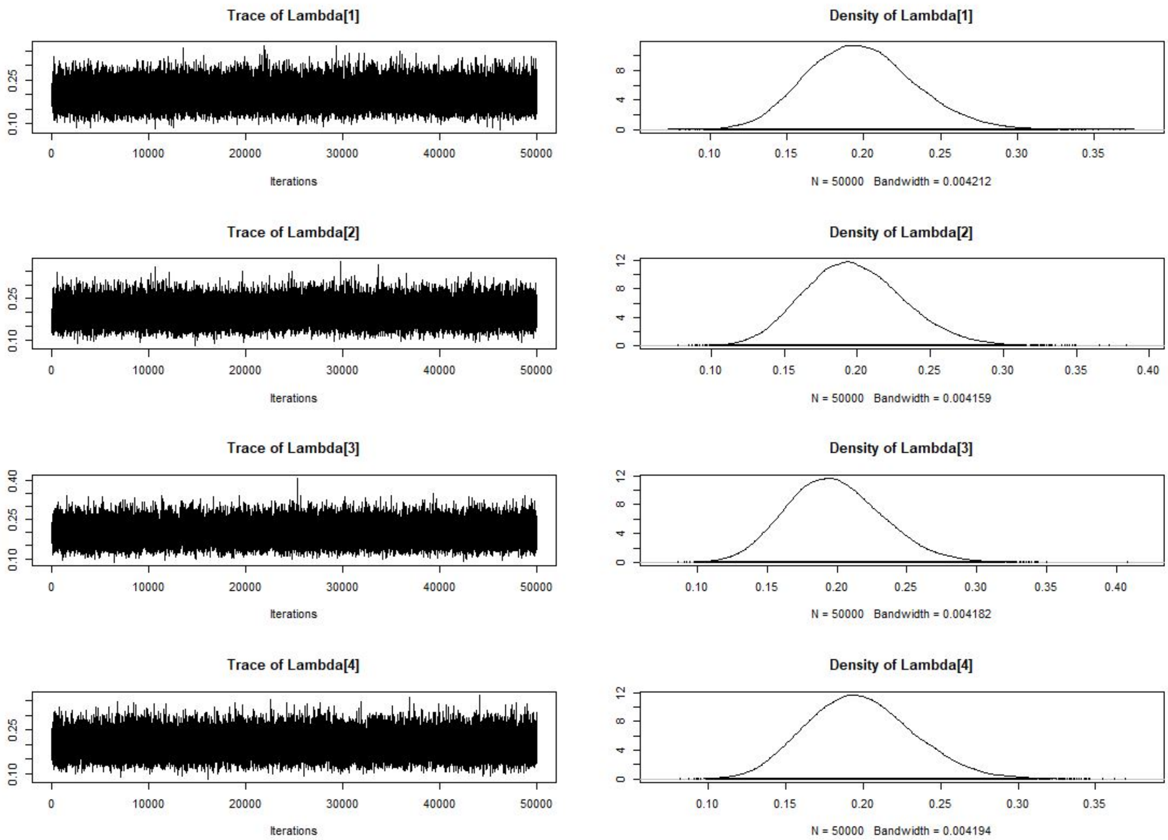

Figure 5.4: Trace plot and the posterior distribution of λ for $a = 1, 2, 3, 4$

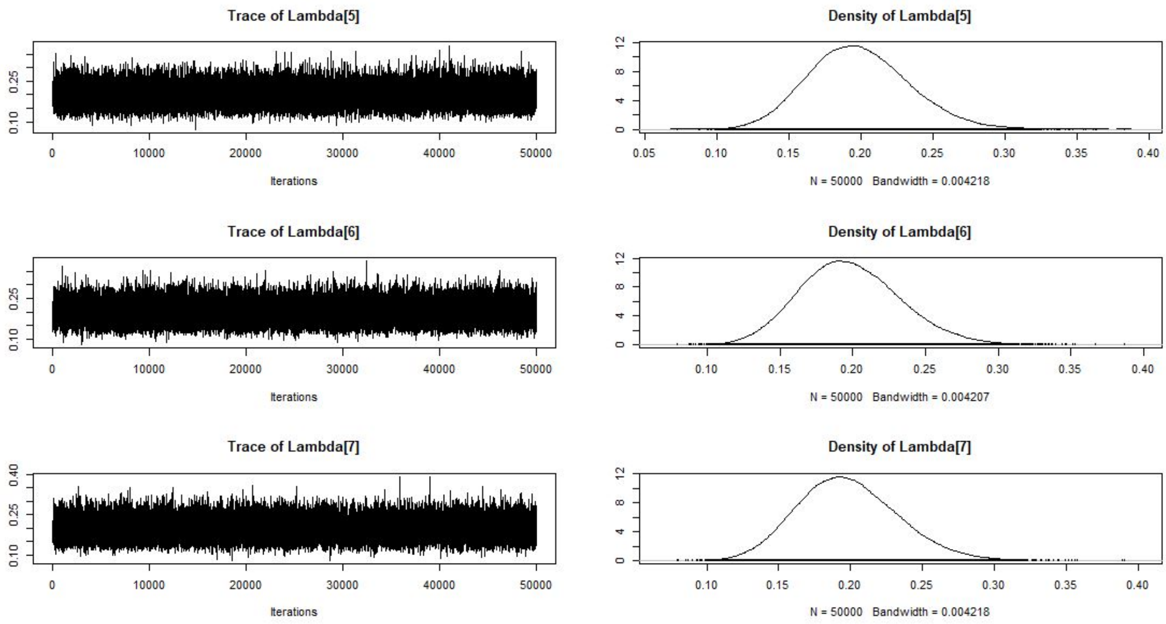

Figure 5.5: Trace plot and the posterior distribution of λ for $a = 5, 6, 7$

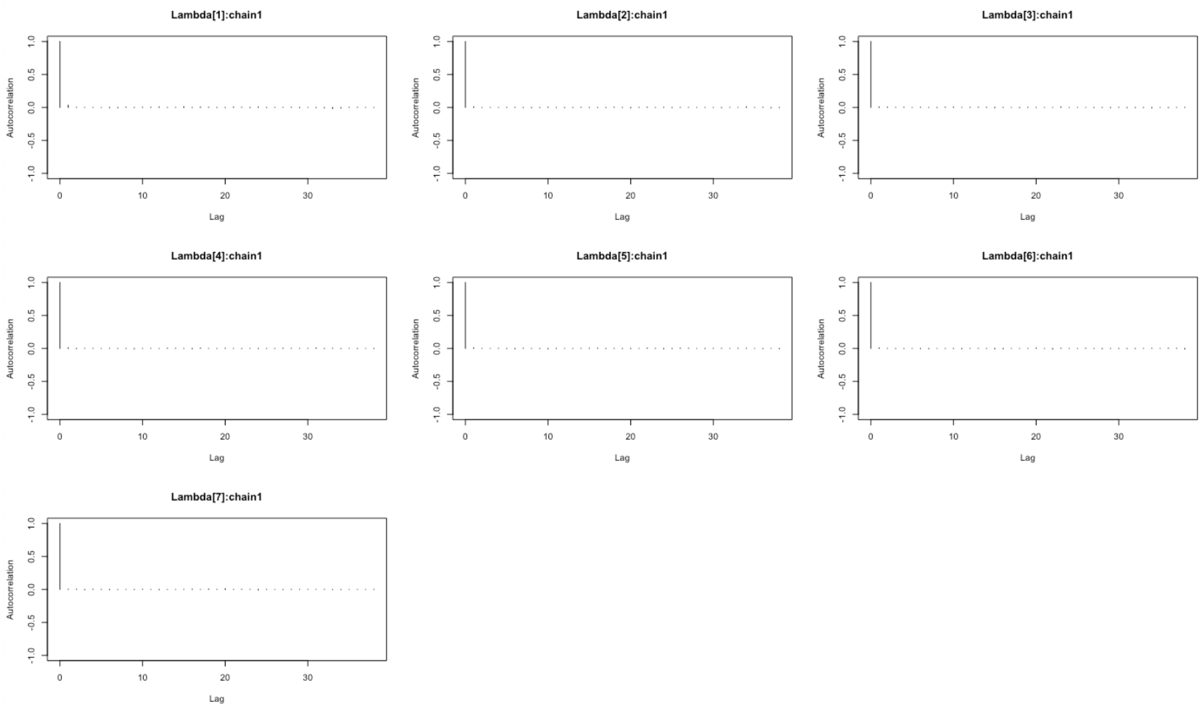

Figure 5.6: Autocorrelation of λ, for $a = 1, .., 7$

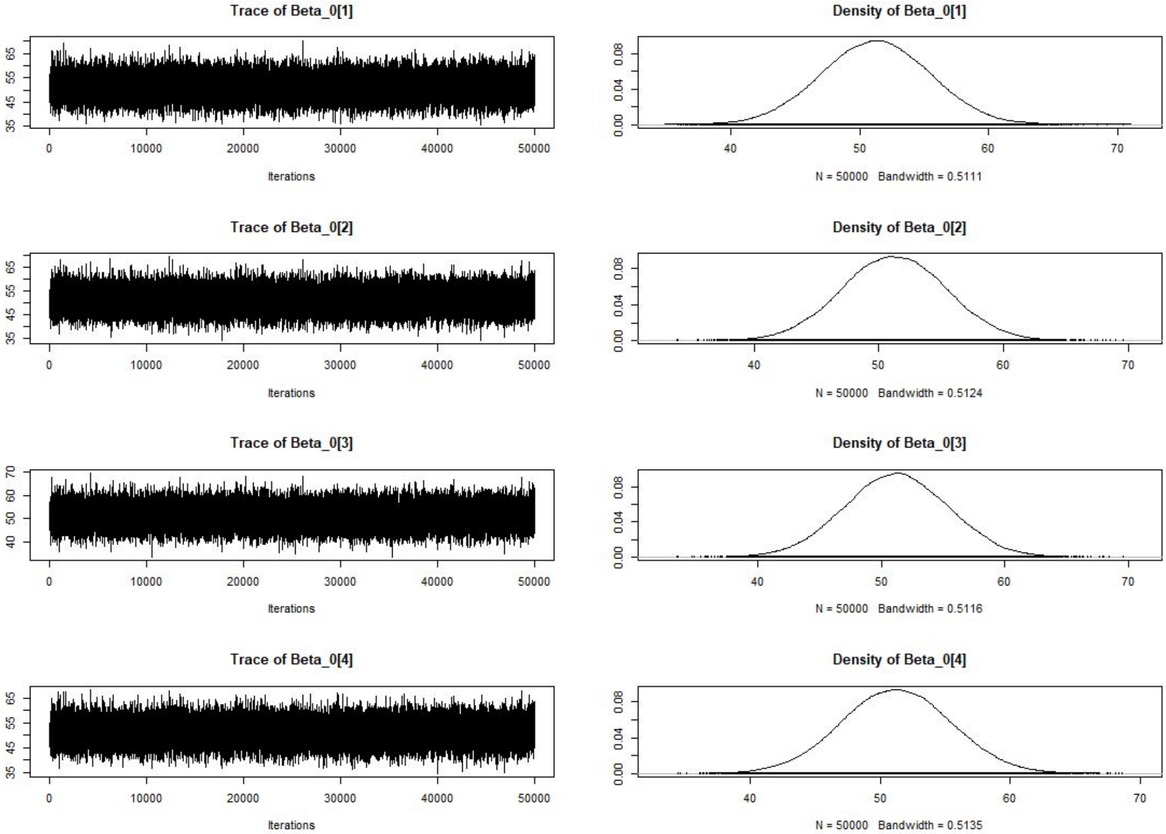

Figure 5.7: Trace plot and the posterior distribution of β_0 for $a = 1, 2, 3, 4$

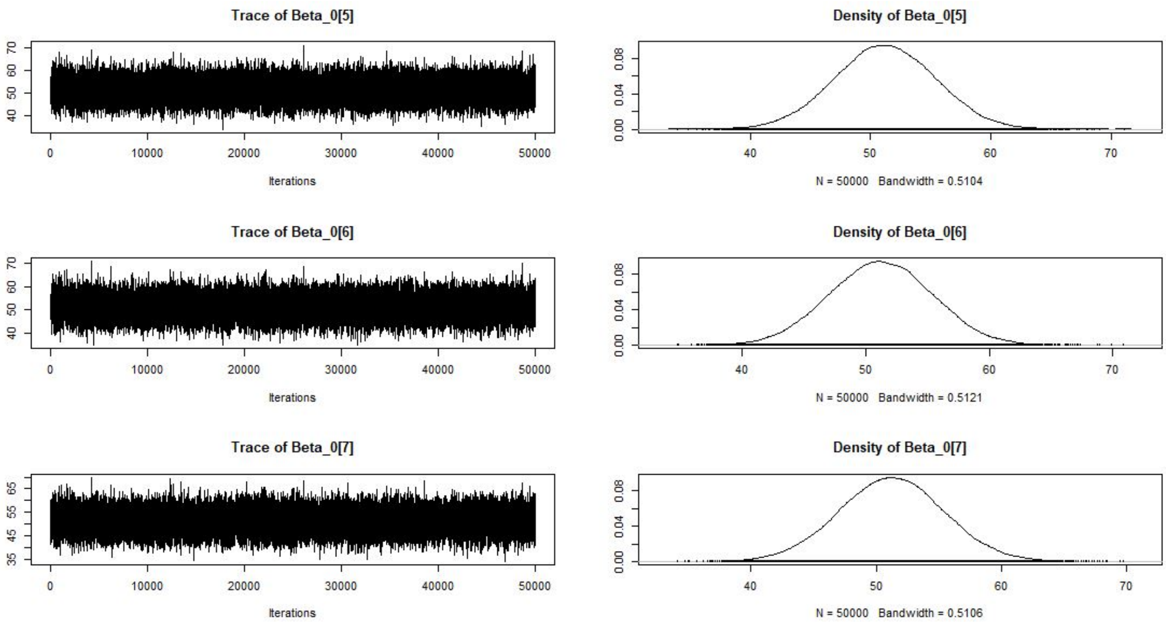

Figure 5.8: Trace plot and the posterior distribution of β_0 for $a = 5, 6, 7$

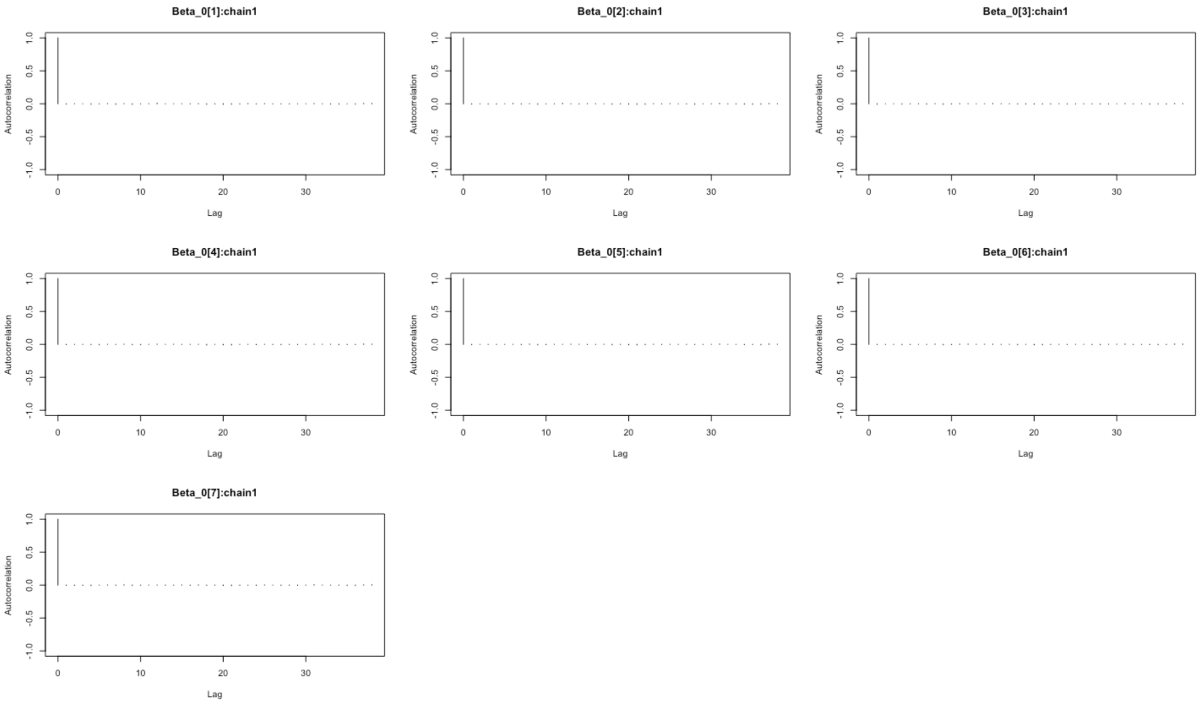

Figure 5.9: Autocorrelation of β_0, for $a = 1, .., 7$

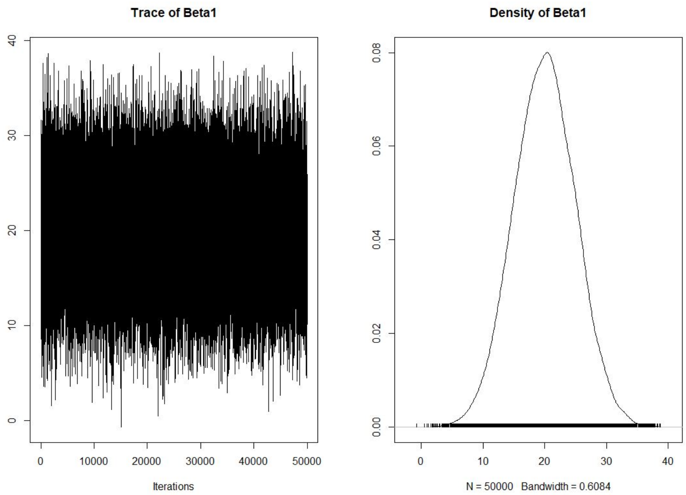

Figure 5.10: Trace plot and the posterior distribution of β_1

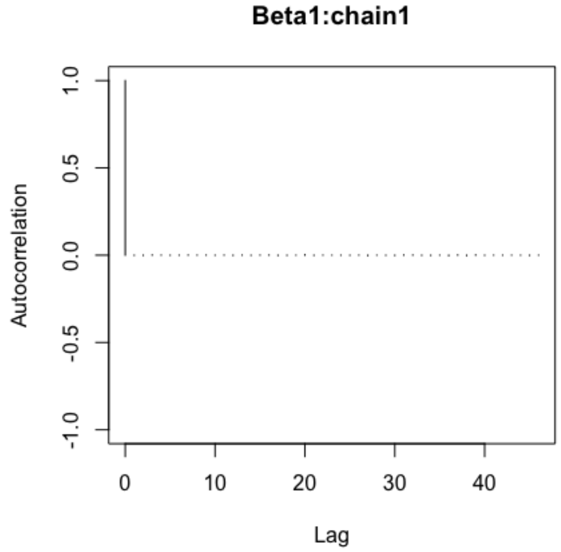

Figure 5.11: Autocorrelation of β_1

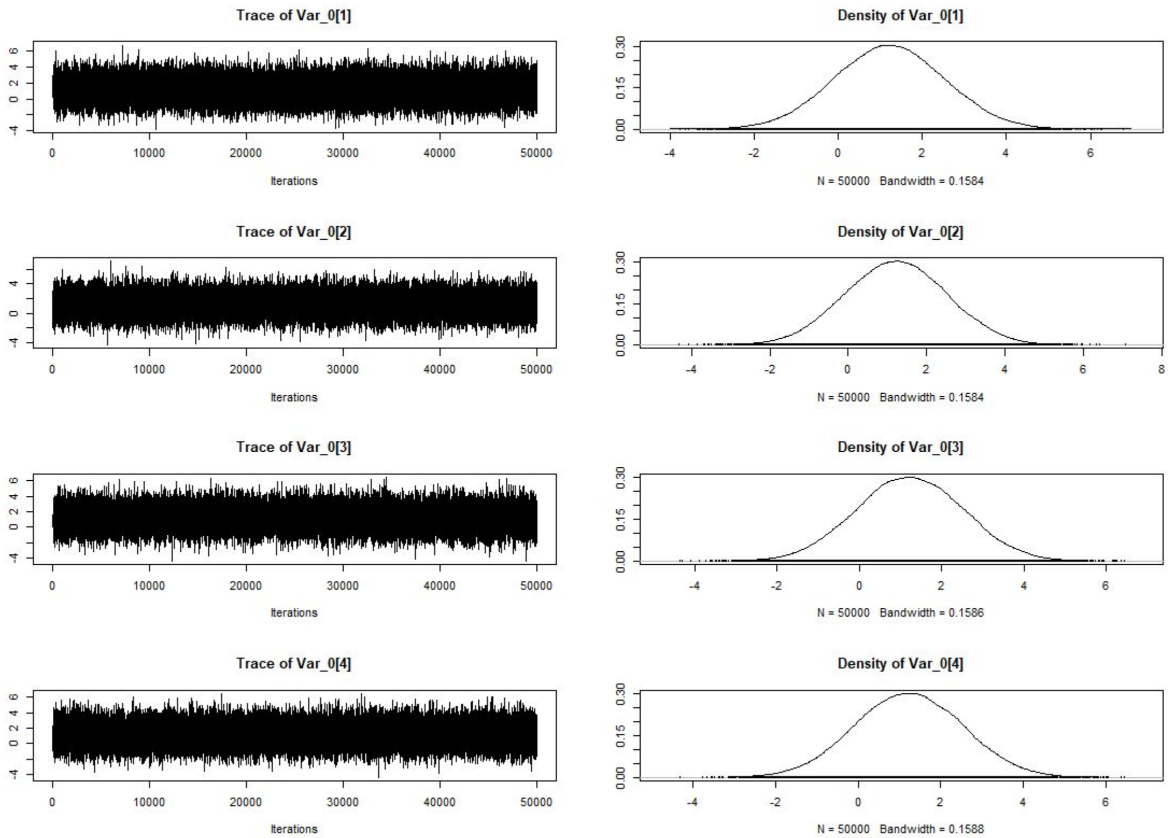

Figure 5.12: Trace plot and the posterior distribution of ε^0 for $a = 1, 2, 3, 4$

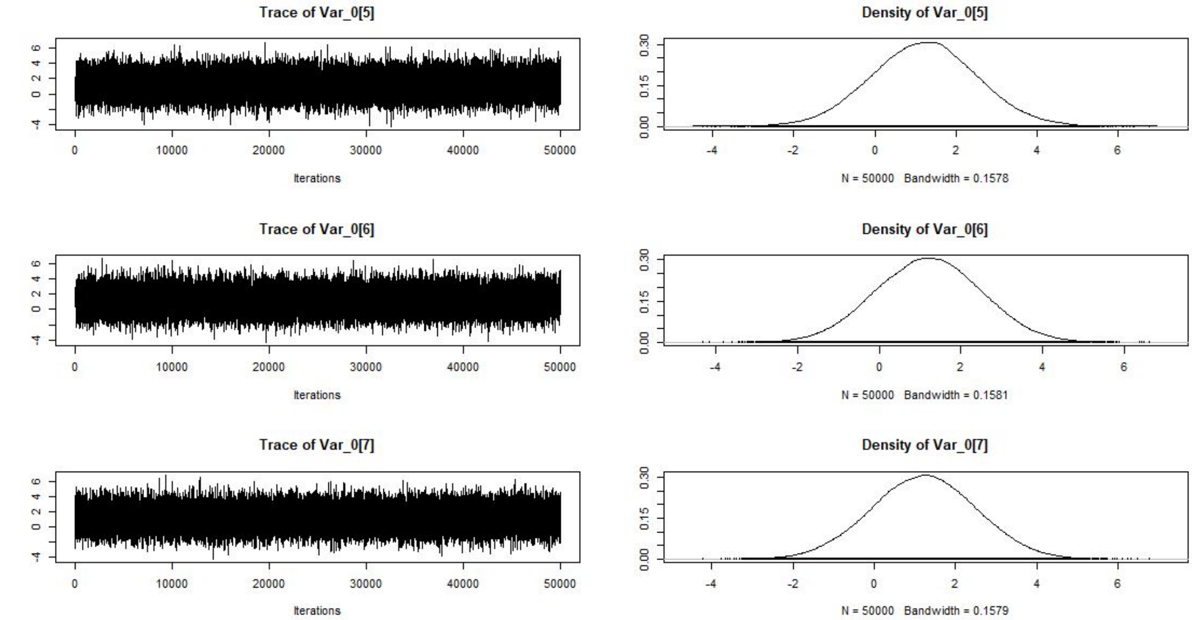

Figure 5.13: Trace plot and the posterior distribution of ε^0 for $a = 5, 6, 7$

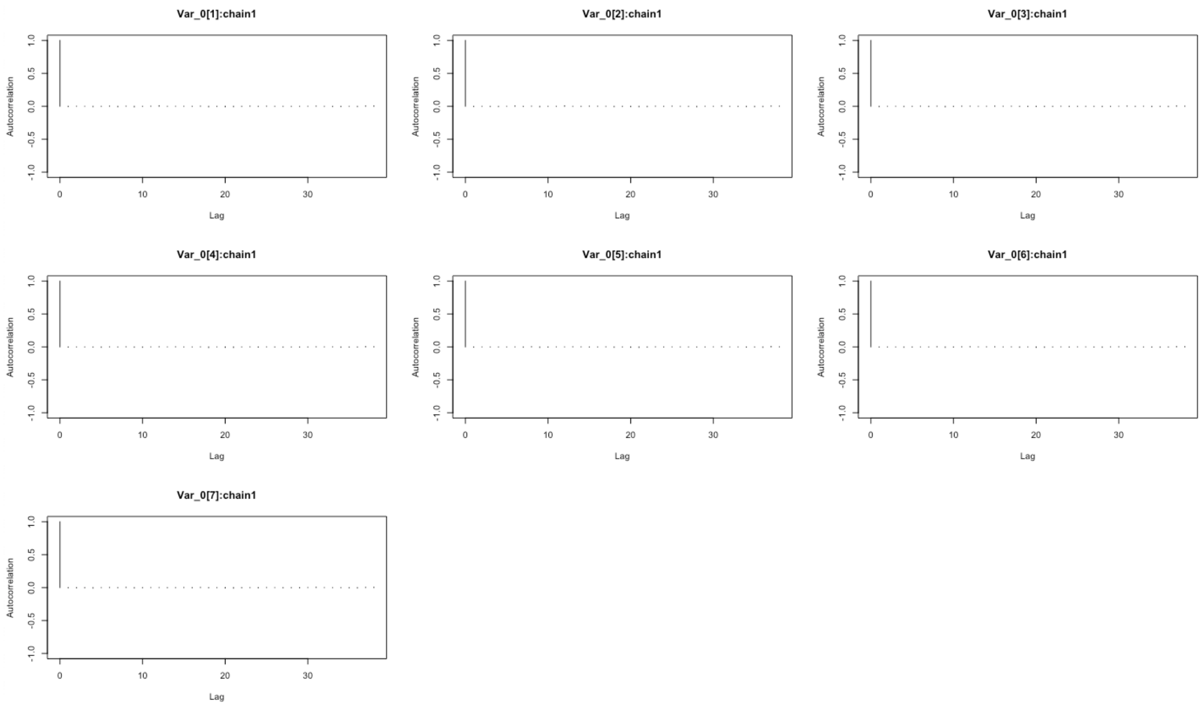

Figure 5.14: Autocorrelation of ε^0, for $a = 1, .., 7$

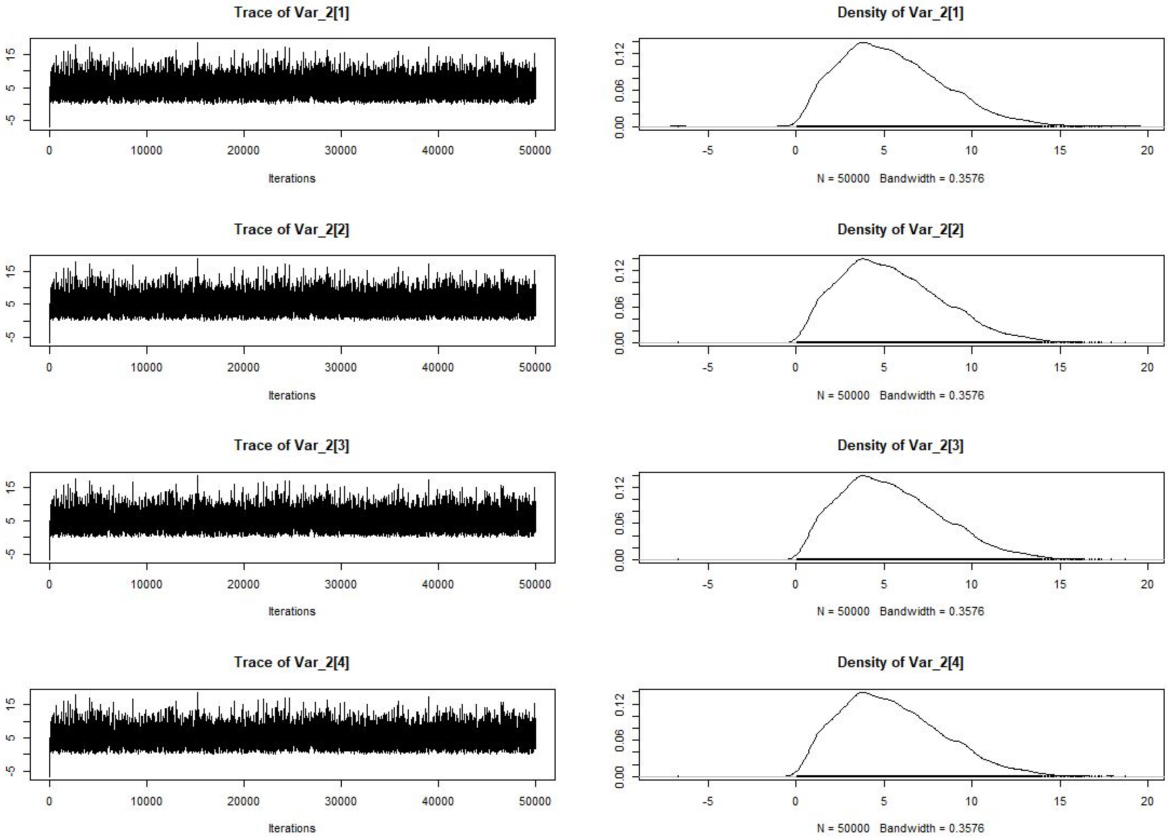

Figure 5.15: Trace plot and the posterior distribution of ε^2 for $a = 1, 2, 3, 4$

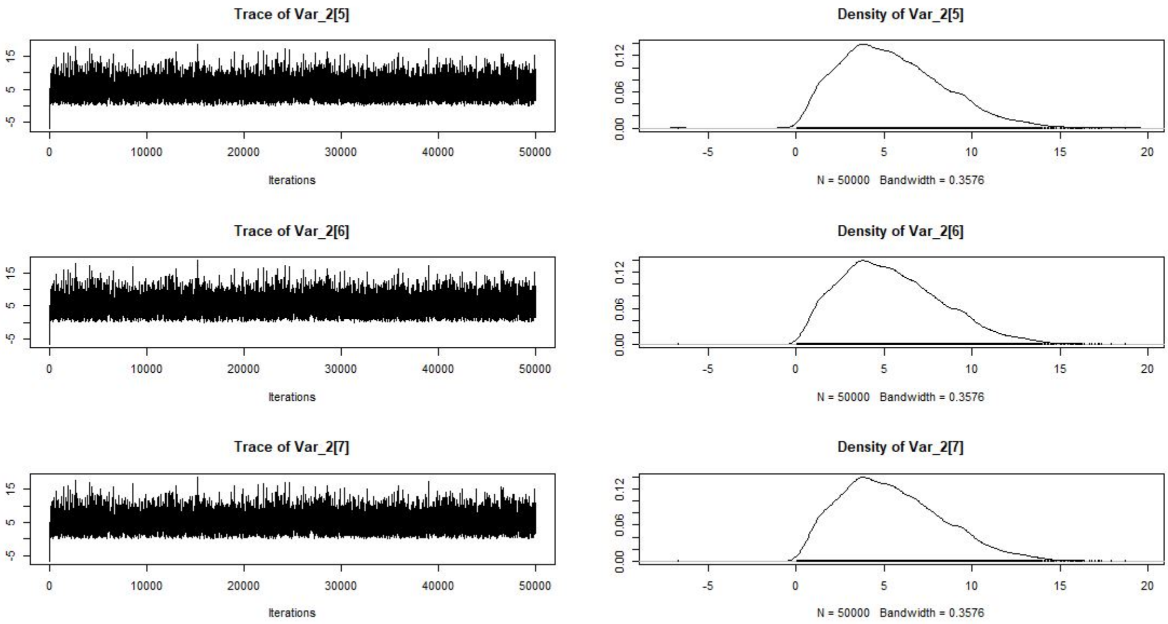

Figure 5.16: Trace plot and the posterior distribution of ε^2 for $a = 5, 6, 7$

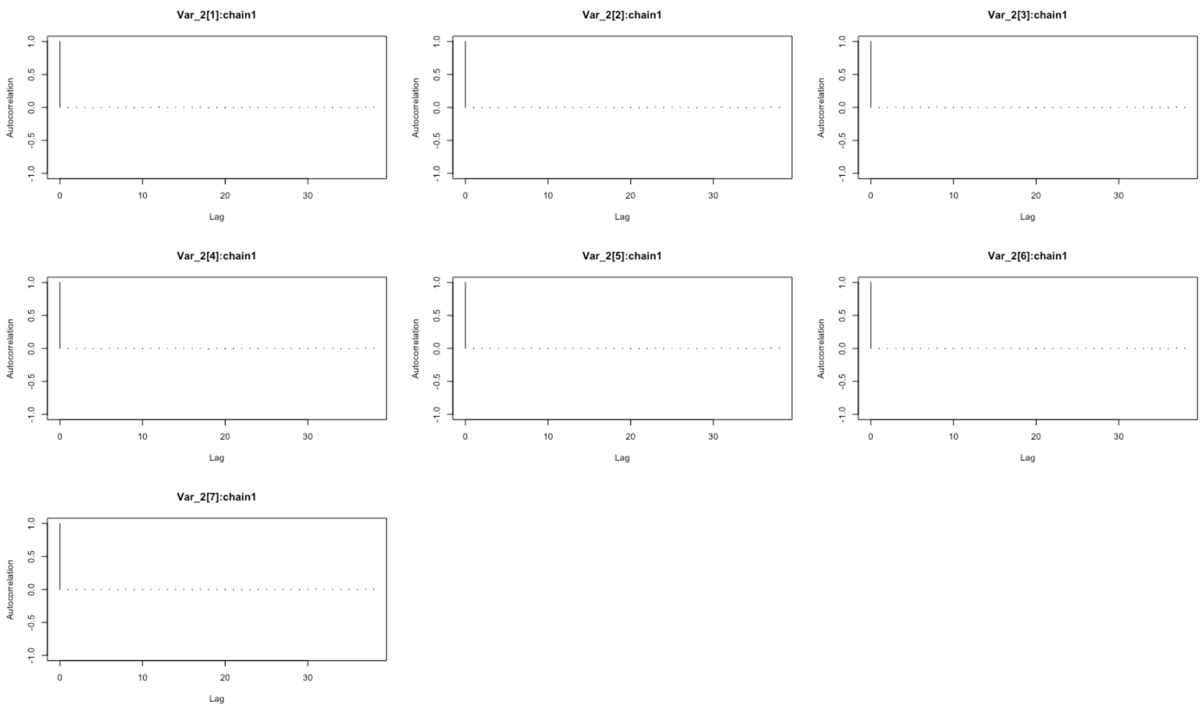

Figure 5.17: Autocorrelation of ε^2, for $a = 1, .., 7$

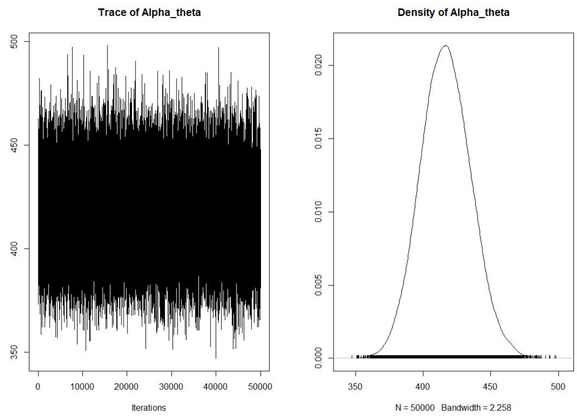

Figure 5.18: Trace plot and the posterior distribution of α_θ

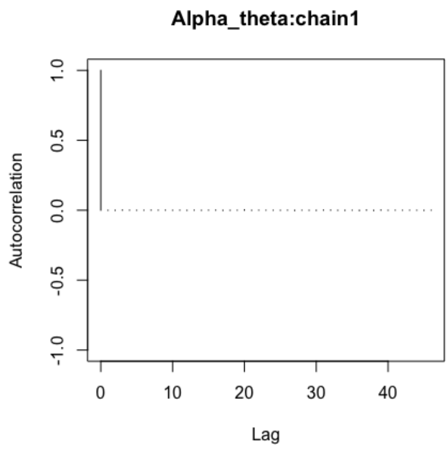

Figure 5.19: Autocorrelation of α_θ

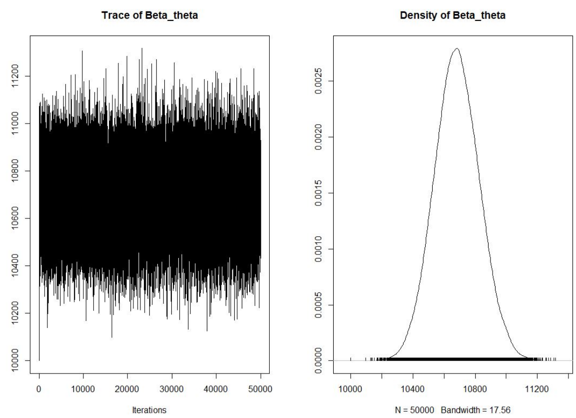

Figure 5.20: Trace plot and the posterior distribution of β_θ

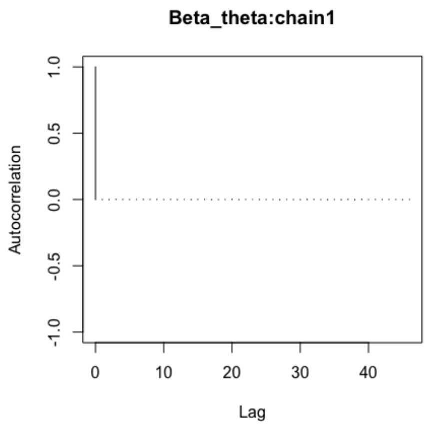

Figure 5.21: Autocorrelation of β_θ

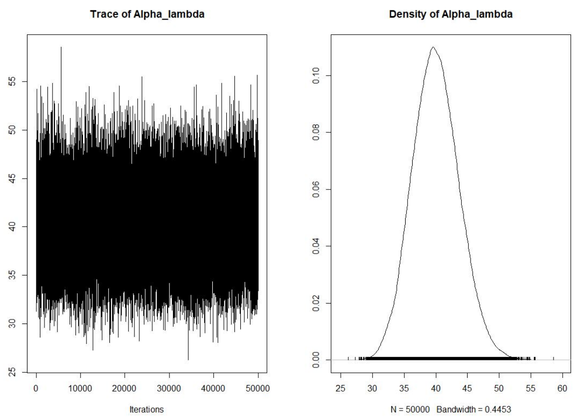

Figure 5.22: Trace plot and the posterior distribution of α_λ

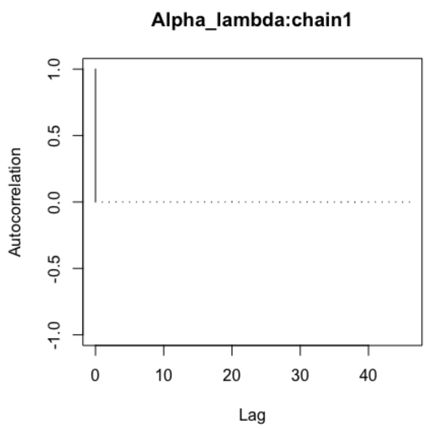

Figure 5.23: Autocorrelation of α_λ

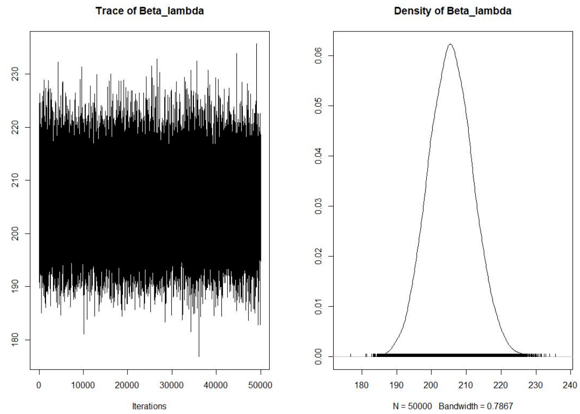

Figure 5.24: Trace plot and the posterior distribution of β_λ

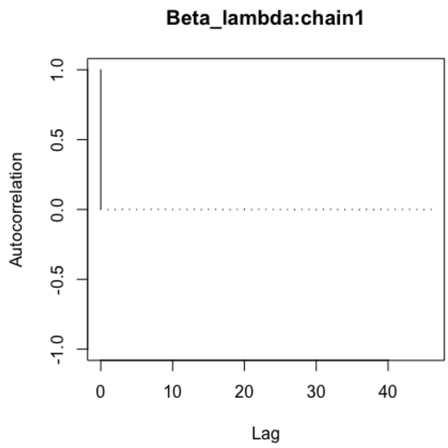

Figure 5.25: Autocorrelation of β_λ

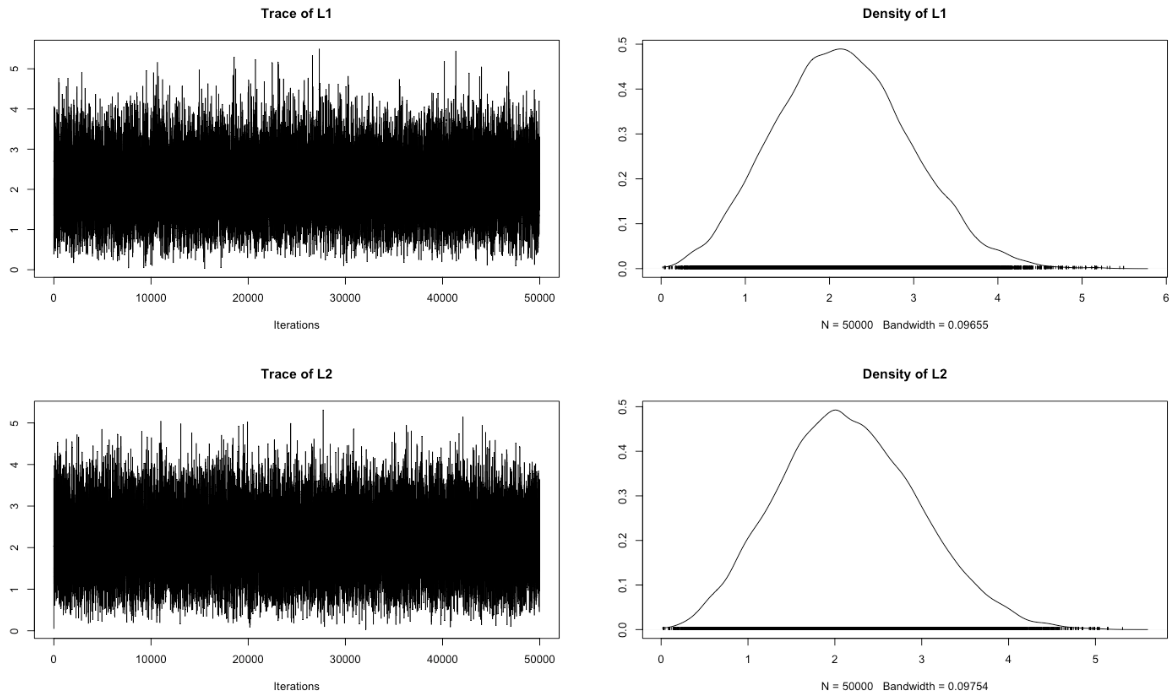

Figure 5.26: Trace plot and the posterior distribution of L_i for $i = 1, 2$

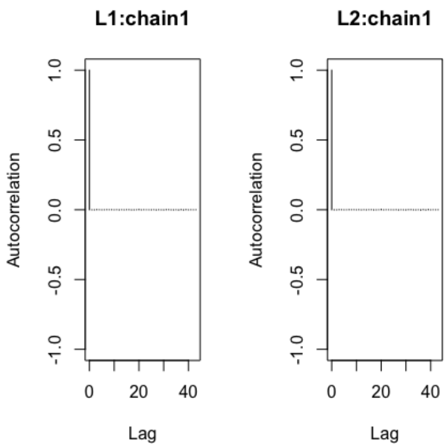

Figure 5.27: Autocorrelation of L_i for $i = 1, 2$

Par.	Mean	SD	2.5% Perc.	Median	97.5% Perc.	Station. Test	Halfwidth Test	P-value
L_1	2.1571	0.7929	0.6867	2.1244	3.7868	Passed	Passed	0.0829
L_2	2.1549	0.793	0.694	2.1243	3.8035	Passed	Passed	0.8184
θ_1	0.0392	0.0024	0.0346	0.0391	0.0439	Passed	Passed	0.26
θ_2	0.0389	0.0024	0.0345	0.039	0.0439	Passed	Passed	0.46
θ_3	0.0391	0.0024	0.0345	0.039	0.0438	Passed	Passed	0.333
θ_4	0.0396	0.0024	0.0345	0.039	0.0439	Passed	Passed	0.323
θ_5	0.0388	0.0024	0.0346	0.0391	0.0438	Passed	Passed	0.978
θ_6	0.0393	0.0024	0.0345	0.039	0.0438	Passed	Passed	0.409
θ_7	0.0398	0.0024	0.0346	0.039	0.0438	Passed	Passed	0.517
λ_1	0.1984	0.0346	0.1363	0.1967	0.2706	Passed	Passed	0.5209
λ_2	0.1982	0.0343	0.1358	0.1963	0.2704	Passed	Passed	0.6615
λ_3	0.1983	0.0343	0.1367	0.1964	0.2706	Passed	Passed	0.0601
λ_4	0.1983	0.0344	0.1363	0.1967	0.2701	Passed	Passed	0.6865
λ_5	0.1986	0.0346	0.1358	0.1967	0.2715	Passed	Passed	0.8604
λ_6	0.1984	0.0345	0.1361	0.1964	0.2715	Passed	Passed	0.7659
λ_7	0.1982	0.0346	0.1358	0.1962	0.2711	Passed	Passed	0.3383
β_{1_0}	51.2266	4.1978	43.0326	51.2199	59.4164	Passed	Passed	0.366
β_{2_0}	51.2216	4.2086	42.9885	51.2122	59.4634	Passed	Passed	0.114
β_{3_0}	51.2297	4.2017	43.0167	51.2166	59.4519	Passed	Passed	0.739
β_{4_0}	51.2364	4.2173	42.9202	51.2218	59.493	Passed	Passed	0.388
β_{5_0}	51.2232	4.1916	42.9841	51.2126	59.4474	Passed	Passed	0.233
β_{6_0}	51.2165	4.2059	42.9748	51.2125	59.4371	Passed	Passed	0.218
β_{7_0}	51.2265	4.2015	42.976	51.2233	59.476	Passed	Passed	0.241
β_{1_2}	2.9555	5.8578	-9.0123	4.3064	12.3846	Passed	Passed	0.831
β_{2_2}	2.9555	5.8578	-9.0123	4.3064	12.3846	Passed	Passed	0.831
β_{3_2}	2.9555	5.8578	-9.0123	4.3064	12.3846	Passed	Passed	0.831
β_{4_2}	2.9555	5.8578	-9.0123	4.3064	12.3846	Passed	Passed	0.831
β_{5_2}	2.9555	5.8578	-9.0123	4.3064	12.3846	Passed	Passed	0.831
β_{6_2}	2.9555	5.8578	-9.0123	4.3064	12.3846	Passed	Passed	0.831
β_{7_2}	2.9555	5.8578	-9.0123	4.3064	12.3846	Passed	Passed	0.831
β_0	50.0077	4	42.1858	50.0029	57.8717	Passed	Passed	0.0832
β_1	20.0063	4.9968	10.2338	20.0168	29.7967	Passed	Passed	0.236
ε_1^0	1.2189	1.3011	-1.3322	1.2156	3.7756	Passed	Passed	0.0564
ε_2^0	1.214	1.301	-1.3229	1.2142	3.76	Passed	Passed	0.057
ε_3^0	1.222	1.3024	-1.3158	1.2205	3.7715	Passed	Passed	0.0515
ε_4^0	1.2288	1.3045	-1.3149	1.2246	3.7806	Passed	Passed	0.1227
ε_5^0	1.2155	1.2961	-1.3157	1.2213	3.7756	Passed	Passed	0.5356
ε_6^0	1.2088	1.2988	-1.3187	1.2101	3.76	Passed	Passed	0.2034
ε_7^0	1.2188	1.2967	-1.3175	1.2219	3.7479	Passed	Passed	0.1416
ε_1^2	5.4847	2.9366	0.9596	5.1249	12.0573	Passed	Passed	0.278
ε_2^2	5.4847	2.9366	0.9596	5.1249	12.0573	Passed	Passed	0.278

Table 5.2 continued from previous page

ε_3^2	5.4847	2.9366	0.9596	5.1249	12.0573	Passed	Passed	0.278
ε_4^2	5.4847	2.9366	0.9596	5.1249	12.0573	Passed	Passed	0.278
ε_5^2	5.4847	2.9366	0.9596	5.1249	12.0573	Passed	Passed	0.278
ε_6^2	5.4847	2.9366	0.9596	5.1249	12.0573	Passed	Passed	0.278
ε_7^2	5.4847	2.9366	0.9596	5.1249	12.0573	Passed	Passed	0.278
τ	35.015	0.0043	34.9001	35.0157	35.1312	Passed	Passed	0.854
$\tau_{\varepsilon 0}$	0.4903	0.0021	0.3977	0.4888	0.5915	Passed	Passed	0.668
$\tau_{\varepsilon 2}$	0.1811	0.0033	0.0720	0.1815	0.2950	Passed	Passed	0.347

Table 5.2: Statistical summary of major parameters

Finally, the statistics of posteriors are summarized in Table 5.2. This contains, the mean, standard deviation, median, credible intervals for the parameters, stationary test with its respective p-value and the Halfwidth Test, the latter is a convergence test. The two parameters of the region, L_1 and L_2, centered on approximately 2, which is consistent with the fact that there is not enough information in terms of the effect of the region in this example; β_{a0} a have averages around 50 and standard deviations equal to 4.2 approximately; β_1 is centered on approximately 20; β_{2a} have average the same mean, approximately 2.9, among all age classes; the claim frequency index, λ_a, reaches the highest values in the 1, 6 and 7 age groups, with average values exceeding 0.19; The severity of the claim, θ_a, does not differ much between the age classes.

Chapter 6

Predictions and Premium Determination

It is of the utmost importance for insurers to know if retained premiums are plenty to cover total losses. For this reason, the application of the values of the underlying parameters in the prediction of the insured population, number of claims and the total amount of the claim for the next unit of time is of interest. The premium can be determined using different premium principles once they can predict the total amount of the claim.

This chapter first contains the Bayesian theory of the prediction algorithm, followed by the predictive results for the numerical example presented in Chapter 5. Finally, it demonstrates ways to determine the premium based on the claim amounts provided under certain premium principles.

6.1 Predictive Results

In order to perform the prediction in Bayesian inferences, the *codamenu* function was used, which is a simple menu-based interface for the functions of the Coda package in the R language. The values to be predicted and the data already simulated are treated as the input values in the menu. The prediction is made based on the values of the parameters, which are determined by the observed data. The general purpose is to determine the prediction of the total claim amount for the next period of time. In order for insurers to determine the premium based on their risk tolerance index measured, for example, by standard deviation or Value at Risk (VaR), it is preferred to have a prediction distribution. This model implemented under the Bayesian paradigm offers the advantage of obtaining a posterior predictive distributions, which provide important information for the experience qualification process.

This project presents only the prediction for a unit of time, which is based on data from the last 20 units of time. The prediction of time 21 is presented in terms of the insured population (see figures 6.1, 6.2, 6.5 and 6.4), the total number of claims (see figures 6.16,6.17, 6.18 and 6.19) and the total amount of claims (see figures 6.37, 6.38, 6.39 and 6.40) are presented below.

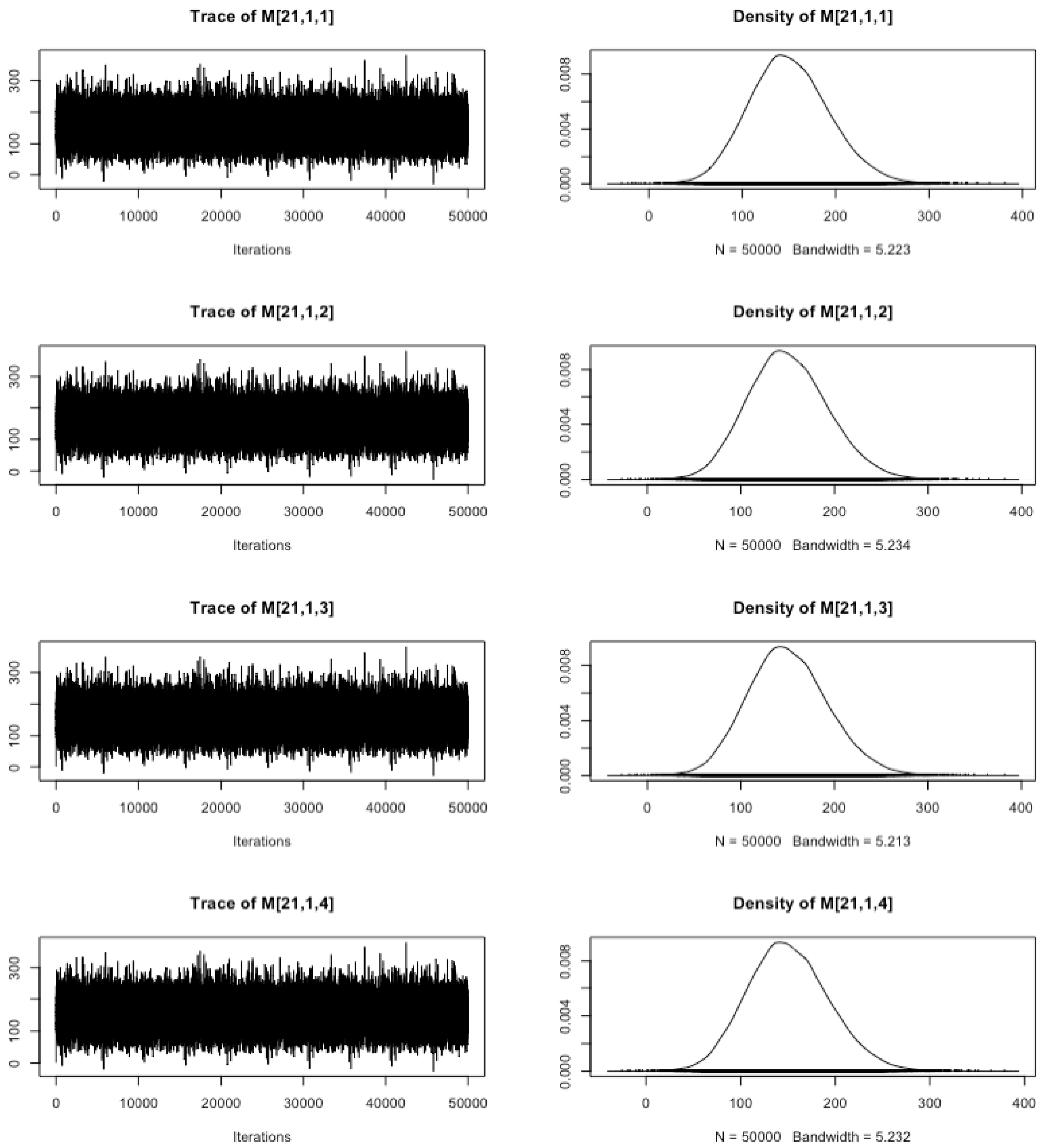

Figure 6.1: Prediction of Insured Population for $21st$ Time Unit - Region 1 - Age classes $1, 2, 3$ and 4

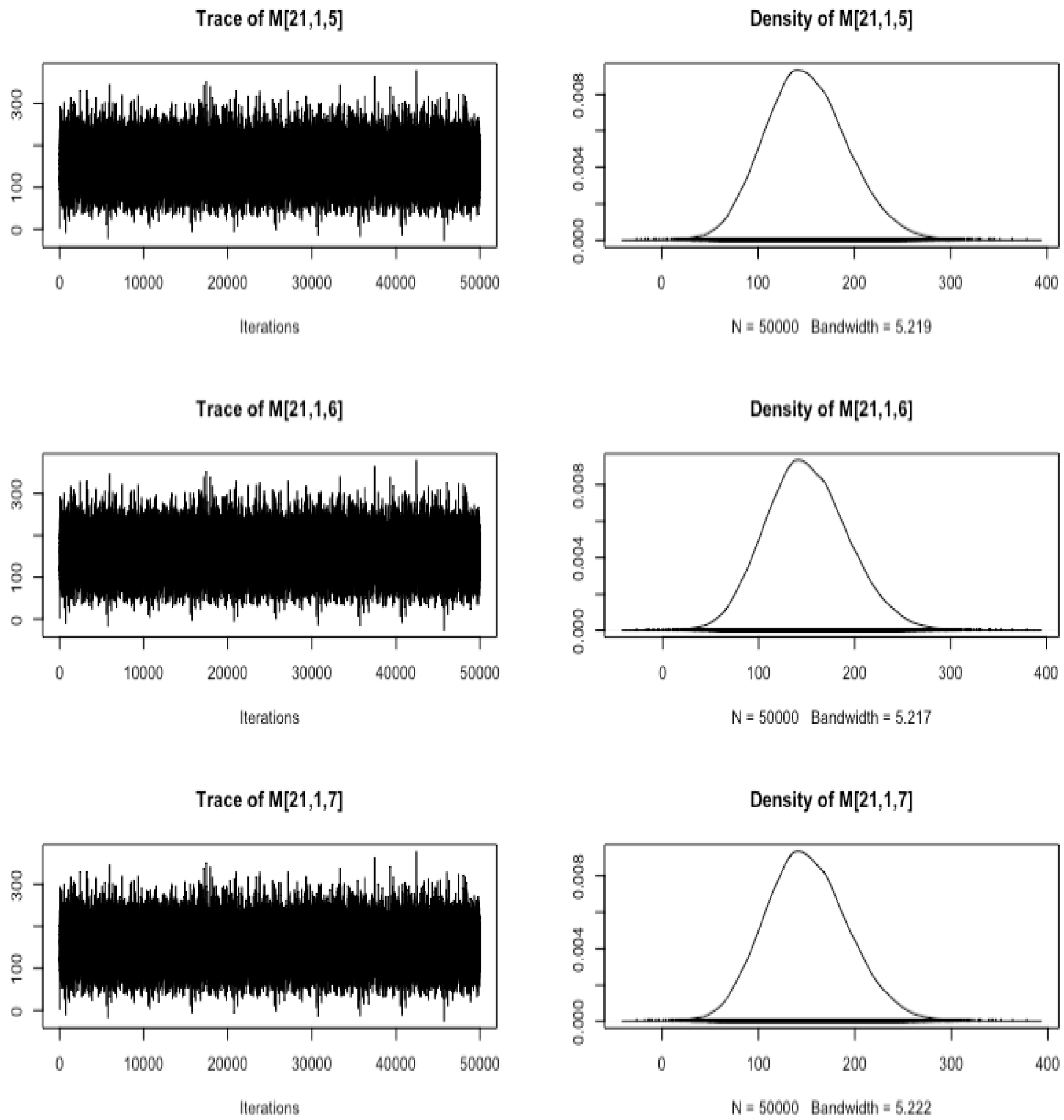

Figure 6.2: Prediction of Insured Population for $21st$ Time Unit - Region 1 - Age classes $5, 6$ and 7

The predicted Insured Population for the 21st time unit by all age classes in region 1 is shown in Figures 6.1 and 6.2. As in the previous chapter, the plot chart has 50,000 iterations for each age class. The domain of the planned insured population could take any value that varies between 50 and 300. This prediction is presented in continuous format with its respective probability density function. In addition, the convergence is highlighted for each age class in both figures, since each trace cyclically alternates up and down and the average lines of the three traces overlap.

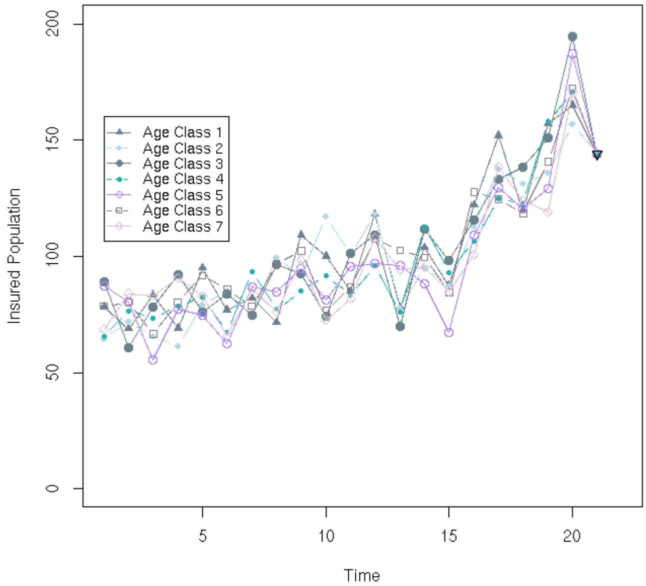

Figure 6.3: Insured Population with Prediction for all Age Groups - Region 1

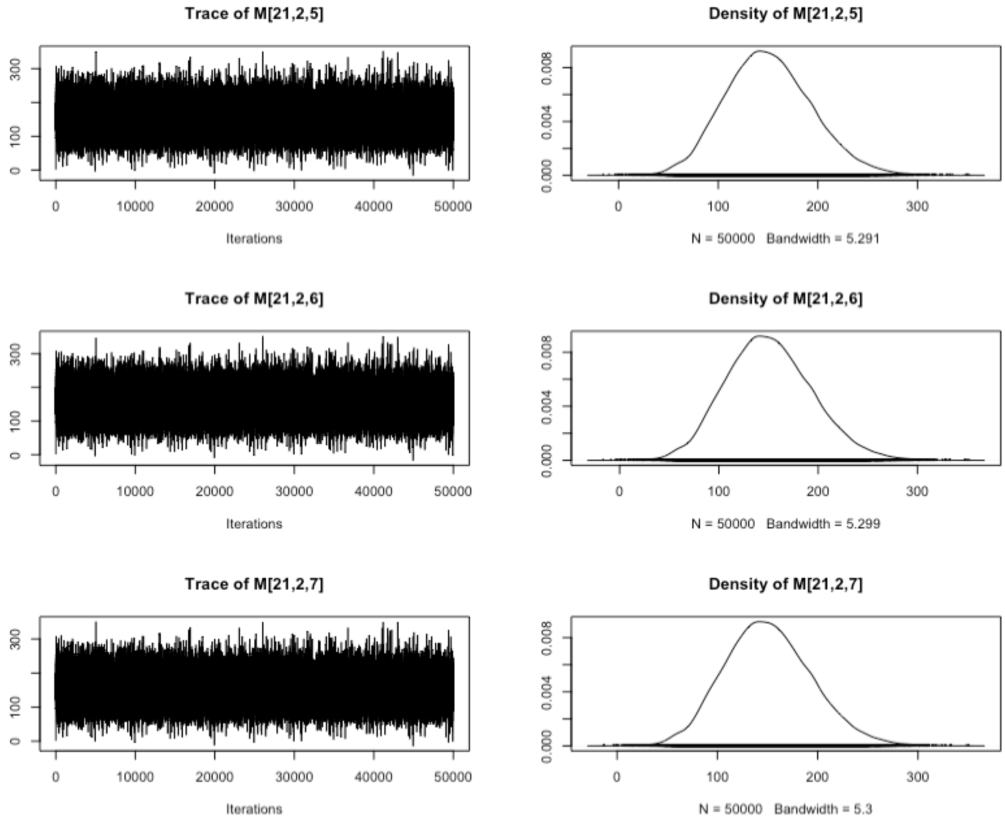

Figure 6.4: Prediction of Insured Population for 21*st* Time Unit - Region 2 - Age classes 5, 6 and 7

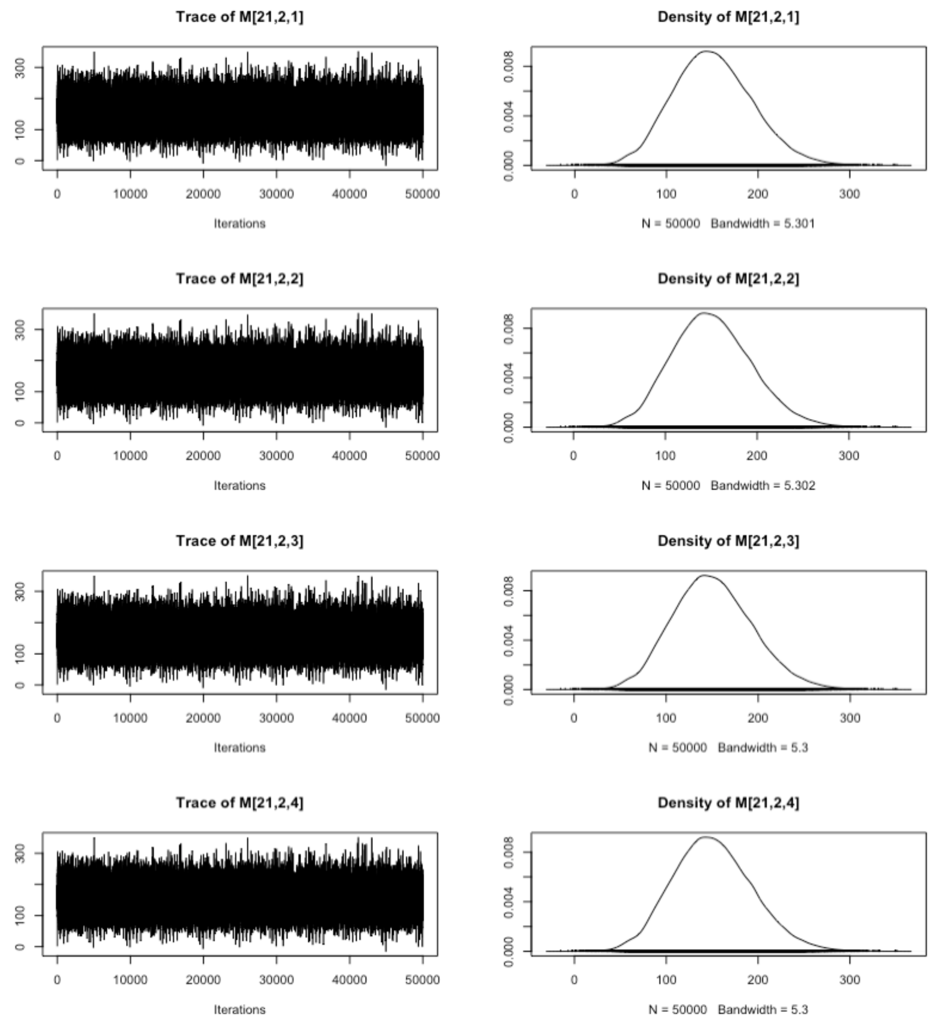

Figure 6.5: Prediction of Insured Population for $21st$ Time Unit - Region 2 - Age classes $1, 2, 3$ and 4

Figure 6.3 shows a comparison of the average insured population among all age classes in region 1. The points up to the unit of time 20 are the historical data of the average insured population, while the triangles located in the unit of time 21 represent the expected average value for this unit of time. The last 20 units of time data reveal an increasing trend which is significantly higher than the value of the initial population. The predicted average values are less than the previous unit of time for all age classes and tend to similar values; If historical values are large, as shown in the previous unit of time, the predicted average values tend to decrease. In general, the planned insured population reflects the characteristics of population growth and, therefore, can be considered as a reasonable estimate.

On the other hand, the traces and their respective probability density of the insured population by each age class in region 2 are also shown in Figures 6.5 and 6.4. The predicted insured population in region 2 is not very different from region 1. Therefore, the same analysis as in Figures 6.1 and 6.2 can be inferred.

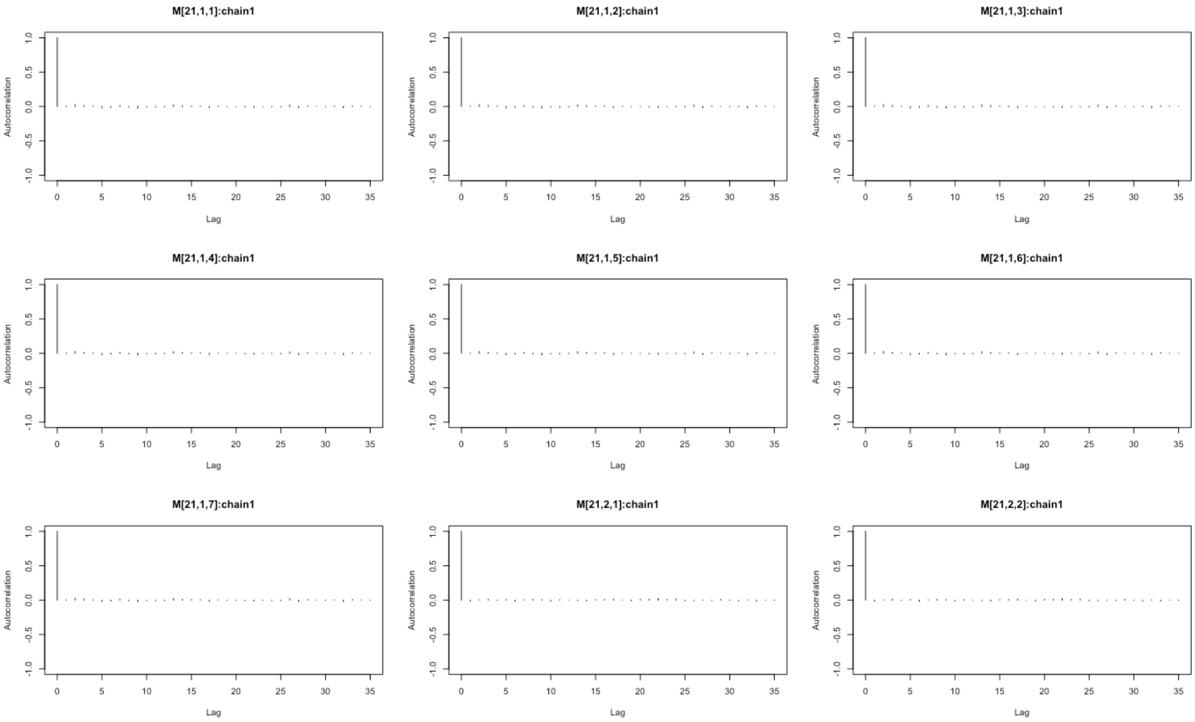

Figure 6.6: Autocorrelation of Prediction of Insured Population for 21st Time Unit - Region 1 and 2- All Age classes

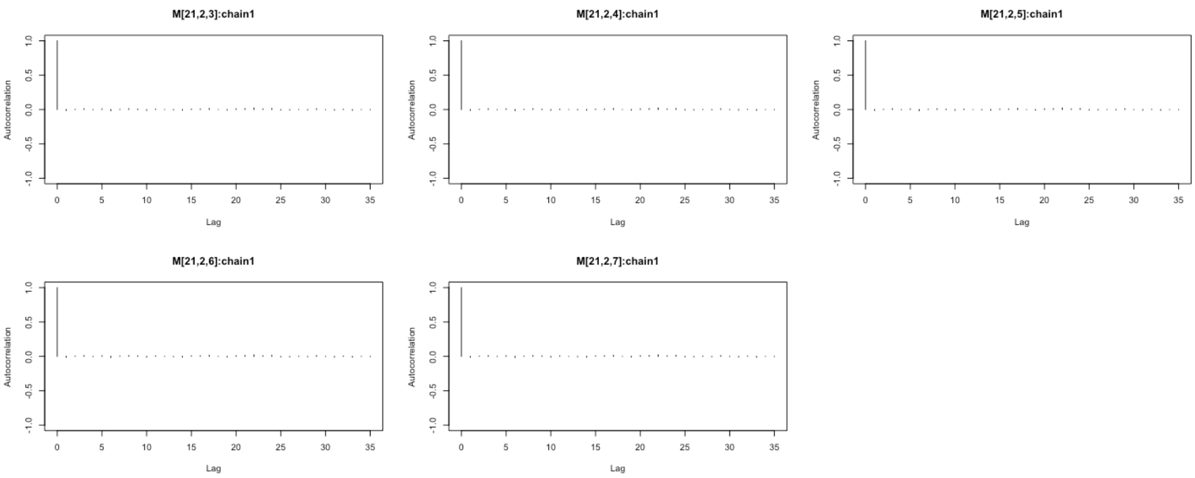

Figure 6.7: Autocorrelation of Prediction of Insured Population for 21st Time Unit - Region 2- All Age classes

Additionally, in Figure 6.6 can be seen the autocorrelation of Prediction of Insured Population for the 21st-time unit by all age classes for region 1 and age classes 1 and 2 for region 2. In addition, figure 6.7 contains the autocorrelation of Prediction of Insured Population for the 21st-time unit by age classes 3, 4, 5, 6 and 7 for region 2. In these figures it can be seen that there is independence in each sample as it was effectively observed in their respective traces.

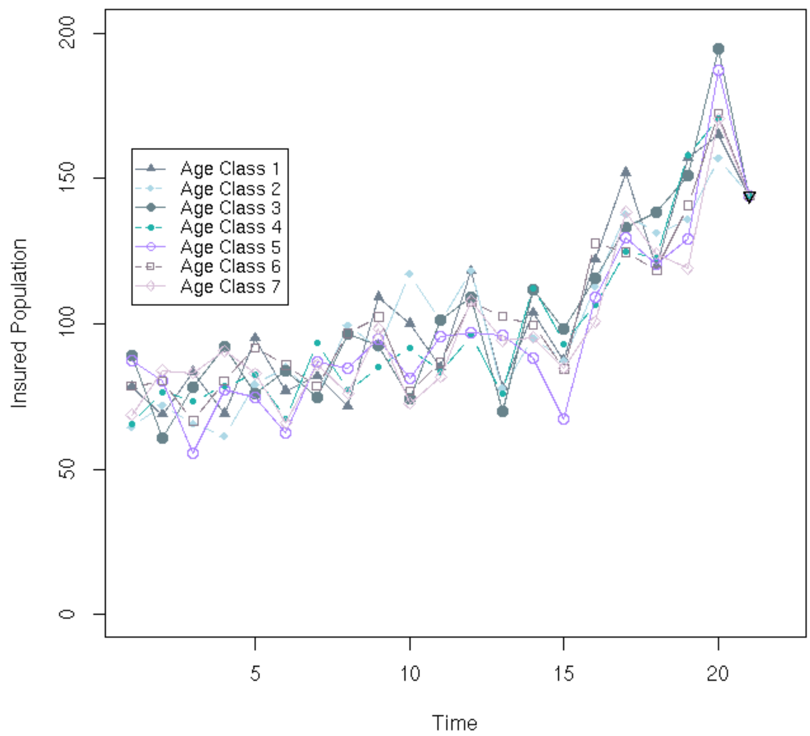

Figure 6.8: Insured Population with Prediction for all Age Groups - Region 2

The comparison of the average insured population among all age classes in region 2 is shown in Figure 6.8. Again, the triangles located in the unit of time 21 represent the expected average value for this unit of time while the points up to the unit of time 20 are the historical data of the average insured population. The tendency of the average insured population in region 2 is similar to that of the region 1, therefore, the same analysis can be applied to this Figure.

Figures 6.9, 6.10, 6.11, 6.12, 6.13, 6.14 and 6.15 present the comparison of predicted insured population by region 1 and 2 in age class 1, 2, 3, 4, 5, 6 and 7 respectively. The black triangles represent the average of the predicted value for the unit of time 21, while the points in the first 20 units of time are historical data. It can be seen that the predicted values by region in each age class do not differ much.

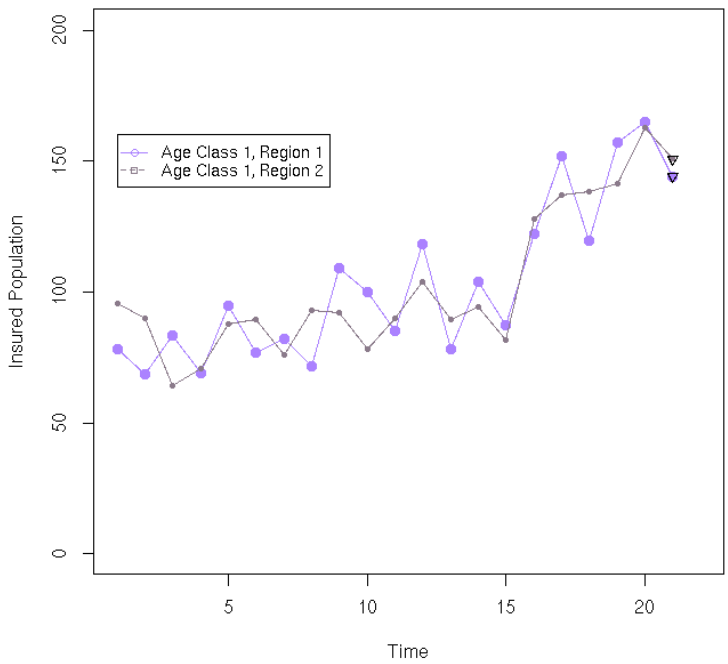

Figure 6.9: Insured Population with Prediction for Region 1 and 2 - Age Group 1

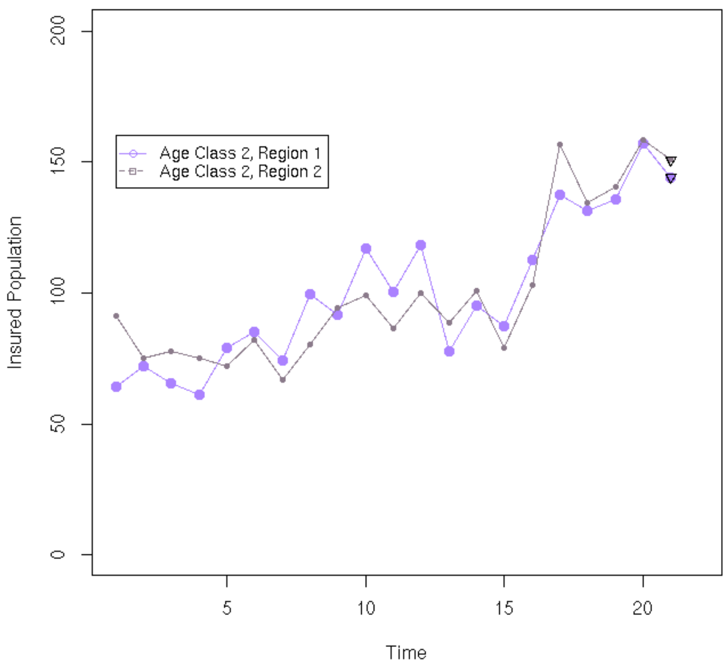

Figure 6.10: Insured Population with Prediction for Region 1 and 2 - Age Group 2

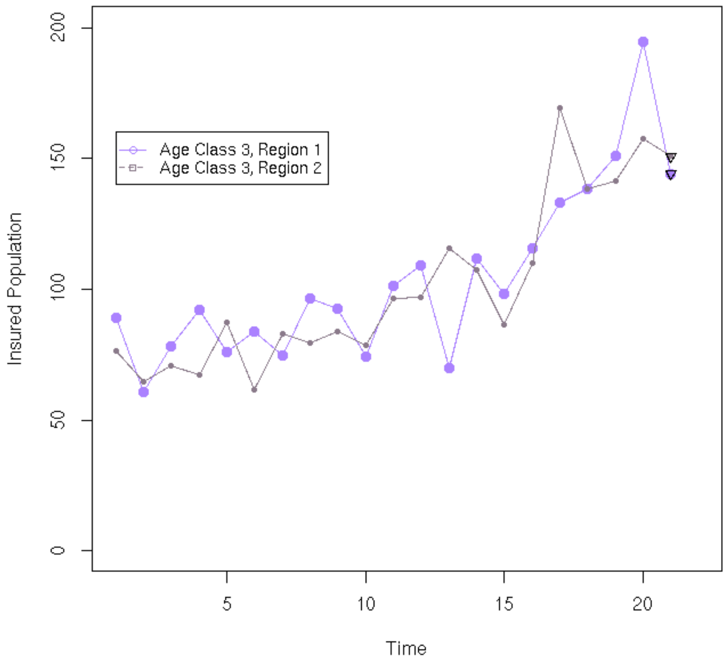

Figure 6.11: Insured Population with Prediction for Region 1 and 2 - Age Group 3

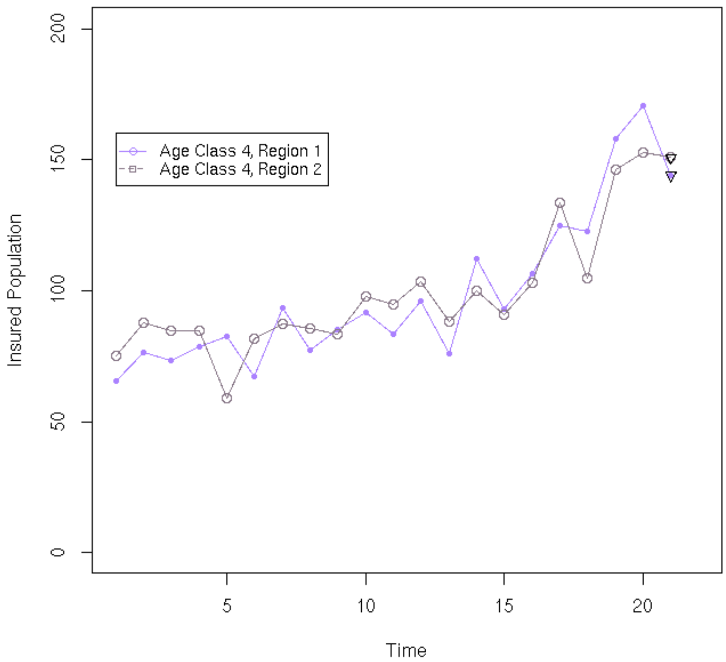

Figure 6.12: Insured Population with Prediction for Region 1 and 2 - Age Group 4

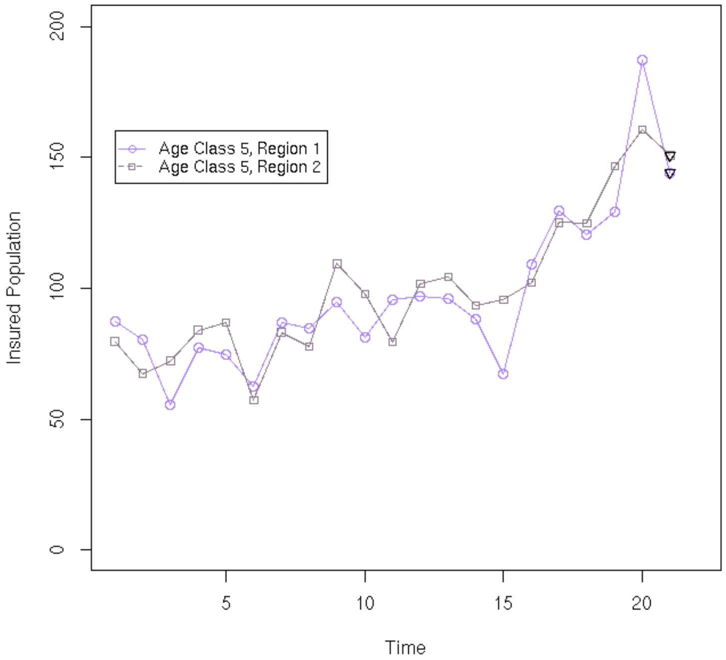

Figure 6.13: Insured Population with Prediction for Region 1 and 2 - Age Group 5

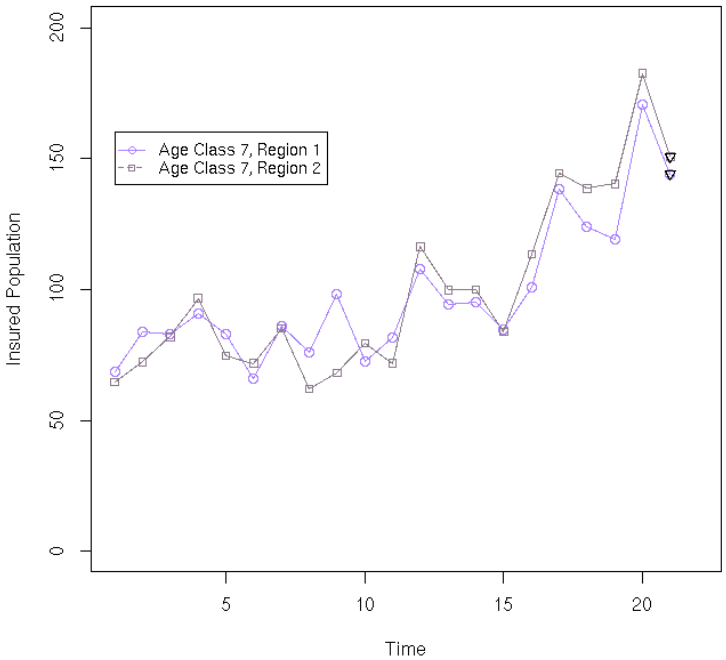

Figure 6.14: Insured Population with Prediction for Region 1 and 2 - Age Group 6

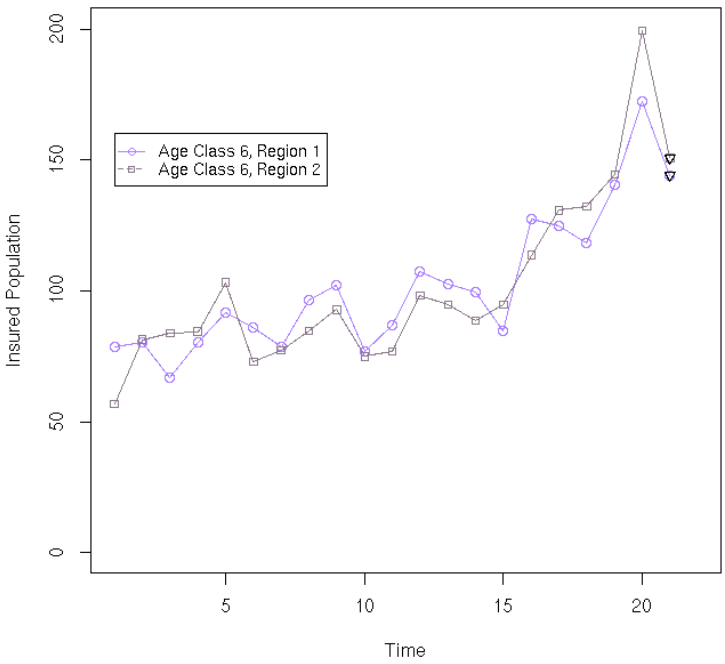

Figure 6.15: Insured Population with Prediction for Region 1 and 2 - Age Group 7

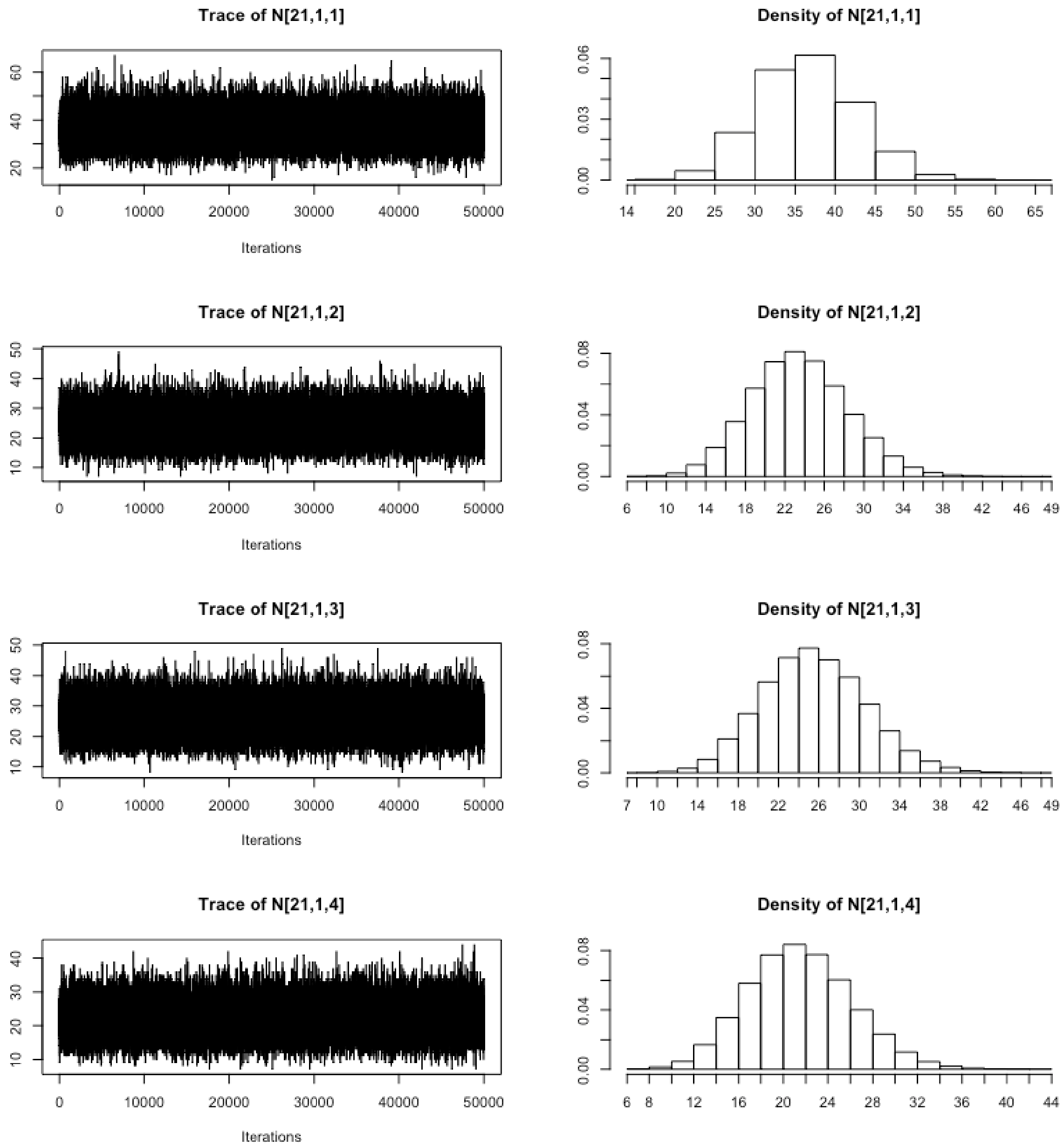

Figure 6.16: Prediction of Number of Claims for $21st$ Time Unit - Region 1 - Age classes $1, 2, 3$ and 4

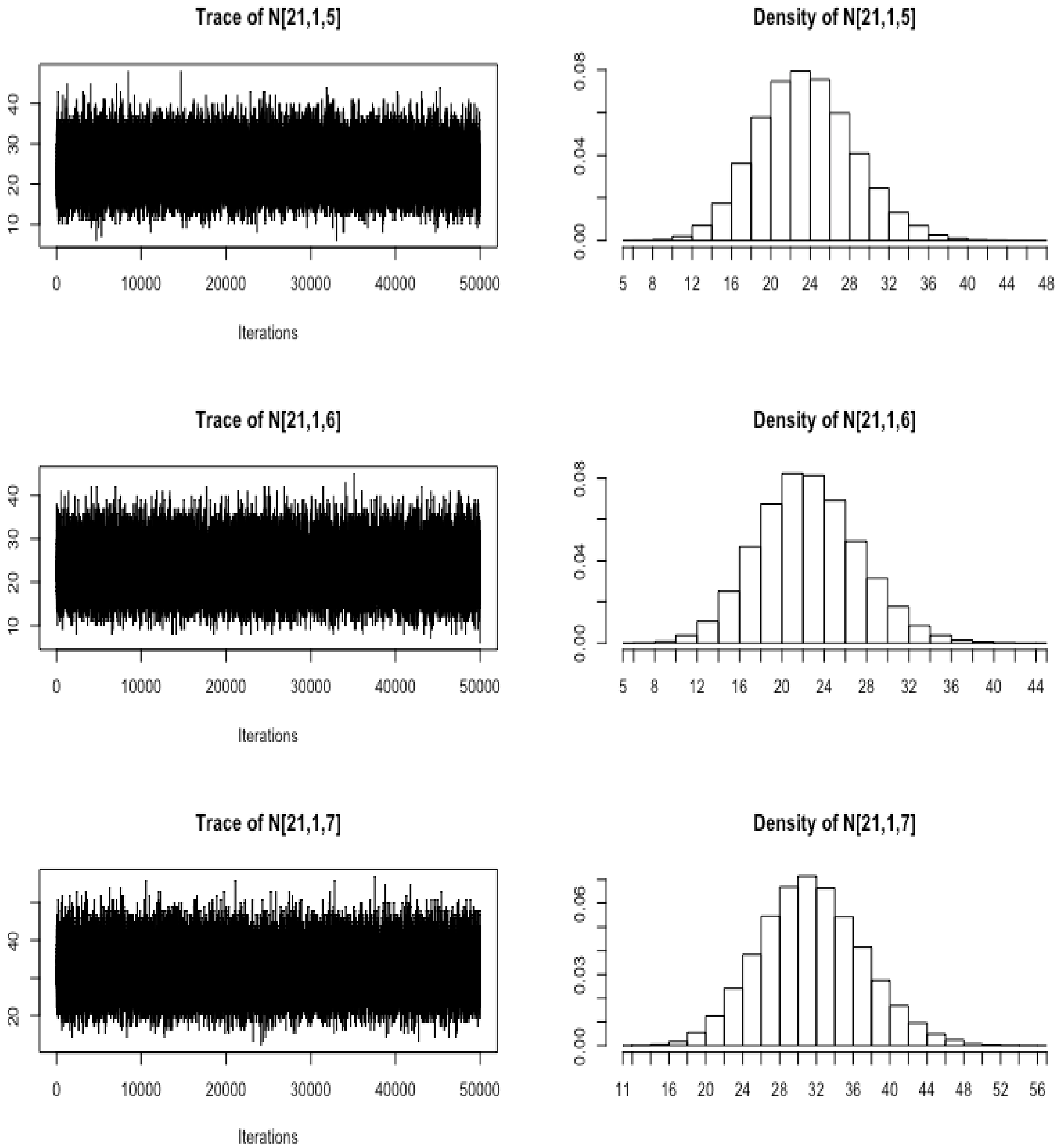

Figure 6.17: Prediction of Number of Claims for $21st$ Time Unit - Region 1 - Age classes $5, 6$ and 7

The posterior prediction for the number of claims for the unit of time 21 in region 1 is shown in Figures 6.16 and 6.17. Traces show strong evidence of convergence for each age class. On the other hand, due to the number of claims is discrete with a small domain, the density graphs are displayed discreetly. In these figures it can be seen that the distribution for each age class has a symmetrical bell shape, but the average number of claims of the 1 age class differs greatly with the rest of the age classes since the average claims for the 1 age class is approximately 30, while in the other age classes the value is lower.

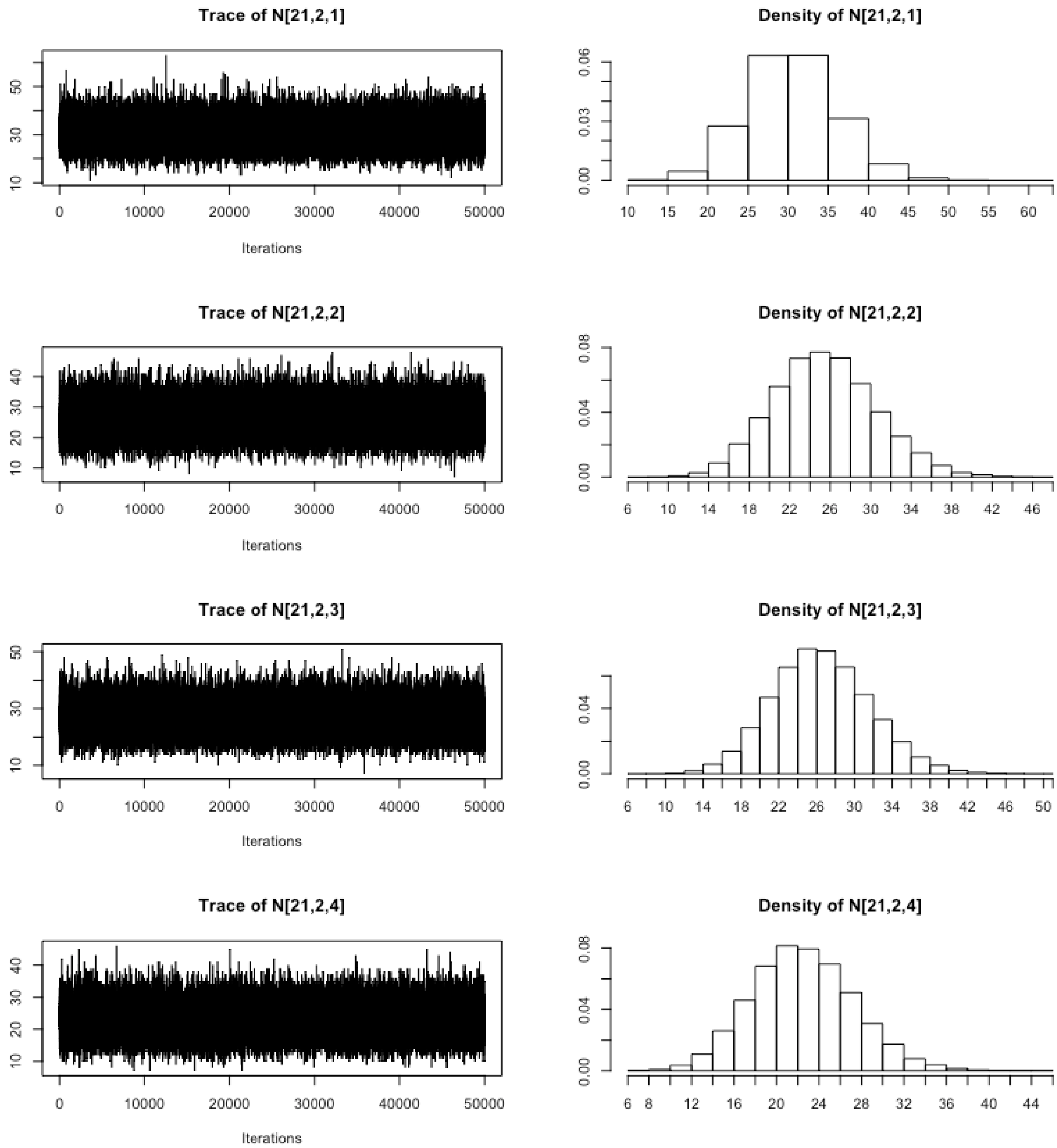

Figure 6.18: Prediction of Number of Claims for $21st$ Time Unit - Region 2 - Age classes $1, 2, 3$ and 4

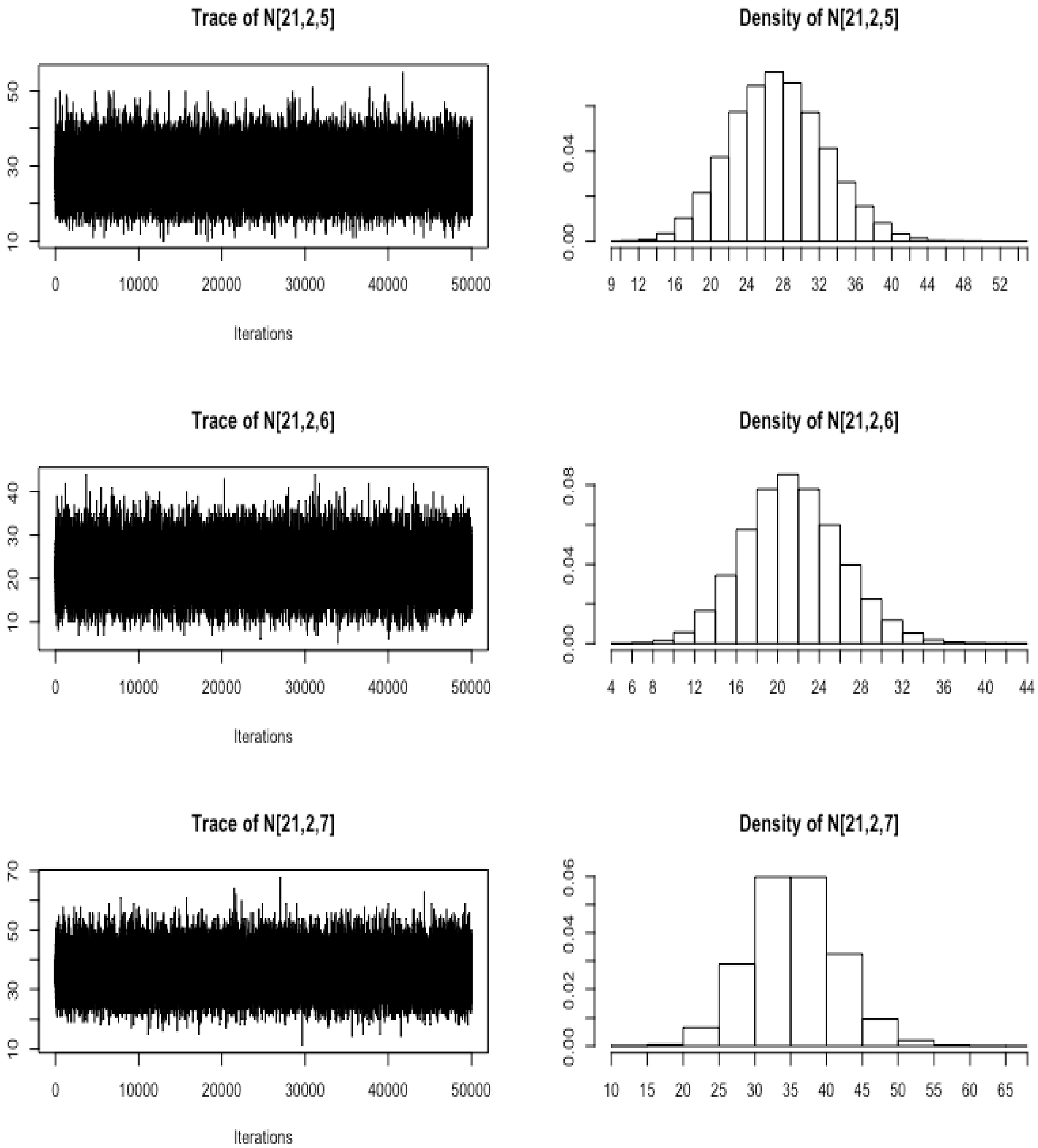

Figure 6.19: Prediction of Number of Claims for $21st$ Time Unit - Region 2 - Age classes $5, 6$ and 7

In Figures 6.18 and 6.19 the posterior prediction for the number of claims for the unit of time 21 in region 2 are shown. Again, there is strong evidence of convergence for each age class by looking at the trace plots. The displayed density plots are shown discretely obtaining a symmetrical bell shape. Since the average claims for the age classes 1 and 7 is greater or equal than 30 and the average number of claims of the the rest age classes is clearly smaller.

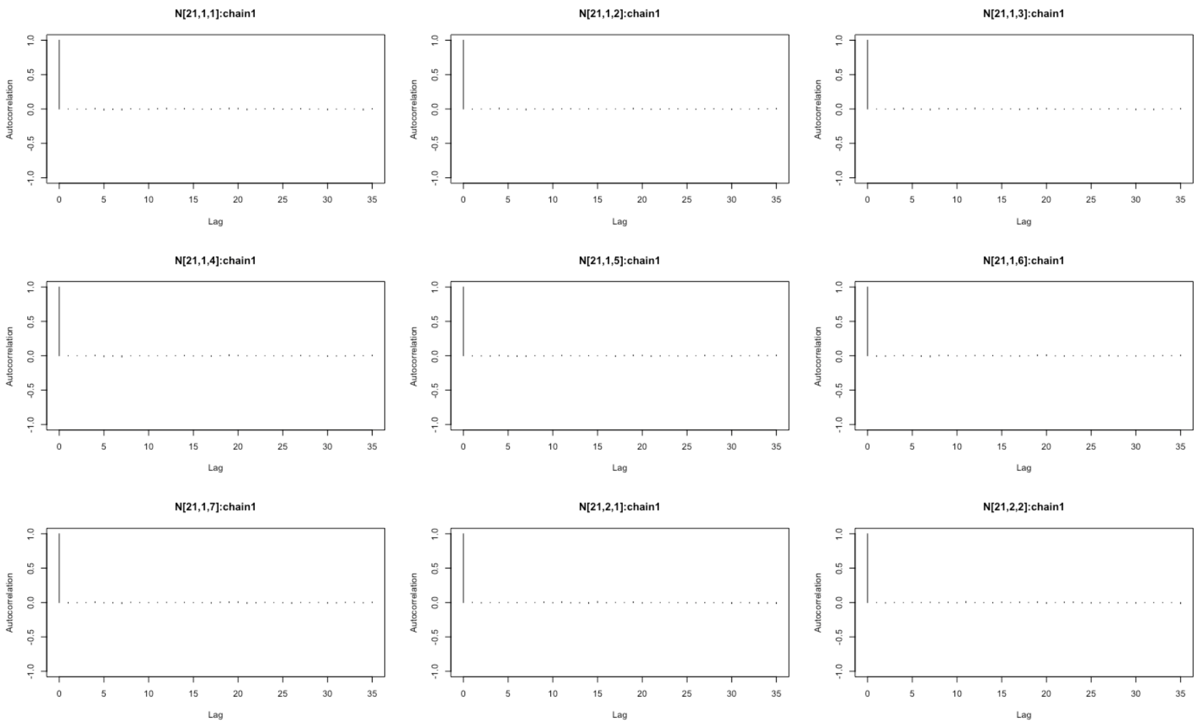

Figure 6.20: Autocorrelation of Prediction of Number of Claims for 21*st* Time Unit - Region 1 and 2- All Age classes

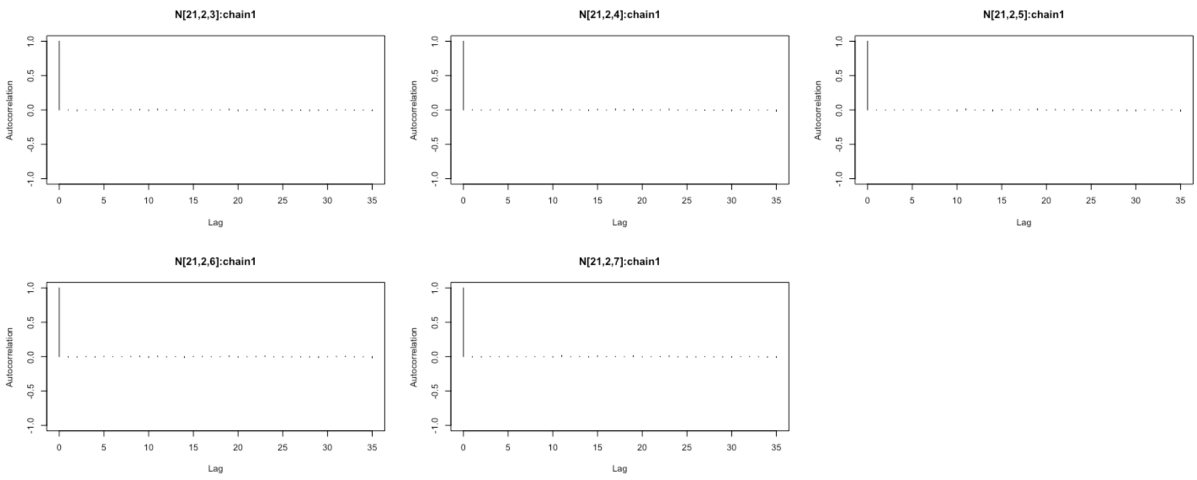

Figure 6.21: Autocorrelation of Prediction of Number of Claims for 21*st* Time Unit - Region 2- All Age classes

In addition, autocorrelation of prediction of number of claims for 21*st*-time Unit for region 1 by all age classes and for region 2 by age classes 1 and 2 is shown in Figure 6.20. On the other hand, 6.21 contains the autocorrelation of prediction of number of claims for 21*st* time Unit for region 2 by age classes $3, 4, 5, 6$ and 7. In both figures clearly exist independence between the samples.

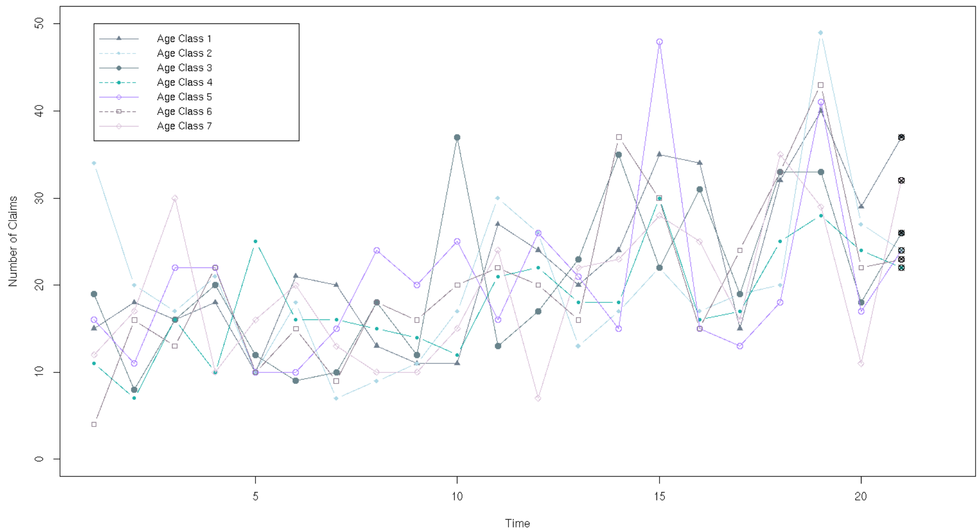

Figure 6.22: Claim Number with Prediction for all Age Classes - Region 1

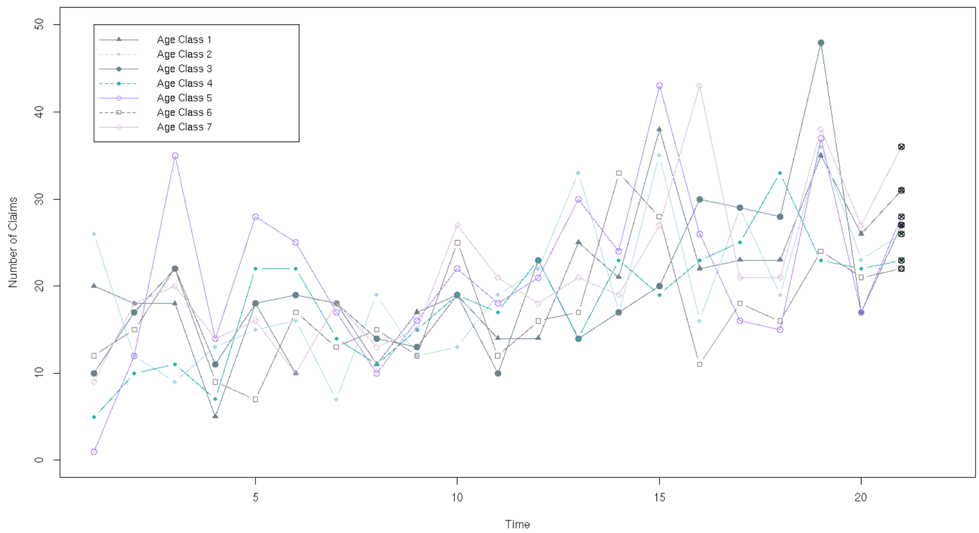

Figure 6.23: Claim Number with Prediction for all Age Classes - Region 2

On the other hand, the comparison of the average claim number among all age classes in region 1 and 2 are shown in Figures 6.22 and 6.23 respectively. The black triangles represent the average prediction value for 21-time unit in each age classes while the points up to the unit of time 20 are the historical data of the average claim number. In both regions the growth tendency is similar. The predicted values for age class 1 and 7 are approximately one another, while the average values for the number of claims in the other age classes are significantly lower. This makes perfect sense since the number of claims of minors and adults, which are in age classes 1 and 7, is greater than that of people with intermediate age. In the same way, this is evidenced in real life, therefore the results obtained make a lot of sense.

The comparison of predicted claim number by region 1 and 2 in each age class are shown in Figures 6.24, 6.25, 6.26, 6.27, 6.28, 6.29 and 6.30. It can be seen that the predicted values for each age class do not differ much between the two regions.

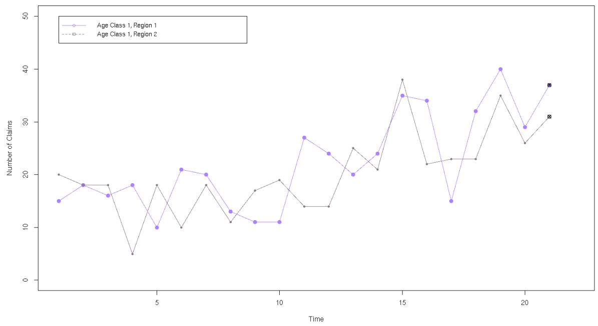

Figure 6.24: Claim Number with Prediction for Region 1 and 2 - Age Group 1

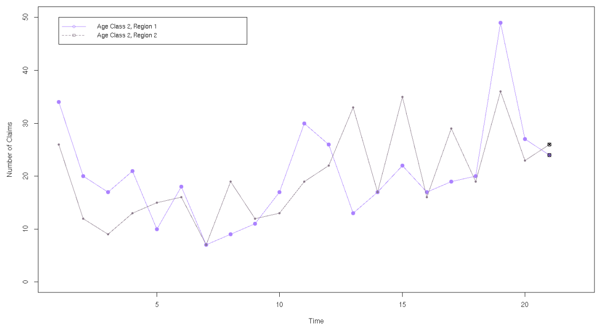

Figure 6.25: Claim Number with Prediction for Region 1 and 2 - Age Group 2

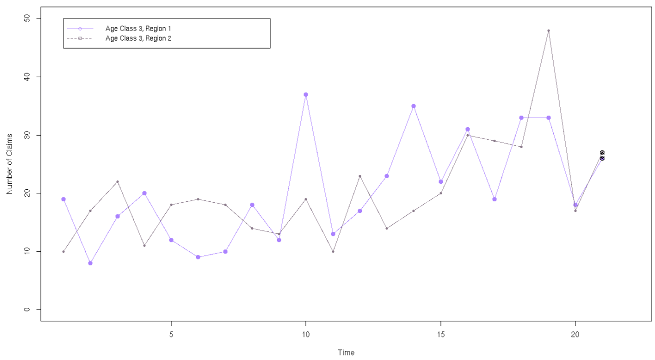

Figure 6.26: Claim Number with Prediction for Region 1 and 2 - Age Group 3

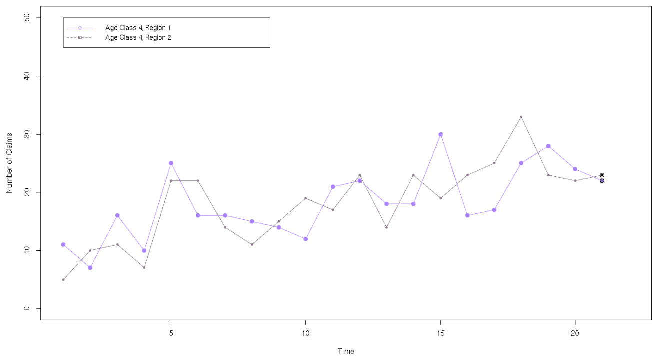

Figure 6.27: Claim Number with Prediction for Region 1 and 2 - Age Group 4

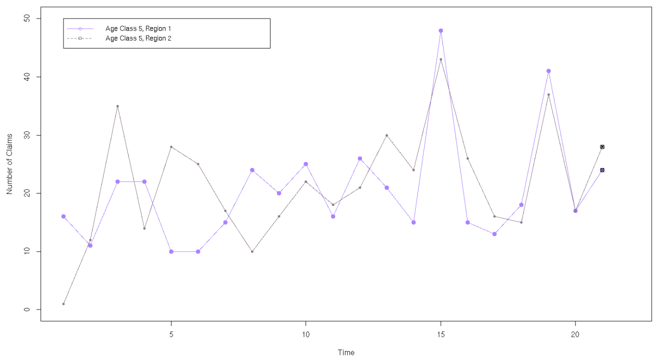

Figure 6.28: Claim Number with Prediction for Region 1 and 2 - Age Group 5

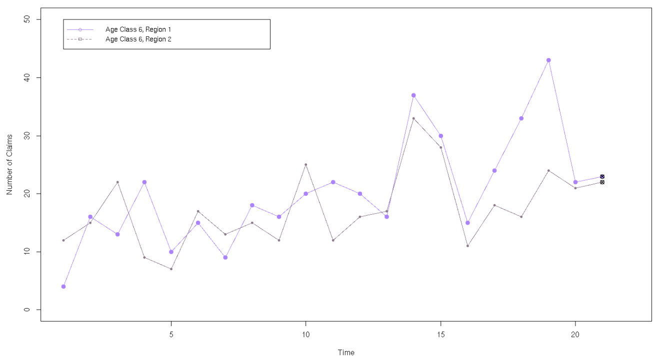

Figure 6.29: Claim Number with Prediction for Region 1 and 2 - Age Group 6

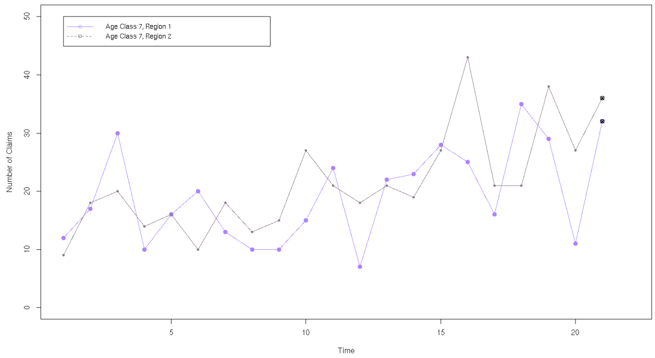

Figure 6.30: Claim Number with Prediction for Region 1 and 2 - Age Group 7

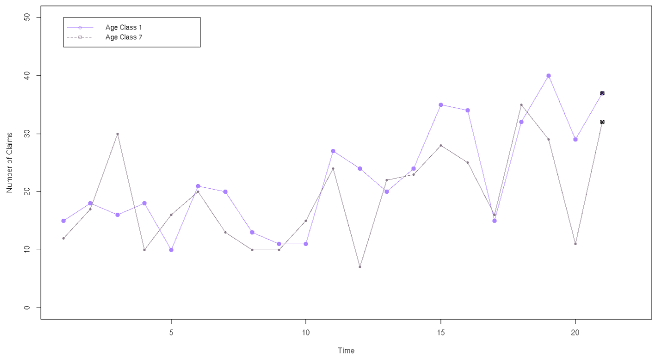

Figure 6.31: Claim Number with Prediction for Region 1 - Age Groups 1 and 7

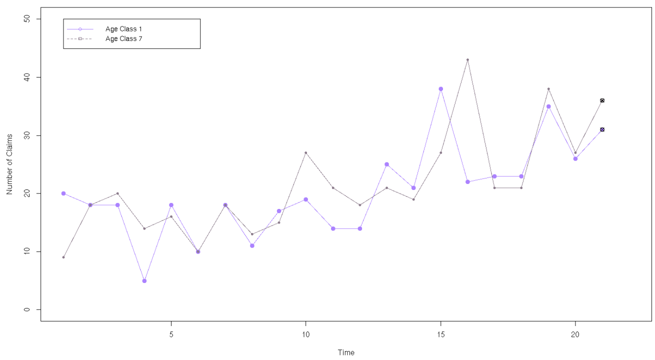

Figure 6.32: Claim Number with Prediction for Region 2 - Age Groups 1 and 7

To more easily observe the discrepancies in the number of claims between the age groups, specific groups of relevance are compared, such as age classes 1 and 7 since they have the highest number of claims in both regions. First, the average of claims of groups 1 and 7 for regions 1 and 2 are compared in Figures 6.31 and 6.32 respectively. It should be taken into account that the number of claims for the age group 1 greatly increases from the time unit 18 to 19 and decreases at the moment 20. Therefore, for a unit of time 21, the average number of claims is expected to be in this range. The predictions in both regions are following the trend of historical values and both prediction values are similar.

In Figures 6.33, 6.34, 6.35 and 6.36 both age groups 3 and 4 present the average predicted values consistent with the existing trend of increase. For all age groups, the claims numbers continue the general pattern increasing with minor corrections. Similarly, the obvious difference between the age classes ((3 and 7) and (1 and 4)) can be observed. Therefore, the expected average number of claims satisfies the knowledge gathered about the frequency of claims since the predicted average values are consistent with the analysis of the values of the λ's parameters.

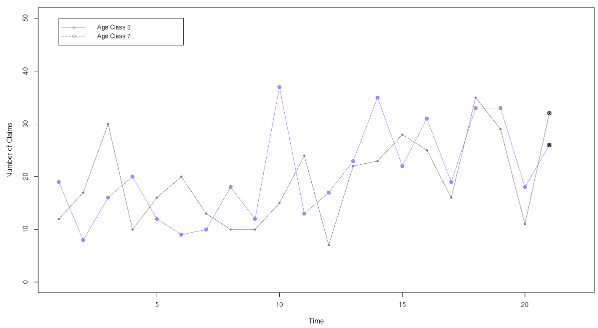

Figure 6.33: Claim Number with Prediction for Region 1 - Age Groups 3 and 7

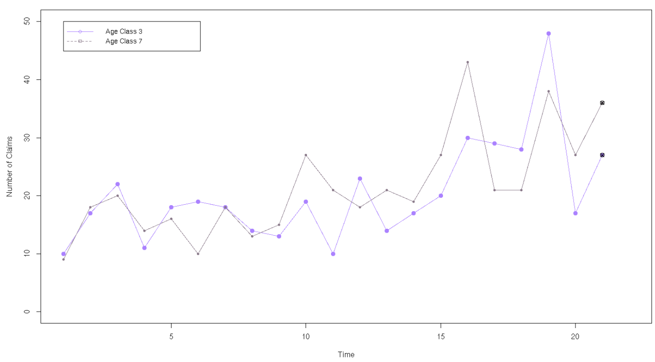

Figure 6.34: Claim Number with Prediction for Region 2 - Age Groups 3 and 7

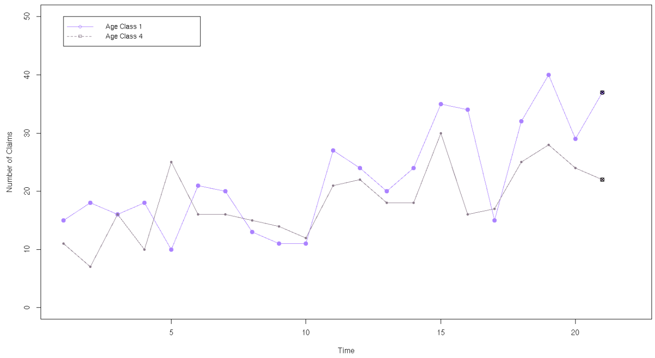

Figure 6.35: Claim Number with Prediction for Region 1 - Age Groups 1 and 4

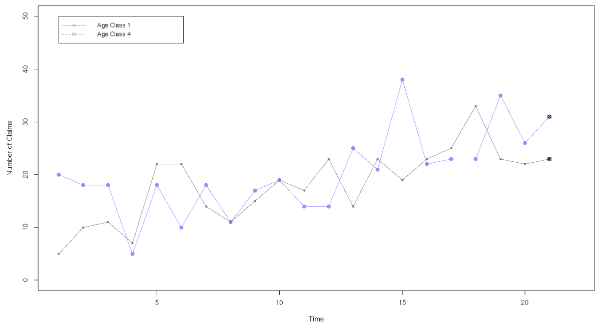

Figure 6.36: Claim Number with Prediction for Region 2 - Age Groups 1 and 4

From the insurers' point of view, total claims amounts may be of greater interest. Finally, the amounts of the claims, without considering the insurance deductible, indicate the real responsibility of the claimants due to it directly affects how the insurers determine the reserves for a corresponding period of time. Figures 6.37 and 6.38 show the result obtained from the predicted claim amounts for all age classes in region 1. On the other hand, Figures 6.39 and 6.40 show the predicted claim amounts for all age classes in region 2. The traces in all the figures mentioned above provide good evidence that the forecast reaches a state of convergence. Due to the wide range of values for a claim amount, the program automatically chooses to present continuous density functions with high bandwidth. It means that the smooth effect imposed on the density curves is strong.

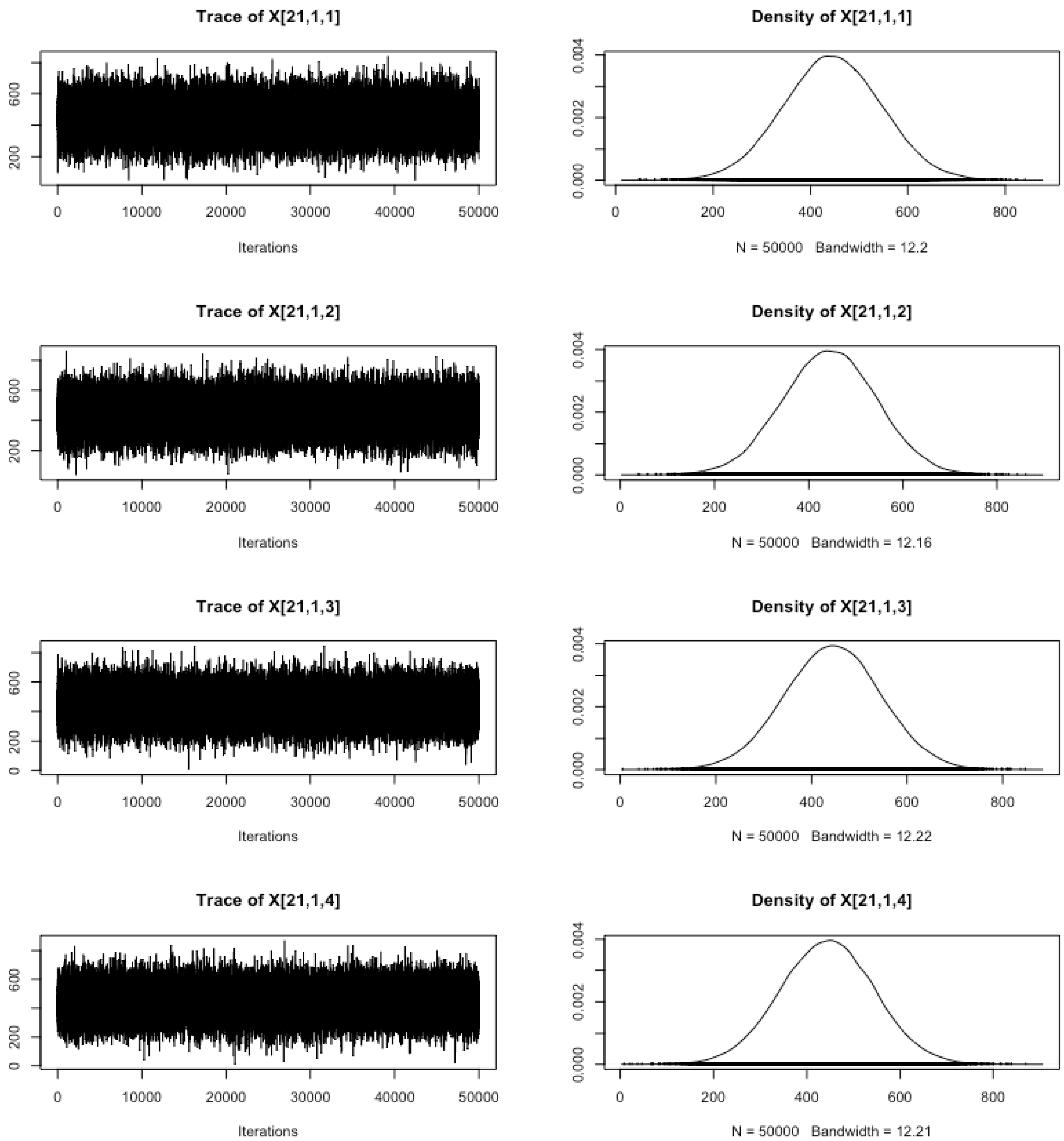

Figure 6.37: Prediction of Total Claim Amount for $21st$ Time Unit - Region 1 - Age classes $1, 2, 3$ and 4

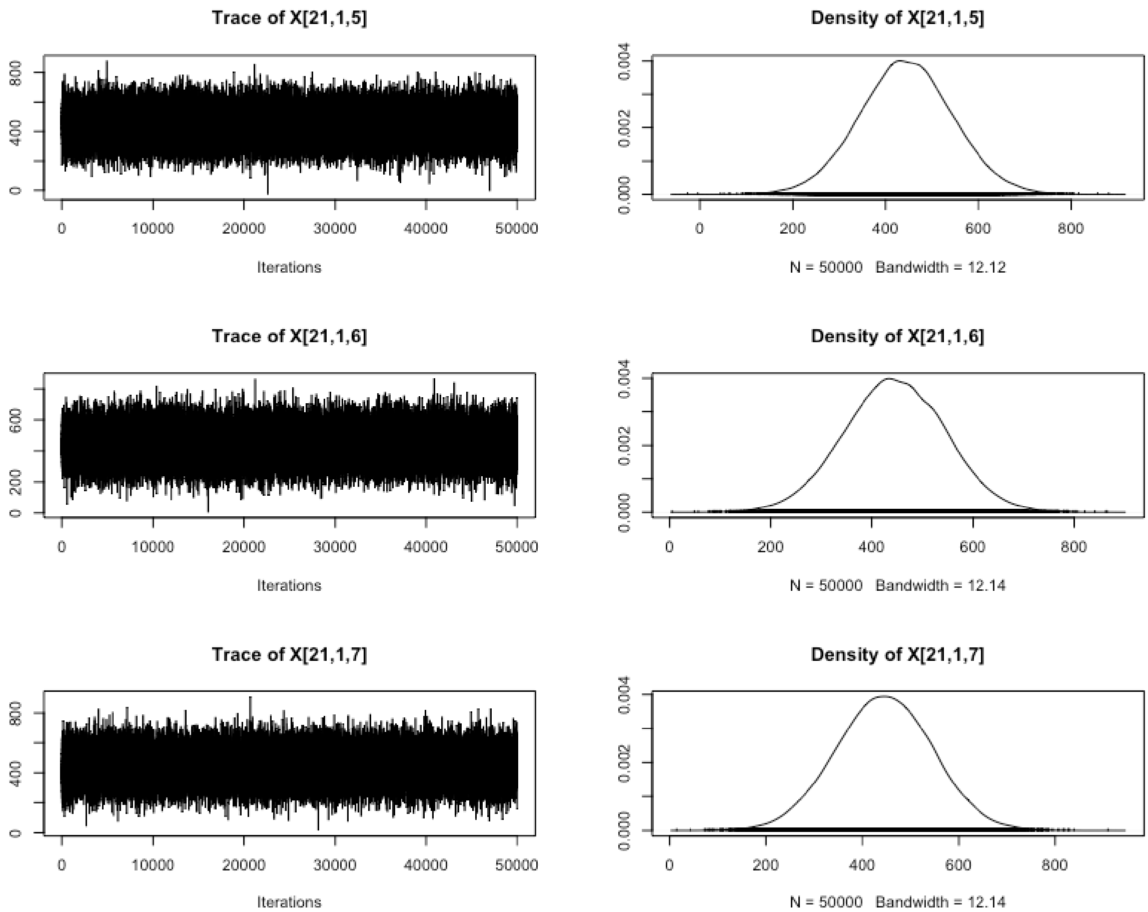

Figure 6.38: Prediction of Total Claim Amount for $21st$ Time Unit - Region 1 - Age classes $5, 6$ and 7

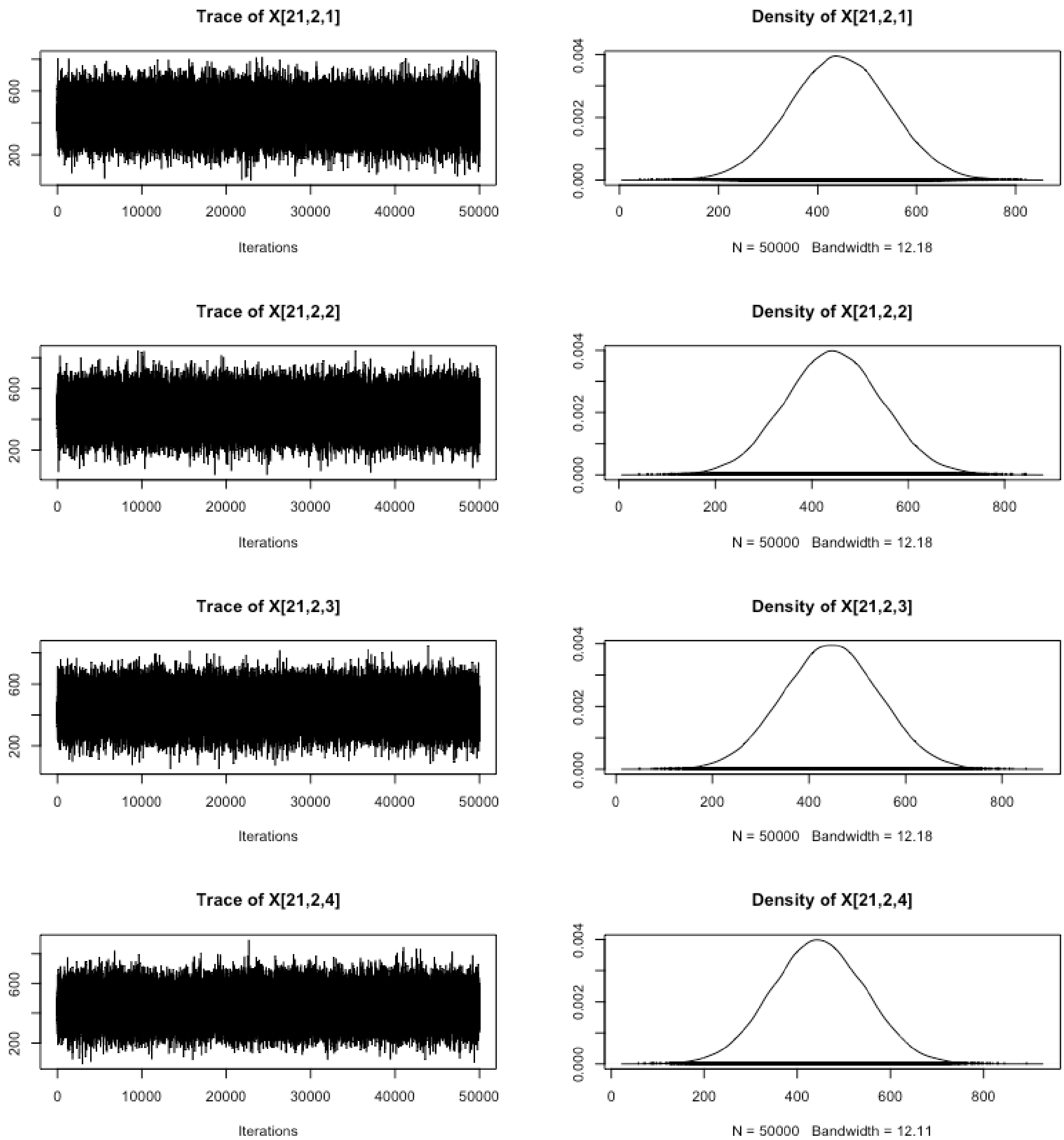

Figure 6.39: Prediction of Total Claim Amount for $21st$ Time Unit - Region 2 - Age classes $1, 2, 3$ and 4

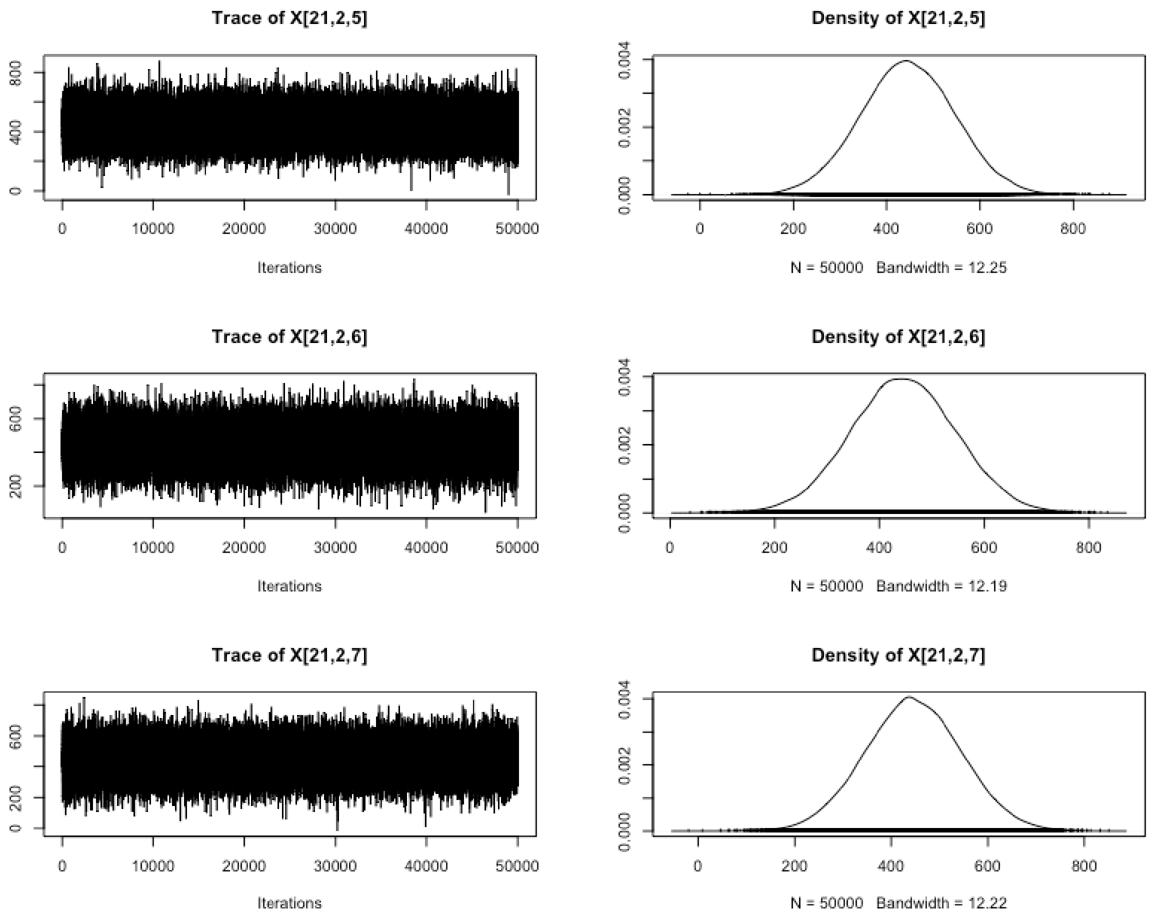

Figure 6.40: Prediction of Total Claim Amount for $21st$ Time Unit - Region 2 - Age classes $5, 6$ and 7

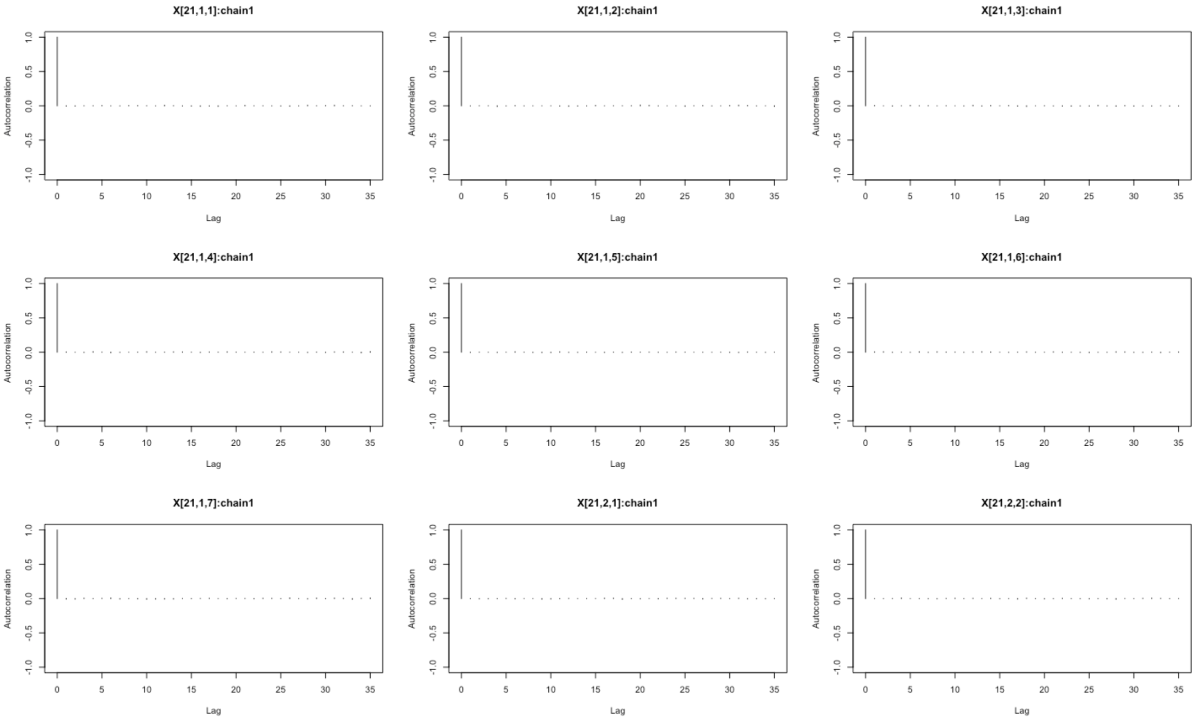

Figure 6.41: Autocorrelation of Prediction of Total Claim Amount for 21st Time Unit - Region 1 and 2- All Age classes

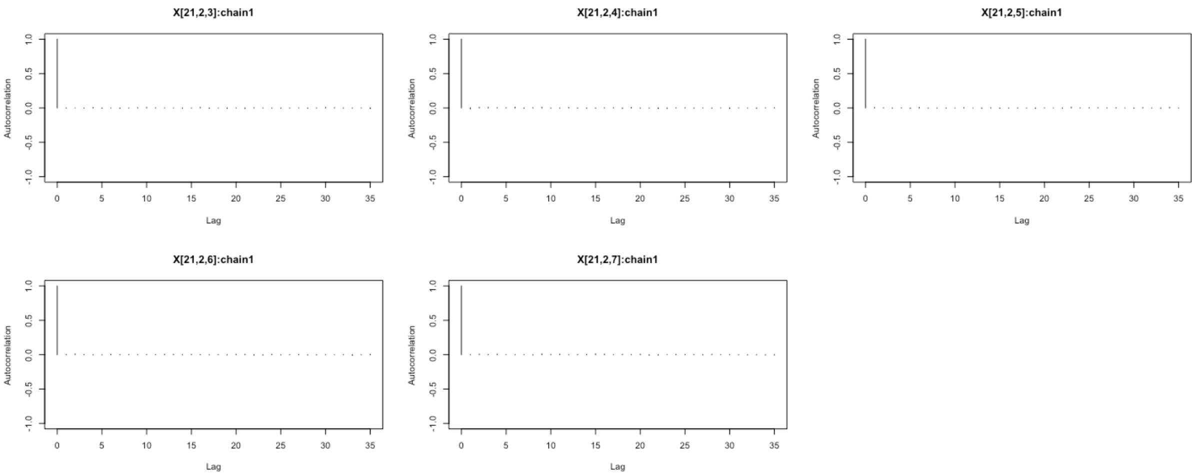

Figure 6.42: Autocorrelation of Prediction of Total Claim Amount for 21st Time Unit - Region 2- All Age classes

Like the previous cases, the predicted values for each age class in both regions form an independent sample as indicated by each of the autocorrelation functions in Figures 6.41 and 6.42.

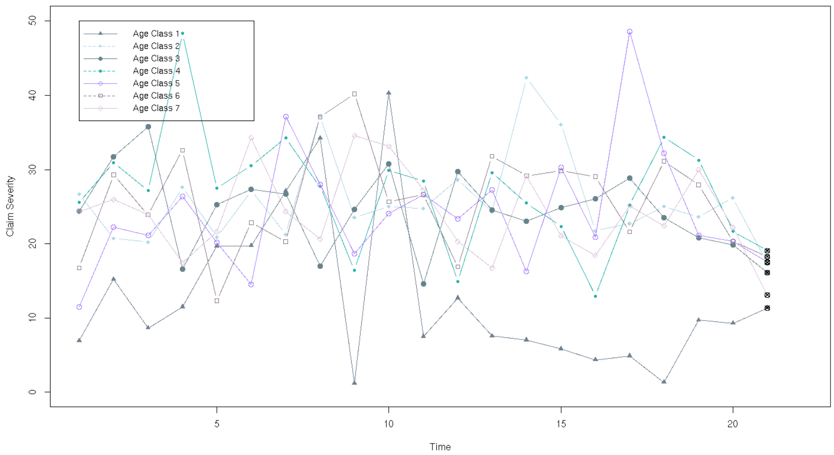

Figure 6.43: Average Amount Per Claim with Prediction in Region 1 - All Age Groups

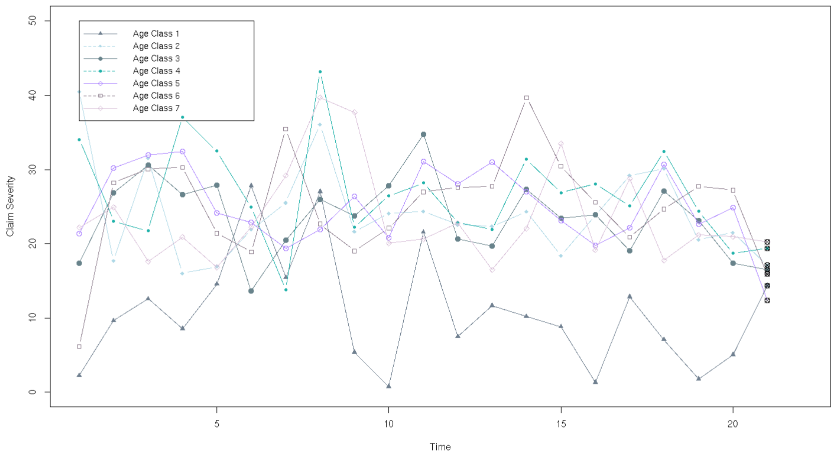

Figure 6.44: Average Amount Per Claim with Prediction in Region 2 - All Age Groups

Figures 6.43 and 6.44 present the severity of the average claim, which is calculated by dividing the total predicted amounts for the corresponding number of claims for each unit of time. The prediction calculated values are given for the unit of time 21 by all age groups in region 1 and 2 respectively. As can be seen in these figures, for region 1 values contain a great similarity to region 2.

From Figure 6.45 until Figure 6.51 the average severity of claim for age class 1 through 7 respectively is shown. The average severity of claim for group 1 varies substantially for the first 11 units of time and then stabilizes thereafter. The expected average severity of claim for age group 1 reaches approximately \$15. Age groups 2, 3 and 4 have a similar pattern in the average severity of the claim with predicted values between \$15 to \$20. In general, the predictions They are consistent with the assumption of \$20 per claim. Throughout time units, some age groups show a slight increase (or decrease) in the severity of the average claim, and the variations are within reasonable ranges.

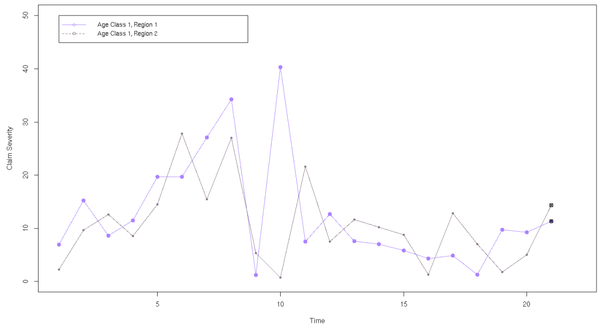

Figure 6.45: Claim Number with Prediction for Region 1 and 2 - Age Group 1

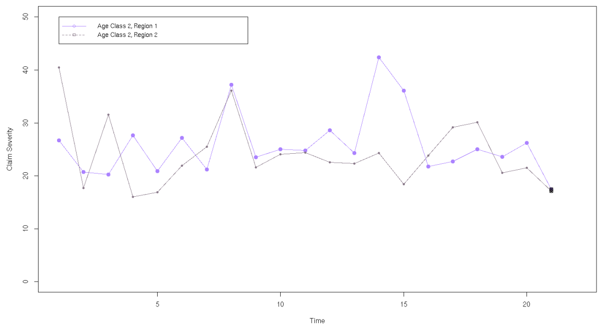

Figure 6.46: Average Amount Per Claim with Prediction for Region 1 and 2 - Age Group 2

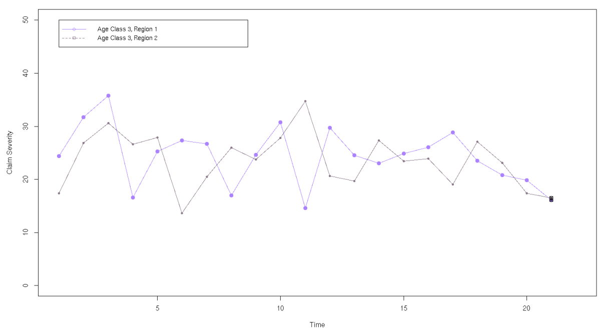

Figure 6.47: Average Amount Per Claim with Prediction for Region 1 and 2 - Age Group 3

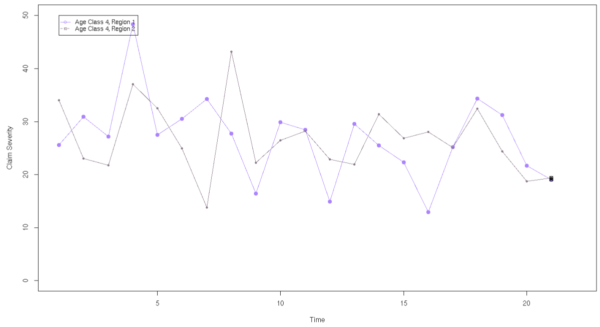

Figure 6.48: Average Amount Per Claim with Prediction for Region 1 and 2 - Age Group 4

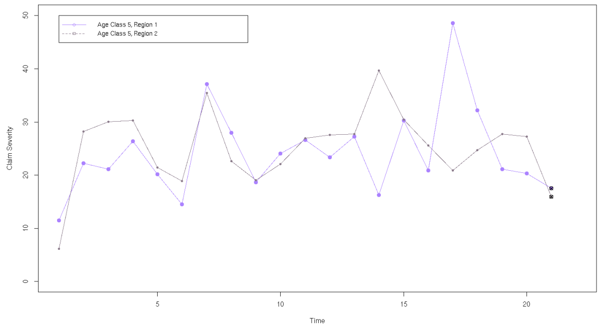

Figure 6.49: Average Amount Per Claim with Prediction for Region 1 and 2 - Age Group 5

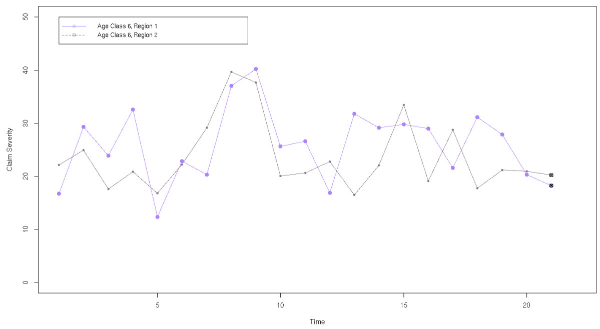

Figure 6.50: Average Amount Per Claim with Prediction for Region 1 and 2 - Age Group 6

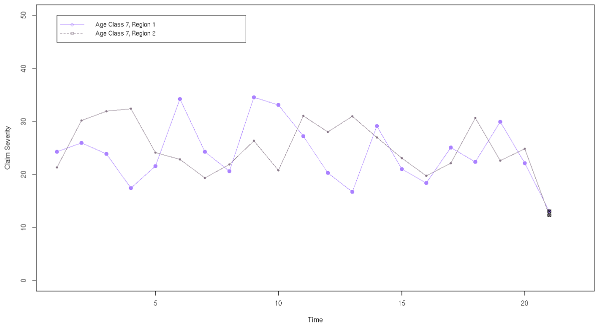

Figure 6.51: Average Amount Per Claim with Prediction for Region 1 and 2 - Age Group 7

$Data_{t,i,a}$	**Mean**	**SD**	**2.5% Perc.**	**Median**	**97.5% Perc.**	**Stationarity**	**Halfwidth Test**	**P-value**
$M_{21,1,1}$	144.08	200.29	247.76	140.73	553.9	Passed	Passed	0.982
$M_{21,1,2}$	144.09	200.29	248.57	140.81	553	Passed	Passed	0.977
$M_{21,1,3}$	144.08	200.3	248.55	140.74	553.95	Passed	Passed	0.979
$M_{21,1,4}$	144.05	200.3	248.65	140.8	553.12	Passed	Passed	0.985
$M_{21,1,5}$	144.08	200.26	248.37	140.78	552.44	Passed	Passed	0.963
$M_{21,1,6}$	144.08	200.29	248.53	140.87	553.71	Passed	Passed	0.985
$M_{21,1,7}$	144.03	200.3	249.13	140.82	552.25	Passed	Passed	0.968
$M_{21,2,1}$	150.83	200.81	233.74	148	554.78	Passed	Passed	0.973
$M_{21,2,2}$	150.84	200.81	235.21	147.97	555.08	Passed	Passed	0.982
$M_{21,2,3}$	150.84	200.8	234.87	147.79	554.81	Passed	Passed	0.962
$M_{21,2,4}$	150.81	200.77	234.49	147.84	555.12	Passed	Passed	0.982
$M_{21,2,5}$	150.83	200.78	234.65	147.88	555.55	Passed	Passed	0.976
$M_{21,2,6}$	150.84	200.79	234.89	147.95	555.17	Passed	Passed	0.982
$M_{21,2,7}$	150.79	200.78	234.26	147.7	554.79	Passed	Passed	0.978
$N_{21,1,1}$	36.9	11.54	34.94	36.90	38.88	Passed	Passed	0.801
$N_{21,1,2}$	23.6	11.44	21.65	23.60	25.56	Passed	Passed	0.396
$N_{21,1,3}$	24.37	11.52	22.40	24.37	26.35	Passed	Passed	0.564
$N_{21,1,4}$	21.88	11.47	19.91	21.87	23.84	Passed	Passed	0.699
$N_{21,1,5}$	24.92	11.51	22.95	24.37	26.88	Passed	Passed	0.364
$N_{21,1,6}$	23.83	11.46	21.87	21.87	25.79	Passed	Passed	0.965
$N_{21,1,7}$	33.85	11.49	31.88	24.92	35.81	Passed	Passed	0.779
$N_{21,2,1}$	32.4	11.63	30.44	23.83	34.37	Passed	Passed	0.326
$N_{21,2,2}$	24.87	11.52	22.90	33.86	26.84	Passed	Passed	0.916
$N_{21,2,3}$	25.46	11.6	23.49	32.40	27.42	Passed	Passed	0.842
$N_{21,2,4}$	22.89	11.55	20.96	24.87	24.87	Passed	Passed	0.991
$N_{21,2,5}$	27.97	11.59	26.01	25.46	29.93	Passed	Passed	0.831
$N_{21,2,6}$	24.62	11.56	22.65	22.89	26.58	Passed	Passed	0.788
$N_{21,2,7}$	36.85	11.5	34.89	36.85	38.80	Passed	Passed	0.338

Table 6.1 continued from previous page

$Data_{t,i,a}$	**Mean**	**SD**	**2.5% Perc.**	**Median**	**97.5% Perc.**	**Stationarity**	**Halfwidth Test**	**P-value**
$X_{21,1,1}$	563.79	99.95	367.22	563.89	664.97	Passed	Passed	0.722
$X_{21,1,2}$	470.43	100.34	276.83	470.57	672.38	Passed	Passed	0.3561
$X_{21,1,3}$	475.06	99.97	280.68	474.69	795.70	Passed	Passed	0.4817
$X_{21,1,4}$	600.68	99.45	404.24	601.57	722.96	Passed	Passed	0.2725
$X_{21,1,5}$	525.81	100.28	328.83	525.56	616.75	Passed	Passed	0.743
$X_{21,1,6}$	468.61	100.01	273.10	468.76	666.00	Passed	Passed	0.3898
$X_{21,1,7}$	556.70	100.34	360.97	557.63	750.89	Passed	Passed	0.4611
$X_{21,2,1}$	516.47	100.04	318.60	516.82	710.95	Passed	Passed	0.9018
$X_{21,2,2}$	556.04	100	360.69	555.88	751.75	Passed	Passed	0.1335
$X_{21,2,3}$	583.00	100.01	387.59	583.12	777.85	Passed	Passed	0.5545
$X_{21,2,4}$	507.49	99.44	313.72	506.62	704.54	Passed	Passed	0.0749
$X_{21,2,5}$	592.31	100.61	393.49	592.33	789.72	Passed	Passed	0.3298
$X_{21,2,6}$	515.40	100.1	319.69	514.89	712.73	Passed	Passed	0.0581
$X_{21,2,7}$	575.63	100.33	377.68	573.23	770.17	Passed	Passed	0.0913

Table 6.1: Predicted Insured Population, Claim Number and Total Claim Amount for 21*st* Time Unit

The statistical summary is presented in Table 6.1. The first column represents the type of data with the subscripts in the order of time, region and age class. The other columns list the mean, standard deviation, the quantile of 2.5%, the median, the quantile of 97.5%, the stationarity test, the Halfwidth Test and the p-value of the predicted distribution. Plotting the predicted values together with the historical insured population already give us a better idea of whether the predictions were legitimate. Note that the number of claims N across all age groups show lower standard deviations compared to the predicted total claim amounts X and the predicted insured population M. It can be noted also that the number of claims of age classes 1 and 7 in both regions is higher compared to the rest of the classes. Finally, all the data parameters passed the test of stationarity and convergence.

6.2 Premium Determination under Various Premium Principles

This section present, based on the available information, several methods for calculating premiums. Each method has its own characteristics and advantages. Some approaches only require the mean and variance of the predicted variables, while others require more details of the predictive distribution, such as percentiles. Some methods are more conservative, with a high premium scheme designed to meet extreme demands, while others are moderate, which makes the products competitive in the market. Insurers are free to choose the one that corresponds to their level of risk tolerance. For more information you can consult [22], [10], [23], [9] and [24], 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 $X|D_T$. We use X instead to represent the loss random variable in order to simplify the notation for premium principle illustrations.

6.2.1 Net Premium Principle

Net premium is the expected present value of a policy's benefits less the expected present value of future premiums. The net premium calculation does not take into account future expenses associated with maintaining the policy [25]. The net premium principle is one of the commonly applied principles in the literature.

The premium would just to cover the claims only due to the risk which is eventually eliminated after selling a great many identical and independently distributed policies. This is the fundamental theory under this principle. Thus, the net premium principle is defined as

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

The net premium in this project for each age group in both regions is simply the predicted average of the total amount of the claim in unit of time 21 which is found in Table 6.1. This principle requires the least amount of information from the planned subsequent distribution with a practical calculation process. In reality, it is almost impossible to sell infinite independent and identical policies. By not having to carry risks, premiums are exposed to extreme events and fluctuations such as very large claims amounts. Therefore, it is not recommended to apply the principle of net premium in practice, but treat it as an estimated measure.

6.2.2 Expected Value Premium Principle

The value premium has become arguably as important as the equity premium in asset allocation, investment management, capital budgeting, security analysis, and many other applications. Most studies use average realized returns as the proxy for expected returns. But average returns are noisy and do not necessarily converge to expected returns in finite samples [26]. The expected value premium principle is often regarded as the extension of the net premium principle, and it is expressed as

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

where ξ is the loading factor. Note that if $\xi = 0$ then it is the same as the net premium principle. Clearly, the premium under this principle is greater than the expected loss. The difference between the expected loss and the premium can be referred to as the premium burden that provides protection against unexpected losses. According to the literature if the charge is not applied, the ruin would eventually occur with certainty. Therefore, if ξ has a great value this produces a large margin of protection.

6.2.3 Variance Premium Principle

The variance premium principle is another extension 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. Note that if the more variability the loss then the premium will be higher. The variance premium principle is proportional to the variance of the loss. It can be expressed as

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

Again, note that if $\omega = 0$ then the variance premium principle is the same as the net premium principle. The insurers have the freedom to determine the risk load based on their risk tolerance, like the expected value premium principle. 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.

6.2.4 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. This principle is expressed as

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

It takes the variability of the loss into the premium determination.
Summarising the different principles mentioned above, 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. More information can be found in [27] and [28].

6.2.5 Value-at-Risk

Value at risk (VaR) is a statistic that measures and quantifies the level of financial risk within a firm, portfolio or position over a specific time frame. This metric is most commonly used by investment and commercial banks to determine the extent and occurrence ratio of potential losses in their institutional portfolios [29]. VaR was popularized during the last ten or fifteen years, presenting applications on stocks, bonds, interest and exchange rate forward contracts, and swaps. Confidence level α is required in order to calculate VaR. Assume that the loss random variable X has the cumulative function $F_X(z) = P(X \leq z)$ then the VaR with confidence level α is defined as

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

i.e., VaR_α is a lowe $\alpha-$percentile of the random variable X. For example, α can take values of 90%, 95% and 98%. The VaR is essentially measuring the percentile of the loss distribution function, providing a minimum value of the loss based on the confidence level.

6.2.6 Tail Value-at-Risk

The Tail Value-at-Risk, TVaR, of the loss random variable $TVaR_\alpha$ is defined as the expected outcome (loss), conditional on the loss exceeding the Value-at-Risk (VaR), of the distribution [30]. Where the support of the distribution is continuous the VaR with confidence level α is usually defined as

$$P(X \leq -VaR_\alpha) = 1 - \alpha$$

Then the corresponding Tail Value-at-Risk would be defined as

$$TVaR_\alpha = -\frac{1}{1-\alpha}\int_{-\infty}^{-VaR_\alpha} xf(x)dx$$

In this project to calculate the TVaR it is used the following expression

$$T\hat{Va}R_\alpha = \frac{1}{N(1-\alpha)}\sum_{j=N\alpha+1}^{N} X_{(j)},$$

where N is the total sample size, $X_{(j)}$ is the j^{th} smallest value (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. For further information you can read Hardy (2006), Sarykalin et al. (2008) and Peng (2009).

6.2.7 Numerical Premium Analysis

It is not complicated to determine the premium according to the previously mentioned principles since the predicted distributions on the total claim amounts are available. In Tables 6.2 and 6.3, the premium for 21 unit time in regions 1 and 2 can be observed, respectively, using the four premium principles and the VaR and $TVaR$ risk measures. The total premium in $ according to the net premium principle is the average amount of total expected claim, as in the second column of Table 1. On the other hand, for the expected premium principle, the variance premium principle and the standard deviation premium principle, risk weights $\xi = 0.2$, $\omega = 0.012$ and $\nu = 1.3$ are assumed respectively. High loads in risk weights provide stronger protection against the uncertainty of the claim. Due to the high value of the variance, the premium principle under variance is very sensitive to the value of ω and, therefore, it should take extra precautions when applying this premium principle.

Since the total predicted distributions claims have approximately the form of a normal distribution the risk charge $\nu = 1.3$, in the standard deviation premium principle, can almost guarantee that the company will pay all claims with a probability of 97.5%. In Figures 6.37, 6.38, 6.39 and 6.40, the predicted distributions are slightly correct (or positively) biased with a fat right tail. That means that the premium with charge risk $\nu = 1.3$ can cover the total claims with a little less than the 97.5% probability. In fact, this is the case, since according to the VaR measure with a confidence level of 97.5%, the premium is slightly higher than the standard deviation premium principle for all age classes. Conditioning the premium greater than $VaR_{97.5\%}$, $TVaR_{97.5\%}$ grants even higher premiums.

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. Tables 6.4 and 6.5 show the premium per policyholder for the next time unit in regions 1 and 2 respectively. Having the total premium and total insured population available, the premium per policyholder can be obtained by averaging premium over population with ease. A crude estimation of the premium per policyholder is to use the premium under the net premium principle over the mean predicted population for each age class. The premium per policyholder is higher for age classes 1, 4 and 7 but generally between $3 and 6$ which is consistent with the assumptions in Chapter 5. 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%\$25 = \$5. This premium determination does not involve any weights and hence can only cover the average cost per policyholder.

Total Premium ($)	**Weight**	**Age Class**						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Net Premium Principle	-	563.79	470.43	475.06	600.68	525.81	468.61	556.7
Expected Premium Principle	0.2	676.55	564.52	570.07	720.82	630.98	562.33	668.04
Variance Premium Principle	0.012	683.66	589.89	595.03	720.79	646.78	589.17	675.64
Standard Deviation Premium Principle	1.3	693.72	600.14	605.04	730.74	656.33	598.91	686.13
$VaR_{97.5\%}$	-	759.55	664.97	672.38	795.7	722.96	666	750.89
$TVaR_{97.5\%}$	-	801.56	785.34	796.91	864.34	823.14	752.63	825.98

Table 6.2: Total Premiums for 21^{st} Time Unit in Region 1 Using Predicted Results

Total Premium ($)	**Weight**	**Age Class**						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Net Premium Principle	-	516.47	556.04	583	507.49	592.31	515.4	573.63
Expected Premium Principle	0.2	619.77	667.24	699.61	608.98	710.77	618.47	688.35
Variance Premium Principle	0.012	637.18	676.08	702.43	627.5	713.28	636.71	693.88
Standard Deviation Premium Principle	1.3	646.86	686.06	712.69	637.49	722.83	646.1	703.76
$VaR_{97.5\%}$	-	710.95	751.75	777.85	704.54	789.72	712.73	770.17
$TVaR_{97.5\%}$	-	796.81	825.62	871.4	803.24	882.97	812.74	847.51

Table 6.3: Total Premiums for 21^{st} Time Unit in Region 2 Using Predicted Results

Premium per person ($)		Age Class						
Premium Measure	**Predicted Insured Population Measure**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Net Premium Principle	Average Insured Population	3.73	3.12	3.15	3.98	3.48	3.1	3.69
Expected Premium Principle	Average Insured Population	4.48	3.74	3.78	4.77	4.18	3.72	4.42
Variance Premium Principle	Average Insured Population	4.53	3.91	3.94	4.77	4.28	3.9	4.47
Standard Deviation Premium Principle	Average Insured Population	4.6	3.98	4.01	4.84	4.35	3.97	4.54
$VaR_{97.5\%}$	Average Insured Population	5.03	4.4	4.45	5.27	4.79	4.41	4.97
$TVaR_{97.5\%}$	Average Insured Population	5.31	5.2	5.28	5.72	5.45	4.98	5.47

Table 6.4: Premium Per Policyholder for 21^{st} 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	3.43	3.69	3.87	3.37	3.94	3.42	3.81
Expected Premium Principle	Average Insured Population	4.12	4.43	4.65	4.05	4.72	4.11	4.57
Variance Premium Principle	Average Insured Population	4.23	4.49	4.67	4.17	4.74	4.23	4.61
Standard Deviation Premium Principle	Average Insured Population	4.3	4.56	4.73	4.23	4.8	4.29	4.68
$VaR_{97.5\%}$	Average Insured Population	4.72	5	5.17	4.68	5.25	4.73	5.12
$TVaR_{97.5\%}$	Average Insured Population	5.29	5.49	5.79	5.34	5.87	5.4	5.63

Table 6.5: Premium Per Policyholder for 21^{st} Time Unit in Region 2

Chapter 7

Conclusions and Future work

In general term, this project has as its main objective the prediction of the total amount of claims under a Bayesian hierarchical framework. The prediction of future claims plays an important role in the measurement of risk for health insurance providers therefore it is a primary issue. Migon and Moura (2005) ensure that the total claims are related to the number of claims and the insured population during a certain period of time. Different patterns are presented in the frequency and severity of the claim reported in different age classes of the insured. For this reason, it is justifiable to classify the insured by age class by predicting the status of the claim in each unit of time. The proposed model introduces one more category, the spatial factor, to describe the regions of the insured's residence. This factor may represent the combined random effect of the elements that influence the behavior of the claim, such as the ability to access the medical service, the level of wealth, education and even climatic conditions. These elements can influence the behavior of the claim. Therefore, it is accessible and understandable the introduction of a spatial factor independent of the existing age classification.

To achieve the aim, first, the model proposed by Migon and Moura (2005) was modified by adding the spatial factor, arguing that the spatial factor, the age class and the measurement time affect the average insured population. For subsequent estimates, given prior knowledge of the parameters and historical information, an MCMC algorithm was used. Also, the Gibbs and Metropolis-Hastings sampling algorithm were used since the complete conditional distributions of the parameters are not in closed forms. A simulation study was carried out on the insured population, the number of claims and the total amount of claims, for 20-time units, 7 age groups and 2 regions to test if the model can effectively detect the true value of the parameters, with some of the default parameter values while others randomly simulated using R and relevant packages. Due to the previously imprecise propose, the subsequent distribution suggests at 50,000 iterations, the traces show signs of convergence. Predictions are made for the unit of time 21, which includes the insured population, the number of claims and the amounts of the claims after generating the values of true parameters. Finally, premiums can be calculated under various premium principles, depending on the expected claim amounts. The premium to be charged per policyholder can also be easily obtained by dividing the total premiums over the predicted insured population.

As future work, in order to achieve the purpose of investigating the effectiveness of the predictions, the comparison of the simulated data with real data will be made, tested by an insurance company in the country.

References

[1] H. Migon and E. Penna, "Bayesian analysis of a health insurance model," *Journal of Actuarial Practice*, vol. 13, 2006.

[2] M. F. Souza, D. and H. Migon, *Small area population prediction via hierarchical models.* Statistics Canada, vol. 39, ch. In Survey Methodology, Catalogue No. 12-001-X, pp. 203–214.

[3] S. Pai, "Bayesian analysis of compound loss distributions," *Journal of Econometrics*, vol. 79, no. 1, 1997.

[4] M. A. Rosenberg, "A statistical method for monitoring a change in the rate of nonacceptable inpatient claims," *North American Actuarial Journal*, vol. 5, no. 4, 2001.

[5] H. Migon and D. Gamerman, *Statistical Inference: An Integrated Approach.* London, United Kingdom: Arnold, 1999.

[6] S. Gschlobl and C. Czado, "Spatial modeling of claim frequency and claim size in non-life insurance," *Scandinavian Actuarial Journal*, pp. 202–225.

[7] X. Dimakos and A. Frigessi, "Bayesian premium rating with latent structure," *Scandinavian Actuarial Journal*, vol. 3, pp. 162–184, 2002.

[8] A. N. Pettitt, I. S. Weir, and A. G. Hart, "A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data," no. 1998, pp. 353–367, 2002.

[9] V. Young, "Premium Principles," *Encyclopedia of Actuarial Science.*, 2004.

[10] M. R. Hardy, "by An Introduction to Risk Measures for Actuarial Applications," *SOA Syllabus Study Note*, 2006.

[11] S. G. Sarykalin, S. and S. Uryasev, *Value-at-risk vs. conditional value-at-risk in risk management and optimization.*, ch. Tutorials in Operations Research, pp. 270–294.

[12] J. Yohai, "Notas de Probabilidades y Estadística," 2008.

[13] S. A. Klugman, "Bayesian statistics in actuarial science: with emphasis on credibility," *Springer Science and Business Media*, vol. 15, 1992.

[14] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* CRC Press, 1997.

[15] H. S. Migon and F. A. S. Moura, “Hierarchical Bayesian collective risk model : an application to health insurance,” vol. 36, pp. 119–135, 2005.

[16] S. Gschl and C. Czado, “Spatial modelling of claim frequency and claim size in non-life insurance,” pp. 1–26, 2007.

[17] J. Correa and C. Barrera, *Introduccion a la Estadistica Bayesiana*, 21st ed. Medellin: Instituto Tecnologico Metropolitano -ITM-, 2018.

[18] R. S. . S. D. J. Gilks, W. R., *Markov Chain Monte Carlo in Practice. na.* London: Chapman and Hall-, 1996.

[19] M. Stephens. (2018) The metropolis hastings algorithm. [Online]. Available: https://stephens999.github.io/fiveMinuteStats/MH_intro.html

[20] G. Q. Yu, “Hierarchical bayesian modelinf of health insurance claims,” 2011.

[21] M. Rosenblatt, “Remarks on some nonparametric estimates of a density function.” *The Annals of Mathematical Statistics*, vol. 3, 1956.

[22] J. Goovaerts, Marc and J. Dhaene, “On the Characterization of Wang’s Class of Premium Principles,” pp. 2–12, 1997.

[23] R. J. Laeven and M. J. Goovaerts, “Premium calculation and insurance pricing,” *Encyclopedia of Quantitative Risk Analysis and Assessment*, vol. 3, 2008.

[24] K. C. Embrechts, P. and T. Mikosch, “Modelling Extremal Events,” *Springer Science Business Media*, vol. 33, 1997.

[25] J. Kagan. (2018) Net premium. [Online]. Available: https://www.investopedia.com/terms/n/net-premium.asp

[26] L. Chen, R. Petkova, and L. Zhang, “The expected value premium $,” vol. 87, pp. 269–280, 2008.

[27] H. U. Gerber, *An introduction to mathematical risk theory.* Philadelphia: SS Huebner Foundation for Insurance Education, Wharton School, University of Pennsylvania., 1979.

[28] H. Bühlmann, *Métodos matemáticos en la teoría del riesgo.* Springer Science Business Media, 1970.

[29] W. Kenton. (2019) Value at risk (var). [Online]. Available: investopedia.com/terms/v/var.asp

[30] ——. (2020) Tail value-at-risk. [Online]. Available: http://www.nematrian.com/TailValueAtRisk

Appendices

Appendix A

Simulated Data

	Region 1							Region 2						
	Age Class							Age Class						
Time	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	62.9	82.12	74.8	64.54	67.98	81.17	52.43	80.41	65.23	67.74	82.36	66.55	87.24	67.49
2	86.38	77.64	72.25	63.25	82.39	79.66	72.8	93.63	71.59	89.65	74.73	67.46	75.38	81.11
3	93.08	83.33	67.7	73.42	83.92	57.61	98.02	89.75	80.87	77.63	61.65	99.11	91	63.71
4	96.2	93.24	80.77	73.78	84.81	85.64	69.72	74.08	76.43	80.87	71.52	69.14	76.11	82.77
5	69.18	76.15	63.52	93.74	68.33	79.77	85.7	83.98	95.71	69.95	88.83	87.74	56.99	65.86
6	80.97	85.42	74.22	64.25	84.56	82.46	86.12	87.27	72.48	87.04	80.3	112.46	73.52	76.75
7	81.26	71.26	61.76	73.36	72.69	80.79	68.92	65.53	74.02	67.19	81.95	80.82	82.84	79.37
8	88.09	82.61	95.73	86.68	85.59	92.09	94.02	90.94	86.43	77.8	94.61	76.67	96.27	91.14
9	91.24	87.9	80.9	79.33	85.93	72.32	77.5	81.16	84.14	83.77	82.42	89.57	76.44	89.75
10	83.73	95.31	106.49	77.99	97.14	86.88	91.41	95.56	91.83	99.34	97.07	90.29	90.31	96.06
11	132.35	96.76	101.41	103.9	96.01	91.21	96.16	98.51	95.64	102.58	105.52	94.06	83.39	108.16
12	89.82	81.8	91.13	79.64	93.93	79.78	91.34	82.73	86.38	86.41	80.66	85.85	74.03	73.63
13	103.11	100.32	110.18	98.9	101.97	106.19	85.47	119.87	102.91	110.17	107.53	113.12	116.62	101.02
14	122.53	110.91	120.09	127.79	123.4	106.69	110.77	121.39	121.2	123.37	121.13	123.11	118.94	101.3
15	172.15	166.14	149.59	155.68	137.16	184.66	162.45	154.29	185.88	115.04	154.7	161.7	174.36	166.59
16	126.77	139.16	139.01	115.16	127.71	111.28	103.38	134.61	103.86	141.56	139.05	140.83	113.29	138.88
17	116.06	126.26	123.9	115.56	113.66	100.91	112.2	137.96	116.64	132.07	114.65	105	122	109.63
18	146.55	140.76	131.26	145.68	141.74	146.64	146.79	137.01	134.32	144.73	150.83	133.49	138.07	130.45
19	177.79	155.3	136.1	136.41	145.82	173.29	150.12	158.5	154.26	181.32	166.05	143.98	160.27	177.65
20	129.18	115.93	132.77	126.24	114.87	130.14	124.17	124.84	112.98	126.57	124.13	127.9	129.29	129.91

Table A.1: $\mu_{t,i,a}$ simulated values

	Region 1							Region 2						
	Age Class							Age Class						
Time	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	85.43	114.49	99	56.78	65.22	20.64	62.08	94.53	118.49	56.57	37.8	6.61	63.35	44.78
2	86.38	77.66	72.27	63.27	82.45	79.65	72.77	93.65	71.62	89.64	74.72	67.45	75.37	81.13
3	93.06	83.31	67.73	73.45	83.9	57.63	98.06	89.68	80.89	77.61	61.67	99.11	90.96	63.65
4	96.21	93.27	80.78	73.84	84.81	85.65	69.74	74.08	76.43	80.86	71.56	69.17	76.08	82.79
5	69.21	76.17	63.51	93.79	68.33	79.77	85.73	83.95	95.71	69.96	88.84	87.76	57.01	65.9
6	81.02	85.45	74.26	64.21	84.54	82.45	86.13	87.3	72.46	87.04	80.29	112.42	73.51	76.75
7	81.25	71.29	61.75	73.38	72.68	80.78	68.96	65.54	74.02	67.24	81.93	80.83	82.84	79.37
8	88.08	82.61	95.73	86.67	85.62	92.12	94.09	90.96	86.41	77.8	94.59	76.65	96.27	91.15
9	91.24	87.92	80.94	79.36	85.93	72.27	77.51	81.17	84.12	83.73	82.43	89.55	76.45	89.78
10	83.72	95.31	106.52	77.96	97.12	86.84	91.4	95.56	91.84	99.3	97.12	90.26	90.33	96.02
11	132.39	96.81	101.39	103.89	96.01	91.2	96.13	98.51	95.64	102.57	105.53	94.06	83.46	108.16
12	89.79	81.82	91.15	79.65	93.95	79.76	91.33	82.77	86.39	86.43	80.66	85.9	74.03	73.63
13	103.13	100.29	110.22	98.89	101.92	106.21	85.45	119.85	102.93	110.12	107.53	113.09	116.65	101.04
14	122.51	110.91	120.05	127.77	123.38	106.7	110.77	121.42	121.2	123.34	121.18	123.11	118.98	101.3
15	172.15	166.11	149.6	155.65	137.1	184.65	162.47	154.3	185.9	115.02	154.65	161.69	174.38	166.57
16	126.76	139.16	139.05	115.19	127.75	111.28	103.4	134.59	103.87	141.51	139.04	140.83	113.29	138.82
17	116.08	126.28	123.87	115.6	113.69	100.87	112.16	137.99	116.63	132.06	114.68	104.98	122.02	109.58
18	146.55	140.76	131.31	145.7	141.72	146.6	146.8	136.97	134.31	144.73	150.86	133.48	138.06	130.46
19	177.79	155.32	136.12	136.44	145.81	173.35	150.1	158.47	154.26	181.32	166.03	143.99	160.33	177.66
20	129.22	115.92	132.75	126.22	114.86	130.12	124.18	124.79	113.04	126.56	124.16	127.9	129.32	129.87

Table A.2: $M_{t,i,a}$, Simulated Insured Population

	Region 1							Region 2						
	Age Class							Age Class						
Time	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	15	34	19	11	16	4	12	20	26	10	5	1	12	9
2	18	20	8	7	11	16	17	18	12	17	10	12	15	18
3	16	17	16	16	22	13	30	18	9	22	11	35	22	20
4	18	21	20	10	22	22	10	5	13	11	7	14	9	14
5	10	10	12	25	10	10	16	18	15	18	22	28	7	16
6	21	18	9	16	10	15	20	10	16	19	22	25	17	10
7	20	7	10	16	15	9	13	18	7	18	14	17	13	18
8	13	9	18	15	24	18	10	11	19	14	11	10	15	13
9	11	11	12	14	20	16	10	17	12	13	15	16	12	15
10	11	17	37	12	25	20	15	19	13	19	19	22	25	27
11	27	30	13	21	16	22	24	14	19	10	17	18	12	21
12	24	26	17	22	26	20	7	14	22	23	23	21	16	18
13	20	13	23	18	21	16	22	25	33	14	14	30	17	21
14	24	17	35	18	15	37	23	21	17	17	23	24	33	19
15	35	22	22	30	48	30	28	38	35	20	19	43	28	27
16	34	17	31	16	15	15	25	22	16	30	23	26	11	43
17	15	19	19	17	13	24	16	23	29	29	25	16	18	21
18	32	20	33	25	18	33	35	23	19	28	33	15	16	21
19	40	49	33	28	41	43	29	35	36	48	23	37	24	38
20	29	27	18	24	17	22	11	26	23	17	22	17	21	27

Table A.3: $N_{t,i,a}$, Simulated Total Claim Frequency

	Region 1							Region 2						
	Age Class							Age Class						
Time	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	104.36	907.33	463.51	281.34	183.51	67.04	291.88	44.34	1053.03	173.53	170.32	6.14	266.14	192.33
2	273.89	413.91	253.82	216.72	244.7	468.67	441.62	173.52	212.3	457.2	230.29	338.77	373.84	543.22
3	138.5	343.75	572.9	435.42	464.14	310.61	718.46	226.5	283.96	672.79	239.27	1052.24	386.91	639.55
4	206.75	580.2	331.57	483.41	581.09	717.53	174.59	42.72	208.22	292.88	259.39	424.11	188.21	453.95
5	186.63	208.87	303.68	687.06	201.56	123.34	345.96	261.87	253.78	502.43	714.59	599.55	117.81	386.07
6	414.34	489.54	245.97	488.49	144.99	342.7	685.51	278.31	350.21	259.88	548.93	472.88	378.49	228.7
7	542.69	148.7	267.42	547.69	557.54	182.65	316.52	297.46	178.51	369.33	193.2	603.05	379.68	349.07
8	445.07	334.61	305.87	415.62	671.17	667.64	206.14	91.31	685.32	363.59	475.49	226.67	595.37	285.31
9	130.78	258.94	295.93	230.52	372.46	643.44	345.47	140.02	259.62	308.92	333.46	304.53	452.33	395.3
10	443.19	425.09	1137.22	358.98	600.98	513.35	496.99	302.28	312.95	528.56	502.88	486.55	501.95	560.93
11	202.73	743.25	189.89	597.99	426.21	585.69	653.87	104.97	462.92	347.65	479.36	485.66	247.48	653.54
12	304.54	744.13	505.13	328.88	607.81	337.89	142.12	291.37	496.45	475.12	526.23	578.74	364.19	505.12
13	151.25	315.99	564.68	532.61	572.54	508.77	367.64	214.08	737.03	276.09	306.46	832.02	280.33	651.81
14	169.07	720.26	806.27	458.49	243.55	1078.94	670.72	335.04	413.49	465.13	722.5	952.82	728.67	512.98
15	204.93	793.86	546.67	670.15	1453.2	894.78	590.33	290.59	643.79	468.18	510.58	1309.17	938.06	623.5
16	148.38	369.56	808.06	207.34	312.99	435.57	461.04	295.81	381.17	716.92	646.08	665.04	210.85	851.61
17	73.37	431.78	547.74	428.62	631.4	518.79	402.32	162.43	846.25	552.72	628	333.9	518.35	464.75
18	142.59	501.08	775.23	858.8	579.64	1027.16	784.23	62.26	572.45	758.09	1071.01	370.25	284.03	644.59
19	389.18	1156.93	686.95	875.16	865.63	1200.76	869.47	189.36	739.61	1108.58	560.87	1025.51	508.25	859.71
20	269.61	707.17	357.38	520.26	345.73	446.6	244.1	131.23	494.28	295.18	411.79	462.96	440	672.43

Table A.4: $X_{t,i,a}$, Simulated Aggregate Loss

Appendix B

Algorithm Code

In this Section, the corresponding source code for Gibbs sampler algorithm is presented below. The code was made on R language.

```
#LOADING LIBRARIES------------------------------------------------

library( actuar )
library( random )
library( coda )
library( MASS )
library( mvtnorm )
library( corpcor )
library( coda )
library( R.utils )

#PARAMETERS MATRIX--------------------------------------------------

n=20

#METROPOLIS-HASTINGS------------------------------------------------

#biased by the initial value
proposalfunction <- function( par1, par2 ){
  dist <- rnorm( 1, par1, par2 )
  prob_dist <- pnorm( 1, par1, par2  )
  return( list( d = dist, p = prob_dist ) )
}
metropolis_hasting <- function( iterations, parameter, mu, sigma ){
  chain <- matrix( 1, n )
  chain[ 1 ] <- parameter[ 1 ]
  for ( i in 2:iterations-1  ) {
    proposal <- proposalfunction( mu, sigma  )
    probab <- min( 1, proposal$p / chain[ i-1 ]  )
```

```
    if ( runif( 1 ) < probab ){
      chain[ i+1 ] <- proposal$d
    } else{
      chain[ i+1 ] <- chain[ i ]
    }
  }
  return(chain)
}
proposalfunction1 <- function( par1, par2 ){
  dist <- rgamma( 1, par1, par2 )
  prob_dist <- rgamma( 1, par1, par2  )
  return( list( d = dist, p = prob_dist ) )
}
metropolis_hasting1 <- function( iterations, parameter, mu, sigma ){
  chain <- matrix( 1, n )
  chain[ 1 ] <- parameter[ 1 ]
  for ( i in 2:iterations-1  ) {
    proposal <- proposalfunction1( mu, sigma  )
    probab <- min( 1, proposal$p / chain[ i-1 ]  )
    if ( runif( 1 ) < probab ){
      chain[ i+1 ] <- proposal$d
    } else{
      chain[ i+1 ] <- chain[ i ]
    }
  }
  return(chain)
}
proposalfunction2 <- function( par1, par2 ){
  dist <- rmvnorm( 1, par1, par2 )
  return( list( d = dist ) )
}

metropolis_hasting2 <- function( iterations, alpha, beta ){
  chain <- matrix( 0, T, I,T )
  chain[ 1, ] <- L_matrix[ 1, ]
  for ( i in 2:iterations-1  ) {
    proposal <- proposalfunction2( alpha, beta)
    probab <- min( 1, proposal$d / chain[i-1, 1], proposal$d / chain[i-1, 2] )
    if ( runif( 1 ) < probab ){
      chain[ i+1, ] <- proposal$d
    } else{
      chain[ i+1, ] <- chain[ i, ]
    }
```

```
    #list( random_chain = chain )
  }
  return(chain)
}
proposalfunction3 <- function( par1, par2 ){
  dist <- rnorm( 1, par1, par2 )
  return( list( d = dist ) )
}

metropolis_hasting3 <- function( iterations, matrix, alpha, beta ){
  chain <- matrix( 0, T, A  )
  chain[ 1, ] <- matrix[ 1, ]
  for ( i in 2:iterations-1  ) {
    proposal <- proposalfunction3( alpha, beta )
    for (j in ncol( matrix ) ) {
      probab <- min( 1, proposal$d / chain[ i-1, j ] )
      if ( runif( 1 ) < probab ){
        chain[ i+1, ] <- proposal$d
      } else{
        chain[ i+1, ] <- chain[ i, ]
      }
    }
  }
  return(chain)
}

#KNOWN VALUES-------------------------------------------------------

T = 20
I = 2
A = 7
mu_0 <- 1
mu_1 <- 0.1
mu_2 <- 0.5
tau_0 <- 0.98
tau_1 <- 0.45
tau_2 <- .65
#for tau
alpha_tau <- 400
beta_tau <- 10000
#for sigma
alpha_sigma <- 0.8
beta_sigma <- 3.8
#for tau_varepsilon_0
```

```
alpha_tau_varepsilon_0 <- 100
beta_tau_varepsilon_0 <- 200
#for tau_varepsilon_2
alpha_tau_varepsilon_2 <- 50
beta_tau_varepsilon_2 <- 200
#for alpha_theta
alpha_alpha_theta <- 500
beta_alpha_theta <- 1.2
#for beta_theta
alpha_beta_theta <- 10000
beta_beta_theta <- 1.17
#for alpha_kappa
alpha_alpha_kappa <- 0.32
beta_alpha_kappa <- 0.13
#for beta_kappa
alpha_beta_kappa <- 0.32
beta_beta_kappa <- 0.13
#for alpha_lambda
alpha_alpha_lambda <- 120
beta_alpha_lambda <- 3
#for beta_lambda
alpha_beta_lambda <- 1200
beta_beta_lambda <- 7

#A PRIORI DISTRIBUTIONS-----------------------------------------------

Beta_kappa <- matrix( 1, n )
Beta_kappa[ 1 ] <- rgamma( 1, alpha_beta_kappa, beta_beta_kappa )
Alpha_kappa <- matrix( 1, n )
Alpha_kappa[ 1 ] <- rgamma( 1, alpha_alpha_kappa, beta_alpha_kappa )
Beta_lambda <- matrix( 1, n )
Beta_lambda[ 1 ] <- 200
Alpha_lambda <- matrix( 1, n )
Alpha_lambda[ 1 ] <- 40
Beta_theta <- matrix( 1, n )
Beta_theta[ 1 ] <- 10000
Alpha_theta <- matrix( 1, n )
Alpha_theta[ 1 ] <- 400
Tau_var0 <- matrix( 1, n )
Tau_var0[ 1 ] <- rgamma( 1, alpha_tau_varepsilon_0, beta_tau_varepsilon_0 )
Tau_var2 <- matrix( 1, n )
Tau_var2[ 1 ] <- rgamma( 1, alpha_tau_varepsilon_2, beta_tau_varepsilon_2 )
Sigma <- matrix( 1, n )
Sigma[ 1 ] <- rgamma( 1, alpha_sigma, beta_sigma )
```

```
Tau <- matrix( 1, n )
Tau[ 1 ] <- rgamma( 1, alpha_tau, beta_tau )
Beta_2 <- matrix( 0, n )
Beta_2[ 1 ] <- 0.075
Beta_1 <- matrix( 1, n )
Beta_1[ 1 ] <- 20
Beta_0 <- matrix( 1, n )
Beta_0[ 1 ] <- 50
Eta <- rpareto( 1, 1, 1)
P <- ( Sigma[ 1 ]^{-1} * ( 1 + Eta ) )/( 1 + 2*Eta )
S <- ( Sigma[ 1 ]^{-1} * Eta )/( 1 + 2*Eta )
sigma_matrix <- matrix( c( P, S, S, P ), 2, 2,T)
cor_coeff <- S / P #Datos estan correlacionados rho = 0.2028092
#PRIORI
L_matrix <- matrix( 1, T, I )
L_matrix[ 1, ] <- rmvnorm( 1 , rep(0,2), Tau[ 1 ]^{-1} * solve(sigma_matrix) )
L_matrix[1,]

sigma <- matrix( c(1,-.6361739 ,-0.6361739 ,1),2,2,T)
mean <-as.vector(c(2,2))
L_post <- metropolis_hasting2( T, mean, sigma)
plot( L_post[ ,1 ], type = "l")
plot( L_post[ ,2 ], type = "l")
L_matrix <- L_post

Var_a2 <- matrix( 1, n, A )
Var_a2[ 1, ] <- rnorm( 1, 0, 1/ Tau_var2[ 1 ] )
Var_a0 <- matrix( 1, n, A )
Var_a0[ 1, ] <- rnorm( 1, 0, 1/ Tau_var0[ 1 ] )
Beta_a2 <- matrix( 1, n, A )
Beta_a2[ 1, ] <- Beta_2[ 1 ] + Var_a2[ 1, ]
Beta_a0 <- matrix( 1, n, A )
Beta_a0[ 1, ] <- Beta_0[ 1 ] + Var_a0[ 1, ]
Lambda <- matrix( 1, n, A )
Lambda[ 1, ] <- rgamma( 1, Alpha_lambda[ 1 ], Beta_lambda[ 1 ] )
Kappa <- matrix( 1, n, A )
Kappa[ 1, ] <- 1
Theta <- matrix( 1, n, A )
Theta[ 1, ] <- rgamma( 1, Alpha_theta, Beta_theta )

Mu <- matrix( 0, T, I*A )
Mu[ 1, 1:A ] <- Beta_a0[ 1 ] + L_matrix[ 1, 1 ] + Beta_1[ 1 ]
                * exp( 1 * Beta_a2[ 1 ] )
Mu[ 1, (A+1):(I*A) ] <- Beta_a0[ 1 ] + L_matrix[ 1, 2 ] + Beta_1[ 1 ]
```

```
                * exp( 1 * Beta_a2[ 1 ] )
Mu[1,]

M <- matrix( 1, T, I*A )
for ( i in 1:ncol( Mu ) ) {
  M[ 1, i ] <- rnorm( 1, Mu[ 1, i ], 1/Tau[ 1 ] )
}
M[1,]
N <- matrix( 1, T, I*A )
for ( i in 1:ncol( M ) ) {
  N[ 1, i ] <- rpois( 1,  M[ 1, i ] * Lambda[ 1, 1 ] )
}
N[1,]
X <- matrix( 0, T, I*A )
for ( i in 1:ncol( N ) ) {
  X[ 1, i ] <- rgamma( 1, N[ 1, i ] * Kappa[ 1, 1 ], Theta[ 1, 1 ] )
}
X[1,]

#POSTERIOR--------------------------------------------------------------
Resta <- matrix( 1, T, I*A )
par1 <- matrix( 1, T )
system.time( for ( i in 2:n ) {
  #THETA
  for ( j in 1:ncol( Theta ) ) {
    Theta[ i, j ] <- rgamma( 1, Alpha_theta[ i - 1 ] + sum( N[ i-1, j ] ),
    sum( X[ i-1, j ] ) + Beta_theta[ i -1 ]  )
  }
  #LAMBDA
  for ( j in 1:ncol( Lambda ) ) {
    Lambda[ i, j ] <- rgamma( 1, Alpha_lambda[ i - 1 ] + sum( N[ i-1, j ] ),
    sum( M[ i-1, j ] ) + Beta_lambda[ i -1 ] )
  }
  #RESTA M-L
  Resta[ i-1, 1:A ] <- M[ i-1, 1:A ] - L_matrix[ i-1, 1 ]
  Resta[ i-1, (A+1):(I*A) ] <- M[ i-1, (A+1):(I*A) ] - L_matrix[ i-1, 2 ]
  #VAR_a2
  Var_a2 <- metropolis_hasting3( n, Var_a2, 0, 1 /sqrt( Tau ) - runif( n, 0, 1)  )
  #VAR_a0
  for ( j in 1:ncol( Var_a0 ) ) {
    Var_a0[ i, j ] <- rnorm( 1, ( Tau[ i-1 ] * sum( Resta[ i-1, ]
                            - Beta_0[ i-1 ] - Beta_1[ i-1 ]
                            * exp( i * ( Beta_2[ i-1 ]+ Var_a2[ i, j ] ) ) ) )/
                            ( Tau[ i-1 ] * T * I + Tau_var0[ i-1 ] ),
```

```
                        1/( Tau[ i-1 ] * T * I + Tau_var0[ i-1 ] ) )
  }
  #BETA_0
  Beta_0[ i ] <- rnorm( 1, ( tau_0 * mu_0 + Tau[ i-1 ] * sum( Resta[ i-1, ]
                      - Var_a0[ i ] - Beta_1[ i-1 ] * exp( i
                      * Beta_a2[ i-1 ] ) ) )/( Tau[ i-1 ]
                      * T * I * A + tau_0 ), 1 / (Tau[ i-1 ] * T * I * A + tau_0 ) )
  #BETA_1
  Beta_1[ i ] <- rnorm( 1, ( tau_1 * mu_1 + Tau[ i-1 ] * sum( Resta[ i-1, ]
                      - Var_a0[ i ] - Beta_0[ i ] ) * exp( i * ( Beta_2[ i-1 ]
                      + Var_a2[ i ] ) ) )/( Tau[ i-1 ] * exp( 2 * i
                      * Beta_a2[ i-1 ] ) + tau_1 ), 1 / ( Tau[ i-1 ]
                      * exp( 2 * i * ( Beta_2[ i-1 ] + Var_a2[ i ] ) ) + tau_1 ) )
  #BETA_2
  Beta_2 <- metropolis_hasting( 100, Beta_2, 0, 1/ sqrt( Tau ) +3/5 )
  #BETA_a0
  Beta_a0[ i, ] <- Beta_0[ i ] + Var_a0[ i, ]
  #BETA_a2
  Beta_a2[ i, ] <- Beta_2[ i ] + Var_a2[ i, ]
  #RESTA M-Mu
  Resta2 <- (M-Mu)^2
  #TAU
  Tau[ i ] <- rgamma( 1, 1/2 * T * I * A + alpha_tau, 1/2 * sum( Resta2[ i, ] )
                    + beta_tau  )
  #TAU_var0
  Tau_var0[ i ] <- rgamma( 1, A/2 + alpha_tau_varepsilon_0, beta_tau_varepsilon_0
                          + 1/2 * sum( ( Var_a0[ i, ] )^2 ) )
  #TAU_var2
  Tau_var2[ i ] <- rgamma( 1, A/2 + alpha_tau_varepsilon_2, beta_tau_varepsilon_2
                          + 1/2 * sum( ( Var_a2[ i, ] )^2 ) )
  #ALPHA_THETA
  Alpha_theta <- metropolis_hasting1( n, Alpha_theta, alpha_alpha_theta,
          beta_alpha_theta )
  #ALPHA_LAMBDA
  Alpha_lambda <- metropolis_hasting1( n, Beta_theta, alpha_alpha_lambda,
                             beta_alpha_lambda )
  #BETA_THETA
  Beta_theta[ i ] <- rgamma( 1, A * Alpha_theta[ i ] + alpha_beta_theta,
                             beta_beta_theta + sum( Theta[ i ] ) )
  #BETA_LAMBDA
  Beta_lambda[ i ] <- rgamma( 1, A * Alpha_lambda[ i ] + alpha_beta_lambda,
                             beta_beta_lambda + sum( Lambda[ i ] ) )
}
)
```

```
#FINAL MATRIXES-------------------------------------------------------
Mu1 <- Mu[ , 1:A ]
Mu2 <- Mu[ , (A+1):ncol( Mu ) ]
for (i in 2:nrow(Mu) ) {
  for (j in 1:ncol(Mu1 ) ) {
    Mu1[ i, j ] <- Beta_a0[ i ] + L_matrix[ i, 1 ] + Beta_1[ i ]
                    * exp( -i * Beta_a2[ i ] )
    Mu2[ i, j ] <- Beta_a0[ i ] + L_matrix[ i, 2 ] + Beta_1[ i ]
                    * exp(- i * Beta_a2[ i ] )
  }
}
Mu <- cbind( Mu1, Mu2 )
Mu <- round(Mu,2)
write.csv( Mu, "Mu_matrix" )
for ( i in 2:nrow(M) )  {
  for (j in 1:ncol(M)) {
    M[ i, j ] <- rnorm( 1, Mu[ i, j ], 1/Tau[ i ] )
  }
}
M <- round( M, 2)

N1 <- N[ , 1:A  ]
N2 <- N[ , (A+1):ncol(N)  ]
M1 <-  M[ , 1:A ]
M2 <- M[ , (A+1):ncol(N) ]
for ( i in 2:nrow( N ) ) {
  for ( j in 1:ncol( N1 ) ) {
    N1[ i, j ] <- rpois( 1,  M1[ i, j ] * Lambda[ i,  ] )
    N2[ i, j ] <- rpois( 1, M2[ i, j ] * Lambda[ i,  ] )
  }
}
N <- cbind( N1, N2 )

X1 <- X[ , 1:A ]
X2 <- X[ , (A+1):ncol(X)  ]
for ( i in 2:nrow( X ) ) {
  for ( j in 1:ncol( X1 ) ) {
    X1[ i, j ] <- rgamma( 1, N1[ i, j ] * Kappa[ 1, 1 ], Theta[ i, j ] )
    X2[ i, j ] <- rgamma( 1, N2[ i, j ] * Kappa[ 1, 1 ], Theta[ i, j ] )
  }
}
X <- cbind( X1, X2 )
X <- round( X, 2 )
```