# INTEGER-VALUED AUTOREGRESSIVE PROCESSES WITH DYNAMIC HETEROGENEITY AND THEIR APPLICATIONS IN AUTOMOBILE INSURANCE

by

Ting Zhang

Bachelor of Science, Peking University, 2007

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Statistics and Actuarial Science


SIMON FRASER UNIVERSITY
Fall 2009

## APPROVAL

**Name:** Ting Zhang

**Degree:** Master of Science

**Title of Project:** Integer-Valued Autoregressive Processes with Dynamic Heterogeneity and their Applications in Automobile Insurance

**Examining Committee:** Dr. BINGHAM, Derek
Chair

Dr. LU, Yi
Senior Supervisor,
Simon Fraser University

Dr. TSAI, Cary Chi-Liang
Supervisor,
Simon Fraser University

Dr. PARKER, Gary
External Examiner,
Simon Fraser University

**Date Approved:**

# Abstract

Bonus-malus systems in automobile insurance describe how the past claim frequencies determine the future insurance premiums. The potential risks of the policyholders vary due to differences in driving behavior, which leads to the unobserved heterogeneity in individual average claim counts. While the Poisson distribution has been used as a simple model for discrete count data, the negative binomial distribution is suggested for modeling the claim counts with unobserved heterogeneity by letting the mean parameter of the Poisson distribution follow a Gamma distribution. In this project, we introduce an integer-valued autoregressive process with dynamic heterogeneity to model the random fluctuations and correlations of the heterogeneity from year to year. Some properties of the model are studied, and a bonus-malus system is built and illustrated using the Gibb's Sampler algorithm. Finally, comparisons with other existing models are provided in terms of the extent to which they use the claim history.

**Keywords: Heterogeneity, Bonus-Malus System, Mixed Poisson Model, Bayesian Posterior Mean, Lognormal Distribution, Integer-Valued Autoregressive (INAR) Model**

*This project is dedicated to my beloved parents, who have been unconditionally granting love and support for twenty-five years.*

# Acknowledgments

I would like to express my deep and sincere gratitude to my supervisor Dr. Yi Lu for her encouraging and personal guidance during my graduate studies at Simon Fraser University. Her wide knowledge and the extensive discussions about my work have been of great value for me and provided a sound basis for my project.

I am deeply grateful to my co-supervisor Dr. Cary Chi-Liang Tsai for his detailed constructive comments and support for my work. I would also like to thank my external examiner Dr. Gary Parker for his detailed review. As actuarial professors, they have provided a solid knowledge of actuarial science and embedded a deep appreciation for scholarly rigor.

My sincere thanks to my colleagues at the WPLS who provided tremendous support throughout this challenging project. Special thanks to Ms. Deanna Napen, Ms. Coralie Breen, and Ms. Sonya Tomin, for encouraging me to never give up and providing enormous flexibility to permit me to concurrently complete the project and earn a living! Most importantly, they are all extremely amazing to work with and thus have made my work a perfect balance to my academic studies!

And, last but not least, writing my project was never a lonely journey because of my many friends. A warm thank you to my friends and fellow graduate students, Jillian Falkenberg, Matthew Pratola, Conghui Qu, Lingzhi Jiang, Li Chen, Yan Zhang, and Qipin He. Without them, it would have been much less fun and more difficult. Special thanks to Jill, it has been my luck to have her going through the program at almost exactly the same pace.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Automobile Insurance

Risks, as defined in actuarial science, refer to the uncertain events in the future that will cause damages or losses. When policyholders sign their contracts and pay premiums, they are covered and hence take a small portion of the potential losses that would have been significant. In an insurance portfolio, the potential risks exposed by policyholders vary; specifically for automobile insurance, the likelihood of having accidents varies among the insured drivers.

One of the main tasks of actuaries is to fairly allocate the burden of baring the potential losses among policyholders, which is materialized by quantitative analysis to specify individual risks and thus to determine the premiums. This procedure is called pricing or rate-making. There are two main phases involved. A base premium is determined when the policy is issued, and then the premium will be adjusted by discounts or surcharges as the policy is carried out. The reasons for separating the two phases will be discussed in the next two sections.

As a key actuarial rate-making principle, the price charged to a policyholder is the expected future costs related to insurance coverage. For automobile insurance, rate-making can be based on claim frequency and severity. The claim frequency is defined as the number of incurred claims per unit of earned exposure (usually a year). Traditionally, the claim frequency is used as the only factor to differentiate the policyholders with higher risks from the ones with lower risks. That is, the number of claims a policyholder may have during a year is assumed to be independent of the cost of these claims. The severity distribution is identical among all the policyholders in a portfolio. This assumption may not be realistic; nevertheless it is adopted by most ratemaking systems and in this project. New claim frequency models and relevant pricing issues are where most efforts had been put into.

## 1.2 *a priori* Rating System

In automobile insurance practice, actuaries design a tariff structure which partitions policies when issued according to their risk characteristics, which is often called *a priori* classification. The policyholders in the same class would pay the same base premium. Such a classification and assigned premium to each class compose an *a priori* rating system. The risk characteristics (or risk variables as used in the statistical models) include age, gender and occupation of the policyholders, the type and use of their cars, the places where they reside and sometimes even the number of cars in the household, marital status, and the color of the vehicles. Generalized linear regression is an ideal candidate to model the claim counts for *a priori* classification, and the average claim frequency for a class can be expressed as

$$\lambda_i = \exp(\mathbf{x_i}'\boldsymbol{\beta}), \tag{1.1}$$

where the subscript $i$ identifies the policyholder $i$, $\mathbf{x}_i = (x_i^1, x_i^2, \ldots, x_i^q)'$ denotes a vector of $q$ risk variables, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)'$ is the vector of regression coefficients.

The logic behind this tariff structure is that the risk characteristics are correlated to the underlying risk of a policyholder, and thus the future numbers of claims can be predicted by them. However, not all of these variables are necessary for prediction purpose if viewed from a statistical standpoint. It is the competition among insurers that forces actuaries to refine the partition to match other competitors; otherwise they would lose some drivers with respect to the factors they do not consider while their competitors do.

## 1.3 Bonus-Malus Systems

In spite of the trend towards more classification factors considered in the *a priori* classification, the *a priori* rating system is still not adequate to achieve the goal of distributing the burden fairly among policyholders. Firstly, some important risk factors may be banned by law because they are beyond the control of the insured, such as age and gender. More significantly, factors that could be incorporated in an *a priori* classification have to be observable and easy to be quantified; however, they are not necessarily all the factors that would affect the risk presented by a driver. Other factors are ruled out of the *a priori* rating system, some probably more crucial, such as the respectfulness towards the law and the aggressiveness when driving. Because they are not easy to observe or quantify, we can call them hidden factors.

The effect of the hidden factors can be revealed by the driving records of an insured. Thus, to compensate the inadequacies of the *a priori* rating system, insurers use experience rating to adjust premiums according to individual past claim histories. Such systems, penalizing insured drivers responsible for one or more accidents by premium surcharges (or malus), and rewarding claim-free policyholders by premium discounts (or bonus), are called no-claim discounts, experience ratings, merit ratings, or bonus-malus systems.

Due to the effect of hidden factors, heterogeneity exists in the individual expected claim frequencies for the policyholders in the same *a priori* class. However, we assume that the heterogeneity could be averaged out when one considers the whole group, i.e. the expected number of claims for the group remains the same as the one identified in the *a priori* rating system. This property is also called financial balance.

Let $\lambda$ denote the overall expected claim frequency for an *a priori* class. A random effect, often denoted by $\Theta$, is introduced to model the heterogeneity. Then the individual expected claim frequency could be expressed as $\lambda\Theta$. It is obvious that $\Theta$ is a positive random variable, and in order to fulfill the financial balance requirement, the mean of $\Theta$ is assumed to be one. The estimation of $\Theta$ given the past numbers of claims leads to premium discount or surcharge rates in a bonus-malus system.

## 1.4 Outline

Insurers implement both the *a priori* rating and bonus-malus systems at the same time to better asses individual risks. The *a priori* rating system determines the base premium for each class, which is based on the average claim frequency for the class, whereas the bonus-malus system reflects the deviation of individual mean claim frequency from the overall average for the class, and leads to discounts/surcharges on the base premium. Since both the rating systems are driven by the inherent nature of individual risk, they affect each other. Often a tougher *a priori* rating system results in a milder bonus-malus system, and vice versa.

The main aim of this project is to propose an integer-valued autoregressive process for the individual annual claim frequency integrated with a dynamic heterogeneity model. Associated pricing techniques are also demonstrated.

In Chapter 2, we review the basic annual claim count models based on a Poisson process, which includes the mixed Poisson model as a benchmark for further discussions, and the classical credibility approach for the premium adjustments. One of the potential problems of the credibility approach is that the age of a claim is not considered in the resulting premium, although it may affect the predictive ability of the claim. In Chapter 3, we review the integer-valued autoregressive (INAR) process accommodating unobserved heterogeneity, followed by integrating the INAR process with a dynamic heterogeneity in order to sufficiently exploit the information contained in a driver's claim history. In Chapter 4, a numerical approach is employed to obtain the pricing results for the INAR(1) process with dynamic heterogeneity, illustrating the superiorities over the formerly reviewed models. Chapter 5 concludes the project.

# Chapter 2

# Literature Review

Most of the rate-making systems in automobile practice are based only on claim counts. In this chapter, we review the basic claim count models and the associated pricing techniques. In Section 2.1, we will introduce the mixed Poisson distribution, which is based on the regular Poisson distribution and has the mean parameter as a random variable to reflect the heterogeneity of the portfolio. In Section 2.2, the pricing formula is derived under the Bayesian statistics framework, followed by two specific cases where the distributions of the heterogeneity prior to any obtained experience claim counts (the prior distribution) are assumed to be Gamma and Lognormal, respectively. In the last section, we will briefly review the relevant researches on other topics in the context of bonus-malus system. Denuit et al. (2007) and Lemaire (1995) provide comprehensive reviews on the relevant topics of automobile insurance rating models discussed in this Chapter.

## 2.1 Claim Counts Model

In this section, we first review the Poisson distribution, Poisson process, and mixed Poisson distribution, followed by the application of the mixed Poisson model in bonus-malus system for a non-homogeneous portfolio, which was first introduced by Dionne and Vanasse (1989).

### 2.1.1 Poisson Distribution

The Poisson distribution plays a prominent role in modeling claim counts due to some desirable properties. Assume $N$ is a counting random variable, following a Poisson distribution with parameter $\lambda$, denoted as $N \sim \text{Poi}(\lambda)$, $\lambda > 0$. Then $N$ takes non-negative integer values, and has probability mass function

$$\Pr[N = k] = \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Its mean and variance are given by

$$E[N] = \lambda,$$

and

$$V[N] = E[N^2] - E[N]^2 = \lambda.$$

The probability generating function of N is

$$\varphi_N(z) = E[z^N] = \sum_{k=0}^{+\infty} \mathrm{e}^{-\lambda}\,\frac{(\lambda z)^k}{k!} = \mathrm{e}^{\lambda(z-1)}, \qquad -1 < z < 1.$$

In addition, it is well known that the sum of $n$ ($n \geq 2$) independent Poisson variables is also Poisson distributed, with the parameter given by the summation of the n Poisson parameters. This property of Poisson variables can be easily proven with the probability generating function.

The property that the mean coincides with the variance makes Poisson distribution an appropriate model for data sets without overdispersion. It is a good fit for individual claim count in a homogeneous portfolio.

### 2.1.2 Poisson Process

A counting process is a collection of random variables $\{N(t), t \geq 0\}$, with $N(0) \geq 0$, indexed by a positive real-valued parameter $t$, where $N(t)$ represents the number of events that have occurred up to time $t$. According to Denuit et al. (2007), such a process must fulfill the following properties:

- $N(t) \geq 0$,
- $N(t)$ is integer-valued,
- if $s < t$, then $N(s) \leq N(t)$, and
- for $s < t$, $N(t) - N(s)$ equals the number of events that occurred in the interval $(s, t]$.

Furthermore, $N(t)$ is said to be a Poisson process with rate $\lambda > 0$ if

- the process has stationary increments, that is,

$$\Pr[N(t+\triangle) - N(t) = k] = \Pr[N(s+\triangle) - N(s) = k]$$

  for any integer $k \geq 0$, instants $s \leq t$, and increment $\triangle > 0$;

- the process has independent increments, that is, for any integer $k > 0$ and instants $0 \leq t_0 < t_1 < t_2 < \cdots < t_k$, the random variables $N(t_1) - N(t_0)$, $N(t_2) - N(t_1)$, ..., $N(t_k) - N(t_{k-1})$ are mutually independent; and

- for small $h$,

$$\Pr[N(h) = k] = \begin{cases} 1 - \lambda h + o(h), & \text{if } k = 0, \\ \lambda h + o(h), & \text{if } k = 1, \\ o(h), & \text{if } k \geq 2, \end{cases}$$

  where $o(h)$ is a function of $h$ such that $\lim_{h \to 0} o(h)/h = 0$.

The Poisson process is intimately related with the Poisson distribution. The number of events in any interval of length $t > 0$ of a Poisson process with rate $\lambda$, is Poisson distributed with mean $\lambda t$, i.e., for all $s, t > 0$

$$\Pr[N(s+t) - N(s) = n] = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!}, \qquad n = 0, 1, 2, \ldots$$

If the number of claims reported since a policy is issued is a Poisson process, then during any time increment $\triangle t$, the number of claims that newly arrived is Poisson distributed with mean $\lambda \triangle t$. Although Poisson process, which implies the aggregate claim counts at any time point, is continuous in time, in insurance practice the observations are only taken at equally spaced discrete time points. Therefore, we are interested in the claim counts during the time increments of equal length. Without loss of generality, we assume that the policies are being updated with one-year cycles (insurance years), and consequently the number of claims in any policy year is Poisson distributed with rate $\lambda$. Based on Poisson distribution, more advanced models for annual claim counts will be discussed in the following sections.

### 2.1.3 Mixed Poisson Distribution

Let $\Theta$ denote a random variable with distribution function $F$ and density function $f$, then $N$ is said to follow a mixed Poisson distribution if

$$\begin{aligned} N|\Theta = \theta &\sim \text{Poi}(\theta\lambda), \quad \theta > 0, \\ \Theta &\sim F \end{aligned} \tag{2.1}$$

With the consideration of financial balance discussed in Chapter 1, we apply an additional condition on the distribution of $\Theta$, that is,

$$E[\Theta] = 1.$$

The deployment of the mixed Poisson distribution in this project is motivated by two considerations. First, it is more common to have a non-homogeneous portfolio, in which individual policies have various mean claim counts, than to have a homogeneous one. The random variable $\Theta$ is introduced to reflect the heterogeneity of the individual mean. Second, in spite of the mathematical conciseness, the equality of mean and variance tends to constrain the attempts of modeling with Poisson distribution, as real data often features over-dispersion, and one could be underestimating the risk if using variance as the measure. The heterogeneity $\Theta$ brings more volatility into the model, and therefore increases the variance. The conditional mean and variance are easily obtained by the properties of Poisson distribution, as

$$E[N|\Theta] = V[N|\Theta] = \lambda\Theta,$$

which implies that the unconditional mean and variance are

$$E[N] = E[E[N|\Theta]] = E[\lambda\Theta] = \lambda, \tag{2.2}$$

and

$$\begin{aligned} V[N] &= E[V[N|\Theta]] + V[E[N|\Theta]] \\ &= E[\lambda\Theta] + V[\lambda\Theta] \\ &= \lambda + \lambda^2 V[\Theta]. \end{aligned}$$

The results above show that the mixed Poisson distribution, with the heterogeneity $\Theta$ assumed to be a non-negative random variable with mean one, maintains the same mean and has a larger variance. Thus it is superior to the Poisson distribution in terms of modeling the over-dispersion.

The unconditional distribution of $N$ can be obtained by integrating the joint distribution of $N$ and $\Theta$ over all the possible values of $\Theta$. An illustration is provided below with the heterogeneity $\Theta$ following a Gamma distribution, denoted as $\Theta \sim \text{Gam}(a, a)$, (the two parameters are set to the same value to fulfill the condition $E[\Theta] = 1$); the density function of $\Theta$ is given by

$$f(\theta) = \frac{a^a \theta^{a-1} \mathrm{e}^{-a\theta}}{\Gamma(a)}, \qquad \theta > 0. \tag{2.3}$$

Then

$$\begin{aligned}
\Pr[N = k] &= \int_0^\infty \Pr[N = k|\Theta = \theta] f(\theta) d\theta \\
&= \int_0^\infty \mathrm{e}^{-\lambda\theta} \frac{(\lambda\theta)^k}{k!} \frac{a^a \theta^{a-1} \mathrm{e}^{-a\theta}}{\Gamma(a)} d\theta \\
&= \frac{\Gamma(k+a)}{\Gamma(a) k!} \frac{a^a \lambda^k}{(\lambda + a)^{(k+a)}} \int_0^\infty \frac{(\lambda + a)^{(k+a)} \theta^{k+a-1} \mathrm{e}^{-(\lambda+a)\theta}}{\Gamma(k+a)} d\theta \\
&= \binom{k+a-1}{k} \left(\frac{a}{\lambda + a}\right)^a \left(1 - \frac{a}{\lambda + a}\right)^k, \quad k = 0, 1, 2, \ldots
\end{aligned}$$

which indicates that the unconditional distribution of $N$ is a negative binomial distribution with parameter $a$ and $a/(\lambda + a)$.

## 2.2 Bayesian Credibility Premium with the Quadratic Loss Function

With the mixed Poisson distribution assumed for annual claim frequency, methods in Bayesian statistics framework can be used to determine the discount or the surcharge rates after the driving experience is obtained. Assume there are $T_i$ years of past numbers of claims available for policy $i$ in the portfolio. We define three vectors to represent this policy, $\mathbf{N}_i = (N_{i,1}, N_{i,2}, \ldots, N_{i,t}, \ldots)$, $\boldsymbol{\lambda}_i = (\lambda_{i,1}, \lambda_{i,2}, \ldots, \lambda_{i,t}, \ldots)$, and $\boldsymbol{\Theta}_i = (\Theta_{i,1}, \Theta_{i,2}, \ldots, \Theta_{i,t}, \ldots)$, where $N_{i,t}$ denotes the number of claims reported, $\lambda_{i,t}$ is the mean claim frequency of the *a priori* class the policyholder belongs to, and $\Theta_{i,t}$ represents the heterogeneity of the policy in policy year $t$. A great deal of effort has been devoted into modeling the heterogeneity $\boldsymbol{\Theta}_i$ and addressing the associated questions regarding prediction and pricing. Of various models that have been found favorable, we start from the classical one introduced by Dionne and Vanasse (1989). The assumptions for this model are that

- for policy i, the heterogeneity is a random variable which does not vary over time, i.e., $\mathbf{\Theta}_i$ can be simplified to $\Theta_i \sim F_{\Theta_i}$; and
- for all policies in the portfolio, $\Theta_i, i = 1, 2, \ldots,$ are independently and identically distributed, and the sequences $(\Theta_i, \mathbf{N}_i), i = 1, 2, \ldots$ are independent.

Given the past numbers of claims, an estimate for the heterogeneity $\Theta_i$ can be obtained. This estimation would be used to determine the surcharge or discount rate. When the observations of claim counts for $T_i$ years, denoted by $n_{i,1}, n_{i,2}, \ldots, n_{i,T_i}$, are available, the estimation is a function of $n_{i,1}, n_{i,2}, \ldots, n_{i,T_i}$, denoted by $\Psi^*(n_{i,1}, n_{i,2}, \ldots, n_{i,T_i})$, which is the closest value to $\Theta_i$ over all the measurable functions $\Psi$. This can be obtained by minimizing the conditional expected value of the so-called quadratic loss function

$$E[(\Theta_i - \Psi(N_{i,1}, N_{i,2}, \ldots, N_{i,T_i}))^2 | N_{i,1} = n_{i,1}, N_{i,2} = n_{i,2}, \ldots, N_{i,T_i} = n_{i,T_i}]. \tag{2.4}$$

Under the Bayesian statistics framework, the solution is given by the posterior mean of $\Theta_i$ given the past $T_i$ years' experience of policy $i$. Let $\mathbf{N}_i^{T_i}$ denote $(N_{i,1}, N_{i,2}, \ldots, N_{i,T_i})$ and $\mathbf{n}_i^{T_i}$ denote $(n_{i,1}, n_{i,2}, \ldots, n_{i,T_i})$. According to Carlin and Louis (2000), equation (2.4) can be expanded as

$$\begin{aligned}
& E\left[\left(\Theta_i - \Psi(\mathbf{N}_i^{T_i})\right)^2 | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}\right] \\
= \; & E\left[\left(\Theta_i - E[\Theta_i | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}] + E[\Theta_i | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}] - \Psi(\mathbf{N}_i^{T_i})\right)^2 | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}\right] \\
= \; & E\left[\left(\Theta_i - E[\Theta_i | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}]\right)^2 | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}\right] \\
& +2E\left[(\Theta_i - E[\Theta_i | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}])(E[\Theta_i | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}] - \Psi(\mathbf{n}_i^{T_i})) | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}\right] \\
& +E\left[\left(E[\Theta_i | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}] - \Psi(\mathbf{n}_i^{T_i})\right)^2 | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}\right] \\
= \; & E\left[\left(\Theta_i - E[\Theta_i | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}]\right)^2 | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}\right] + \left(E[\Theta_i | \mathbf{N}_i^{T_i} = \mathbf{n}_i^{T_i}] - \Psi(\mathbf{n}_i^{T_i})\right)^2,
\end{aligned}$$

which is minimized when $\Psi(N_{i,1}, \ldots, N_{i,T_i}) = E[\Theta | N_{i,1}, \ldots, N_{i,T_i}]$, as the cross term essentially equals to zero.

Therefore, given $N_{i,1} = n_{i,1}, N_{i,2} = n_{i,2}, \ldots, N_{i,T_i} = n_{i,T_i}$, the optimal solution of (2.4) $\Psi^*$ has the expression

$$\begin{aligned}\Psi^*(n_{i,1}, n_{i,2}, \ldots, n_{i,T_i}) &= E[\Theta_i | N_{i,1} = n_{i,1}, N_{i,2} = n_{i,2}, \ldots, N_{i,T_i} = n_{i,T_i}] \\ &= \frac{\int_0^{+\infty} \theta_i (\prod_{t=1}^{T_i} \Pr[N_{i,t} = n_{i,t} | \Theta_i = \theta_i]) dF_{\Theta_i}(\theta_i)}{\int_0^{+\infty} (\prod_{t=1}^{T_i} \Pr[N_{i,t} = n_{i,t} | \Theta_i = \theta_i]) dF_{\Theta_i}(\theta_i)} \\ &= \frac{\int_0^{+\infty} \mathrm{e}^{-\lambda_{i\cdot}\theta_i}\, \theta_i^{n_{i\cdot}+1} dF_{\Theta_i}(\theta_i)}{\int_0^{+\infty} \mathrm{e}^{-\lambda_{i\cdot}\theta_i}\, \theta_i^{n_{i\cdot}} dF_{\Theta_i}(\theta_i)},\end{aligned} \tag{2.5}$$

where

$$\lambda_{i\cdot} = \sum_{t=1}^{T_i} \lambda_{i,t}$$

is the intensity rate of the Poisson distribution for the total claim count over $T_i$ years, and

$$n_{i\cdot} = \sum_{t=1}^{T_i} n_{i,t},$$

is the total number of claims over $T_i$ years. In the following, we illustrate two commonly used mixed Poisson models.

### 2.2.1 Poisson-Gamma Model

Assume that $\Theta_i$ is Gamma distributed with density function given by (2.3); the conditional (posterior) distribution of $\Theta_i$ given all the historical numbers of claims is

$$f_{\Theta_i|\mathbf{N}_i}(\theta_i | n_{i,1}, \ldots, n_{i,T_i}) = \frac{\mathrm{e}^{-\theta(a+\lambda_{i\cdot})}\, \theta_i^{a+n_{i\cdot}-1}}{\int_0^{+\infty} \mathrm{e}^{-\theta_i(a+\lambda_{i\cdot})}\, \theta_i^{a+n_{i\cdot}-1} d\theta_i}, \qquad a > 0, \quad \theta_i > 0,$$

which is the density function of a Gamma distribution with parameters $a + n_{i\cdot}$ and $a + \lambda_{i\cdot}$. That is, we have

$$\Theta_i | (N_{i,1} = n_{i,1}, \ldots, N_{i,T_i} = n_{i,T_i}) \sim \mathrm{Gam}(a + n_{i\cdot}, a + \lambda_{i\cdot}),$$

and the estimation of $\Theta_i$ is given by

$$\begin{aligned}\Psi^*(n_{i,1}, n_{i,2}, \ldots, n_{i,T_i}) &= E[\Theta | N_{i,1} = n_{i,1}, N_{i,2} = n_{i,2}, \ldots, N_{i,T_i} = n_{i,T_i}] \\ &= \frac{a + n_{i\cdot}}{a + \lambda_{i\cdot}}.\end{aligned} \tag{2.6}$$

Furthermore, the pure premium for year $T_i + 1$ derived from (2.5) is

$$\begin{aligned} & E[N_{i,T_i+1}|N_{i,1} = n_{i,1}, \ldots, N_{i,T_i} = n_{i,T_i}] \\ =& \lambda_{i,T_i+1} E[\Theta_i|N_{i,1} = n_{i,1}, N_{i,2} = n_{i,2}, \ldots, N_{i,T_i} = n_{i,T_i}] \\ =& \lambda_{i,T_i+1} \frac{a + n_{i\cdot}}{a + \lambda_{i\cdot}}. \end{aligned} \tag{2.7}$$

The Poisson-Gamma model has been widely used in the literature that was developed for Bonus-Malus systems. It has several appealing properties. First, it is concise in mathematical expression. Both the prior distribution and the posterior distribution of $\Theta_i$ are Gamma. This property of the heterogeneity $\Theta_i$ is called conjugation in the Bayesian statistics framework. Second, if employed in insurance practice, the Bonus-Malus system defined in (2.6) is fair in the sense that:

- The insurance company does not make profit or lose money by launching such a system. Considering all the possible sequences of past numbers of claims, the resulting average pure premium is equal to the premium charged based only on the *a priori* mean claim frequency. For a detailed proof, please refer to Lemaire (1995). In other words, if there are sufficiently many policies such that the sequences of the past numbers of claims match the exact pattern of claims under the theoretical distribution, then the surcharges should pay off the discounts granted.

- In the long run, everyone will pay a premium corresponding to his own risk. We consider the results above in the extreme situation. As $T_i$ approaches infinity, $E[\Theta_i|N_{i\cdot} = n_{i\cdot}]$ converges to $n_{i\cdot}/\lambda_{i\cdot}$, which is the actual risk of the policy. The posterior variance

  $$V[\Theta_i|N_{i,1} = n_{i,1}, N_{i,2} = n_{i,2}, \ldots, N_{i,T_i} = n_{i,T_i}] = \frac{a + n_{i\cdot}}{(a + \lambda_{i\cdot})^2}$$

  tends to 0 as $T_i \to \infty$. It indicates that the discrimination among the insureds is asymptotically perfect.

Finally, the Bayesian credibility premium contains a weighted average when (2.7) is rewritten as

$$E[N_{i,T_i+1}|N_{i,1} = n_{i,1}, \ldots, N_{i,T_i} = n_{i,T_i}] = \left( \frac{a}{a + \lambda_{i\cdot}} E[\Theta_i] + \frac{\lambda_{i\cdot}}{a + \lambda_{i\cdot}} \frac{n_{i\cdot}}{\lambda_{i\cdot}} \right) E[N_{i,T_i+1}],$$

which is the product of the mean of the marginal distribution of $N_{i,T_i+1}$ and the weighted average consisting of the expectation of the prior distribution of $\Theta_i$ with weight $a/(a + \lambda_{i\cdot})$, and the average claim frequency for that policy $n_{i\cdot}/\lambda_{i\cdot}$ with a weight $\lambda_{i\cdot}/(a + \lambda_{i\cdot})$.

In the following we give two numerical examples for the Poisson-Gamma model. Tables 2.1 and 2.2 show the pure premiums after various combination of the number of the observed years $T_i$ and the number of the past claims $n_{i\cdot}$; each shows the results for a good driver (with smaller $\lambda_i$) and a moderate driver (with larger $\lambda_i$), respectively.

**Table 2.1:** *Values of $E[\Theta|N_{i\cdot} = n_{i\cdot}]$ for different combination of the observed periods $T_i$ and the numbers of the past claims $n_{i\cdot}$ for a good driver (average claim frequency 9.28%)*

| $T_i$ | Number of claims $n_{i\cdot}$ | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| 1 | 92.0% | 178.4% | 264.7% | 351.1% | 437.5% | 523.8% |
| 2 | 85.2% | 165.1% | 245.1% | 325.0% | 405.0% | 485.0% |
| 3 | 79.3% | 153.7% | 228.2% | 302.6% | 377.0% | 451.5% |
| 4 | 74.2% | 143.8% | 213.4% | 283.0% | 352.7% | 422.3% |
| 5 | 69.7% | 135.1% | 200.5% | 265.9% | 331.3% | 396.7% |
| 6 | 65.7% | 127.3% | 189.0% | 250.6% | 312.3% | 374.0% |
| 7 | 62.1% | 120.4% | 178.9% | 237.1% | 295.4% | 353.7% |
| 8 | 58.9% | 114.3% | 169.6% | 224.9% | 280.2% | 335.6% |
| 9 | 56.0% | 108.7% | 161.3% | 213.9% | 266.6% | 319.2% |
| 10 | 53.4% | 103.6% | 153.8% | 204.0% | 254.1% | 304.3% |

To compare the severity of the two scales, let us consider the case where one claim is reported in the first year. We see the good driver will receive a surcharge of 178.4% at the beginning of the second year, whereas the surcharge for the moderate driver in the second year is only 171.2%. The scale for good driver is more severe than the one for the moderate driver when they have reported claims. On the other hand, the second column of these two tables show the discount rates for keeping a claim-free record, and we see that moderate drivers receive better discounts. This is the consequence of the nature of financial balance, which requires that the surcharges compensate the discounts. Given the fact that the scale for good drivers, compared to the one for moderate drivers, is applied on a population which is more likely to have small numbers of claims, it is reasonable that this scale poses more severe surcharges and less preferable discounts.

### 2.2.2 Poisson-Lognormal Model

In this section, we derive the estimation of the heterogeneity with the past numbers of claims under the assumption that the prior distribution is a Lognormal distribution.

**Table 2.2:** *Values of $E[\Theta|N_{i\cdot} = n_{i\cdot}]$ for different combination of the observed periods $T_i$ and the numbers of the past claims $n_{i\cdot}$ for a moderate driver (average claim frequency 14.09%)*

| $T_i$ | Number of claims $n_{i\cdot}$ | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| 1 | 88.3% | 171.2% | 254.2% | 337.1% | 420.0% | 503.0% |
| 2 | 79.1% | 153.3% | 227.6% | 301.8% | 376.1% | 450.4% |
| 3 | 71.6% | 138.8% | 206.0% | 273.3% | 340.5% | 407.7% |
| 4 | 65.4% | 126.8% | 188.2% | 249.6% | 311.0% | 372.5% |
| 5 | 60.2% | 116.7% | 173.2% | 229.8% | 286.3% | 342.8% |
| 6 | 55.8% | 108.1% | 160.5% | 212.8% | 265.2% | 317.5% |
| 7 | 51.9% | 100.7% | 149.4% | 198.2% | 247.0% | 295.7% |
| 8 | 48.6% | 94.2% | 139.8% | 185.5% | 231.1% | 276.7% |
| 9 | 45.7% | 88.5% | 131.4% | 174.3% | 217.1% | 260.0% |
| 10 | 43.1% | 83.5% | 123.9% | 164.3% | 204.8% | 245.2% |

The heterogeneity $\Theta_i$ is said to follow a Lognormal distribution, denoted by $\Theta_i \sim \mathrm{LN}(\mu, \sigma^2)$, if its density function is

$$f_{\Theta_i}(\theta_i) = \frac{1}{\sqrt{2\pi}x\sigma}\, \mathrm{e}^{-\frac{(\ln\theta_i-\mu)^2}{2\sigma^2}}, \quad \theta_i > 0, \quad \mu \in \mathbb{R}, \quad \sigma \geq 0. \tag{2.8}$$

The mean of $\Theta_i$ is $\mathrm{e}^{\mu+\sigma^2/2}$, and the financial balance requirement yields $\mu = -\sigma^2/2$. The optimal estimation of $\Theta_i$ can be obtained by applying (2.8) to (2.5), that is,

$$\begin{aligned} \Psi^*(n_{i,1}, n_{i,2}, \ldots, n_{i,T_i}) &= E[\Theta_i|N_{i,1} = n_1, N_{i,2} = n_2, \ldots, N_{i,T_i} = n_{T_i}] \\ &= \frac{\int_0^{+\infty} \mathrm{e}^{-\lambda_{i\cdot}\theta_i}\, \theta_i^{n_{\cdot}+1} \frac{1}{\sqrt{2\pi}\theta_i\sigma}\, \mathrm{e}^{-\frac{(\ln\theta_i-\mu)^2}{2\sigma^2}}\, d\theta_i}{\int_0^{+\infty} \mathrm{e}^{-\lambda_{i\cdot}\theta_i}\, \theta_i^{n_{\cdot}} \frac{1}{\sqrt{2\pi}\theta_i\sigma}\, \mathrm{e}^{-\frac{(\ln\theta_i-\mu)^2}{2\sigma^2}}\, d\theta_i}. \end{aligned} \tag{2.9}$$

There is no explicit form for the integrations on the numerator and denominator, yet the estimation can be obtained by numerical approaches.

## 2.3 The Deficiency of the Classical Models

There are two drawbacks of the classical model. One is that the heterogeneity $\Theta_i$ is assumed not to vary over time. This is not realistic since all the hidden factors may change; certainly a driver could grow more careful and responsible. It is reasonable to assume different random variables correlated with each other for the heterogeneity in different years $\{\Theta_{i,t}\}_{t=1}^{\infty}$. As pointed out by Brouhns et al. (2003), a driver will tend to be more prudent after an accident but this effect naturally decreases with the age of the claim. Pinquet et al. (2001) presented a dynamic heterogeneity model, and obtained the estimates of the parameters and prediction on premiums based on the longitudinal data inference.

Another shortcoming of the classical model which can be observed from (2.5) is that the estimation of the heterogeneity reflects only the sum of all the past numbers of claims. How the claims are distributed among the past years does not affect the estimation. This limitation reduces the efficiency of the classical model. The reasons are as follows. First, not all the information of the past claim frequencies are being used. Second, different distributions of claims indeed imply different potential risks; thus, the burden of claims is not distributed fairly. To solve this problem Gourieroux and Jasiak (2004) developed a claim counting process with the first order Integer-Valued Autoregressive (INAR(1)) process, and derived the pricing formulas with experience claim counts up to three years. They obtained the desirable results where the more claims distributed in the latter years, the higher the premium for the following year would tend to be.

## 2.4 Other Modeling Approaches

A great deal of efforts had been devoted in modeling the claim counts and developing mechanisms to discriminate good drivers from the bad ones by charging premiums differently. Apart from the classical model for bonus-malus systems reviewed earlier, here are some other models and aspects for the automobile insurance in the literature.

### 2.4.1 Bonus-Malus Scales

Instead of directly using all the past numbers of claims to approximate the surcharge or discount rates, bonus-malus scales are designed to store the information of the past claim counts of the policyholders and determine the relative premiums. This pricing mechanism is widely adopted in the insurance industry. A bonus-malus scale consists of a finite number of levels, and each of them is assigned a surcharge or discount rate (or so called relativity). Policyholders are transferred by a transition rule within these levels every year according to the current level occupied and the number of claims filed in the current year.

The process of a policyholder moving in the bonus-malus scale fulfills the properties of a

Markov Chain. The calculation of the relativities is based on the theory of stationary status of the Markov Chain and the Bayesian theory. This had been studied by Pitrebois et al. (2003b), following on from Taylor (1997) for the extension of Norberg's (1976) pioneering work on segmented tariffs. Pitrebois et al. (2004) developed the linear relativities more easily accepted by the insurance agents and customers. Pitrebois et al. (2003a) addressed the problem of the bonus-malus scale used in Belgium, which is not a Markov Chain due to some special transition rules, by introducing fictitious levels. In Brouhns et al. (2003), a more advanced claim counting process model, where the *a priori* class and the heterogeneity factor of a certain policy are allowed to change over years, was implemented into the bonus-malus scale; the authors proposed a numerical simulation to estimate the relativities. Details on the theory and application of bonus-malus scales can be found in Chapter 4 of Denuit et al. (2007).

### 2.4.2 Hierarchical Models

The hierarchical models had been introduced to model the automobile claim frequencies in order to better reflect the over-dispersion based on mixed Poisson distribution, in which the Poisson mean is a random variable (called heterogeneity); another layer of randomness is incorporated by letting one of the parameters of the heterogeneity also be a random variable. For the latter random variable, all the parameters are fully specified as deterministic constants. Gomez-Deniz et al. (2008) studied the properties of a Poisson-Gamma-Gamma model as

$$\begin{aligned} & N_{i,t}|(\lambda_{i,t}, \Theta_i = \theta_i) \sim \text{Poi}(\lambda_{i,t}\theta_i), \\ & \Theta_i \sim Gamma(a, b), \\ \text{and} \qquad & b \sim Gamma(\alpha, \beta) \end{aligned}$$

where $a, \alpha, \beta > 0$. They derived the marginal distribution of the claim counts, and applied the central moments matching method to estimate the parameters. In Gomez-Deniz and Vaquez-Polo (2006) the hierarchical model was used to construct the bonus-malus system, and the premiums were calculated under the Bayesian statistics framework.

### 2.4.3 Bonus Hunger

In the mechanism of bonus-malus systems, on one side, the insurer updates the premium based on the reported claims and applies surcharges as penalty for causing accidents; on the other side, as rational customers, the policyholders would be intuitively seeking a balance between having damages covered with premium increased the next year and bearing the damage themselves to keep the premium low. It is reasonable to assume that the policyholders would avoid reporting some minor accidents, which leads to the fact that the data sets possessed by insurers are highly likely censored. Regarding this issue, Lemaire (1976) and Lemaire (1977) studied the so-called bonus hunger and proposed a dynamic programming algorithm to determine the optimal claiming behavior from the perspective of the policyholders. Following that, Walhin and Paris (2000) applied the Lemaire's algorithm and a non-parametric Poisson fit to a censored motor insurance portfolio.

### 2.4.4 Models Incorporating Claim Severities

Traditionally, the modeling of bonus-malus systems takes only the numbers of claims into account, regardless of the magnitude of the losses these accidents may have caused. This is also a common case in practice. It is equivalent to assuming that the number of accidents an insured may have per year is independent of the severity of these accidents. This assumption certainly contributes to the necessary simplicity of the model, yet it is closer to reality to incorporate the claim severity into the risk measure. With this consideration, Picard (1976) first proposed a model to distinguish the claims that caused only property damage from those that caused both bodily injury and property damage. Lemaire (1995) established a credibility model based on the Poisson-Gamma mixture and the assumption that given the predicted average claim frequency, the portion of the claims with bodily injury involved follows a Beta distribution. Pinquet (1998) applied the multi-equation Poisson model with random effects and designed an optimal credibility model that takes into account the types of claims. Pitrebois et al. (2006) designed a bonus-malus scale that imposes different penalties for different types of claims.

# Chapter 3

# Heterogeneous INAR Model

In this chapter, the Integer-Valued Autoregressive (INAR) process is introduced to model the numbers of claims for consecutive policy years. Al-Osh and Alzaid (1987) proposed what they have called an integer-valued first order autoregressive (INAR(1)) model. Later, Gourieroux and Jasiak (2004) applied it in bonus-malus system design. Gourieroux and Jasiak (2004) also incorporated a static Gamma heterogeneity into the Poisson error terms of the process, and derived the associated pricing formulas.

We will review their work in the first two sections. In the last section, a dynamic model is employed for modeling the heterogeneity, which brings reasonable fluctuation and correlation into the heterogeneity over years, followed by corresponding pricing formulas.

## 3.1 Integer-Valued Autoregressive Process

The integer-valued autoregressive process has been developed under the time series literature to fit the integer valued processes, such as the individual annual claim counts. Its expression is similar to a regular autoregressive process, which consists of an autoregressive component and the error term(s). Keeping all the notations consistent with the last chapter, the recursive expression of the INAR(1) process is

$$N_{i,t} = \mathcal{B}_t(p) \circ N_{i,t-1} + \epsilon_{i,t}, \quad 0 < p < 1, \quad t = 1, 2, \dots \tag{3.1}$$

In this formula, $\{\epsilon_{i,t}\}_{t=1}^{\infty}$ is a sequence of random variables taking nonnegative integer values, and $\mathcal{B}_t(p)$ is the so-called Binomial thinning operator, which is independent of the error terms and defined by

$$\mathcal{B}_t(p) \circ N_{i,t-1} = \sum_{j=1}^{N_{i,t-1}} U_j \tag{3.2}$$

where $\{U_j\}$ is a sequence of iid Bernoulli random variables with parameter $p$.

As an intuitive interpretation of the INAR(1) model above, the number of claims for year $t$ consists of two parts. First, the error terms are determined by both of the observable and the hidden factors of the insured, which are represented by $\lambda_{i,t}$ and $\Theta_{i,t}$ respectively. The results from the previous chapter are adopted in the INAR(1) process and embedded in the error terms, i.e., the error terms have a mixed Poisson distribution

$$\epsilon_{i,t}|\Theta_{i,t} = \theta_{i,t} \sim \text{Poi}(\theta_{i,t}\lambda_{i,t}).$$

The other part is a random portion of the number of claims for the previous policy year. Each of the claims from the previous year has a probability $p$ to contribute a claim to the claim count for the next year. Al-Osh and Alzaid (1987) introduced an alternative expression of the INAR(1), as stated in the following proposition, which is more convenient for further derivations.

**Proposition 3.1.** *The INAR(1) process $\{N_{i,t}\}_{t=1}^{\infty}$ defined by (3.1) has the following form as an infinite summation:*

$$N_{i,t} = \sum_{j=0}^{\infty} \mathcal{B}_t(p^j) \circ \epsilon_{i,t-j}. \tag{3.3}$$

**Proof.** $N_{i,t}$ can be expanded recursively by (3.1) as

$$\begin{aligned}
N_{i,t} &= \epsilon_{i,t} + \mathcal{B}_t(p) \circ (\mathcal{B}_{t-1}(p) \circ N_{i,t-2} + \epsilon_{i,t-1}) \\
&= \epsilon_{i,t} + \mathcal{B}_t(p) \circ \mathcal{B}_{t-1}(p) \circ N_{i,t-2} + \mathcal{B}_t(p) \circ \epsilon_{i,t-1} \\
&= \epsilon_{i,t} + \mathcal{B}_t(p) \circ \epsilon_{i,t-1} + \mathcal{B}_t(p) \circ \mathcal{B}_{t-1}(p) \circ \epsilon_{i,t-2} + \dots \\
&\quad + \mathcal{B}_t(p) \circ \mathcal{B}_{i,t-1}(p) \circ \dots \circ \mathcal{B}_{t-h+1}(p) \circ \epsilon_{i,t-h} + \dots \\
&= \sum_{j=0}^{\infty} \mathcal{B}_t(p^j) \circ \epsilon_{i,t-j}.
\end{aligned}$$

□

Al-Osh and Alzaid (1987) provided an interpretation of the INAR(1) process in the form of (3.3) in the context of queuing system. The summation represents the total number of customers in the system. A new queue forms as the time goes forward by one unit, and the length of the queue is Poisson distributed. In the meantime, whether an existing customer stays in the system is a Bernoulli random variable. The likelihood of an existing customer staying in the system decreases exponentially with time. The customers are assumed to make their decisions independently. The next proposition gives the mean, variance, and auto-correlation of the INAR(1) process.

**Proposition 3.2.** *Let j, k, and m be non-negative integers, then the mean and the auto-correlation of the INAR(1) process are*

$$E[N_{i,t}] = \sum_{j=0}^{\infty} p^j E[\epsilon_{i,t-j}] \tag{3.4}$$

*and*

$$Cov[N_{i,t}, N_{i,t-k}] = \sum_{j=k}^{\infty} \{p^j E[\epsilon_{i,t-j}] + p^{2j-k}(V[\epsilon_{i,t-j}] - E[\epsilon_{i,t-j}])\} + \sum_{\substack{m=0,j=k \\ j\neq m}}^{\infty} p^{j+m-k} Cov[\epsilon_{i,t-m}, \epsilon_{i,t-j}] \tag{3.5}$$

**Proof.** By (3.3) in Proposition 3.1, we have the expression for the expected value

$$E[N_{i,t}] = E\left[\sum_{j=0}^{\infty} \mathcal{B}_t(p^j) \circ \epsilon_{i,t-j}\right] = \sum_{j=0}^{\infty} E[\mathcal{B}_t(p^j) \circ \epsilon_{i,t-j}], \tag{3.6}$$

and the autocovariance

$$\begin{aligned} Cov[N_{i,t}, N_{i,t-k}] &= Cov\left[\sum_{m=0}^{\infty} \mathcal{B}(p^i) \circ \epsilon_{i,t-m}, \sum_{j=0}^{\infty} \mathcal{B}(p^j) \circ \epsilon_{i,t-k-j}\right] \\ &= Cov\left[\sum_{m=0}^{\infty} \mathcal{B}(p^i) \circ \epsilon_{i,t-m}, \sum_{j=k}^{\infty} \mathcal{B}(p^{j-k}) \circ \epsilon_{i,t-j}\right]. \end{aligned} \tag{3.7}$$

Equation (3.7) can be split into two groups of covariances, one group with two components having the same Poisson variables, and the other group having different Poisson variables. Therefore,

$$\begin{aligned} Cov[N_{i,t}, N_{i,t-k}] &= \sum_{j=k}^{\infty} Cov[\mathcal{B}(p^j) \circ \epsilon_{i,t-j}, \mathcal{B}(p^{j-k}) \circ \epsilon_{i,t-j}] \\ &\quad + \sum_{\substack{m=0,j=k \\ m\neq j}}^{\infty} Cov[\mathcal{B}(p^m) \circ \epsilon_{i,t-m}, \mathcal{B}(p^{j-k}) \circ \epsilon_{i,t-j}]. \end{aligned} \tag{3.8}$$

Note that $\mathcal{B}(p^j) \circ \epsilon_{i,t-j}$ follows a compound Poisson distribution with Bernoulli severity random variables. The following results regarding its mean and variance are straightforward.

$$E[\mathcal{B}(p^j) \circ \epsilon_{i,t-j}] = p^j E[\epsilon_{i,t-j}], \tag{3.9}$$

and

$$Var[\mathcal{B}(p^j) \circ \epsilon_{i,t-j}] = E[\epsilon_{i,t-j}]p^j(1-p^j) + Var[\epsilon_{i,t-j}]p^{2j}. \tag{3.10}$$

Now in order to derive the covariance involved in (3.8), define four non-negative integers $j, k, l$, and $m$ with $m \neq j$, and $k \geq l$. Let $\{U_{m,a}\}$ and $\{U_{j,b}\}$ be two sets of independent Bernoulli random variables with parameters $p^k$ and $p^l$, respectively.

For two same Poisson variables under the thinning operators, the covariance is

$$\begin{aligned} Cov[\mathcal{B}(p^k) \circ \epsilon_{i,t-m}, \mathcal{B}(p^l) \circ \epsilon_{i,t-m}] &= Cov[\mathcal{B}(p^{k-l}) \circ \mathcal{B}(p^l) \circ \epsilon_{i,t-m}, \mathcal{B}(p^l) \circ \epsilon_{i,t-m}] \\ &= p^{k-l} Var[\mathcal{B}(p^l) \circ \epsilon_{i,t-m}] \end{aligned} \tag{3.11}$$

and for two different Poisson variables under the thinning operators, the covariance derived by using the compound Poisson expression is.

$$\begin{aligned} & Cov[\mathcal{B}(p^k) \circ \epsilon_{i,t-m}, \mathcal{B}(p^l) \circ \epsilon_{i,t-j}] \\ =& Cov\left[\sum_{a=1}^{\epsilon_{i,t-m}} U_{m,a}, \sum_{b=1}^{\epsilon_{i,t-j}} U_{j,b}\right] \\ =& E\left[\sum_{a=1}^{\epsilon_{i,t-m}} U_{m,a} \cdot \sum_{b=1}^{\epsilon_{i,t-j}} U_{j,b}\right] - E\left[\sum_{a=1}^{\epsilon_{i,t-m}} U_{m,a}\right] E\left[\sum_{b=1}^{\epsilon_{i,t-j}} U_{j,b}\right] \\ =& E\left[E\left[\sum_{a=1}^{\epsilon_{i,t-m}} U_{m,a} \cdot \sum_{b=1}^{\epsilon_{i,t-j}} U_{j,b} \mid \epsilon_{i,t-m}, \epsilon_{i,t-j}\right]\right] - p^{k+l} E[\epsilon_{i,t-m}] E[\epsilon_{i,t-j}] \\ =& p^{k+l} E[\epsilon_{i,t-m}\epsilon_{i,t-j}] - p^{k+l} E[\epsilon_{i,t-m}] E[\epsilon_{i,t-j}] \\ =& p^{k+l} Cov[\epsilon_{i,t-m}, \epsilon_{i,t-j}]. \end{aligned} \tag{3.12}$$

All of these above combined, bring the moments of the Compound Poisson variables stated in (3.9) and (3.10), and the covariance obtained in (3.11) and (3.12), into (3.6) and (3.8). Then we have

$$E[N_{i,t}] = \sum_{j=0}^{\infty} E[\mathcal{B}_t(p^j) \circ \epsilon_{i,t-j}] = \sum_{j=0}^{\infty} p^j E[\epsilon_{i,t-j}],$$

and

$$\begin{aligned} Cov[N_{i,t}, N_{i,t-k}] &= \sum_{j=k}^{\infty} \left\{ p^j E[\epsilon_{i,t-j}] + p^{2j-k}(V[\epsilon_{i,t-j}] - E[\epsilon_{i,t-j}]) \right\} \\ &+ \sum_{\substack{m=0,j=k \\ m \neq j}}^{\infty} p^{m+j-k} Cov[\epsilon_{i,t-m}, \epsilon_{i,t-j}]. \end{aligned} \tag{3.13}$$

□

Now consider the pure premium for year $T_i+1$ with $T_i$ years' experience claim frequency available. Note that $\mathcal{B}_t(p) \circ N_{i,t-1}$ is the sum of $N_{i,t-1}$ independent and identically distributed Bernoulli random variables with the common parameter $p$, and that $\epsilon_{i,t}$ follows the mixed Poisson distribution with $\Theta_{i,t}$ as the heterogeneity. For year $T_i + 1$, under the Bayesian statistics frame, the pure premium is

$$\begin{aligned} P_{i,T_i+1} &= E[N_{i,T_i+1}|N_{i,1}, \ldots, N_{i,T_i}] \\ &= E[\mathcal{B}_{T_i}(p) \circ N_{i,T_i} + \epsilon_{i,T_i+1}|N_{i,1}, \ldots, N_{i,T_i}] \\ &= E[\mathcal{B}_{T_i}(p) \circ N_{i,T_i}|N_{i,1}, \ldots, N_{i,T_i}] + E[E[\epsilon_{i,T_i+1}|\Theta_{i,T_i+1}, N_{i,1}, \ldots, N_{i,T_i}]] \\ &= pN_{i,T_i} + E[\Theta_{i,T_i+1}\lambda_{i,T_i+1}|N_{i,1}, \ldots, N_{i,T_i}] \\ &= pN_{i,T_i} + \lambda_{i,T_i+1} E[\Theta_{i,T_i+1}|N_{i,1}, \ldots, N_{i,T_i}]. \end{aligned} \tag{3.14}$$

In the following, we first start with the assumption that the heterogeneity for one policyholder $\Theta_{i,t}$ does not vary over time and follows a static random distribution. However, as discussed in Chapter 2 the sequence $\{\Theta_{i,t}\}_{t=1}^{\infty}$ may be different random variables with correlations. Instead of assuming $\Theta_{i,t}$ a constant that follows a static distribution, it is more realistic to apply a dynamic process with certain level of correlation. The INAR(1) model with these two types of heterogeneities for the error terms are examined in the next two sections followed by the associated pricing results. As we concentrate on modeling the heterogeneity, for simplicity, we assume that the *a priori* rate $\lambda_{i,t}$ does not vary over time and is a constant $\lambda_i$.

## 3.2 INAR(1) Process with Static Heterogeneity

In this section, we first assume that the heterogeneity for one particular policyholder $\Theta_{i,t}$ is a time independent random variable and follows a Gamma distribution; thus $\boldsymbol{\Theta}_i$ degenerates to $\Theta_i$. Given $\Theta_i = \theta_i$, the error terms for the INAR(1) process $\{\epsilon_{i,t}\}_{t=1}^{\infty}$ are independent and identically Poisson distributed with parameter $\lambda_i\theta_i$. We further assume $\Theta_i$ follows a Gamma distribution with parameter $a$; then the corresponding INAR(1) process can be described as

$$\begin{aligned} &N_{i,t} = \mathcal{B}_t(p) \circ N_{i,t-1} + \epsilon_{i,t} \\ &\epsilon_{i,t}|\Theta_i = \theta_i \sim \text{Poi}(\theta_i\lambda_i) \\ &\Theta_i \sim \text{Gam}(a, a), \qquad a > 0. \end{aligned} \tag{3.15}$$

As pointed out by Gourieroux and Jasiak (2004), under the assumption of Poisson distributed error terms, the INAR(1) defines a counting process which has a marginal Poisson distribution with a modified parameter $\lambda/(1-p)$, and by setting the parameter $p = 0$ it reduces to the classical mixed Poisson model discussed in Chapter 2. More specific results of this INAR(1) process are obtained below.

**Corollary 3.1.** *With the assumptions listed in (3.15), the conditional mean and autocorrelation of the INAR(1) process are*

$$E[N_{i,t}|\Theta_i = \theta_i] = \frac{\theta_i\lambda_i}{1-p} \tag{3.16}$$

*and*

$$Cov[N_{i,t}, N_{i,t-k}|\Theta_i = \theta_i] = \frac{p^k}{1-p}\theta_i\lambda_i. \tag{3.17}$$

**Proof.** By equation (3.4) in Proposition 3.2,

$$E[N_{i,t}|\Theta_i = \theta_i] = \sum_{j=0}^{\infty} p^j\theta_i\lambda_i = \frac{\theta_i\lambda_i}{1-p}.$$

Given $\Theta_i = \theta_i$, the Poisson error terms are naturally independent of, and the thinning operators are independent of all the error terms. The error terms conditioning on $\Theta_i = \theta_i$ follow Poisson distribution with parameter $\theta_i\lambda_i$. Therefore, the second summation of equation (3.5) equals zero, and

$$\begin{aligned} Cov[N_{i,t}, N_{i,t-k}|\Theta_i = \theta_i] =& p^k(Var[\epsilon_{i,t}|\Theta_i = \theta_i] - E[\epsilon_{i,t}|\Theta_i = \theta_i]) \sum_{j=k}^{\infty} p^{2(j-k)} \\ &+ p^k E[\epsilon_{i,t}|\Theta_i = \theta_i] \sum_{j=k}^{\infty} p^{j-k} \\ =& \frac{p^k}{1-p}\theta_i\lambda_i \end{aligned}$$

□

For the pricing formulas with the assumptions listed in (3.15), for various scenarios where different numbers of years' experience are available, Gourieroux and Jasiak (2004) derived the following results and stated that this model addresses one major deficiency of the classical mixed Poisson model. With the classical Poisson-Gamma model, the past numbers of claims affect the premium only by the total number of claims for all the past years, whereas in this model, the arrival time of claims matters, and the earlier a claim arrives, the less it would impact the premium. The resulting posterior distribution of the random effect is a mixture of gamma distributions with weights and degrees of freedom depending on the claim history. Their results are listed below as a proposition. For more detailed proof, please refer to Gourieroux and Jasiak (2004).

**Proposition 3.3.** *The pure premium for year $T_i + 1$ is*

$$P_{i,T_i+1} = pN_{i,T_i} + \lambda E[\Theta_i | N_{i,1}, \ldots, N_{i,T_i}].$$

*i) For $T_i = 0$, the conditional distribution of $\Theta_i$ is* $\mathrm{Gam}(a, a)$.

*ii) For $T_i = 1$, the conditional distribution of $\Theta_i$ given $N_{i,1}$ is*

$$\mathrm{Gam}\left(a + n_{i,1}, a + \frac{\lambda_i}{1-p}\right).$$

*iii) For $T_i = 2$, the conditional distribution of $\Theta_i$, given $N_{i,1}, N_{i,2}$ is*

$$\frac{\sum_{z_2=0}^{min(n_{i,1},n_{i,2})} \pi(z_2, n_{i,1}, n_{i,2}) \gamma(a + n_{i,1} + n_{i,2} - z_2, a + \lambda_i + \frac{\lambda_i}{1-p})}{\sum_{z_2=0}^{min(n_{i,1},n_{i,2})} \pi(z_2, n_{i,1}, n_{i,2})},$$

*where*

$$\pi(z_2, n_{i,1}, n_{i,2}) = C_{n_{i,1}}^{z_2} \left(\frac{p}{1-p}\right)^{z_2} \frac{1}{\lambda_i^{z_2}} \frac{1}{(y_2 - z_2)!} \frac{\Gamma(a + n_{i,1} + n_{i,2} - z_2)}{[a + \lambda_i + \frac{\lambda_i}{1-p}]^{a+n_{i,1}+n_{i,2}-z_2}}$$

*iv) For $T_i = 3$, the conditional distribution of $\Theta_i$ given $N_{i,1}, N_{i,2}, N_{i3}$ is*

$$\frac{\sum_{z_3=0}^{min(n_{i,2},n_{i,3})} \sum_{z_2=0}^{min(n_{i,1},n_{i,2})} \pi(z_2, z_3, n_{i,1}, n_{i,2}, n_{i,3}) \gamma(a + n_{i,1} + n_{i,2} - z_2 - z_3, a + 2\lambda + \frac{\lambda}{1-p})}{\sum_{z_3=0}^{min(n_{i,2},n_{i,3})} \sum_{z_2=0}^{min(n_{i,1},n_{i,2})} \pi(z_2, z_3, n_{i,1}, n_{i,2}, n_{i,3})},$$

*where*

$$\pi(z_2, z_3, n_{i,1}, n_{i,2}, n_{i,3}) = C_{n_{i,2}}^{z_3} \quad C_{n_{i,1}}^{z_2} \left(\frac{p}{1-p}\right)^{z_2+z_3} \frac{1}{\lambda^{z_2+z_3}} \frac{1}{(y_2-z_2)!} \frac{1}{(y_3-z_3)!} \\ \times \quad \frac{\Gamma(a+n_{i,1}+n_{i,2}+n_{i3}-z_2-z_3)}{[a+2\lambda_i+\frac{\lambda_i}{1-p}]^{a+n_{i,1}+n_{i,2}+n_{i,3}-z_2-z_3}}.$$

*The estimate of the random effect and the pure premium follow directly from (3.14) as follows:*

*i) For* $T_i = 0$, $\hat{\Theta}_i = E(\Theta_i) = 1$, $P_1 = \lambda_i$.

*ii) For* $T_i = 1$, $\hat{\Theta}_i = E[\Theta|n_{i,1}] = \frac{a+n_{i,1}}{a+\frac{\lambda_i}{1-p}}$, *and* $P_2 = pn_{i,1} + \lambda_i \frac{a+n_{i,1}}{a+\frac{\lambda_i}{1-p}}$.

*iii) For* $T_i = 2$, $P_3 = pn_{i,2} + \lambda_i \hat{\Theta}_i$, *where*

$$\hat{\Theta}_i = \frac{\sum_{z_2=0}^{min(n_{i,1},n_{i,2})} \pi(z_2, n_{i,1}, n_{i,2}) \frac{a+n_{i,1}+n_{i,2}-z_2}{a+\lambda_i+\frac{\lambda_i}{1-p}}}{\sum_{z_2=0}^{min(n_{i,1},n_{i,2})} \pi(z_2, n_{i,1}, n_{i,2})}.$$

*iv) For* $T_i = 3$, $P_4 = pn_{i3} + \lambda_i \hat{\Theta}_i$, *where*

$$\hat{\Theta}_i = \frac{\sum_{z_3=0}^{min(n_{i,2},n_{i,3})} \sum_{z_2=0}^{min(n_{i,1},n_{i,2})} \pi(z_2, z_3, n_{i,1}, n_{i,2}, n_{i,3}) \frac{a+n_{i,1}+n_{i,2}-z_2-z_3}{a+2\lambda_i+\frac{\lambda_i}{1-p}}}{\sum_{z_3=0}^{min(n_{i,2},n_{i,3})} \sum_{z_2=0}^{min(n_{i,1},n_{i,2})} \pi(z_2, z_3, n_{i,1}, n_{i,2}, n_{i,3})}.$$

## 3.3 INAR(1) Process with Dynamic Heterogeneity

It is natural for the heterogeneity $\{\Theta_{i,t}\}_{t=1}^{\infty}$ to vary from year to year and to be correlated with each other. Therefore, in this section, a time series model is applied to model $\{\Theta_{i,t}\}_{t=1}^{\infty}$. The INAR(1) model with the heterogeneity embedded in the error terms is discussed in the form of the Generalized Linear Model (GLM). At the end of this section, the mean, variance, and pricing formulas of the underlying INAR(1) process are derived.

### 3.3.1 Dynamic Heterogeneity

As introduced in chapter 2, $\mathbf{\Theta}_i = (\Theta_{i,1}, \ldots, \Theta_{i,T_i})$ represents the heterogeneity for policy $i$ over $T_i$ years. We assume $\{\Theta_{i,1}\}_{t=1}^{\infty}$ on the logarithm scale to be a first order autoregressive (AR(1)) process, and $E[\Theta_{i,t}] = 1$ for all $t > 0$. That is,

$$\ln \Theta_{i,t} = \rho \ln \Theta_{i,t-1} + \varepsilon_{i,t}, \qquad t = 2, 3, \ldots, \tag{3.18}$$

where $0 < \rho < 1$ reflects how strong the heterogeneities for different years are correlated, and $\{\varepsilon_{i,t}\}$ are independent Gaussian errors, for which the mean and variance are chosen to be $(\rho - 1)\frac{\sigma^2}{2}$ and $\sigma^2(1 - \varrho^2)$, respectively, to fulfill the financial balance requirement on $\{\Theta_{i,t}\}_{t=0}^{\infty}$. The AR(1) model in (3.18) implies that $\{\Theta_{i,t}\}_{t=0}^{\infty}$ follow a Lognormal distribution with density function given by (2.8), and that $(\Theta_{i,1}, \Theta_{i,2}, \ldots, \Theta_{i,T_i})$ is a random vector following a multivariate Lognormal distribution with density function

$$f(\theta_{i,1}, \cdots, \theta_{i,T_i}) = \frac{1}{(2\pi)^{\frac{T_i}{2}} \sigma |\mathbf{\Sigma}_i|^{\frac{1}{2}} \theta_{i,1} \cdots \theta_{i,T_i}} \exp\left\{ -\frac{1}{2\sigma^2} (\ln \boldsymbol{\theta}_i - \boldsymbol{\mu})' \mathbf{\Sigma}_i^{-1} (\ln \boldsymbol{\theta}_i - \boldsymbol{\mu}) \right\}, \tag{3.19}$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} -\frac{\sigma^2}{2} \\ \vdots \\ -\frac{\sigma^2}{2} \end{pmatrix}_{T_i \times 1}, \qquad \mathbf{\Sigma}_i = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T_i-1} \\ \rho & 1 & \rho & \cdots & \rho^{T_i-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T_i-1} & \rho^{T_i-2} & \rho^{T_i-3} & \cdots & 1 \end{pmatrix}_{T_i \times T_i} \tag{3.20}$$

are the mean vector and correlation matrix of $\ln \mathbf{\Theta}_i$, respectively.

This approach has been followed by Chan and Ledolter (1995) in analyzing the well-known polyo incidence data first studied by Zeger (1988). Pinquet et al. (2001) explored its application in actuarial science, followed by Brouhns et al. (2003) discussing a bonus-malus scale based on this model, in which the relativities were calculated by simulation.

### 3.3.2 Generalized Linear Model Interpretation

The error terms of the INAR(1) process $\{\epsilon_{i,t}\}_{t=1}^{\infty}$ with the dynamic heterogeneity embedded has the nature of a generalized Linear Model (GLM). The research on Bonus-Malus systems based on the INAR(1) process benefits from adopting the existing theorems in the context of GLM, such as the generalized estimating equation (GEE) to estimate the involved parameters, and various correlation structures to reflect the heterogeneity. In this section, we discuss the process in the GLM form.

Recall that, in (1.1), the *a priori* rate $\lambda_{i,t}$ is determined by all the observable factors with a linear regression model. We assume that there are $q$ observable variables. Let

$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)'$ denote the regression coefficients, and $\mathbf{x}_{i,t} = (x^1_{i,t}, x^2_{i,t}, \ldots, x^q_{i,t})'$ denote the values of the $q$ observable factors for policy $i$ in year $t$. Define $e_{i,t} = \ln \Theta_{i,t}$ as a Gaussian error with mean $-\frac{\sigma^2}{2}$ and variance $\sigma^2$. The mean of the error term of the INAR(1) process in year t can be written as

$$\begin{aligned} E[\epsilon_{i,t}|\mathbf{x}_{i,t}] &= \lambda_{i,t}\Theta_{i,t} \\ &= \exp(\boldsymbol{\beta}'\mathbf{x}_{i,t})\exp(e_{i,t}) \\ &= \exp(\boldsymbol{\beta}'\mathbf{x}_{i,t} + e_{i,t}), \qquad t = 1, 2, \ldots, T_i. \end{aligned} \tag{3.21}$$

Further let

$$\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \\ \vdots \\ \epsilon_{i,T_i} \end{pmatrix}, \quad \mathbf{x}_i = \begin{pmatrix} x^1_{i,1} & x^1_{i,2} & \cdots & x^1_{i,T_i} \\ x^2_{i,1} & x^2_{i,2} & \cdots & x^2_{i,T_i} \\ \vdots & \vdots & \ddots & \vdots \\ x^q_{i,1} & x^q_{i,2} & \cdots & x^q_{i,T_i} \end{pmatrix}, \quad \mathbf{e}_i = \begin{pmatrix} e_{i,1} \\ e_{i,2} \\ \vdots \\ e_{i,T_i} \end{pmatrix}.$$

The mean of the error terms of the INAR(1) process for $T_i$ years as an vector can be expressed as

$$E[\boldsymbol{\epsilon}_i|\mathbf{x}_i] = \exp\{\boldsymbol{\beta}'\mathbf{x}_i + \mathbf{e}_i\}. \tag{3.22}$$

which is the typical expression of a generalized linear model with exponential link function. The vector of Gaussian errors $\mathbf{e}_i$ follows the multivariate normal distribution. The density function is

$$f(e_{i,1}, \cdots, e_{i,T_i}) = \frac{1}{(2\pi)^{\frac{T_i}{2}}\sigma|\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{e}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}_i^{-1}(\mathbf{e}_i - \boldsymbol{\mu})\right\}$$

where the covariance matrix $\boldsymbol{\Sigma}_i$ is defined in (3.20).

As an advantage of the GLM form, GLM accommodates a big range of correlation structures for variables being analyzed. $\boldsymbol{\Sigma}_i$ defined by (3.20) is called AR(1) correlation structure, which results in the correlation between the heterogeneities for different years to exponentially decrease as the time lag increases. The correlation structure can also be designed in other types. Some commonly used ones include (correlation matrices are as below)

$$\begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & \cdots & 0 \\ 0 & \rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Compound Symmetry One-Dependent

$$\begin{pmatrix} 1 & \rho^{|t_{i,1}-t_{i,2}|} & \rho^{|t_{i,1}-t_{i,3}|} & \dots & \rho^{|t_{i,1}-t_{i,n}|} \\ \rho^{|t_{i,2}-t_{i,1}|} & 1 & \rho^{|t_{i,2}-t_{i,3}|} & \dots & \rho^{|t_{i,2}-t_{i,n}|} \\ \rho^{|t_{i,3}-t_{i,1}|} & \rho^{|t_{i,3}-t_{i,2}|} & 1 & \dots & \rho^{|t_{i,3}-t_{i,n}|} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{|t_{i,n}-t_{i,1}|} & \rho^{|t_{i,n}-t_{i,2}|} & \rho^{|t_{i,n}-t_{i,3}|} & \dots & 1 \end{pmatrix}$$

Markov

- Compound Symmetry Structure: equal correlation is imposed between observations at all time points.
- One-Dependent Structure: Correlation is expected to be largest in magnitude among observations that are adjacent in time; relative to this magnitude, correlation between non-consecutive observations may be negligible[1].
- Markov Structure: Adjusted to reflect the correlation for observations with uneven time intervals, based on AR(1) structure. Correlation decreases with the time lag.

For details on the correlation structure design, refer to Laird and Ware (2004, Chapter 7).

### 3.3.3 INAR(1) Model with Dynamic Heterogeneity

Under the assumption of the AR(1) correlation structure, the error terms $\{\epsilon_{i,t}\}_{t=1}^{\infty}$ incorporate much more complex correlations, which also contribute to the total auto-correlation of the INAR(1) model for the claim counts in the consecutive policy years. Starting with probing the correlation between the error terms $\{\epsilon_{i,t}\}_{t=1}^{\infty}$, we derive the auto-correlation of $\{N_{i,t}\}_{t=1}^{\infty}$ below.

[1]Although theoretically this structure cannot be a correlation matrix, it is artificially designed this way to simplify analysis when $\rho$ is small.

**Proposition 3.4.** *Let $h$ be a positive integer, the mean and covariance of the sequence of the error terms $\{\epsilon_{i,t}\}_{t=1}^{\infty}$ are*

$$E[\epsilon_{i,t}] = \lambda_{i,t}, \tag{3.23}$$

$$V[\epsilon_{i,t}] = \lambda_{i,t}[1 + \lambda_{i,t}\,\mathrm{e}^{\sigma^2} - 1)], \tag{3.24}$$

$$Cov[\epsilon_{i,t}, \epsilon_{i,t+h}] = \lambda_{i,t}\lambda_{i,t+h}(\mathrm{e}^{\sigma^2\rho^h} - 1). \tag{3.25}$$

**Proof.** By the assumptions introduced in the previous section, we know that the Gaussian errors for the generalized linear model $e_{i,t} = \ln \Theta_{i,t}$, $t = 1, 2, \ldots$, follow a normal distribution with

$$E[e_{i,t}] = -\frac{\sigma^2}{2}, \quad V[e_{i,t}] = \sigma^2, \quad \text{and} \quad Cov[e_{i,t}, e_{i,t+h}] = \sigma^2\rho^h.$$

Therefore, by the properties of Lognormal distribution,

$$E[\Theta_{i,t}] = \mathrm{e}^{E[e_{i,t}]+\frac{1}{2}V[e_{i,t}]} = \mathrm{e}^{-\frac{\sigma^2}{2}+\frac{1}{2}\sigma^2} = 1,$$
$$V[\Theta_{i,t}] = E[\Theta_{i,t}^2] - E[\Theta_{i,t}]^2 = \mathrm{e}^{2E[e_{i,t}]+2V[e_{i,t}]} - 1 = \mathrm{e}^{\sigma^2} - 1,$$

and

$$Cov[\Theta_{i,t}, \Theta_{i,t+h}] = \mathrm{e}^{\sigma^2\rho^h} - 1,$$

then we have

$$E[\epsilon_{i,t}] = \lambda_{i,t}E[\Theta_{i,t}] = \lambda_{i,t},$$

$$\begin{aligned} V[\epsilon_{i,t}] &= E[V[\epsilon_{i,t}|\Theta_{i,t}]] + V[E[\epsilon_{i,t}|\Theta_{i,t}]] \\ &= \lambda_{i,t} + \lambda_{i,t}^2 V[\Theta_{i,t}] = \lambda_{i,t}[1 + \lambda_{i,t}(\mathrm{e}^{\sigma^2} - 1)], \end{aligned}$$

and

$$\begin{aligned} Cov[\epsilon_{i,t}, \epsilon_{i,t+h}] &= E[Cov[\epsilon_{i,t}, \epsilon_{i,t+h}|\mathbf{\Theta}_i]] + Cov[E[\epsilon_{i,t}|\mathbf{\Theta}_i], E[\epsilon_{i,t+h}|\mathbf{\Theta}_i]] \\ &= \lambda_{i,t}\lambda_{i,t+h}Cov[\Theta_{i,t}, \Theta_{i,t+h}] \\ &= \lambda_{i,t}\lambda_{i,t+h}(\mathrm{e}^{\sigma^2\rho^h} - 1). \end{aligned}$$

□

By applying the results of Proposition 3.4 into (3.5), more specific results of the autocovariance of the lagged claim counts can be obtained. For simplicity, we assume the *a priori* average $\lambda_{i,t}$ does not change over time, i.e., $\lambda_{i,t} = \lambda_i$, for any $t > 0$. Then, we have

$$
\begin{aligned}
Cov[N_{i,t}, N_{i,t-k}] =& \sum_{j=k}^{\infty} \{p^j E[\epsilon_{i,t-j}] + p^{2j-k}(V[\epsilon_{i,t-j}] - E[\epsilon_{i,t-j}])\} \\
&+ \sum_{\substack{m=0,j=k \\ j\neq m}}^{\infty} p^{j+m-k} Cov[\epsilon_{i,t-m}, \epsilon_{i,t-j}] \\
=& \frac{p^k}{1-p}\lambda_i + \frac{p^k}{1-p^2}\lambda_i^2(\mathrm{e}^{\sigma^2} - 1) \\
&+ \lambda_i^2 \sum_{\substack{m=0,j=k \\ j\neq m}}^{\infty} p^{j+m-k}(\mathrm{e}^{\sigma^2\rho^{|j-m|}} - 1).
\end{aligned}
\tag{3.26}
$$

It is easy to prove that the summation involved in the expression above converges. The auto-correlation with time lag 0 to 5 are calculated in Table 3.1, which shows a descending trend in k. The parameters have been chosen as $\sigma = 1.4632$, $\rho = 0.5499$, $p = 0.3$, and $\lambda = 0.3$, to be consistent with the numerical results in Chapter 4. From (3.26), we know that the variance of $N_{i,t}'s$ is same for all $t > 0$, which implies that the correlation of the lagged claim counts $Cor[N_{i,t}, N_{i,t-k}] = Cov[N_{it}, N_{i,t-k}]/\sqrt{V[N_{i,t}]V[N_{i,t-k}]}$ decreases with $k$, i.e. the predictive ability of a claim decreases with its age.

**Table 3.1:** *Auto-correlation of claim counts with different time lags*

| | Time lag $k$ | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| $Cor_k$ | 1 | 0.402 | 0.2226 | 0.1688 | 0.1526 | 0.1478 |

To determine the pure premium with $T_i$ years of experience, we continue the work under the Bayesian framework and use the quadratic loss function. The pricing formula given by (3.14) still holds, for which we will derive the posterior mean of $\Theta_{i,T_i+1}$ given $N_{i,1}, N_{i,2}, \ldots, N_{i,T_i}$ in the following proposition.

**Proposition 3.5.** *The predicted heterogeneity for policyholder $i$ in year $T_i + 1$, $\hat{\Theta}_{i,T_i+1}$, is*

$$
E[\Theta_{i,T_i+1}|N_{i,1}, \ldots, N_{i,T_i}] = \frac{\int \cdots \int \theta_{i,T_i+1} f(n_{i,1}, \ldots, n_{i,T_i}, \theta_{i,1}, \ldots, \theta_{i,T_i+1}) d\theta_{i,1} \cdots d\theta_{i,T_i+1}}{\int \cdots \int f(n_{i,1}, \ldots, n_{i,T_i}, \theta_{i,1}, \ldots, \theta_{i,T_i}) d\theta_{i,1} \cdots d\theta_{i,T_i}},
\tag{3.27}
$$

*where*

*i) when* $T_i = 1$

$$f(n_{i,1},\theta_{i,1},\theta_{i,2}) = \mathrm{e}^{-\frac{\lambda_i\theta_{i,1}}{1-p}} \frac{\left(\frac{\lambda_i\theta_{i,1}}{1-p}\right)^{n_{i,1}}}{n_{i,1}!} f(\theta_{i,1},\theta_{i,2})$$

*and*

$$f(n_{i,1},\theta_{i,1}) = \mathrm{e}^{-\frac{\lambda_i\theta_{i,1}}{1-p}} \frac{\left(\frac{\lambda_i\theta_{i,1}}{1-p}\right)^{n_{i,1}}}{n_{i,1}!} f(\theta_{i,1}),$$

*ii) when* $T_i \geq 2$

$$\begin{aligned} & f(n_{i,1},\ldots,n_{i,T_i},\theta_{i,1},\ldots,\theta_{i,T_i+1}) \\ =& f(\theta_{i,1},\ldots,\theta_{i,T_i+1})\, \mathrm{e}^{-\frac{\lambda_i\theta_{i,1}}{1-p}} \frac{\left(\frac{\lambda_i\theta_{i,1}}{1-p}\right)^{n_{i,1}}}{n_{i,1}!} \\ & \prod_{j=2}^{T_i}\left(\sum_{z_j=0}^{\min(n_{i,j},n_{i,j-1})} \binom{n_{i,j-1}}{z_j} p^{z_j}(1-p)^{n_{i,j-1}-z_j}\,\mathrm{e}^{-\lambda_i\theta_{i,j}} \frac{(\lambda_i\theta_{i,j})^{n_{i,j}-z_j}}{(n_{i,j}-z_j)!}\right) \end{aligned} \tag{3.28}$$

*and*

$$\begin{aligned} & f(n_{i,1},\ldots,n_{i,T_i},\theta_{i,1},\ldots,\theta_{i,T_i}) \\ =& f(\theta_{i,1},\ldots,\theta_{i,T_i})\, \mathrm{e}^{-\frac{\lambda_i\theta_{i,1}}{1-p}} \frac{\left(\frac{\lambda_i\theta_{i,1}}{1-p}\right)^{n_{i,1}}}{n_{i,1}!} \\ & \prod_{j=2}^{T_i}\left(\sum_{z_j=0}^{\min(n_{i,j},n_{i,j-1})} \binom{n_{i,j-1}}{z_j} p^{z_j}(1-p)^{n_{i,j-1}-z_j}\,\mathrm{e}^{-\lambda_i\theta_{i,j}} \frac{(\lambda_i\theta_{i,j})^{n_{i,j}-z_j}}{(n_{i,j}-z_j)!}\right) \end{aligned} \tag{3.29}$$

*with* $f(\theta_{i,1},\ldots,\theta_{i,T_i+1})$ *being the density function of a* $(T_i+1)$*-dimension multivariate Log-normal distribution.*

**Proof.** By the definition of conditional expected value we have

$$\begin{aligned} E[\Theta_{i,T_i+1}|N_{i,1},\ldots,N_{i,T_i}] &= \textstyle\int \theta_{i,T_i+1} f(\theta_{i,T_i+1}|n_{i,1},\ldots,n_{i,T_i})d\theta_{i,T_i+1} \\ &= \textstyle\int \theta_{i,T_i+1}\frac{f(\theta_{i,T_i+1},n_{i,1},\ldots,n_{i,T_i})}{f(n_{i,1},\ldots,n_{i,T_i})}d\theta_{i,T_i+1} \\ &= \frac{\int\cdots\int \theta_{i,T_i+1} f(n_{i,1},\ldots,n_{i,T_i},\theta_{i,1},\ldots,\theta_{i,T_i+1})d\theta_{i,1}\cdots d\theta_{i,T_i+1}}{\int\cdots\int f(n_{i,1},\ldots,n_{i,T_i},\theta_{i,1},\ldots,\theta_{i,T_i})d\theta_{i,1}\cdots d\theta_{i,T_i}}. \end{aligned} \tag{3.30}$$

i) When $T_i = 1$, the distribution of $N_{i,t}$ given the heterogeneity $\Theta_{i,1} = \theta_{i,1}$ for the first year, is a Poisson distribution with parameter $\frac{\lambda_i\theta_{i,1}}{1-p}$. Therefore,

$$f(n_{i,1},\theta_{i,1},\theta_{i,2}) = f(n_{i,1}|\theta_{i,1},\theta_{i,2})f(\theta_{i,1},\theta_{i,2}) = \mathrm{e}^{-\frac{\lambda_i\theta_{i,1}}{1-p}} \frac{\left(\frac{\lambda_i\theta_{i,1}}{1-p}\right)^{n_{i,1}}}{n_{i,1}!} f(\theta_{i,1},\theta_{i,2}),$$

and

$$f(n_{i,1},\theta_{i,1}) = f(n_{i,1}|\theta_{i,1})f(\theta_{i,1}) = \mathrm{e}^{-\frac{\lambda_i\theta_{i,1}}{1-p}}\frac{\left(\frac{\lambda_i\theta_{i,1}}{1-p}\right)^{n_{i,1}}}{n_{i,1}!}f(\theta_{i,1}).$$

ii) When $T_i \geq 2$, the joint distribution is the product of a series of conditional distributions and a Lognormal density function.

$$\begin{aligned}
&f(n_{i,1},\dots,n_{i,T_i},\theta_{i,1},\dots,\theta_{i,T_i+1})\\
=&f(n_{i,T_i}|n_{i,1},\dots,n_{i,T_i-1},\theta_{i,1},\dots,\theta_{i,T_i+1})\\
&f(n_{i,T_i-1}|n_{i,1},\dots,n_{i,T_i-2},\theta_{i,1},\dots,\theta_{i,T_i+1})\\
&\dots\\
&f(n_{i,1}|\theta_{i,1},\dots,\theta_{i,T_i+1})f(\theta_{i,1},\dots,\theta_{i,T_i+1})
\end{aligned} \tag{3.31}$$

The conditional distribution functions involved in the formula above can be written as follows.

The first year claim count given the heterogeneity $\Theta_{i,1} = \theta_{i,1}$ has Poisson mass probability function with parameter $\frac{\lambda_i\theta_{i,1}}{1-p}$, so that

$$f(n_{i,1}|\theta_{i,1},\dots,\theta_{T_i+1}) = \mathrm{e}^{-\frac{\lambda_i\theta_{i,1}}{1-p}}\frac{\left(\frac{\lambda_i\theta_{i,1}}{1-p}\right)^{n_{i,1}}}{n_{i,1}!} \tag{3.32}$$

Note that $f(n_{i,j}|n_{i,1},\dots,n_{i,j-1},\theta_{i,1},\dots,\theta_{i,T_i+1})$, $j = 2,\dots,T_i$, is the density function of the sum of a Poisson variable and a Binomial variable, which could be expressed as convolutions. Therefore,

$$\begin{aligned}
&f(n_{i,j}|n_{i,1},\dots,n_{i,j-1},\theta_{i,1},\dots,\theta_{i,T_i+1})\\
=&\sum_{z_j=0}^{\min(n_{i,j},n_{i,j-1})}\binom{n_{i,j-1}}{z_j}p^{z_j}(1-p)^{n_{i,j-1}-z_j}\,\mathrm{e}^{-\lambda_j\theta_j}\,\frac{(\lambda_i\theta_{i,j})^{n_{i,j}-z_j}}{(n_{i,j}-z_j)!}.
\end{aligned} \tag{3.33}$$

(3.28) and (3.29) can be obtained by applying (3.32) and (3.32) into (3.31)

□

With the results given by Proposition 3.5, the updated premiums can be calculated by

$$P_{T_i+1} = pN_{i,T_i} + \lambda_i E[\Theta_{i,T_i+1}|N_{i,1},\dots,N_{i,T_i}].$$

The explicit results of the pricing formula involves high dimensional integrations and a large amount of calculation. In fact, an efficient numerical algorithm for the integration would be appealing. In the next chapter, Gibb's sampling method is used to obtain numerical results.

# Chapter 4

# Numerical Illustration

In the previous two chapters, we considered claims count models and their associated pricing techniques with various assumptions for the heterogeneity. In Chapter 3, an integer-valued autoregressive process is introduced with dynamic heterogeneity embedded in the error terms. The pricing formulas are derived in Proposition 3.5. The problem becomes how to evaluate the multi-dimensional integrations involved in the pricing formula (3.27).

For convenience, let us restate the integrands on the numerator and denominator in (3.27). For the numerator, from (3.28), the integrand can be written as

$$
\begin{aligned}
&\theta_{i,T_i+1} f(n_{i,1},\ldots,n_{i,T_i},\theta_{i,1},\ldots,\theta_{i,T_i+1})\\
=&f(\theta_{i,1},\ldots,\theta_{i,T_i+1})\theta_{i,T_i+1}\,\mathrm{e}^{-\frac{\lambda_i\theta_{i,1}}{1-p}}\,\frac{\left(\frac{\lambda_i\theta_{i,1}}{1-p}\right)^{n_{i,1}}}{n_{i,1}!}\\
&\prod_{j=2}^{T_i}\left(\sum_{z_j=0}^{\min(n_{i,j},n_{i,j-1})}\begin{pmatrix} n_{i,j-1}\\ z_j\end{pmatrix} p^{z_j}(1-p)^{n_{i,j-1}-z_j}\,\mathrm{e}^{-\lambda_i\theta_{i,j}}\,\frac{(\lambda_i\theta_{i,j})^{n_{i,j}-z_j}}{(n_{i,j}-z_j)!}\right)\\
=&f(\theta_{i,1},\ldots,\theta_{i,T_i+1})g_1(\theta_{i,1},\ldots,\theta_{i,T_i+1};n_{i,1},\ldots,n_{i,T_i}).
\end{aligned}
$$

For the denominator, by (3.29), the integrand can be expressed as

$$
\begin{aligned}
& f(n_{i,1},\ldots,n_{i,T_i},\theta_{i,1},\ldots,\theta_{i,T_i}) \\
=& f(\theta_{i,1},\ldots,\theta_{i,T_i})\,\mathrm{e}^{-\frac{\lambda_i\theta_{i,1}}{1-p}}\,\frac{\left(\frac{\lambda_i\theta_{i,1}}{1-p}\right)^{n_{i,1}}}{n_{i,1}!} \\
& \prod_{j=2}^{T_i}\left(\sum_{z_j=0}^{\min(n_{i,j},n_{i,j-1})}\binom{n_{i,j-1}}{z_j}p^{z_j}(1-p)^{n_{i,j-1}-z_j}\,\mathrm{e}^{-\lambda_i\theta_{i,j}}\,\frac{(\lambda_i\theta_{i,j})^{n_{i,j}-z_j}}{(n_{i,j}-z_j)!}\right) \\
=& f(\theta_{i,1},\ldots,\theta_{i,T_i})g_2(\theta_{i,1},\ldots,\theta_{i,T_i};n_{i,1},\ldots,n_{i,T_i}).
\end{aligned}
$$

Therefore, the pricing formula is the ratio of two expected values of functions of the random variables $g_1$ and $g_2$, respectively, which can be expressed as

$$
\frac{E[g_1(\Theta_{i,1},\ldots,\Theta_{i,T_i+1};n_{i,1},\ldots,n_{i,T_i})]}{E[g_2(\Theta_{i,1},\ldots,\Theta_{i,T_i};n_{i,1},\ldots,n_{i,T_i})]},
$$

where $\boldsymbol{\Theta}_i = (\Theta_{i,1},\ldots,\Theta_{i,T_i+1})$ is a random vector which, under the assumptions given in Section 3.3.3, follows the multivariate Lognormal distribution with density function given by (3.19). The integrations have no closed form solutions, and it is quite time consuming trying to solve them with numerical calculation packages such as SAS or Matlab. The simulated average can be obtained by sampling $\boldsymbol{\Theta}_i$. With the Gibb's Sampler presented in Section 4.1, the complicated calculations can be avoided and replaced by simpler ones, and therefore the process of simulation became fairly time efficient.

## 4.1 Gibb's Sampler

The Gibb's sampler is a computer-intensive algorithm that has been found useful in many practical problems in statistics. While applications of the Gibb's sampler have been found mostly in Bayesian models, it is also extremely useful in the classical (likelihood) calculations, as pointed out in Casella and George (1992). The authors also explained the mechanism of the scheme that drives the sampling process using elementary properties of Markov Chains.

### The Algorithm

We are going to illustrate the Gibb's sampling algorithm with the joint distribution of the multivariate Lognormal distribution $f(\theta_{i,1},\ldots,\theta_{i,T_i})$, where the random variables $\{\Theta_{i,t}\}_{t=1}^{T_i}$

denote the heterogeneity for one policyholder for $T_i$ policy years. Suppose $M$ iterations are going to be obtained, and let $\boldsymbol{\theta}_i^{(j)} = (\theta_{i,1}^{(j)}, \ldots, \theta_{i,T_i}^{(j)})$ denote the values for the $j^{th}$ iteration. Then the sampling procedure is described as follows.

i) Specify the initial value of $\theta_{i,1}^{(0)}, \ldots, \theta_{i,T_i-1}^{(0)}$.

ii) Repeat for $j = 0, 1, \ldots, M-1$,

generate $\theta_{i,1}^{(j+1)}$ from $f(\theta_{i,1}|\theta_{i,2}^{(j)}, \ldots, \theta_{i,T_i}^{(j)})$,

generate $\theta_{i,2}^{(j+1)}$ from $f(\theta_{i,2}|\theta_{i,1}^{(j+1)}, \theta_{i3}^{(j)}, \ldots, \theta_{i,T_i}^{(j)})$,

$\vdots$

generate $\theta_{i,T_i}^{(j+1)}$ from $f(\theta_{i,T_i}|\theta_{i,1}^{(j+1)}, \ldots, \theta_{i,T_i-1}^{(j+1)})$.

iii) Return the values $\{\boldsymbol{\theta}_i^{(1)}, \boldsymbol{\theta}_i^{(2)}, \ldots, \boldsymbol{\theta}_i^{(M)}\}$.

## 4.2 Numerical Illustration

With the simulation procedure described above, we generate $100,000$ four-dimensional random vectors that follow the multivariate Lognormal distribution with $\sigma = 1.4632$, and $\rho = 0.5499$, which were estimated by Brouhns et al. (2003) based on the data set of a Belgium automobile insurance portfolio. The values of other parameters are chosen to be $p = 0.3$ and $\lambda = 0.3$ for comparison between different claim patterns. The sensitivity of the results against $p$ and $\lambda$ is discussed in the later section. We consider the scenarios where three claims were reported during a period of three years. The arrival time of the claims varies in ten different ways, as shown by the first column of Table 4.1 (the three integers in the brackets, separated by commas, orderly denote the numbers of claims reported in years 1, 2, and 3). The predicted heterogeneity and the resulting premium at the beginning of years 2, 3, and 4 are calculated for these different patterns of claims and comparisons are made between various models.

### 4.2.1 Predicted Heterogeneity

Table 4.1 shows the predicted heterogeneity at the beginning of years 2, 3, and 4, denoted by $\hat{\Theta}_{i,2} = E[\Theta_{i,2}|N_{i,1}]$, $\hat{\Theta}_{i,3} = E[\Theta_{i,3}|N_{i,1}, N_{i,2}]$, and $\hat{\Theta}_{i,4} = E[\Theta_{i,3}|N_{i,1}, N_{i,2}, N_{i,3}]$, respectively. The rows are sorted by $\hat{\Theta}_{i,4}$ in a descending order. The same results are also graphed in

Figures 4.1, 4.2, and 4.3. For convenience, the formula of the INAR(1) procedure is restated as

$$N_{i,t} = \mathcal{B}_t(p) \circ N_{i,t-1} + \epsilon_{i,t}.$$

The error term serves as an adjustment to the random portion of the number of reported claims in the previous year when the number of claims for the following year is projected.

**Table 4.1:** *Predicted heterogeneity with various patterns of the past numbers of claims*

| Claim History | $\hat{\Theta}_{i,2}$ | $\hat{\Theta}_{i3}$ | $\hat{\Theta}_{i4}$ |
|---|---|---|---|
| (0,0,3) | 0.7213 | 0.6954 | 3.1502 |
| (1,0,2) | 1.3846 | 0.9274 | 2.4158 |
| (0,1,2) | 0.7213 | 1.3763 | 2.0207 |
| (2,0,1) | 2.0990 | 1.0965 | 1.6225 |
| (0,2,1) | 0.7213 | 2.1917 | 1.3475 |
| (1,1,1) | 1.3846 | 1.1667 | 1.0940 |
| (0,3,0) | 0.7213 | 3.0173 | 1.3190 |
| (1,2,0) | 1.3846 | 1.9773 | 1.0832 |
| (3,0,0) | 2.7962 | 1.2180 | 0.9160 |
| (2,1,0) | 2.0990 | 1.2791 | 0.9092 |

To examine how the different claim patterns impact the predicted heterogeneity, let us start with Figure 4.1, which shows the estimation of $\Theta_{i,2}$ with one year's experience. It presents a strict descending trend with the number of claims in year 1 decreasing from three to zero. As mentioned in Chapter 3, the number of claims for year 1 without any gained experience is Poisson distributed with parameter $\Theta_{i,1}\lambda/(1-p)$, and $\Theta_{i,2}$ is positively correlated with $\Theta_{i,1}$, which explains the trend in Figure 4.1 that $\hat{\Theta}_{i,2}$ is positively driven by the claim count for year 1.

We now proceed to look at Figure 4.2, which shows the predicted heterogeneity for year 3 with two years' experience. Note that $\Theta_{i,3}$ is positively correlated with both $\Theta_{i,1}$ and $\Theta_{i,2}$. $\Theta_{i,2}$ drives the error term $\epsilon_{i,2}$ and therefore part of the increment (positive or negative) of the claim counts from year 1 to year 2. Thus, the prediction $\hat{\Theta}_{i,3}$ carries information of not only the absolute number of claims in year 1, but also the increment from year 1 to year 2. Shown by Figure 4.2, the claim pattern (2,0) indicates a decrease in the number of claims. As a result, the predicted heterogeneity in this case is lower than the one for the

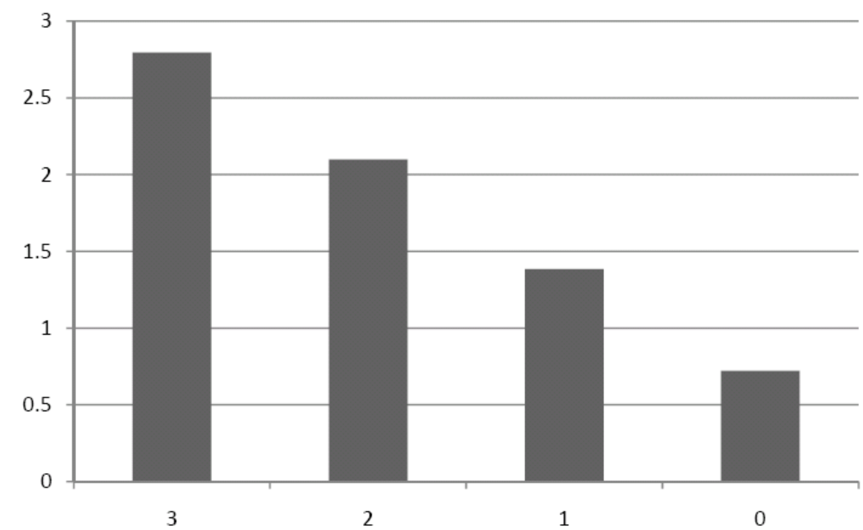


**Figure 4.1:** *Predicted heterogeneity for the second year with various patterns of claim counts*

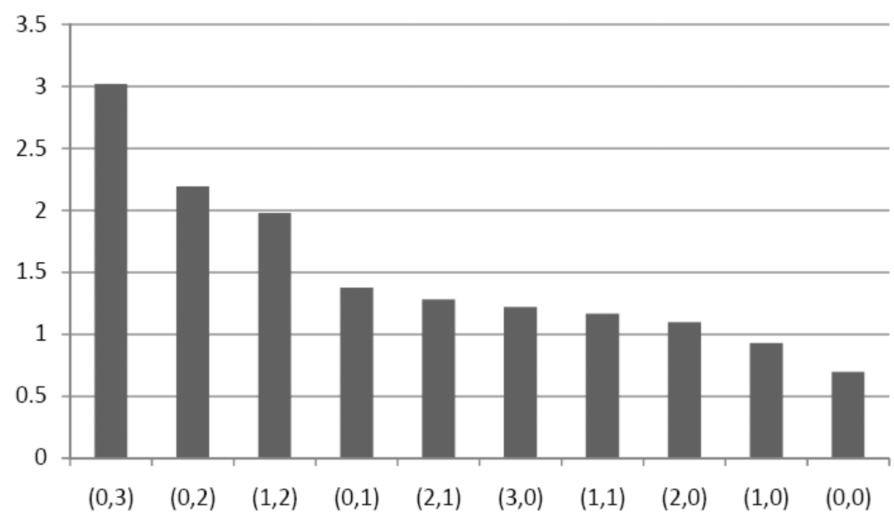


**Figure 4.2:** *Predicted heterogeneity for the third year with various patterns of claim counts*

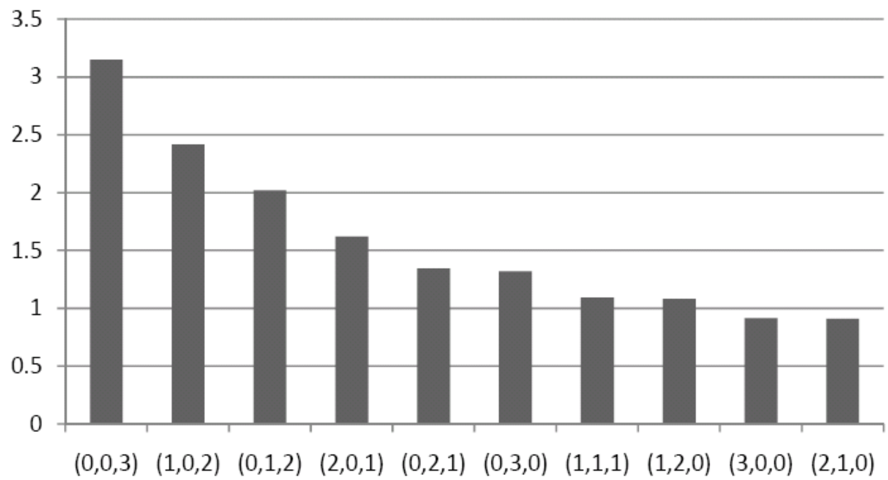


**Figure 4.3:** *Predicted heterogeneity for the fourth year with various patterns of claim counts*

pattern (0,2), which presents a steep growth in the second year. The pattern (1,1) shows a stable trend around one claim per year. Consequently, the predicted heterogeneity is in between the two cases discussed above. However, the pattern (3,0) has a larger predicted heterogeneity than the one for (2,0) and a steeper decreasing trend, which is due to the higher claim counts for year 1 that also contribute to the estimation.

To expand this reasoning to $\hat{\Theta}_{i,4}$, let us look at Figure 4.3, the predicted heterogeneity with three years of experience where two pairs of consecutive numbers of claims are available. The estimation of the heterogeneity for year 4 carries the information given by the mixture of three pieces: the absolute number of claims for year 1, the increment from year 1 to year 2, and the increment from year 2 to year 3. Since the AR(1) correlation structure is adopted to model the heterogeneity, the most recent increment has the largest weight to impact the prediction. Therefore, it is reasonable that the pattern (0,0,3) with the steepest increase during the last two years results in the highest predicted heterogeneity. However, which one of the three pieces of information would dominate the trend also relies on the specific claim patterns as well as the value of the parameters. Shown by Figure 4.3, compared to the pattern (2,1,0), (1,2,0) results in a higher prediction, which indicates that the increase from one claim in year 1 to two claims in year 2 shadows the steeper decrease from year 2 to year 3 and the smaller claim count in year 1. This could also be explained by the fact that, with this set of parameters, the system sees a driver with a steady descending claim history less likely to have a positive increment in year 4. With another set of parameters, some local orders in Figure 4.3 may change.

### 4.2.2 The Premiums and Comparison of Models

With the predicted heterogeneities discussed in the previous section, the premiums for years 2, 3, and 4, denoted by $P_2 = E[N_{i,2}|N_{i,1}]$, $P_3 = E[N_{i,3}|N_{i,2}]$, and $P_4 = E[N_{i,3}|N_{i,2}]$, are calculated and shown in columns three to five, respectively in Table 4.2. For comparison, from column six to column eleven, we state the premiums with the same patterns of claim counts generated by two formerly discussed models, the mixed Poisson with a Lognormal heterogeneity and the INAR(1) process with a static Gamma heterogeneity. At last, to test the AR(1) correlation structure, the premiums are calculated with $\rho = 0$, which means $\{\Theta_{i,t}\}_{t=1}^{\infty}$ on logarithm degenerates to a white noise process, and the results are listed in the last three columns of Table 4.2.

From Table 4.2 and Figure 4.4, we see that the Bonus-Malus system driven by the INAR(1) process with a dynamic heterogeneity has two features worth mentioning. First, the system sees the increment of claim counts from year to year being carried forward.

- As shown by the first four rows of Table 4.2, the four highest premiums result from the claim patterns that respectively have increments +3, +2, +1, and +1, respectively in the last two years . The system punishes those who had sudden increased numbers of claims in recent years the most severely.

Second, the system is sensitive to how the claims are distributed. We compare the results between two pairs of patterns to illustrate the property.

**Table 4.2:** *Premium comparison for different models*

| | Claim History | INAR(1) with Dynamic Heterogeneity $P_2$ | $P_3$ | $P_4$ | Poisson - Lognormal $P_2$ | $P_3$ | $P_4$ | INAR(1)with Static Gamma Heterogeneity $P_2$ | $P_3$ | $P_4$ | INAR(1) with White Noise Heterogeneity $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | (0,0,3) | 0.216 | 0.209 | 1.845 | 0.156 | 0.120 | 0.725 | 0.286 | 0.278 | 1.259 | 0.293 | 0.307 | 1.174 |
| 2 | (1,0,2) | 0.715 | 0.278 | 1.325 | 0.495 | 0.330 | 0.725 | 0.618 | 0.308 | 0.959 | 0.595 | 0.305 | 0.887 |
| 3 | (0,1,2) | 0.216 | 0.713 | 1.206 | 0.156 | 0.330 | 0.725 | 0.286 | 0.608 | 0.933 | 0.293 | 0.604 | 0.885 |
| 4 | (2,0,1) | 1.230 | 0.329 | 0.787 | 1.048 | 0.639 | 0.725 | 0.950 | 0.339 | 0.659 | 0.899 | 0.301 | 0.597 |
| 5 | (0,2,1) | 0.216 | 1.258 | 0.704 | 0.156 | 0.639 | 0.725 | 0.286 | 0.939 | 0.633 | 0.293 | 0.901 | 0.602 |
| 6 | (1,1,1) | 0.715 | 0.650 | 0.628 | 0.495 | 0.639 | 0.725 | 0.618 | 0.616 | 0.613 | 0.595 | 0.603 | 0.602 |
| 7 | (0,3,0) | 0.216 | 1.805 | 0.396 | 0.156 | 1.004 | 0.725 | 0.286 | 1.270 | 0.359 | 0.293 | 1.199 | 0.297 |
| 8 | (1,2,0) | 0.715 | 1.193 | 0.325 | 0.495 | 1.004 | 0.725 | 0.618 | 0.944 | 0.333 | 0.595 | 0.897 | 0.306 |
| 9 | (3,0,0) | 1.739 | 0.365 | 0.275 | 1.747 | 1.004 | 0.725 | 1.282 | 0.370 | 0.359 | 1.196 | 0.295 | 0.305 |
| 10 | (2,1,0) | 1.230 | 0.684 | 0.273 | 1.048 | 1.004 | 0.725 | 0.950 | 0.644 | 0.333 | 0.899 | 0.599 | 0.307 |

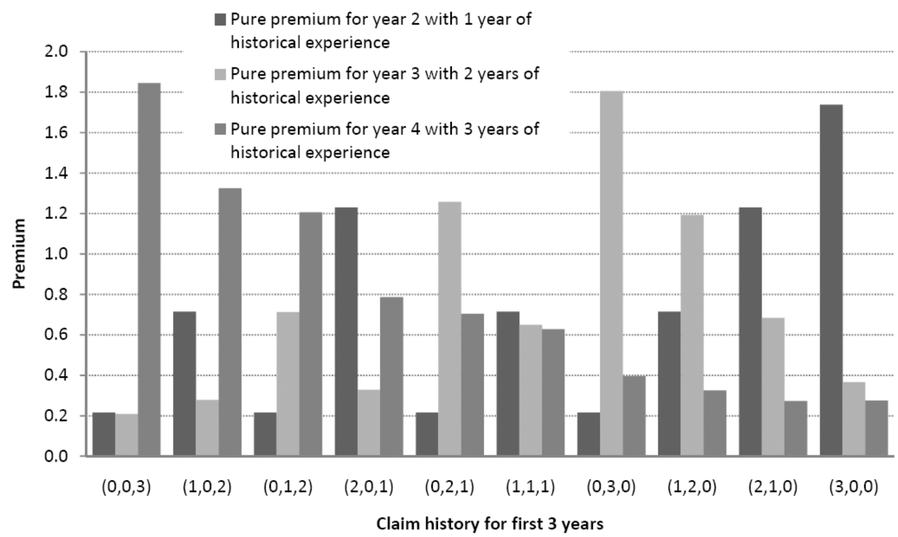


**Figure 4.4:** *Three years' premium trend with various patterns of claim counts*

- We first compare the results between the patterns (1,2,0) and (2,1,0). The two patterns result in similar premiums for year 4, which are among the lowest ones, but through very different paths. $P_3$ for the pattern (1,2,0) is slower than $P_2$ for the pattern (2,1,0). This can be explained by the fact that $P_3$ for (1,2,0) is the premium after the first year with 1 claim and the second year with 2 claims, the system treats the two claims in year 2 as a step-upward from the one claim in year 1, and a part of the penalty had already been imposed by the surcharge incorporated in $P_2$ for the pattern (1,2,0), thus it is lower than $P_2$ for causing 2 claims in the year immediately after the issuance of the policy.
- Similarly, $P_3$ for the pattern (2,1,0) is lower than $P_2$ for the pattern (1,2,0).

However, the premium trend shows the opposite order when a claim-free year is involved in the first two years. As an example, we compare the results between the patterns (1,0,2) and (0,1,2).

- The two patterns result in similar premiums for year 4, which are among the highest ones. $P_3$ for the pattern (1,0,2) is higher than $P_2$ for the pattern (0,1,2). It shows that the system remembers the claim in year 1 and tends to be more conservative for a discount.

- Similarly, $P_3$ for the pattern (0,1,2) is lower than $P_2$ for the pattern (1,0,2), which shows that the system remembers the claim-free year immediately after the issuance of the policy, and sees the driver with the pattern (0,1,2) less risky at the end of year 2 than a driver that caused one claim in year 1.

Furthermore, both pairs have similar results for the premium for year 4. Since within both pairs, the numbers of claims for year 3 are the same, the subtle difference in $P_4$ depends only on the predicted heterogeneity for year 4, which is discussed in the last section.

Table 4.2 also lists the results with three other models for parallel comparison. We have the following observations.

- For the mixed Poisson model with a Lognormal heterogeneity, whose results are shown in columns 6 - 8, we let $\lambda$ and $\sigma$ take the same values as in the INAR(1) model discussed above. The results demonstrate the property of the mixed Poisson model stated in Chapter 2 that the premiums rely only on the total number of past claims, whereas the results for the INAR(1) model with dynamic heterogeneity are different for the possible 10 patterns in all three years. Compared to the classical mixed Poisson model, the INAR(1) model with dynamic heterogeneity brings more variation to the premium, and the premium trends reflect the claim history more closely.

- The results for INAR(1) process with a static Gamma heterogeneity are given in columns 9 - 11. The parameter $a$ is set equal to 9, and all other parameters are kept the same. Although not all the predicted heterogeneities are distinguishable with different claim patterns, all the premiums are different due to the autoregressive part of the process up to year 3. However, for year 4, two pairs of patterns lead to the same premiums. The Gamma heterogeneity results in a smaller range for $P_4$ than the dynamic heterogeneity, because it has less fluctuation. Yet it has more fluctuation than the mixed Poisson model and therefore a larger range for $P_4$ due to the allowance of the regressive part in the INAR(1) process.

- In the INAR(1) process with a white noise heterogeneity, $\{\Theta_{i,t}\}_{t=1}^{\infty}$ identically and independently follow the Lognormal distribution. The predicted heterogeneity carries no information from the past. As a result, $P_4$ clusters around 1.2, 0.89, 0.6, and 0.3, in which cases, the last claim counts are 3, 2, 1, and 0, respectively.

### 4.2.3 Effect of Parameters $p$ and $\lambda$

Figures 4.5 - 4.24 show the effect of the autoregressive factor $p$ and the Poisson parameter $\lambda$ for the error terms to the predicted heterogeneities and premiums for year 4. In each figure, five curves are drawn for $\lambda = 0.2$, 0.4, 0.6, 0.8, and 1, respectively. Parameter $p$ is measured on the x-axis. The effect of the two parameters are as follows.

- The predicted heterogeneity decreases with $\lambda$. The Poisson mean of the error terms is the product of $\lambda$ and the heterogeneity. With a larger $\lambda$, a given claim experience indicates a smaller predicted heterogeneity. In contrast, the premium increases with $\lambda$. Both the trends are consistent for all ten different claim patterns.

- $p$, as the autoregressive parameter, is the probability that each of the incurred claims in the previous year contributes one claim into the claim count for the next year. The predicted heterogeneity decreases with $p$ consistently for all claim patterns. With any given claim experience, the more likely that a claim from the previous year contributes a claim to the claim count for the next year, the smaller the heterogeneity for the error term would be indicated.

- The effect of $p$ on premium varies in different claim patterns. The premium decreases with $p$ for the claim patterns that have no claims in the third year, showing the same trend as the predicted heterogeneity. This is easy to see from the pricing formula (3.14). In this case, the premium is the product of the predicted heterogeneity and $\lambda$. In contrast, for the claim patterns that have claim count larger than 0 for the third year, the premium increases with $p$. In this case it is the autoregressive component of the INAR(1) process that dominates the projected premium.

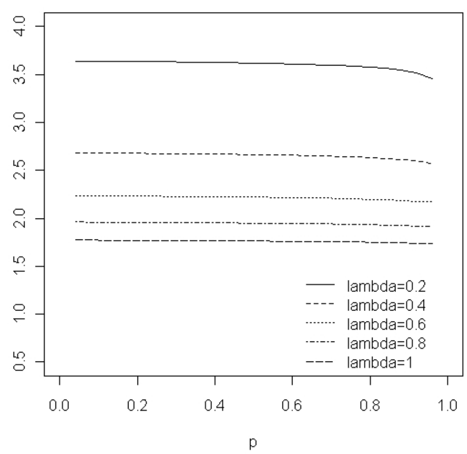


**Figure 4.5:** *Predicted heterogeneity* $\hat{\Theta}_{i,4}$ *with claim pattern (0,0,3)*

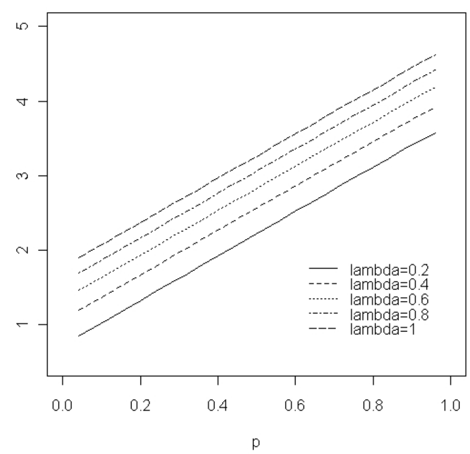


**Figure 4.6:** *Premium for year 4 with claim pattern (0,0,3)*

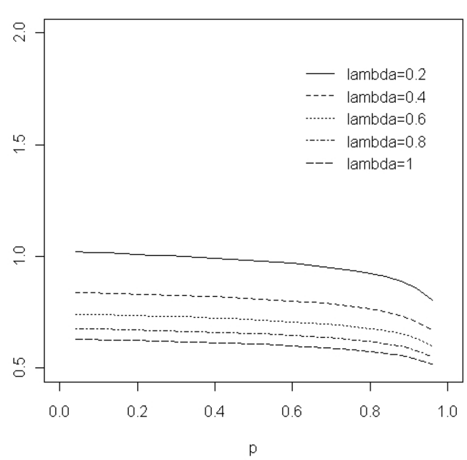


**Figure 4.7:** *Predicted heterogeneity* $\hat{\Theta}_{i,4}$ *with claim pattern (3,0,0)*

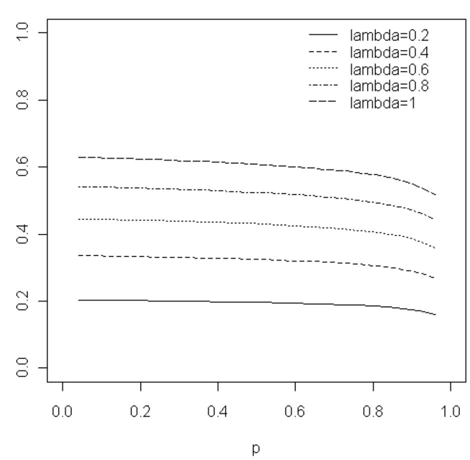


**Figure 4.8:** *Premium for year 4 with claim pattern (3,0,0)*

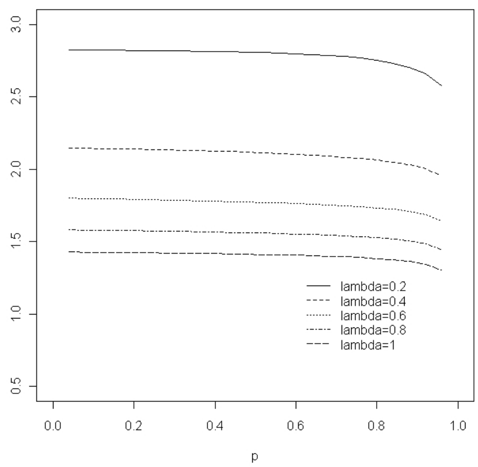


**Figure 4.9:** *Predicted heterogeneity $\hat{\Theta}_{i,4}$ with claim pattern (1,0,2)*

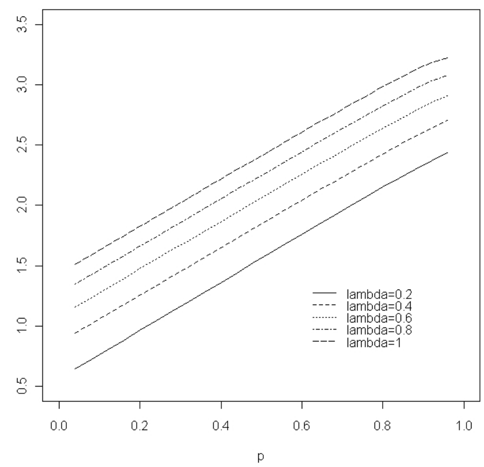


**Figure 4.10:** *Premium for year 4 with claim pattern (1,0,2)*

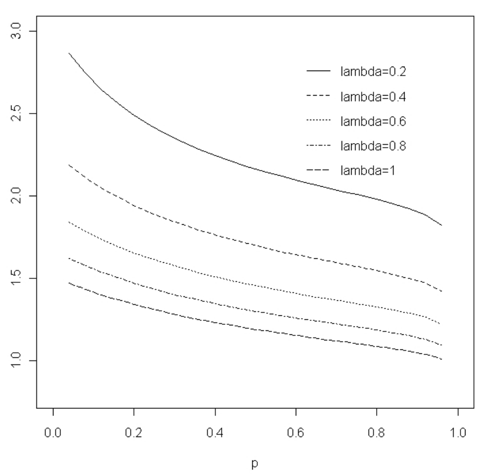


**Figure 4.11:** *Predicted heterogeneity $\hat{\Theta}_{i,4}$ with claim pattern (0,1,2)*

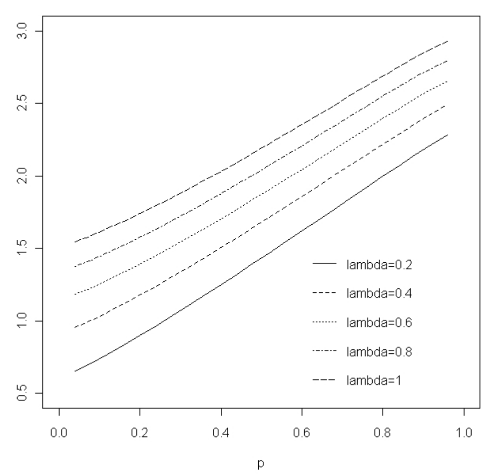


**Figure 4.12:** *Premium for year 4 with claim pattern (0,1,2)*

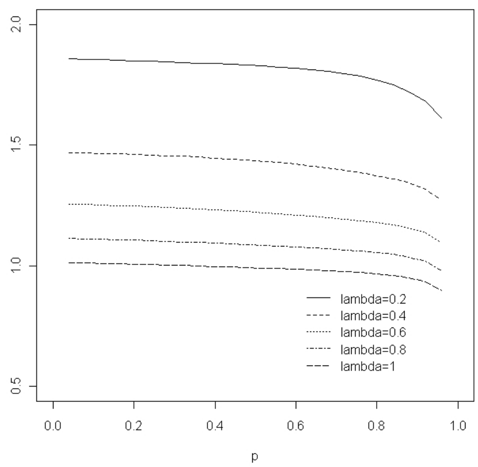


**Figure 4.13:** *Predicted heterogeneity* $\hat{\Theta}_{i,4}$ *with claim pattern (2,0,1)*

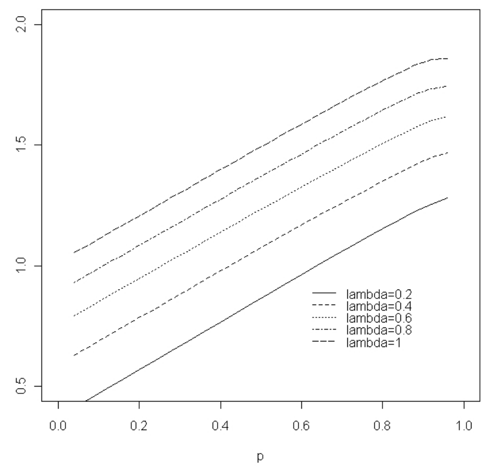


**Figure 4.14:** *Premium for year 4 with claim pattern (2,0,1)*

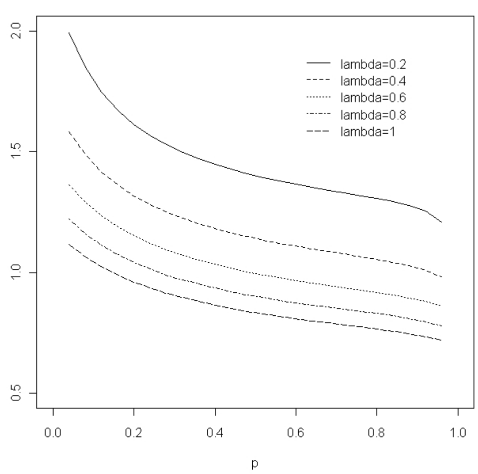


**Figure 4.15:** *Predicted heterogeneity* $\hat{\Theta}_{i,4}$ *with claim pattern (0,2,1)*

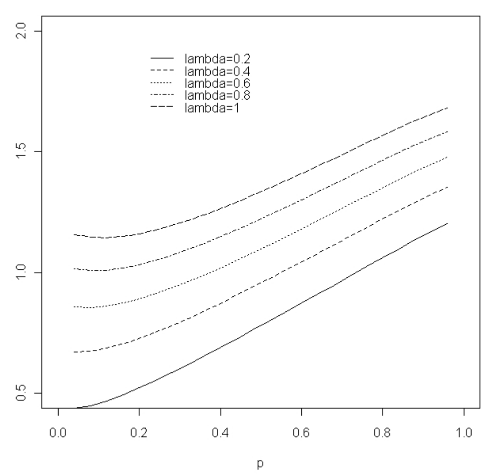


**Figure 4.16:** *Premium for year 4 with claim pattern (0,2,1)*

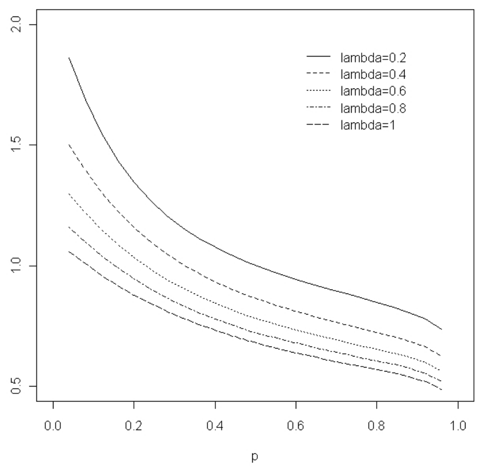


**Figure 4.17:** *Predicted heterogeneity* $\hat{\Theta}_{i,4}$ *with claim pattern (1,1,1)*

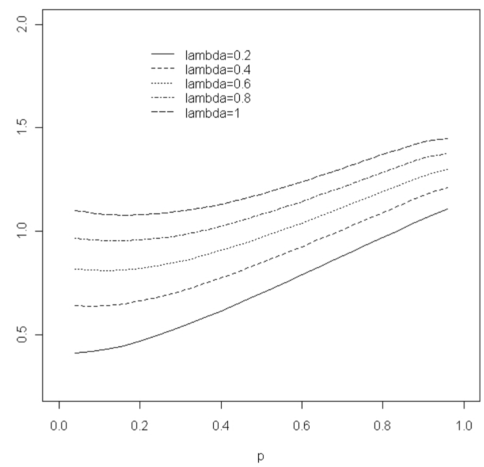


**Figure 4.18:** *Premium for year 4 with claim pattern (1,1,1)*

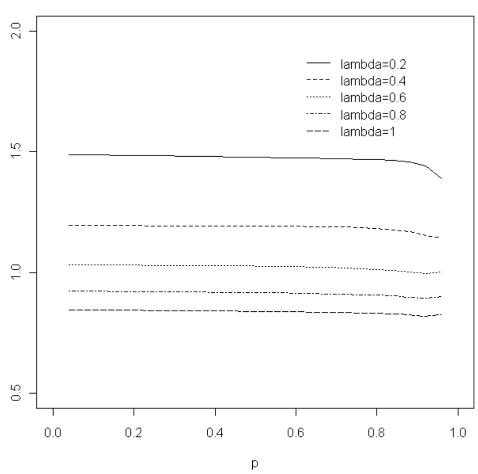


**Figure 4.19:** *Predicted heterogeneity* $\hat{\Theta}_{i,4}$ *with claim pattern (0,3,0)*

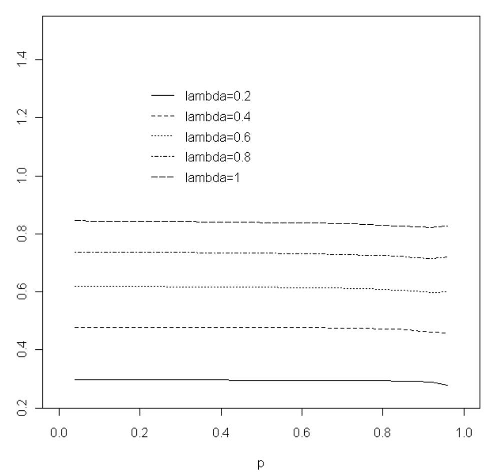


**Figure 4.20:** *Premium for year 4 with claim pattern (0,3,0)*

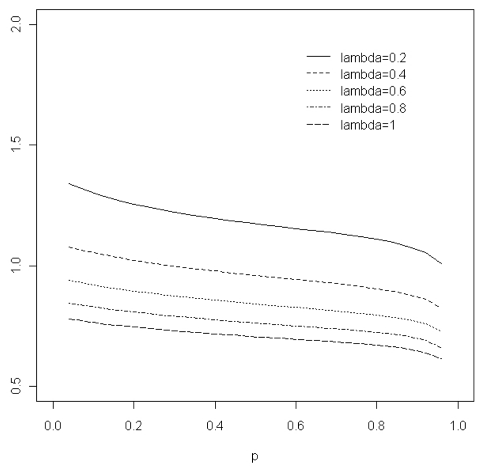


**Figure 4.21:** *Predicted heterogeneity* $\hat{\Theta}_{i,4}$ *with claim pattern (1,2,0)*

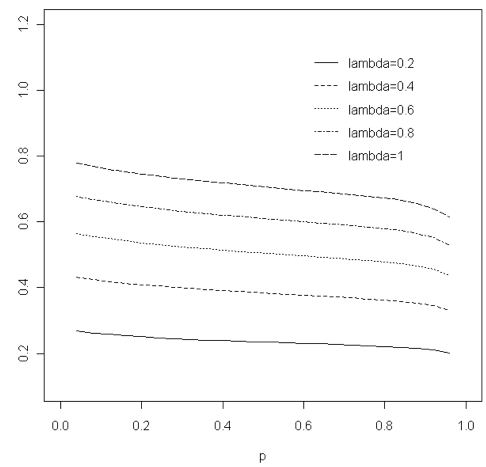


**Figure 4.22:** *Premium for year 4 with claim pattern (1,2,0)*

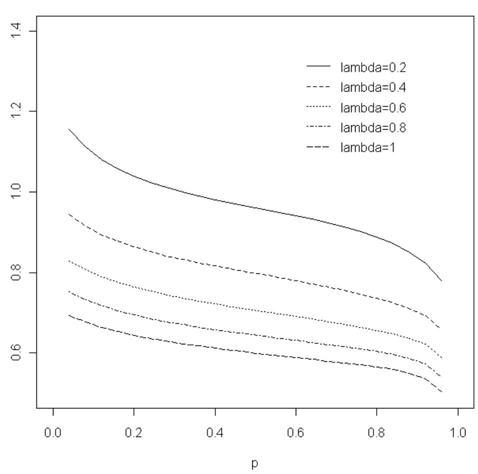


**Figure 4.23:** *Predicted heterogeneity* $\hat{\Theta}_{i,4}$ *with claim pattern (2,1,0)*

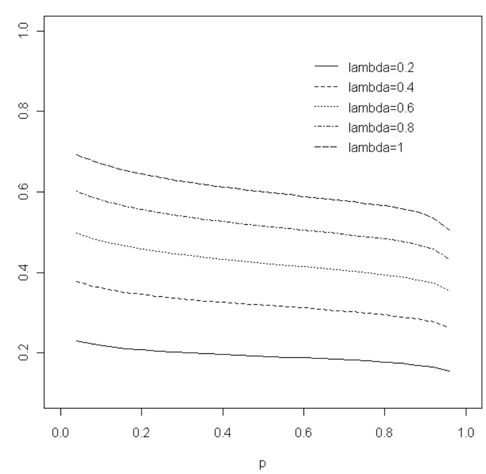


**Figure 4.24:** *Premium for year 4 with claim pattern (2,1,0)*

# Chapter 5

# Conclusion

Since the pioneer work of Dionne and Vanasse (1989) on the application of mixed Poisson model in automobile insurance, numerous efforts have been put into claim count modeling for a non-homogeneous portfolio and the associated pricing techniques. Seeking the Bayesian posterior mean has been a well established and accepted approach to obtain the estimation of the individual heterogeneity given past numbers of claims. Gourieroux and Jasiak (2004) discussed the INAR(1) process for automobile insurance modeling, where the claim counts regressively rely on the lagged claim counts, and the heterogeneity of the portfolio is reflected in the error terms. Brouhns et al. (2003) introduced advanced model for the heterogeneity. They applied an AR(1) model and pointed out that how the individual mean claim frequency departs from the overall average of the portfolio should be dynamic and correlated over time.

We studied an INAR(1) process, formerly discussed in Gourieroux and Jasiak (2004), with a dynamic heterogeneity, introduced by Brouhns et al. (2003), integrated through the error terms. We studied the properties of this model and developed pricing formulas. A simulation is performed to obtain the numerical results.

The claim history affects the prediction of future claim counts and therefore the premium in two ways: the autoregressive part which directly connects the lagged counts, and the estimation of the heterogeneity. For the autocorrelation of the lagged claim counts, we found that the correlation decreases as the time lag increases, which is a natural consequence of the AR(1) assumptions, and corresponds the fact that the predictive ability of a claim declines as it ages.

The numerical illustration is performed with ten different claim patterns where three claims are spread in three years. It is shown from the results that the predicted premium is sensitive to different distributions of claims. Not only are the past numbers of claims reflected in the updated premium, but also the trend of claim counts changing from year to year is carried forward. Compared to other models, the bonus-malus system based on the INAR(1) process with dynamic heterogeneity exploits the past information more efficiently.

For the INAR(1) process with the dynamic heterogeneity, we could explore more applications in automobile insurance. First, as for the modeling of the correlated heterogeneity for years, more correlation structures could be considered as discussed at the end of section 3.3.2. Secondly, based on the first and second moments discussed in Chapter 3, a generalized estimating equation method could be employed to estimate the parameters, and once the observation of a sufficiently large sample is obtained, the bonus-malus system for the population could be established. Lastly, one could also consider a bonus-malus scale with a defined transition rule based on the model, for which the relativities may be obtained by numerical simulations.

# Bibliography


Al-Osh, M. A. and A. A. Alzaid (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis 8*, 261–275.

Brouhns, N., M. Guillén, M. Denuit, and J. Pinquet (2003). Bonus-malus scales in segmented tariffs with stochastic migration between segments. *The Journal of Risk and Insurance 70*, 577–599.

Carlin, B. P. and T. A. Louis (2000). *Bayes and Empirical Bayes Method for Data Analysis Second Edition*. Chapman & Hall/CRC.

Casella, G. and E. I. George (1992). Explaining the Gibb's sampler. *The American Statistician 46*, 167–174.

Chan, K. and J. Ledolter (1995). Monte carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association 90*, 242–252.

Denuit, M., X. Maréchal, S. Pitrebois, and J.-F. Walhin (2007). *Actuarial Modeling of Claim Count*. John Wiley & Sons Ltd.

Dionne, G. and C. Vanasse (1989). A generalization of actuarial automobile insurance rating models: the negative binomial distribution with a regression component. *ASTIN Bulletin 19*, 199–212.

Gomez-Deniz, E., J. M. Sarabia, J. M. Perez-Stanched, and F. J. Vaquez-Polo (2008). Using a bayesian hierarchical model for fitting automobile claim frequency data. *Communications in Statistics-Theory and Methods 38*, 1425–1435.

Gomez-Deniz, E. and F. J. Vaquez-Polo (2006). A note on computing bonus-malus insurance premiums using a hierarchical bayesian framework. *Test 15*, 345–359.

Gourieroux, C. and J. Jasiak (2004). Heterogeneous INAR(1) model with application to car insurance. *Insurance: Mathematics and Economics 34*, 177–192.

Laird, G. M. and J. H. Ware (2004). *Applied Longitudinal Analysis*. John Wiley & Sons Ltd.

Lemaire, J. (1976). Driver versus company: Optimal behaviour of the policyholder. *Scandinavian Actuarial Journal 59*, 209–219.

Lemaire, J. (1977). La soif du bonus. *ASTIN Bulletin 9*, 181–190.

Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers.

Picard, P. (1976). Generalisation de l'étude sur la survenance des sinistres en assurance automobile. *Bulletin Trimestriel de l'Institut des Actuaires Français*.

Pinquet, J. (1998). Designing optimal bonus-malus systems from different types of claims. *ASTIN Bulletin 28*, 205–220.

Pinquet, J., M. Guillen, and C. Bolance (2001). Allowance for the age of claims in bonus-malus systems. *ASTIN Bulletin 31*, 337–348.

Pitrebois, S., M. Denuit, and J.-F. Walhin (2003a). Fitting the belgian bonus-malus system. *Belgian Actuarial Bulletin 3*, 58–62.

Pitrebois, S., M. Denuit, and J.-F. Walhin (2003b). Setting a bonus-malus scale in the presence of other rating factors: Taylor's work revisited. *ASTIN Bulletin 33*, 419–436.

Pitrebois, S., M. Denuit, and J.-F. Walhin (2004). Bonus-malus scales in segmented tariffs: Gilde & Sundt's work revisted. *Australian Actuarial Journal 10*, 107–125.

Pitrebois, S., M. Denuit, and J.-F. Walhin (2006). Multi-event bonus-malus scales. *The Journal of Risk and Insurance 73*, 517–528.

Taylor, G. (1997). Setting a bonus-malus scale in the presence of other rating factors. *ASTIN Bulletin 27*, 319–327.

Walhin, J. F. and J. Paris (2000). The true claim amount and frequency distributions within a bonus-malus system. *ASTIN Bulletin 30*, 391–403.

Zeger, S. (1988). A regression model for time series of counts. *Biometrika 75*, 621–629.