IBNR Claims Reserving Using INAR Processes

by

Yang Bai

B.Sc., Zhejiang University, 2013

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

Approval

Name: **Yang Bai**

Degree: **Master of Science (Actuarial Science)**

Title: ***IBNR Claims Reserving Using INAR Processes***

Examining Committee: **Chair: Dr. Tim Swartz**
Professor

Dr. Yi Lu
Senior Supervisor
Associate Professor

Dr. X. Joan Hu
Supervisor
Professor

Dr. Cary Chi-Liang Tsai
External Examiner
Associate Professor

Date Defended: 15 December 2016

Abstract

This project studies the reserving problem for incurred but not reported (IBNR) claims in non-life insurance. Based on an idea presented in Kremer (1995), we propose a new Poisson INAR (integer-valued autoregressive) model for the unclosed claim counts, which are the number of reported but not enough reported claims. The properties and the prediction of the proposed Poisson INAR model are discussed. We modify the estimation methods proposed in Silva et al. (2005) for the replicated INAR(1) processes to be applied to our model and introduce new algorithms for estimating the model parameters. The performance of three different estimation methods used in this project is compared, and the impact of the sample size to the accuracy of the estimates is examined in the simulation study. To illustrate, we also present the prediction results of our proposed model using a generated sample.

Keywords: IBNR; INAR; Frequency-severity techniques; MSEP; Yuller-Walker estimation; Least squares estimations

Dedication

To my beloved parents, for their love and support

Acknowledgements

First and foremost I would like to thank my senior supervisor Dr. Yi Lu, for her continuous support and patience with me. There were hard times I almost want to give up this topic, but she encouraged me every time I felt disappointed and figured out a way to nicely present what we have done. I thank her deeply for the time she dedicated to me to investigate the problems and difficult issues we met. I have also learned a lot of conventions on academic writing during the process of revising my project report. I would never think I could graduate without her kind consideration and tremendous help.

I would also want to express my sincere gratitude to my committee members Dr. X. Joan Hu and Dr. Cary Chi-Liang Tsai. Thanks for the time they spent to look at my project report and the constructive suggestions they gave on polishing it. I also want to thank Joan for her great ideas on the further research of this project.

Special thanks go to my parents who have come to visit me during the hardest time of writing my project report. Their endless support and unconditional understanding have given me the energy and confidence to finally finish this project.

Lastly, I want to thank my friends and fellow students here for the memorable times we spent together. It is an unforgettable experience to have my graduate study at Simon Fraser University. We worked together through difficult courses, encouraged each other and shared our opinions greatly. Your accompany has made my life in Vancouver enjoyable and fruitful. Thanks all the faculties and staffs at our department who together provide a friendly and effective working environment. I also want to thank my room-mate because they always have the ears to listen to the difficulties and troubles I went through. It is them that make me feel like at home though far away from my parents.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The techniques for loss reserving of property and casualty insurance has been discussed widely by both academia and industry. As a major uncertainty (risk) carried by the non-life insurers, finding more sophisticated methods or superior models which fit to various different conditions of insurance companies is of crucial importance.

Generally, loss reserves or claim reserves refer to the unpaid amount that the insurance company owes to the insured due to an incurred but not yet settled claims situation. For example, in the car insurance case, when a multi-vehicle accident happens and is reported to the insurers, the payments may not be paid at one time because of the complex administration process or even law suits involved. If explained in actuarial terms, according to Friedland (2010), loss reserves are often divided into five components:

(1) case outstanding or unpaid case, which refers to "the estimates of unpaid claims established by the claims department, third-party adjusters, or independent adjusters for known and reported claims only", and does not include future development on reported claims;

(2) provision for future development on known claims;

(3) estimates for reopened claims;

(4) provision for claims incurred but not reported;

(5) provision for claims in transit, reported but not yet recorded.

Usually, insurance companies would estimate the total amount of the loss reserve directly. However, the estimation could be done by each component for more accurate results or diagnostic purposes. A most commonly used category is the separation of IBNyR (incurred but not yet reported) claims and IBNeR (incurred but not enough reported) claims. By this categorization, for the five above-mentioned components, IBNyR includes (4) and (5) and IBNeR includes (1)-(3). In general, IBNyR is also known as pure IBNR or a narrow definition of IBNR.

To estimate the loss reserve, the claim data is usually organized into a so-called claim development triangle, also called a run-off triangle. A most typical claim development triangle is to organize the claim data by the occurrence dates, usually in one year unit by each row and the cumulative amount paid after the events have occurred for each following development year (including the current year) by each column. The loss reserve is just the total ultimate claim amount that should be paid minus the current cumulative claim payments. There are two commonly used claim development triangles in industry. One is the paid loss triangle and the other is the incurred loss triangle. The paid loss triangle records the cumulative paid amount in each cell while the incurred loss triangle records the cumulative paid amount plus the case outstanding reserve suggested by the claim adjusters. As for which claim development triangle to use, it usually depends on the situation of the insurance company, the administrative legislation and the economic environment. Except for the development triangles which record the claim amount, there are also triangles for claim counts, for example, cumulative number of reported claims, number of closed claims and number of unclosed (open) claims. There are various ways to organize the claim data into a development triangle; for more information and details, see Friedland (2010).

The most popular method to estimate the IBNR claims reserve is the classical chain-ladder (CL) method or algorithm, due to its simplicity and its non-parametric distribution-free nature (see, Mack, 1993). A major disadvantage is that the CL method does not consider a priori information for the ultimate claim amount. To combine the observed information with a priori estimate, the Bornhuetter-Freguson (BF) method is considered (see, Bornhuetter, 1972). Borrowing the idea from the credibility theory, the CL method and the BF method are further combined into the Benktander-Hovinen (BH) method (see, Benktander, 1976 and Hovinen, 1981). In addition, there is a method called Cape-Cod method (see, Bühlmann, 1983) which is developed to provide a BF-type yet more robust estimation. However, these popularly used methods are all non-parametric models which do not assume proper underlying claim distributions, nor are they studied under a stochastic framework. It is not possible to use them to test the uncertainty of the estimators. As a result, there are parametric models built up for modelling the claim development pattern. Most parametric models satisfy the general assumptions of the non-parametric models. That is, these parametric models can be treated as specific stochastic realizations of the non-parametric models (see, for example, Wüthrich and Merz, 2008). Within the parametric models, there are models for claims counts as well as claim amount (or severities). Although estimating the ultimate claim amount directly is simpler and more straightforward, it is more reasonable sometimes to estimate the claim counts and claim severities separately. The kind of idea is also referred to as frequency-severity techniques. For details on general stochastic claims reserving methods in insurance, see Wüthrich and Merz (2008).

To model the insurance claim counts, the traditional time series models are not theoretically suitable since they do not model integer valued processes. Alosh and Alzaid (1987) has extended the AR(1) process (autoregressive process of order one) to the INAR(1) process, first order integer-values autoregressive process. Gourieroux and Jasiak (2004) proposed a heterogeneous INAR(1) model for the claim frequency of the policyholders in car insurance and the future insurance premiums could be determined by the past claim frequencies. Zhang (2009) further extended the idea of using INAR process to model claim counts and considered the time dependent heterogeneity for individual policyholders.

Kremer (1995) proposed an idea of modelling IBNR claims using INAR processes; however, there are little literature on this topic after Kremer's idea has been brought up. Because the INAR process successfully mimics the process of the arriving of the newly reported claims with the degeneration of the already reported claims, a further investigation of these kinds of models would be insightful. Kremer's model can be seen as an underlying stochastic framework of the CL method; however, the dynamics underlying the stochastic assumptions and the model parameters are difficult to interpret in practice. By incorporating the idea of using the INAR processes to model the claim numbers, we propose a Poisson INAR model for the unclosed claim numbers, which are the number of reported but not enough reported claims. Although there are some restrictions to the application of the proposed model, it is a simple and reasonable model. The Poisson INAR model has desired properties and we can also write down the non-parametric assumptions for the Poisson INAR model with the introduction of the unclosed claims development pattern. The prediction formulas and the mean square error of prediction under the proposed model are then obtained and discussed. Because of the similarity between the Poisson INAR model and the replicated INAR(1) processes (Silva et al., 2005), we consider the estimation methods studied by Silva et al. (2005) to obtain the estimation of the model parameters. Noting that the relative size of the number of unknown model parameters compared to the number of the observed data points may affect the accuracy of the estimates, we consider different cases with different number of model parameters and different sizes of the development triangle when calculating the estimation results. The numerical results from a simulation study are presented and discussed, which confirms the influence of the number of model parameters and the size of the observed data to the estimation. A random sample is generated to conduct the prediction of the number of the outstanding claims and the prediction error is calculated to measure the accuracy of the prediction results.

The project is organized as following. Chapter 2 provides a more detailed background of the loss reserving problem. Some typical non-parametric and parametric models for loss reserving are presented and compared in details. Kremer's model is also presented for further comparison. Chapter 3 gives a thorough introduction of our Poisson INAR model, including the model assumption, properties, predictions and the estimation methods. Three estimation methods are proposed and estimation formulas of the model parameters and the corresponding algorithms are derived. Because of the high dimension of the parameters that need to be estimated, simplified models are considered under each method in order to evaluate the effectiveness of the parameter estimations. Chapter 4 presents the major discoveries from a simulation study in the aspect of comparing the effectiveness of different estimation methods and the influence of the number of unknown model parameters. A practical prediction analysis is considered for a random sample that is generated from the Poisson INAR model. Chapter 5 concludes the project with some suggestions on further research regarding to the use of INAR processes for the IBNR claims reserving problem.

Chapter 2

Literature Review

As mentioned in the previous chapter, the loss reserve contains five components, but the most commonly used category is the IBNR claims, including IBNyR and IBNeR claims. The claim data is organized into various claim development triangles for the purpose of estimating the loss reserve, the components of the loss reserve and even the claim counts. In order to better understand how the claim development triangle is formed, a brief introduction of the basic concepts of the claim process is presented in Section 2.1. In Section 2.2, we introduce the notations and several examples of the claim development triangles. Sections 2.3 and 2.4 give a review of the classical non-parametric and parametric models, respectively, for the claim reserving problem. Section 2.5 reviews the modelling of IBNR claims with an INAR process involved. The chapter ends with an introduction to the frequency-severity techniques to add on the reasonableness of estimating the claim counts.

2.1 The Claim Process

In general, the loss reserve is the total outstanding payments of all incurred claims, whether reported or not. In other words, it is an aggregation of the outstanding payments for every single claim. The claim process reflects the dynamics of the development of a single claim and is discussed in Wüthrich and Merz (2008). With the definition of the claim process, the loss reserve can be presented accordingly. We briefly summarize the idea in the following.

Suppose that there are N claims in total with an ordered reporting dates $T_{1,0}, T_{2,0}, \ldots,$ $T_{N,0}$. For the nth claim, there exists a payment process defined as $(T_{n,m}, X_{n,m})_{m\geq 0}$, where $T_{n,m}$ is the mth payment time of claim n and $X_{n,m}$ is the amount paid at time $T_{n,m}$. Specifically, we denote that T_{n,M_n} is the final settlement date of this claim. Consequently, we have for any $k \geq 1$, $T_{n,M_n+k} = \infty$, and $X_{n,M_n+k} = 0$.

Now let $C_n(t)$ be the cumulative payments for claim n at time t, and $R_n(t)$ be the outstanding claim payments for this claim at time t. Then we can write

$$\begin{aligned} C_n(t) &= \textstyle\sum_{m\in\{k:T_{n,k}\le t\}} X_{n,m}, \\ C_n(\infty) &= C_n(T_{n,M_n}) = \textstyle\sum_{m=0}^{\infty} X_{n,m}, \\ R_n(t) &= C_n(\infty) - C_n(t) = \textstyle\sum_{m\in\{k:T_{n,k}>t\}} X_{n,m}. \end{aligned}$$

Furthermore, denote $C^N(t)$ as the aggregate claim payments up to time t and $R^N(t)$ as the total outstanding payments for all the claims at time t. If N includes only the claims already known (reported) at time t, $R^N(t)$ is just the IBNeR reserve. If N includes all the incurred claims whether known or not, $R^N(t)$ is the total loss reserve. Clearly,

$$\begin{aligned} C^N(t) &= \textstyle\sum_{n=1}^{N} C_n(t) \\ C^N(\infty) &= \textstyle\sum_{n=1}^{N} C_n(\infty) \\ R^N(t) &= \textstyle\sum_{n=1}^{N} R_n(t) = C^N(\infty) - C^N(t). \end{aligned}$$

Generally, we are interested in knowing $Z_t = \mathsf{E}[C^N(\infty)|\mathcal{F}_t]$, where $\mathcal{F}_t = \sigma(\{(T_{n,m}, X_{n,m});\ 1 \le n \le N, m \ge 0, T_{n,m} \le t\})$ represents the information of the payment process up to time t. However, the true value of Z_t may not be available. Therefore, people usually focus on getting an estimate of Z_t and then discussing the accuracy of the estimate.

2.2 The Claim Development Triangle

The claim development triangle is a commonly used and smart way to organize the claim data so that it can provide more information to better estimate the loss reserve.

Suppose that the current calendar year is denoted as L. We group the claims by the year happened (accident year in automobile insurance case) and the cumulative payments for the claims incurred in a particular accident year are recorded for each development year after the accident year (inclusive). The most recent accident year is denoted as I and the last development year is denoted as J. Usually, $L = I$, however, L can be bigger than I when a business line (an insurance coverage) is no longer operated by the insurance company but still owes some outstanding payments. For simplicity, we only consider $L = I$ in this project. Further assume that there are N_i claims happened in accident year i, including not reported claims. Let $C_{i,j}$ be the cumulative payments for the claims happened in accident year i and developed to development year j. Then

$$C_{i,j} = C^{N_i}(j), \qquad 0 \le i \le I, \quad 0 \le j \le J, \quad i + j \le I. \tag{2.1}$$

Let $X_{i,j}$ be the payments made in development year j for all the claims happened in accident year i (incremental payments). Clearly, $X_{i,0} = C_{i,0}$, $0 \leq i \leq I$ and

$$X_{i,j} = C_{i,j} - C_{i,j-1}, \qquad 0 \leq i \leq I, \;\; 0 < j \leq J, \;\; i + j \leq I.$$

See Table 2.1 for a clear presentation of a simple claim development triangle.

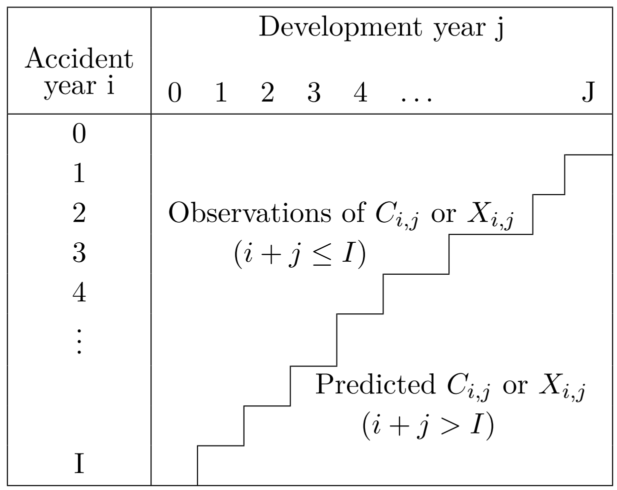

Table 2.1: A simple claim development triangle

From Table 2.1, we can see that the claim development triangle is split into two parts: the upper and lower triangle. The upper triangle contains the observed data while the lower triangle is the outstanding claims to be estimated. Denote that the observed data set $\mathcal{D}_I = \{X_{i,j};\ i+j \leq I, 0 \leq i \leq I, 0 \leq j \leq J\}$ and the outstanding claims $\mathcal{D}_I^c = \{X_{i,j};\ i+j > I, 0 \leq i \leq I, 0 \leq j \leq J\}$; one has to find the estimates of the elements of $\mathcal{D}_I^c$ based on the information available in $\mathcal{D}_I$.

A most typical case to do the estimation is when $I = J$ because in this case both $\mathcal{D}_I^c$ and $\mathcal{D}_I$ are triangles instead of trapezoids. Within other cases, $J > I$ happens when the business has not been operated for a long time or the settlement process for the business (such as bodily injures and third party liabilities) takes quite a long time, or sometimes only because of missing data for early accident years, while $J < I$ is the case when there are enough data provided to do the estimation so that the prediction only needs to be done for the small bottom-right corner of the rectangular. But in this latter case, people usually cut away the first few accident years to make I equal to J to get a smaller development triangle. Although generally there is a rule to use as more observed data as we can, the reason to get rid of the early claim history is because it may no longer be relevant to current year situation or even contradicts the recent development pattern. For simplicity and better understanding of this project, we assume $I = J$ in all the following context.

According to the definition of N_i and Equation (2.1), $C_{i,j}$ is the cumulative payments for all incurred claims, whether reported or not. Alternatively, $C_{i,j}$ could be different types of variable and then the meaning of $X_{i,j}$ would be changed correspondingly. Table 2.2 gives several examples of the claim development triangles and $C_{i,I}$ is the variable that we are interested in estimating in each case.

$C_{i,j}$	$X_{i,j}$	$C_{i,I}$
cumulative payments	incremental payments	ultimate claim amount/load
IBNeR claims	change of reported claim amount	ultimate incurred claims
cumulative reported claims	number of reported claims with delay j	ultimate claim counts

Table 2.2: Categories of the claim development triangles

2.3 Classical Non-parametric Models

The most straightforward idea to use the claim development triangle to do the estimation of the loss reserve is to assume an underlying model for the development triangle, and then the outstanding payments can be estimated with the expectation conditioning on the observed data $\mathcal{D}_I$. There are two types of models for the claim reserving problem, non-parametric and parametric models. The non-parametric models are often more simpler; therefore, they play an important role in the industry. In this section, we go over the most famous non-parametric models (see Wüthrich and Merz, 2008, for details).

Non-parametric models are the models that do not make particular assumptions for the distribution of $X_{i,j}$'s or $C_{i,j}$'s, also known as distribution free models. The most common assumption is that *the cumulative claims $C_{i,j}$ of different accident years i are independent*.

Based on the above assumption, the chain-ladder method (CL method) further assumes that there exist development factors $\{f_j\}_{j=0}^{I-1}$ such that for all $0 \le i \le I$ and all $0 < j \le I$,

$$\mathsf{E}[C_{i,j}|C_{i,0},\ldots,C_{i,j-1}] = \mathsf{E}[C_{i,j}|C_{i,j-1}] = f_{j-1}C_{i,j-1}. \tag{2.2}$$

That is, the expected next year's cumulative claims are proportional only to the current year's cumulative claims. The development factors are usually estimated by

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}}, \quad 0 \le j \le I-1.$$

Denote $\hat{C}_{i,j}^{CL}$ as the estimator of $C_{i,j}$ under the CL method and then

$$\hat{C}_{i,j}^{CL} = \hat{\mathsf{E}}[C_{i,j}|\mathcal{D}_I] = C_{i,I-i}\cdot\hat{f}_{I-i}\cdot\cdots\cdot\hat{f}_{j-1}, \quad 0 < i \le I,\ I-i < j \le I. \tag{2.3}$$

The CL method is an empirical estimate that does not depend on any prior information available for the ultimate claims. Alternatively, the Bornhuetter-Freguson (BF) method assumes that there exist parameters $\{\mu_i\}_{i=0}^{I}$ which can be seen as a set of prior estimates for the ultimate claims. Further assume that there is a development pattern $\{\beta_j\}_{j=0}^{I}$ with $\beta_I = 1$ for the cumulative claims such that for $0 \le i \le I$, $0 \le j \le I-1$ and $1 \le k \le I-j$,

$$\begin{aligned} \mathsf{E}[C_{i,0}] &= \beta_0 \mu_i, \\ \mathsf{E}[C_{i,j+k} | C_{i,0}, \ldots, C_{i,j}] &= C_{i,j} + (\beta_{j+k} - \beta_j)\mu_i. \end{aligned} \tag{2.4}$$

The estimator of the development pattern of the BF method could be obtained from the estimator of the development factors of the CL method by the relationship

$$\beta_j = \prod_{k=j}^{I-1} f_k^{-1}, \quad 0 \le j < I. \tag{2.5}$$

Denote $\hat{C}_{i,j}^{BF}$ as the estimator of $C_{i,j}$ under the BF method and then

$$\hat{C}_{i,j}^{BF} = \hat{\mathsf{E}}[C_{i,j} | \mathcal{D}_I] = C_{i,I-i} + (\hat{\beta}_j - \hat{\beta}_{I-i})\mu_i, \qquad 0 < i \le I, \ I-i < j \le I,$$

where $\hat{\beta}_j = \prod_{k=j}^{I-1} \hat{f}_k^{-1}$. According to Equations (2.3) and (2.5), it can be concluded that $\hat{\beta}_{I-i} \hat{C}_{i,I}^{CL} = C_{i,I-i}, \ 0 < i \le I$. Therefore,

$$\hat{C}_{i,I}^{BF} = \hat{\beta}_{I-i} \hat{C}_{i,I}^{CL} + (1 - \hat{\beta}_{I-i})\mu_i, \quad 0 < i \le I.$$

From the above equation, we know that the BF estimator is a linear combination of the empirical estimate and the prior estimate.

Despite the information of the most recent cumulative claims $C_{i,I-i}$, the CL method depends only on the observations while the BF method uses only the prior information when estimating the outstanding claims. It is natural to combine the two methods and obtain another credibility type of estimator. This idea is considered by the Benktander-Hovinen (BH) method.

Assume that there exists a new estimator $\mu_i(c)$ for the ultimate claims of accident year i which is a linear combination of the CL estimates and the prior estimates μ_i, that is, there exists a parameter $c \in [0, 1]$ so that

$$\mu_i(c) = c\,\hat{C}_{i,I}^{CL} + (1 - c)\mu_i, \qquad 0 < i \le I. \tag{2.6}$$

In (2.6), c should be increased with the increase of the number of development years (corresponding to early accident years) because the more the observed data the more credit should be put on the empirical estimates. An natural choice would be $c = \hat{\beta}_{I-i}$. Hence,

$\mu_i(c) = \hat{C}_{i,I}^{BF}$. The estimator of $C_{i,j}$ under the BH method can be written as:

$$\begin{aligned} \hat{C}_{i,j}^{BH} &= C_{i,I-i} + (\hat{\beta}_j - \hat{\beta}_{I-i})\mu_i(c) \\ &= C_{i,I-i} + (\hat{\beta}_j - \hat{\beta}_{I-i})\hat{C}_{i,I}^{BF}, \quad 0 < i \le I,\ I-i < j \le I. \end{aligned} \tag{2.7}$$

Particularly, we point out that, $\hat{C}_{i,I}^{BH} = \hat{\beta}_{I-i}\hat{C}_{i,I}^{CL} + (1-\hat{\beta}_{I-i})\hat{C}_{i,I}^{BF}$.

Similar to the idea of BF method, there is a more robust estimator provided by Cape-Cod (CC) method. Instead of assuming $\{\mu_i\}_{i=0}^I$ as the prior estimates for the ultimate claims for each accident year, a general overall loss ratio κ is defined, noting that the loss ratio is the total loss incurred divided by the total premium earned. Assume that there exists a premium pattern $\{\Pi_i\}_{i=0}^I$ such that $\mathsf{E}[C_{i,j}] = \kappa\Pi_i\beta_j$, $0 \le i, j \le I$.

To estimate the ultimate claims, one has to first estimate the overall loss ratio. A general estimation formula is proposed as

$$\hat{\kappa}^{CC} = \frac{\sum_{i=0}^I C_{i,I-i}}{\sum_{i=0}^I \hat{\beta}_{I-i}\Pi_i}.$$

Noting that $\beta_I = 1$, we have

$$\hat{C}_{i,I}^{CC} = C_{i,I-i} + (1-\hat{\beta}_{I-i})\hat{\kappa}^{CC}\Pi_i, \quad 0 < i \le I.$$

To summarize, non-parametric models are easy to understand and widely used in practice. However, most of the models do not support further diagnosis of the estimators.

2.4 Parametric Models

As mentioned in the previous section, although non-parametric models for claim reserving is widely accepted and appreciated in practice, there is still a need to consider parametric models. The advantages of the parametric models are: firstly, it is easier to estimate the uncertainty (such as variance) of the estimators under a parametric framework; secondly, the definition of $C_{i,j}$'s or $X_{i,j}$'s can be more specific, either claim counts or claim amount under each model. There are parametric models designed to satisfy the general assumptions of the non-parametric models, and can be treated as the underlying stochastic frameworks of the non-parametric models. Some models for incremental claims are summarized in Table 2.3. There is a general assumption that *the incremental claims $X_{i,j}$ in each cell (i,j) are either independent or conditionally independent.* In Table 2.3, the Poisson-Gamma model and the Negative-Binomial model also belong to the class of the over dispersed Poisson models. For the Poisson-Gamma model, the ultimate claims could be estimated by the posterior mean $\mathsf{E}[\Theta_i|\mathcal{D}_I]$ which is also a Bayesian model similar to the BH method. The Poisson model and the over dispersed Poisson models are normally used for estimating claim counts while the Gamma model and the Log-normal model are the models for estimating claim amount.

Poisson Model	$X_{i,j} \sim Poisson(\mu_i \gamma_j)$, the development pattern $\{\gamma_j\}_{j=0}^{I}$ satisfy $\sum_{j=0}^{I} \gamma_j = 1$.
Poisson-Gamma Model	$X_{i,j}\|\Theta \sim Poisson(\Theta_i \gamma_j)$ where $\Theta_i \sim Gamma(a_i, b_i)$, the development pattern $\{\gamma_j\}_{j=0}^{I}$ satisfy $\sum_{j=0}^{I} \gamma_j = 1$.
Negative-Binomial Model	$X_{i,j}$ has a conditional Negative-Binomial distribution and there exists a development pattern $\{f_j\}_{j=0}^{I-1}$ such that for all $0 \leq i \leq I$ and $1 \leq j \leq I$, $\mathsf{E}[X_{i,j}\|C_{i,0}, \ldots, C_{i,j-1}] = C_{i,j-1}(f_{j-1} - 1)$, and $\mathsf{Var}[X_{i,j}\|C_{i,0}, \ldots, C_{i,j-1}] = C_{i,j-1}(f_{j-1} - 1) f_{j-1}$.
Gamma Model	Individual payment $X_{i,j}^{k} \sim Gamma(\nu, \frac{\nu}{m_{i,j}})$, and $X_{i,j} = \sum_{k=0}^{r_{i,j}} X_{i,j}^{k}$, $r_{i,j}$ denotes the deterministic number of payments, $m_{i,j}$ is the average amount of each payment.
Log-normal Model	$log(X_{i,j}) \sim \mathcal{N}(m_{i,j}, \sigma^2)$.

Table 2.3: Summary of some parametric models for incremental claims

Among the parametric models, there is a compound model that is also worth mentioning. It is extended from the Gamma model. While keeping all the assumptions of the Gamma model unchanged, further assumes that the number of payments $r_{i,j}$ is now a random variable. Let $R_{i,j}$ be the number of payments which are Poisson distributed with mean $r_{i,j}$, independent of individual claim payments. Now $X_{i,j}$ follows a compound Poisson distribution. This model is often referred to as the Tweedie's compound Poisson model, (Wüthrich, 2003 and Peters et al., 2009).

Wüthrich and Merz (2008) has more detailed discussions on various parametric models in addition to the ones mentioned here.

2.5 IBNR and INAR

As the techniques for loss reserving become more sophisticated and well-developed, the models incorporating time series have drawn some attention as well, but the models may not be theoretically suitable for modelling claim counts because they do not model integer valued random variables.

Kremer (1995) has proposed an idea of modelling the development of claim counts based on the integer-valued autoregressive (INAR) process. We introduce briefly in this section the modelling idea, formulas derived, and the estimation method presented in Kremer (1995).

Let $C_{i,j}$ be the cumulative number of reported claims. Suppose that for each $j = 1, \ldots, I$,

$$C_{i,j} = \beta_{j-1} \circ C_{i,j-1} + e_{i,j}, \quad i = 0, \ldots, I,$$

in which

1. β_j, $j = 0, \dots, I-1$, are non-negative;

2. $\beta \circ C = \lfloor \beta \rfloor C + (\beta - \lfloor \beta \rfloor) \circ C$ with $\lfloor \beta \rfloor$ denotes the integer part of β and $(\beta - \lfloor \beta \rfloor) \circ C = \sum_{k=1}^{C} Y_k$, where Y_k's are i.i.d. and follow the Bernoulli distribution with parameter $\beta - \lfloor \beta \rfloor$, which means that $\mathsf{P}(Y_k = 1) = 1 - \mathsf{P}(Y_k = 0) = \beta - \lfloor \beta \rfloor$;

3. $e_{i,j}$'s are integer-valued random variable for any $0 \leq i \leq I$, $1 \leq j \leq I$, and $i + j \leq I$. Each $e_{i,j}$ is independent of $C_{i,j-1}$ with $\mathsf{E}[e_{i,j}|C_{i,j-1}] = 0$ and $\mathsf{Var}[e_{i,j}|C_{i,j-1}] = \sigma_j^2 \cdot f(C_{i,j-1})$, where σ_j^2 is a non-negative finite number and f is a non-negative given function on the non-negative real numbers.

Denote that $\alpha_j = \beta_j - \lfloor \beta_j \rfloor$, we can easily get

$$\begin{aligned} \mathsf{E}[C_{i,j}|C_{i,j-1}] &= \beta_{j-1} \cdot C_{i,j-1}, \\ \mathsf{Var}[C_{i,j}|C_{i,j-1}] &= \alpha_{j-1} \cdot (1 - \alpha_{j-1}) \cdot C_{i,j-1} + \sigma_j^2 \cdot f(C_{i,j-1}). \end{aligned} \tag{2.8}$$

According to the conditional expectation of $C_{i,j}$, for any i and j such as $i + j > I$, $C_{i,j}$ can be estimated by

$$\hat{C}_{i,j} = \beta_j \cdot \hat{C}_{i,j-1}.$$

Noting that the prediction could give non integer-valued results, one can round them to their nearst integer values afterwards. The prediction formula is similar to the Chain-ladder method as seen in Equation (2.3).

The estimation of the model parameters β_j, $j = 0, \dots, I-1$, and σ_j's can be obtained by the iterative weighted conditional least squares (IWCLS) estimation method which minimize the sum of squares of the error between the observation and its conditional mean with the reciprocal of the conditional variance being the weight. The estimates are updated iteratively until it converges. Theoretically, one has to minimize:

$$\sum_{i=0}^{I-j} \frac{(C_{i,j} - \beta_{j-1} \cdot C_{i,j-1})^2}{D_{i,j}(\alpha_{j-1}, \sigma_j^2)}, \qquad 1 \leq j \leq I, \tag{2.9}$$

where

$$D_{i,j}(\alpha_{j-1}, \sigma_j^2) = \alpha_{j-1} \cdot (1 - \alpha_{j-1}) \cdot C_{i,j-1} + \sigma_j^2 \cdot f(C_{i,j-1}).$$

The iterative process of the estimation can be described as follows. Because both α_{j-1} and σ_j^2 are unknown, the minimization of (2.9) can not be directly obtained. First one can find a preliminary estimator $\hat{\beta}_{j-1}$ for β_{j-1} with $D_{i,j}(\alpha_{j-1}, \sigma_j^2)$ replaced simply by 1, and then $\hat{\alpha}_{j-1} = \hat{\beta}_{j-1} - \lfloor \hat{\beta}_{j-1} \rfloor$. The next step is to find the estimates for σ_j^2's by minimize:

$$\sum_{i=0}^{I-j} \left((C_{i,j} - \hat{\beta}_{j-1} \cdot C_{i,j-1})^2 - D_{i,j}(\hat{\alpha}_{j-1}, \sigma_j^2) \right), \qquad 1 \leq j \leq I.$$

After getting the estimates for σ_j^2's, insert $\hat{\alpha}_{j-1}$ and $\hat{\sigma}_j^2$ into (2.9) and get the updated estimates for β_{j-1}, $j = 1, 2, \ldots, I$. Repeat the above steps until the estimation converges to the desired estimates of β_{j-1}'s.

One interesting finding of Kremer (1995) is that when assuming $f(C_{i,j-1}) = C_{i,j-1}$, the estimator of β_j is simplified to the one obtained by the classical Chain-ladder method as discussed in Section 2.3. In this sense, this model can be seen as an extension of the traditional models.

2.6 Frequency-Severity Techniques

Among all the models mentioned above, non-parametric models do not specify a clear interpretation of $C_{i,j}$'s or $X_{i,j}$'s, and work for both estimating claim counts and claim amount. As for parametric models, some are more suitable for modelling claim counts, while the others are more reasonable for modelling claim amount. Especially, the Poisson model is usually refer to as a claim counts model because it assumes that the data are all integer values and Poisson distribution is a commonly used distribution to model the claim frequency. The Gamma model and the Tweedie's compound Poisson model can only be used for modelling claim amount because it requires the information of number of payments during each development period. The INAR model described in Section 2.5 is clearly only suitable for claim counts based on its own set up.

In practice, it is also valuable to look at both the claim counts data and the claim amount data. Sometimes the patten of claim counts development and the average claim payments can be used as a diagnosis tool to examine whether an organization is undergoing change in operations, philosophy, or management (Friedland, 2010).

The Frequency-Severity techniques, in short, refer to the methods that estimate the ultimate claim loads by the product of the estimated ultimate claim counts (frequency times exposure) and the ultimate severities (Friedland, 2010). Hence, we can use our models for claim counts to first estimate the ultimate claim counts and multiply the estimated claim severity to get the ultimate claim loads. However, the estimation of claim severity under a specific claim counts model will not be discussed in this project.

Chapter 3

The Model for Claim Reserving with INAR Processes

In Section 2.5, we have presented the idea of using the integer-valued autoregressive (INAR) process to study the reserving problem for IBNR claims, introduced in Kremer (1995). Based on this idea and the rationale of the frequency-severity techniques, we propose a new parametric model for the unclosed claims (the total number of reported but not yet settled claims, or not enough reported claims), which we refer to as the Poisson INAR model in the rest of the context. This chapter is organized as follows. Section 3.1 lists the model assumptions with some comments. Sections 3.2 and 3.3 present the major characteristics of the model such as moments, conditional mean and variance, and also predictions. Section 3.4 suggests some methods of estimating the model parameters.

3.1 Model Assumptions

Although the modelling idea comes from Kremer (1995), but the Poisson INAR model we discuss here is quite different from the model proposed by Kremer (1995). The general assumption of Kremer's model is that the cumulative number of reported claims $C_{i,j}$ satisfies

$$C_{i,j} = \lfloor \beta_{j-1} \rfloor C_{i,j-1} + (\beta_{j-1} - \lfloor \beta_{j-1} \rfloor) \circ C_{i,j-1} + e_{i,j}, \quad i = 0, \dots, I, j = 1, \dots, I, \tag{3.1}$$

(see Section 2.5 in details). It can be seen from (3.1) that the cumulative claims $C_{i,j}$ is the summation of three terms: 1) the integer part of the development factor times the previous year's cumulative claims $C_{i,j-1}$, 2) the decimal part of the development factor being the probability of having newly reported claims with the previous year's cumulative claims as a base unit and 3) an error term. The second part is kind of an INAR process. In addition, as proved by Equation (2.8), Kremer's model satisfies the classical CL assumptions presented by (2.2). Although Kremer's model satisfies the classical CL assumptions, the assumption

(3.1) itself is quite difficult to be interpreted in practice. Instead of using an INAR process to model part of the incremental claims, we introduce the idea of constant close rate and use the INAR process to model the unclosed claims. As we have mentioned earlier, the unclosed claims (open claims, i.e., the reported but not yet settled claims) triangle is also a commonly used way of organizing the claim data. We present first the assumptions of the Poisson INAR model proposed.

Define that the unclosed claims $C_{i,j}$ in cell (i, j) is the total number of claims that occur in accident year i and have been reported up to development period j but have not yet been settled at the end of development year j.

Assumption 1.

- Unclosed claims $C_{i,j}$ of different accident years i are independent, i.e., $C_{i,j}$ and $C_{l,k}$ are independent for any j and k when $i \neq l$.
- There exist parameters $\mu_0, \ldots, \mu_I$ and $\gamma_0, \ldots, \gamma_I$ such that the newly reported claims $X_{i,j}$ incurred in accident year i but reported with j years of delay are independently Poisson distributed with $\mathsf{E}[X_{i,j}] = \mu_i \gamma_j$, for all $0 \leq i, j \leq I$, and $\sum_{j=0}^{I} \gamma_j = 1$.
- The unclosed claims $C_{i,j}$ of different accident years i follow an INAR process such that

$$C_{i,j} = \rho \circ C_{i,j-1} + X_{i,j}, \qquad 0 \leq i, j \leq I, \tag{3.2}$$

  with $\rho \circ C_{i,j-1} = \sum_{k=1}^{C_{i,j-1}} Y_k$, where $Y_k \sim Bermoulli(\rho)$ and $0 \leq \rho \leq 1$, noting that $C_{i,-1} = 0$.

Remark 1.

- Note that in our Poisson INAR model, $1 - \rho$ can be interpreted as the constant close rate for each cell. Parameter μ_i is the total expected number of claims in accident year i, while parameter γ_j is the proportion of the reported number of claims in development year j. In general, γ_j should be in a decreasing trend in j.
- In practice, it may not be proper to assume a constant close rate because 1) usually there is a large number of claims closed during the first several years and leave along the complicated cases that can not be settled until to the end of the development years and 2) the close rate may be different for different accident years or calendar years as a result of the change of various environment factors. It may be more reasonable to assume a constant close rate at the tail, but for simplicity, we keep the constant close rate as a major assumption in this project.
- Another constraint of this model is that it assumes the newly reported claims can not be closed (settled) within the same year as they are reported, but it is not always

true for short-tail business like property damage. By ignoring the cases which can be settled within a year, the ultimate claim numbers may be underestimated because we do not include these cases in the development triangle.

- In addition, there may be reopened claims in the claim system, but the models treat them as the newly reported claims. As a result, the same claim may be counted twice. However, the number of reopened claims are relatively small in general, so ignoring them should not affect the estimation and prediction much.

3.2 Properties

The proposed Poisson INAR model has the following properties.

Proposition 1. The unclosed claims $C_{i,j}$ can be written as a summation of the not yet settled claims from all the past and current development years $j-k$, $0 \le k \le j$, that is,

$$C_{i,j} = \sum_{k=0}^{j} \rho^k \circ X_{i,j-k}, \qquad 0 \le i, j \le I. \tag{3.3}$$

Proof. According to (3.2), we have

$$\begin{aligned} C_{i,j} &= \rho \circ C_{i,j-1} + X_{i,j} \\ &= \rho \circ (\rho \circ C_{i,j-2} + X_{i,j-1}) + X_{i,j} \\ &= \rho \circ \rho \circ C_{i,j-2} + \rho \circ X_{i,j-1} + X_{i,j} \\ &= \rho^2 \circ C_{i,j-2} + \rho \circ X_{i,j-1} + X_{i,j} \\ &\vdots \\ &= \sum_{k=0}^{j} \rho^k \circ X_{i,j-k}. \end{aligned}$$

□

According to (3.3), the unclosed claims $C_{i,j}$ is the summation of the unclosed (not yet settled) claims from all the reported number of claims $X_{i,j-k}$ with reporting delay $j-k$ years, $k = 0, 1, \ldots, j$. The probability of the reported number of claims $X_{i,j-k}$ being still unclosed is ρ^k, which means that, the earlier the reporting date, the smaller the probability of being unsettled. On the condition that the total number of past unclosed claims $C_{i,j-h}$ is known, we can rewrite $C_{i,j}$ as

$$C_{i,j} = \sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} + \rho^h \circ C_{i,j-h}, \quad 0 \le i \le I, 1 \le j \le I,\ 1 \le h \le j. \tag{3.4}$$

In (3.4), the unclosed claims $C_{i,j}$ can be interpreted as the summation of the unclosed claims from all the reported number of claims $X_{i,j-k}$ with reporting delay $j-k$ years, $k = 0, 1, \ldots, h-1$ and the total number of past unclosed claims $C_{i,j-h}$. The probability of being unclosed for each claim within $C_{i,j-h}$ is ρ^h. Now consider the unclosed number of claims in the lower right of the development triangle. Based on the observed data $\mathcal{D}_I$, $C_{i,I-i}$'s are known for $i = 0, 1, \ldots, I$. Let $h = i + j - I$ in (3.4), we have

$$C_{i,j} = \sum_{k=0}^{i+j-I-1} \rho^k \circ X_{i,j-k} + \rho^{i+j-I} \circ C_{i,I-i}, \qquad 1 \leq i \leq I,\ j > I - i. \tag{3.5}$$

The outstanding claim numbers $R_{i,j}$ at the end of the development year j, $j = 0, 1, \ldots, I-1$, for a particular accident year i, is the summation of the not yet reported claims plus the unclosed claims left from this development period (the summation of IBNyR and IBNeR claims), which is defined by

$$R_{i,j} = \sum_{k=j+1}^{I} X_{i,k} + C_{i,j}. \tag{3.6}$$

When $j = I$, $R_{i,I} = C_{i,I}$, which is the unclosed claims at the end of the development period.

Proposition 2. The mean, variance and the auto-correlation of $C_{i,j}$ can be obtained as

$$\begin{aligned}
\mathsf{E}[C_{i,j}] &= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \mu_i, \\
\mathsf{Var}[C_{i,j}] &= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \mu_i = \mathrm{E}[C_{i,j}], \\
\mathsf{Cov}[C_{i,j}, C_{i,j-h}] &= \rho^h \mathsf{Var}[C_{i,j-h}] = \rho^h \left(\sum_{k=0}^{j-h} \rho^k \gamma_{j-h-k}\right) \mu_i = \left(\sum_{k=h}^{j} \rho^k \gamma_{j-k}\right) \mu_i.
\end{aligned} \tag{3.7}$$

Proof. Because $\rho^k \circ X_{i,j-k}$ follows a Bernoulli distribution when $X_{i,j-k}$ is given,

$$\begin{aligned}
\mathsf{E}[\rho^k \circ X_{i,j-k} | X_{i,j-k}] &= \rho^k \cdot X_{i,j-k}, \\
\mathsf{Var}[\rho^k \circ X_{i,j-k} | X_{i,j-k}] &= \rho^k \cdot (1 - \rho^k) \cdot X_{i,j-k}.
\end{aligned}$$

Taking the expectation of both sides of (3.3) gives

$$\begin{aligned}
\mathsf{E}[C_{i,j}] &= \mathsf{E}\left[\sum_{k=0}^{j} \rho^k \circ X_{i,j-k}\right] \\
&= \sum_{k=0}^{j} \mathsf{E}\left[\rho^k \circ X_{i,j-k}\right] \\
&= \sum_{k=0}^{j} \mathsf{E}[\mathsf{E}[\rho^k \circ X_{i,j-k} | X_{i,j-k}]]
\end{aligned}$$

$$= \sum_{k=0}^{j} \mathsf{E}[\rho^k \cdot X_{i,j-k}]$$
$$= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right)\mu_i.$$

According to Assumption 1, $X_{i,j}$ are independently Poisson distributed, for any $0 \leq i, j \leq I$, and therefore, $\rho^k \circ X_{i,j-k}$, $0 \leq k \leq j$, are also independent from each other. Hence, the variance of $C_{i,j}$ is

$$\begin{aligned}
\mathsf{Var}[C_{i,j}] &= \mathsf{Var}\left[\sum_{k=0}^{j} \rho^k \circ X_{i,j-k}\right] \\
&= \sum_{k=0}^{j} \mathsf{Var}[\rho^k \circ X_{i,j-k}] \\
&= \sum_{k=0}^{j} \left(\mathsf{Var}[\mathsf{E}[\rho^k \circ X_{i,j-k}|X_{i,j-k}]] + \mathsf{E}[\mathsf{Var}[\rho^k \circ X_{i,j-k}|X_{i,j-k}]]\right) \\
&= \sum_{k=0}^{j} \left(\mathsf{Var}[\rho^k \cdot X_{i,j-k}] + \mathsf{E}[\rho^k \cdot (1-\rho^k) \cdot X_{i,j-k}]\right) \\
&= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right)\mu_i.
\end{aligned}$$

As for the covariance of $C_{i,j}$ and $C_{i,j-h}$, when $h \geq 1$, according to (3.4), $C_{i,j}$ can be written as a summation of $\rho^k \circ X_{i,j-k}$, $0 \leq k \leq h-1$, and $C_{i,j-h}$, respectively. Furthermore, according to (3.3), $C_{i,j-h}$ can be written as a summation of $\rho^k \circ X_{i,j-h-k}$, $0 \leq k \leq j-h$, and if let $k = k+h$, $C_{i,j-h}$ can be written as a summation of $\rho^{k-h} \circ X_{i,j-k}$, $h \leq k \leq j$. Noting that $X_{i,j}$'s are independently Poisson distributed for any $0 \leq i, j \leq I$, we can conclude that $C_{i,j-h}$ is independent of $X_{i,j-k}$, $0 \leq k \leq h-1$. Hence,

$$\begin{aligned}
\mathsf{Cov}[C_{i,j}, C_{i,j-h}] &= \mathsf{Cov}\left[\sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} + \rho^h \circ C_{i,j-h}, C_{i,j-h}\right] \\
&= \mathsf{Cov}[\rho^h \circ C_{i,j-h}, C_{i,j-h}] \\
&= \mathsf{Cov}\left[\mathsf{E}[\rho^h \circ C_{i,j-h}|C_{i,j-h}], \mathsf{E}[C_{i,j-h}|C_{i,j-h}]\right] \\
&\quad + \mathsf{E}[\mathsf{Cov}\left[\rho^h \circ C_{i,j-h}, C_{i,j-h}|C_{i,j-h}]\right] \\
&= \mathsf{Cov}[\rho^h \cdot C_{i,j-h}, C_{i,j-h}] \\
&= \rho^h \mathsf{Var}[C_{i,j-h}] \\
&= \rho^h \left(\sum_{k=0}^{j-h} \rho^k \gamma_{j-h-k}\right)\mu_i
\end{aligned}$$

$$= \left(\sum_{k=h}^{j} \rho^k \gamma_{j-k} \right) \mu_i.$$

Note that Appendix A present an alternative proof of the auto-correlation formula. □

From Proposition 2, it is concluded that the Poisson INAR model is a non-dispersed model (variance equals to mean), which is not a desired property for claim counts. One possible solution would be to add a prior distribution for μ_i; see Appendix B for more details.

Moreover, the conditional expectation and variance of $C_{i,j}$ is given by

$$\begin{aligned} \mathsf{E}[C_{i,j}|C_{i,j-h}] &= \left(\sum_{k=0}^{h-1} \rho^k \gamma_{j-k} \right) \mu_i + \rho^h \cdot C_{i,j-h}, \\ \mathsf{Var}[C_{i,j}|C_{i,j-h}] &= \left(\sum_{k=0}^{h-1} \rho^k \gamma_{j-k} \right) \mu_i + \rho^h \cdot (1-\rho^h) \cdot C_{i,j-h}, \\ & \qquad 0 \le i \le I, 1 \le j \le I,\ 1 \le h \le j. \end{aligned} \tag{3.8}$$

Based on the observed data $\mathcal{D}_I$ and using (3.8), the expected number of unclosed claims in the lower right development triangle can be rewritten as

$$\begin{aligned} \mathsf{E}[C_{i,j}|\mathcal{D}_I] &= \mathsf{E}[C_{i,j}|C_{i,I-i}] = \left(\sum_{k=0}^{i+j-I-1} \rho^k \gamma_{j-k} \right) \mu_i + \rho^{i+j-I} \cdot C_{i,I-i}, \\ \mathsf{Var}[C_{i,j}|\mathcal{D}_I] &= \mathsf{Var}[C_{i,j}|C_{i,I-i}] = \left(\sum_{k=0}^{i+j-I-1} \rho^k \gamma_{j-k} \right) \mu_i + \rho^{i+j-I} \cdot (1-\rho^{i+j-I}) \cdot C_{i,I-i}, \\ & \qquad 1 \le i \le I,\ j > I-i. \end{aligned} \tag{3.9}$$

As suggested by Section 2.4, the parametric models are usually stochastic realizations of the traditional non-parametric models, however, the Poisson INAR model does not satisfy the traditional CL assumption (2.2) nor the BF assumption (2.4). The reason is that the CL and BF assumptions are typically for modelling cumulative claims while the Poisson INAR model works for the unclosed claims.

We introduce a claim development pattern $\{\beta_j\}_{j=0}^{I}$ with

$$\beta_j = \sum_{k=0}^{j} \rho^k \gamma_{j-k}. \tag{3.10}$$

According to Proposition 2 and Equation (3.8), for $0 \le i \le I$, $0 \le j \le I-1$,

$$\begin{aligned} \mathsf{E}[C_{i,0}] &= \beta_0 \mu_i, \\ \mathsf{E}[C_{i,j+h}|C_{i,0}, \ldots, C_{i,j}] &= \rho^h \cdot C_{i,j} + \left(\textstyle\sum_{k=0}^{h-1} \rho^k \gamma_{j+h-k} \right) \mu_i \\ &= \rho^h \cdot C_{i,j} + \left(\textstyle\sum_{k=0}^{j+h} \rho^k \gamma_{j+h-k} - \sum_{k=h}^{j+h} \rho^k \gamma_{j+h-k} \right) \mu_i \\ &= \rho^h \cdot C_{i,j} + (\beta_{j+h} - \rho^h \cdot \beta_j) \cdot \mu_i, \quad 1 \le h \le I-j. \end{aligned} \tag{3.11}$$

By (3.10), β_j is in decreasing order as the unclosed claims are getting less and less as to the end of the development year, and β_I should converge to 0 theoretically. Although (3.11) is very similar to the BF assumptions (2.4), the mechanism underlying is totally different. Equation (3.11) can be seen as the general non-parametric assumption for models regarding to the unclosed claims and can be used as a criteria to decide the appropriation of the models similar to our Poisson INAR model.

3.3 Prediction

To predict the outstanding liabilities (loss reserves) with our Poisson INAR model, the frequency-severity techniques could be applied. There are usually two commonly used methods which are described below.

- The first method has already been mentioned in Section 2.6, i.e., predicting the ultimate claim numbers first and then multiplying by the average claim severity to get the ultimate claim loads. Hence then, the outstanding liabilities are just the ultimate claim loads minus the cumulative claim payments.
- The second method is to get the outstanding liabilities directly by the product of the outstanding claim numbers (including both IBNyR and IBNeR claims) and the outstanding average claim severity.

In this project, we focus only on the estimation of the ultimate claim numbers and the outstanding claim numbers of our Poisson INAR model. We do not consider estimating the average claim severity to get the estimation of the outstanding liabilities.

3.3.1 Prediction Based on Known Model Parameters

Now first consider the estimation of the unclosed claims $C_{i,j}$. Based on (3.9), the estimator of $C_{i,j}$ for any $j > I - i$, $i = 1, 2, \ldots, I$, under the Poisson INAR model is given by

$$\hat{C}_{i,j}^{Poi\_INAR} = \mathsf{E}[C_{i,j}|\mathcal{D}_I] = \left(\sum_{k=0}^{i+j-I-1} \rho^k \gamma_{j-k}\right)\mu_i + \rho^{i+j-I} \cdot C_{i,I-i}. \tag{3.12}$$

If we write the prediction using the development pattern for the unclosed claims introduced in (3.11),

$$\begin{aligned}\hat{C}_{i,j}^{Poi\_INAR} &= \rho^{i+j-I} \cdot C_{i,I-i} + (\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}) \cdot \mu_i \\ &= \rho^{i+j-I} \cdot \beta_{I-i} \cdot \frac{C_{i,I-i}}{\beta_{I-i}} + (\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}) \cdot \mu_i.\end{aligned} \tag{3.13}$$

According to Proposition 2, $\mathsf{E}[C_{i,j}] = \beta_j \mu_i$ for any $0 \leq i, j \leq I$, and hence, $C_{i,I-i}/\beta_{I-i}$ can be seen as an empirical estimate for the ultimate claim numbers. Therefore, the prediction formula of $C_{i,j}$ has a linear credibility form that combines the prior and the observed information. In fact, the credibility formula holds for $\hat{C}_{i,j}^{Poi\_INAR}$ with weight $\rho^{i+j-I} \cdot \beta_{I-i}$ on the empirical estimate for the ultimate claim numbers and weight $\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}$ on the prior information.

According to (3.6), the outstanding claim numbers $R_{i,j}$ can be estimated by

$$\hat{R}_{i,j}^{Poi\_INAR} = \mu_i \cdot \left(\sum_{k=j+1}^{I} \gamma_k \right) + \hat{C}_{i,j}^{Poi\_INAR}, \quad I - i \leq j < I,\ 1 \leq i \leq I. \tag{3.14}$$

Particularly, the outstanding claims of the current calendar year for the accidents happened in year i is given by

$$\hat{R}_{i,I-i}^{Poi\_INAR} = \begin{cases} \mu_i \cdot \left(\sum_{k=I-i+1}^{I} \gamma_k \right) + C_{i,I-i}, & 1 \leq i \leq I, \\ C_{i,I-i}, & i = 0. \end{cases} \tag{3.15}$$

There exists a credibility type estimator for the ultimate claim numbers and we denote as $\hat{\mu}_i^{Poi\_INAR}$. For any $j \geq I - i$, $i = 1, 2, \ldots, I$,

$$\begin{aligned} \hat{\mu}_i^{Poi\_INAR} &= \frac{\hat{C}_{i,j}^{Poi\_INAR}}{\beta_j} \\ &= \frac{\rho^{i+j-I} \cdot \beta_{I-i}}{\beta_j} \cdot \frac{C_{i,I-i}}{\beta_{I-i}} + \left(1 - \frac{\rho^{i+j-I} \cdot \beta_{I-i}}{\beta_j} \right) \cdot \mu_i. \end{aligned}$$

Choose $j = I$ for any $1 \leq i \leq I$ in the above equation, and then

$$\hat{\mu}_i^{Poi\_INAR} = \frac{\rho^{i} \cdot \beta_{I-i}}{\beta_I} \cdot \frac{C_{i,I-i}}{\beta_{I-i}} + \left(1 - \frac{\rho^{i} \cdot \beta_{I-i}}{\beta_I} \right) \cdot \mu_i. \tag{3.16}$$

The above equation also holds when $i = 0$, and then $\hat{\mu}_0^{Poi\_INAR}$ depends entirely on the observed information. As shown in (3.16), the ultimate claim numbers can be estimated by the linear combination of $C_{i,I-i}/\beta_{I-i}$ and μ_i.

3.3.2 Prediction Based on Unknown Model Parameters

Section 3.3.1 presents the prediction for various figures when the model parameters are known. However in practice, all of the parameters are normally unknown and need to be estimated first before the prediction can be done. Denote the estimators of the models parameters as $\hat{\rho}$, $\hat{\mu}_i$ and $\hat{\gamma}_j$, respectively, $0 \leq i, j \leq I$. According to (3.12) and (3.13), the unclosed claims $C_{i,j}$ can be estimated by

$$\begin{aligned}\hat{C}_{i,j}^{Poi\_INAR} &= \hat{\mathsf{E}}[C_{i,j}|\mathcal{D}_I] \\ &= \hat{\rho}^{i+j-I} \cdot \hat{\beta}_{I-i} \cdot \frac{C_{i,I-i}}{\hat{\beta}_{I-i}} + (\hat{\beta}_j - \hat{\rho}^{i+j-I} \cdot \hat{\beta}_{I-i}) \cdot \hat{\mu}_i.\end{aligned} \tag{3.17}$$

It can be seen that the predicted unclosed claims $\hat{C}_{i,j}^{Poi\_INAR}$ is again a linear combination of the empirical estimates for the ultimate claim numbers and the estimated prior information $\hat{\mu}_i$.

Similar to (3.15), the outstanding claims of the current year for the accidents happened in year i is given by

$$\hat{R}_{i,I-i}^{Poi\_INAR} = \begin{cases} \hat{\mu}_i \cdot \left(\sum_{k=I-i+1}^{I} \hat{\gamma}_k \right) + C_{i,I-i}, & 1 \le i \le I, \\ C_{i,I-i}, & i = 0. \end{cases} \tag{3.18}$$

The ultimate claim numbers can be estimated simply by $\hat{\mu}_i$, $0 \le i \le I$. A credibility type estimator could also been obtained by

$$\hat{\mu}_i^{Poi\_INAR} = \frac{\hat{\rho}^i \cdot \hat{\beta}_{I-i}}{\hat{\beta}_I} \cdot \frac{C_{i,I-i}}{\hat{\beta}_{I-i}} + \left(1 - \frac{\hat{\rho}^i \cdot \hat{\beta}_{I-i}}{\hat{\beta}_I}\right) \cdot \hat{\mu}_i, \quad 0 \le i \le I. \tag{3.19}$$

Noting that when $i = 0$, $\hat{\mu}_0^{Poi\_INAR}$ depends entirely on the observed information. Moreover if μ_i's are known, the ultimate claim numbers can be estimated by the linear combination of $C_{i,I-i}/\hat{\beta}_{I-i}$ and μ_i.

3.3.3 Mean Square Error of Prediction

The accuracy of the prediction can be measured by the mean square error of prediction (MSEP). The MSEP is defined as the expected square error between the true and the predicted values, for example, $C_{i,j}$ and $\hat{C}_{i,j}^{Poi\_INAR}$ respectively, in our Poisson INAR model. In addition, the conditional MSEP (MSEP conditioning on the observed data) is also an effective way of measuring prediction accuracy. The following lemma presents the conditional MSEP and the unconditional MSEP of the estimator of the unclosed claims $C_{i,j}$ under the Poisson INAR model.

Lemma 1 (MSEP of the Poisson INAR Model)**.** If the model parameters are known, that is, $\hat{C}_{i,j}^{Poi\_INAR} = \mathsf{E}[C_{i,j}|\mathcal{D}_I]$ given by (3.12), denoting $\hat{C}_{i,j}^{Poi\_INAR}$ by $\hat{C}_{i,j}$ for simplicity, we have

$$\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I] = (\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}) \cdot \mu_i + \rho^{i+j-I} \cdot (1 - \rho^{i+j-I}) \cdot C_{i,I-i}.$$

The unconditional MSEP is the same as the conditional one.

Proof. By definition of the conditional MSEP and (3.12), we can easily get

$$\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I] = \mathsf{E}[(C_{i,j} - \hat{C}_{i,j})^2|\mathcal{D}_I] = \mathrm{Var}[C_{i,j}|\mathcal{D}_I].$$

According to (3.9), we can further write $\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I]$ as

$$\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I] = \left(\sum_{k=0}^{i+j-I-1} \rho^k \gamma_{j-k}\right) \mu_i + \rho^{i+j-I} \cdot (1 - \rho^{i+j-I}) \cdot C_{i,I-i}.$$

If it is written using the development pattern β_j, then

$$\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I] = (\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}) \cdot \mu_i + \rho^{i+j-I} \cdot (1 - \rho^{i+j-I}) \cdot C_{i,I-i}.$$

The unconditional MSEP is just the expectation of the conditional MSEP, which gives

$$\mathsf{MSEP}[\hat{C}_{i,j}] = \mathsf{E}[\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I]] = \mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I].$$

□

As for the MSEP of $\hat{R}_{i,I-i}^{Poi\_INAR}$, $1 \leq i \leq I$, when the model parameters are known and according to (3.15),

$$\begin{aligned}
\mathsf{MSEP}[\hat{R}_{i,I-i}^{Poi_INAR}] &= \mathsf{E}\left[(R_{i,I-i} - \hat{R}_{i,I-i}^{Poi_INAR})^2\right] \\
&= \mathsf{E}\left[\left(\left(\sum_{k=I-i+1}^{I} X_{i,k} + C_{i,I-i}\right) - \left(\sum_{k=I-i+1}^{I} \gamma_k \mu_i + C_{i,I-i}\right)\right)^2\right] \\
&= \mathsf{E}\left[\left(\sum_{k=I-i+1}^{I} X_{i,k} - \sum_{k=I-i+1}^{I} \gamma_k \mu_i\right)^2\right] \\
&= \mathsf{E}\left[\left(\sum_{k=I-i+1}^{I} X_{i,k} - \sum_{k=I-i+1}^{I} \mathsf{E}[X_{i,k}]\right)^2\right] \\
&= \mathsf{Var}\left[\sum_{k=I-i+1}^{I} X_{i,k}\right] \\
&= \sum_{k=I-i+1}^{I} \mathsf{Var}[X_{i,k}] \\
&= \left(\sum_{k=I-i+1}^{I} \gamma_k\right) \mu_i.
\end{aligned}$$

Finally, we consider the MSEP of $\hat{\mu}_i^{Poi\_INAR}$. When the parameters are all known and based on (3.16),

$$\begin{aligned}
\mathsf{MSEP}[\hat{\mu}_i^{Poi_INAR}] &= \mathsf{E}[(\hat{\mu}_i^{Poi_INAR} - \mu_i)^2] \\
&= \mathsf{E}\left[\left(\frac{\rho^i \cdot \beta_{I-i}}{\beta_I} \cdot \frac{C_{i,I-i}}{\beta_{I-i}} + \left(1 - \frac{\rho^i \cdot \beta_{I-i}}{\beta_I}\right) \cdot \mu_i - \mu_i\right)^2\right] \\
&= \mathsf{E}\left[\left(\frac{\rho^i \cdot \beta_{I-i}}{\beta_I} \cdot \left(\frac{C_{i,I-i}}{\beta_{I-i}} - \mu_i\right)\right)^2\right] \\
&= \mathsf{E}\left[\left(\frac{\rho^i}{\beta_I} \cdot (C_{i,I-i} - \beta_{I-i}\mu_i)\right)^2\right] \\
&= \left(\frac{\rho^i}{\beta_I}\right)^2 \mathsf{E}[(C_{i,I-i} - \mathsf{E}[C_{i,I-i}])^2] \\
&= \left(\frac{\rho^i}{\beta_I}\right)^2 \mathsf{Var}[C_{i,I-i}] \\
&= \left(\frac{\rho^i}{\beta_I}\right)^2 \beta_{I-i}\mu_i.
\end{aligned}$$

If the model parameters are unknown, that is, $\hat{C}_{i,j}^{Poi\_INAR} = \hat{\mathsf{E}}[C_{i,j}|\mathcal{D}_I]$ given by (3.17). Denote $\hat{C}_{i,j}^{Poi\_INAR}$ by $\hat{C}_{i,j}$ for simplicity, in this case,

$$\begin{aligned}
\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I] &= \mathsf{E}[(C_{i,j} - \hat{C}_{i,j})^2|\mathcal{D}_I] \\
&= \mathsf{E}[((C_{i,j} - \mathsf{E}[C_{i,j}|\mathcal{D}_I]) - (\hat{C}_{i,j} - \mathsf{E}[C_{i,j}|\mathcal{D}_I]))^2|\mathcal{D}_I] \\
&= \mathsf{Var}[C_{i,j}|\mathcal{D}_I] + \mathsf{E}[(\hat{C}_{i,j} - \mathsf{E}[C_{i,j}|\mathcal{D}_I])^2|\mathcal{D}_I]
\end{aligned}$$

where the first term $\mathrm{Var}[C_{i,j}|\mathcal{D}_I]$ is the MSEP of the estimator when the model parameters are known and the second term is the prediction error results from the uncertainty of the parameter estimations (Buchwalder et al., 2006). The first term can be estimated by

$$\widehat{\mathsf{Var}}[C_{i,j}|\mathcal{D}_I] = (\hat{\beta}_j - \hat{\rho}^{i+j-I} \cdot \hat{\beta}_{I-i}) \cdot \hat{\mu}_i + \hat{\rho}^{i+j-I} \cdot (1 - \hat{\rho}^{i+j-I}) \cdot C_{i,I-i},$$

but the explicit expression for the second term is difficult to obtain. According to (3.18), the MSEP of $\hat{R}_{i,i-i}^{Poi\_INAR}$ is given by

$$\begin{aligned}
&\mathsf{MSEP}[\hat{R}_{i,I-i}^{Poi_INAR}] \\
&= \mathsf{E}\left[(R_{i,I-i} - \hat{R}_{i,I-i}^{Poi_INAR})^2\right] \\
&= \mathsf{E}\left[\left(\left(\sum_{k=I-i+1}^{I} X_{i,k} + C_{i,I-i}\right) - \left(\sum_{k=I-i+1}^{I} \hat{\gamma}_k\hat{\mu}_i + C_{i,I-i}\right)\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathsf{E}\left[\left(\sum_{k=I-i+1}^{I} X_{i,k} - \sum_{k=I-i+1}^{I} \hat{\gamma}_k \hat{\mu}_i\right)^2\right] \\
&= \mathsf{E}\left[\left(\left(\sum_{k=I-i+1}^{I} X_{i,k} - \sum_{k=I-i+1}^{I} \mathsf{E}[X_{i,k}]\right) - \left(\sum_{k=I-i+1}^{I} \hat{\gamma}_k \hat{\mu}_i - \sum_{k=I-i+1}^{I} \gamma_k \mu_i\right)\right)^2\right] \\
&= \mathsf{Var}\left[\sum_{k=I-i+1}^{I} X_{i,k}\right] + \mathsf{E}\left[\left(\sum_{k=I-i+1}^{I} \hat{\gamma}_k \hat{\mu}_i - \sum_{k=I-i+1}^{I} \gamma_k \mu_i\right)^2\right].
\end{aligned}
$$

From above equation, the MSEP of $\hat{R}_{i,I-i}^{Poi\_INAR}$ is also a summation of prediction error under known parameters and the parameter estimation error. The MSEP of $\hat{\mu}_i^{Poi\_INAR}$ could be explained similarly as that of $\hat{R}_{i,I-i}^{Poi\_INAR}$.

It can be seen from the above expressions that the explicit formula for MSEP is difficult to obtain under unknown model parameters case, and therefore, we simply ignore the estimation error of parameters when estimating the MSEP in our project. The formulas for the estimated MSEPs are hence given by

$$
\begin{aligned}
\widehat{\mathsf{MSEP}}[\hat{C}_{i,j}|\mathcal{D}_I] &= (\hat{\beta}_j - \hat{\rho}^{i+j-I} \cdot \hat{\beta}_{I-i}) \cdot \hat{\mu}_i + \hat{\rho}^{i+j-I} \cdot (1 - \hat{\rho}^{i+j-I}) \cdot C_{i,I-i}, \\
\widehat{\mathsf{MSEP}}[\hat{R}_{i,I-i}^{Poi\_INAR}] &= \left(\textstyle\sum_{k=I-i+1}^{I} \hat{\gamma}_k\right) \cdot \hat{\mu}_i, \quad 1 \le i \le I, \\
\widehat{\mathsf{MSEP}}[\hat{\mu}_i^{Poi\_INAR}] &= \left(\frac{\hat{\rho}^i}{\hat{\beta}_I}\right)^2 \hat{\beta}_{I-i} \hat{\mu}_i.
\end{aligned}
$$

3.4 Estimation of the Parameters

The traditional INAR(1) process has already been discussed by Alosh and Alzaid (1987) and their results have been extended to the replicated INAR(1) process by Silva et al. (2005). Since the Poisson INAR model is quite similar to the replicated INAR(1) process, similar estimation methods can be adopted. Among all the methods, the Yuller-Walker and the least squares estimations are easier to be understood and computed, and they are discussed in the following sections.

3.4.1 Yuller-Walker Estimation

The main idea of Yuller-Walker estimation is to estimate ρ by the empirical estimates for the auto-covariance. Denote that $\xi_{i,j}(h) = \mathsf{Cov}[C_{i,j}, C_{i,j-h}]$. According to Proposition 2, the following equation holds:

$$
\xi_{i,j}(h) = \rho^h \xi_{i,j-h}(0), \qquad 0 \le i \le I,\ 1 \le j \le I,\ 1 \le h \le j.
$$

If we can get the estimates for all the $\xi_{i,j}(1)$ and $\xi_{i,j-1}(0)$ for $0 \leq i \leq I-1$ and $1 \leq j \leq I-i$, the estimator of ρ can be written as

$$\hat{\rho} = \frac{\sum\limits_{i=0}^{I-1}\sum\limits_{j=1}^{I-i} \hat{\xi}_{i,j}(1)}{\sum\limits_{i=0}^{I-1}\sum\limits_{j=1}^{I-i} \hat{\xi}_{i,j-1}(0)}.$$

However, we only have one observation for each cell (i,j), and $\xi_{i,j}(1)$ and $\xi_{i,j-1}(0)$ depend on both i and j, and therefore, the empirical estimates for $\xi_{i,j}(1)$ and $\xi_{i,j-1}(0)$ could not be obtained.

Now consider a special case of the model when all μ_i's are equal, being μ. It follows that the auto-covariance does not depend on accident year i, which results in $\xi_{i,j}(1) = \xi_j(1)$ and $\xi_{i,j-1}(0) = \xi_{j-1}(0)$ for any $0 \leq i \leq I$. Hence, for any development year j, $1 \leq j \leq I-1$, we have $1 + I - j$ pairs of observations for the previous and current development year, and they could be used to get the empirical estimates for $\xi_j(1)$'s and $\xi_{j-1}(0)$'s. That is,

$$\begin{aligned} \hat{\xi}_j(1) &= \frac{\sum_{i=0}^{I-j}(C_{i,j}-\bar{C}_{\cdot,j})(C_{i,j-1}-\bar{C}_{\cdot,j-1})}{(I-j+1)}, \\ \hat{\xi}_{j-1}(0) &= \frac{\sum_{i=0}^{I-j}(C_{i,j-1}-\bar{C}_{\cdot,j-1})^2}{(I-j+1)}, \end{aligned}$$

where $\bar{C}_{\cdot,j} = \sum_{i=0}^{I-j} C_{i,j}/(I-j+1)$, $\bar{C}_{\cdot,j-1} = \sum_{i=0}^{I-j} C_{i,j-1}/(I-j+1)$. Therefore, the unclosed rate ρ can be estimated by the weighted average of $\hat{\xi}_j(1)/\hat{\xi}_{j-1}(0)$, which is given by

$$\hat{\rho} = \frac{\sum_{j=1}^{I-1}\left[\sum_{i=0}^{I-j}(C_{i,j} - \bar{C}_{\cdot,j})(C_{i,j-1} - \bar{C}_{\cdot,j-1})\right]}{\sum_{j=1}^{I-1}\left[\sum_{i=0}^{I-j}(C_{i,j-1} - \bar{C}_{\cdot,j-1})^2\right]}. \tag{3.20}$$

Next we consider the estimation of μ_i's and γ_j's. Below we first show a proposition about the sum of $\mathsf{E}[C_{i,j}]$'s.

Proposition 3. The sum of $\mathsf{E}[C_{i,j}]$'s of the observed data of each row and each column has the following relationships:

$$\begin{aligned} \frac{\sum_{j=0}^{I-i}\mathsf{E}[C_{i,j}]}{\sum_{j=0}^{I-i}\mathsf{E}[C_{i-1,j}]} &= \frac{\mu_i}{\mu_{i-1}}, & 1 \leq i \leq I, \\ \sum_{i=0}^{I-j}\mathsf{E}[C_{i,j}] - \rho\sum_{i=0}^{I-j}\mathsf{E}[C_{i,j-1}] &= \gamma_j \sum_{i=0}^{I-j}\mu_i, & 0 \leq j \leq I. \end{aligned} \tag{3.21}$$

Proof. We begin with proving the relationship between adjacent rows. First, if summing both sides of the first equation of (3.7) for the ith row over j (from 0 to $I-i$, all the observed data in this row),

$$\sum_{j=0}^{I-i} \mathsf{E}[C_{i,j}] = \sum_{j=0}^{I-i} \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k} \right) \mu_i.$$

For the previous row,

$$\sum_{j=0}^{I-i} \mathsf{E}[C_{i-1,j}] = \sum_{j=0}^{I-i} \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k} \right) \mu_{i-1}.$$

Clearly, based on the above two equations, the first part of (3.21) holds.

For the adjacent columns, letting h=1 in (3.8) gives $\mathsf{E}[C_{i,j}|C_{i,j-1}] = \gamma_j \mu_i + \rho C_{i,j-1}$ for any $0 \leq i \leq I$ and $1 \leq j \leq I$, and thus,

$$\mathsf{E}[C_{i,j}] = \gamma_j \mu_i + \rho \mathsf{E}[C_{i,j-1}].$$

By summing both sides of the equation above over the observed data for column j, we get

$$\left(\sum_{i=0}^{I-j} \mathsf{E}[C_{i,j}] \right) = \gamma_j \left(\sum_{i=0}^{I-j} \mu_i \right) + \rho \left(\sum_{i=0}^{I-j} \mathsf{E}[C_{i,j-1}] \right), \quad 1 \leq j \leq I.$$

When $j = 0$, $\mathsf{E}[C_{i,0}] = \gamma_0 \cdot \mu_i$, so that $\left(\sum_{i=0}^{I} \mathsf{E}[C_{i,0}] \right) = \gamma_0 \left(\sum_{i=0}^{I} \mu_i \right)$. Because $C_{i,-1} = 0$ for all $0 \leq i \leq I$ by definition, the above equation also holds when $j = 0$. Now the proof completes. □

Using the estimator of ρ given by (3.20) and the observation $C_{i,j}$ as an approximation of $\mathsf{E}[C_{i,j}]$ and noting $\mu_i = \mu$, and by Proposition 3, we get

$$\left(\sum_{i=0}^{I-j} C_{i,j} \right) - \hat{\rho} \left(\sum_{i=0}^{I-j} C_{i,j-1} \right) = \gamma_j \cdot (I - j + 1) \cdot \mu, \quad 0 \leq j \leq I. \tag{3.22}$$

Because of the constraint that $\sum_{j=0}^{I} \gamma_j = 1$, the estimator $\hat{\mu}$ can be obtained as

$$\hat{\mu} = \sum_{j=0}^{I} \left(\frac{\sum_{i=0}^{I-j} C_{i,j} - \hat{\rho} \sum_{i=0}^{I-j} C_{i,j-1}}{I - j + 1} \right). \tag{3.23}$$

Plug in the estimator of μ to (3.22), the estimator of γ_j is obtained as

$$\hat{\gamma}_j = \frac{\sum_{i=0}^{I-j} C_{i,j} - \hat{\rho} \sum_{i=0}^{I-j} C_{i,j-1}}{(I - j + 1) \cdot \hat{\mu}}, \qquad j = 0, 1, \ldots, I. \tag{3.24}$$

Remark 2. If μ_i's are not all equal, according to Proposition 3, we could introduce a transformation to the observed data $\mathcal{D}_I$. Let

$$\Pi_i = \prod_{k=1}^{i} \left(\frac{\sum_{j=0}^{I-k} \mathsf{E}[C_{k,j}]}{\sum_{j=0}^{I-k} \mathsf{E}[C_{k-1,j}]} \right), \qquad 1 \leq i \leq I,$$

and $C^*_{i,j} = C_{i,j}/\Pi_i$. The transformed dataset has the same mean, variance and covariance formulae as the Poisson INAR model when all μ_i's are equal. However, the model dynamic is influenced by the transformation, i.e., $C^*_{i,j} \neq \rho \circ C^*_{i,j-1} + X^*_{i,j}$, because $X^*_{i,j} = X_{i,j}/\Pi_i$ does not follow a Poisson distribution. To conclude, Yuller-Walk estimation can only be applied to the case when all μ_i's are equal in our Poisson INAR model. .

Consider an even more simpler case when all μ_i's are equal and γ_j's, $j = 0, 1, \ldots, I$, are known. If all γ_j's are known, we are only left with two parameters to estimate. Not only the estimation is simplified, but also we could examine whether less unknown parameters could improve the estimation accuracy. In this case, the estimator of ρ is still given by (3.20), and (3.23) for $\hat{\mu}$ reduces to

$$\hat{\mu} = \frac{\sum_{i=0}^{I} \sum_{j=0}^{I-i} \left(C_{i,j} - \hat{\rho} C_{i,j-1}\right)}{\sum_{j=0}^{I} (I - j + 1)\gamma_j}. \tag{3.25}$$

The above equation can be obtained by taking a summation of both sides of (3.22) from $j = 0$ to $j = I$.

3.4.2 Conditional Least Squares Estimation

The main idea of the conditional least squares (CLS) estimation is to get the estimated parameters by minimizing the sum of the squared errors between the observed values and its conditional expectation.

According to (3.8), $\mathsf{E}[C_{i,j}|C_{i,j-1}] = \gamma_j \mu_i + \rho C_{i,j-1}$ for $0 \leq i \leq I$ and $1 \leq j \leq I$. This equation also holds for $j = 0$. To get the estimation of ρ, μ_i's and γ_j's, one has to minimize the objective function

$$Q(\theta) = \sum_{i=0}^{I} \sum_{j=0}^{I-i} (C_{i,j} - \rho C_{i,j-1} - \gamma_j \mu_i)^2$$

under the constraint $\sum_{j=0}^{I} \gamma_j = 1$. Parameter θ here denotes the parameters that needed to be estimated in the objective function. Because there is a constraint to the parameters, the Lagrange multiplier method is applied to solve the problem. The new objective function now becomes

$$Q^*(\theta) = \sum_{i=0}^{I} \sum_{j=0}^{I-i} (C_{i,j} - \rho C_{i,j-1} - \gamma_j \mu_i)^2 + \lambda \left(1 - \sum_{j=0}^{I} \gamma_j \right).$$

Taking derivatives with respect to ρ, μ_i's and γ_j's, respectively, we get

$$\begin{aligned}
\frac{\partial Q^*(\theta)}{\partial \rho} &= \textstyle\sum_{i=0}^{I}\sum_{j=0}^{I-i} -2(C_{i,j} - \rho C_{i,j-1} - \gamma_j \mu_i) C_{i,j-1}, \\
\frac{\partial Q^*(\theta)}{\partial \mu_i} &= \textstyle\sum_{j=0}^{I-i} -2(C_{i,j} - \rho C_{i,j-1} - \gamma_j \mu_i)\gamma_j, \\
\frac{\partial Q^*(\theta)}{\partial \gamma_j} &= \left(\textstyle\sum_{i=0}^{I-j} -2(C_{i,j} - \rho C_{i,j-1} - \gamma_j \mu_i)\mu_i\right) - \lambda.
\end{aligned}$$

By setting the three equations above to zero, the estimators of ρ, μ_i's and γ_j's can be obtained as

$$\begin{aligned}
\hat{\rho} &= \frac{\sum_{i=0}^{I}\sum_{j=0}^{I-i}(C_{i,j} - \hat{\gamma}_j \hat{\mu}_i) C_{i,j-1}}{\sum_{i=0}^{I}\sum_{j=0}^{I-i} C_{i,j-1}^2}, \\
\hat{\mu}_i &= \frac{\sum_{j=0}^{I-i}(C_{i,j} - \hat{\rho} C_{i,j-1})\hat{\gamma}_j}{\sum_{j=0}^{I-i}\hat{\gamma}_j^2}, \qquad i = 0, 1, \ldots, I, \\
\hat{\gamma}_j &= \frac{\sum_{i=0}^{I-j}(C_{i,j} - \hat{\rho} C_{i,j-1})\hat{\mu}_i + \frac{\hat{\lambda}}{2}}{\sum_{i=0}^{I-j}\hat{\mu}_i^2}, \qquad j = 0, 1, \ldots, I.
\end{aligned} \tag{3.26}$$

Summing over the expression of $\hat{\gamma}_j$ for j yields

$$\sum_{j=0}^{I}\left(\frac{\sum_{i=0}^{I-j}(C_{i,j} - \hat{\rho} C_{i,j-1})\hat{\mu}_i + \frac{\hat{\lambda}}{2}}{\sum_{i=0}^{I-j}\hat{\mu}_i^2}\right) = 1.$$

After regrouping the terms, we get

$$\hat{\lambda} = \frac{\left(\sum_{j=0}^{I}\left(\frac{\sum_{i=0}^{I-j}(C_{i,j} - \hat{\rho} C_{i,j-1})\hat{\mu}_i}{\sum_{i=0}^{I-j}\hat{\mu}_i^2}\right) - 1\right) \times 2}{-\left(\sum_{j=0}^{I}\frac{1}{\sum_{i=0}^{I-j}\hat{\mu}_i^2}\right)}.$$

Based on the derivations above, we have the following algorithm for estimating the unknown parameters.

Algorithm 1.

- First randomly select a starting value $\hat{\rho}_0$ between 0 and 1.
- Based on this known $\hat{\rho}_0$, denote $\hat{X}_{i,j} = C_{i,j} - \hat{\rho}_0 \cdot C_{i,j-1}$, $0 \le i \le I$, $0 \le j \le I - i$. The starting values of $\hat{\mu}_i$'s and $\hat{\gamma}_j$'s can be estimated by assuming $\hat{X}_{i,j}$'s approximately follow the Poisson model; see Section 2.4 for the definition of Poisson model. Therefore,

$$\begin{aligned}
\hat{\mu}_{0,0} &= \textstyle\sum_{j=0}^{I}(\hat{X}_{i,j}), \\
\hat{\mu}_{i,0} &= \frac{\sum_{j=0}^{I-i}\hat{X}_{i,j}}{\left(1 - \sum_{j=I-i+1}^{I}\hat{\gamma}_{j,0}\right)}, \qquad i = 1, 2, \ldots, I, \\
\hat{\gamma}_{j,0} &= \frac{\sum_{i=0}^{I-j}\hat{X}_{i,j}}{\left(\sum_{i=0}^{I-j}\hat{\mu}_{i,0}\right)}, \qquad j = 0, 1, \ldots, I.
\end{aligned}$$

- After obtaining the starting values, the estimates of the parameters can be updated iteratively using the equations given by (3.26). During each iteration, update $\hat{\rho}$ first, then update $\hat{\mu}_i$'s with the new $\hat{\rho}$ and the previous $\hat{\gamma}_j$'s, calculate the Lagrange multiplier $\hat{\lambda}$ for this iteration, and finally update $\hat{\gamma}_j$'s with the updated $\hat{\rho}$, $\hat{\mu}_i$'s and the calculated $\hat{\lambda}$.

- Repeat the iteration until the estimates coverages. The convergence criteria is to check whether the distance between the old and updated estimates is within a particular tolerance level. For example, choose the tolerance level to be 0.0001 to guarantee the accuracy of the estimates is within three decimal places.

Since the Yuller-Walker estimation works only for the case when all μ_i's are equal, we give the CLS estimation under equal μ_i's as well for comparison. Meanwhile, the case when all the μ_i's are equal and the γ_j's are known is also considered. The estimation formulas and the algorithm both become quite different from that under the general assumption.

Case I: All μ_i's are equal and all γ_j's are unknown

When $\mu_i = \mu$, $i = 0, 1, \ldots, I$, the objective function becomes

$$\tilde{Q}(\theta) = \sum_{i=0}^{I}\sum_{j=0}^{I-i}(C_{i,j} - \rho C_{i,j-1} - \gamma_j \mu)^2 + \lambda\left(1 - \sum_{j=0}^{I}\gamma_j\right).$$

Taking derivatives with respect to ρ, μ and γ_j, we obtain that

$$\begin{aligned}
\hat{\rho} &= \frac{\sum_{i=0}^{I}\sum_{j=0}^{I-i}(C_{i,j} - \hat{\gamma}_j\hat{\mu})C_{i,j-1}}{\sum_{i=0}^{I}\sum_{j=0}^{I-i}C_{i,j-1}^2}, \\
\hat{\mu} &= \frac{\sum_{i=0}^{I}\sum_{j=0}^{I-i}(C_{i,j} - \hat{\rho}C_{i,j-1})\hat{\gamma}_j}{\sum_{i=0}^{I}\sum_{j=0}^{I-i}\hat{\gamma}_j^2}, \qquad\qquad (3.27)\\
\hat{\gamma}_j &= \frac{\sum_{i=0}^{I-j}(C_{i,j} - \hat{\rho}C_{i,j-1})\hat{\mu} + \frac{\hat{\lambda}}{2}}{(I-j+1)\hat{\mu}^2}, \qquad j = 0, 1, \ldots, I.
\end{aligned}$$

The estimation formulae for ρ and μ can be rewritten as

$$\begin{aligned}
\Big(\sum_{i=0}^{I}\sum_{j=0}^{I-i}C_{i,j-1}^2\Big)\hat{\rho} + \Big(\sum_{i=0}^{I}\sum_{j=0}^{I-i}\hat{\gamma}_j C_{i,j-1}\Big)\hat{\mu} &= \sum_{i=0}^{I}\sum_{j=0}^{I-i}C_{i,j}C_{i,j-1}, \qquad (3.28)\\
\Big(\sum_{i=0}^{I}\sum_{j=0}^{I-i}\hat{\gamma}_j C_{i,j-1}\Big)\hat{\rho} + \Big(\sum_{i=0}^{I}\sum_{j=0}^{I-i}\hat{\gamma}_j^2\Big)\hat{\mu} &= \sum_{i=0}^{I}\sum_{j=0}^{I-i}\hat{\gamma}_j C_{i,j}.
\end{aligned}$$

Solve (3.28) by treating them as the equations regarding to $\hat{\rho}$ and $\hat{\mu}$, one can get

$$\begin{aligned}
\hat{\rho} &= \frac{f(\hat{\gamma})\cdot g(\hat{\gamma}) - b\cdot h(\hat{\gamma})}{f(\hat{\gamma})^2 - a\cdot g(\hat{\gamma})}, \qquad (3.29)\\
\hat{\mu} &= \frac{b\cdot f(\hat{\gamma}) - a\cdot g(\hat{\gamma})}{f(\hat{\gamma})^2 - a\cdot g(\hat{\gamma})},
\end{aligned}$$

where $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_I)$, and

$$\begin{aligned}
a &= \sum_{i=0}^{I}\sum_{j=0}^{I-i} C_{i,j-1}^2, \\
b &= \sum_{i=0}^{I}\sum_{j=0}^{I-i} C_{i,j}C_{i,j-1}, \\
f(\hat{\gamma}) &= \sum_{i=0}^{I}\sum_{j=0}^{I-i} \hat{\gamma}_j C_{i,j-1}, \\
g(\hat{\gamma}) &= \sum_{i=0}^{I}\sum_{j=0}^{I-i} \hat{\gamma}_j C_{i,j}, \\
h(\hat{\gamma}) &= \sum_{i=0}^{I}\sum_{j=0}^{I-i} \hat{\gamma}_j^2.
\end{aligned}$$

Under the constraint $\sum_{j=0}^{I} \gamma_j = 1$, the last equation in (3.27) gives

$$\sum_{j=0}^{I} \left(\frac{\sum_{i=0}^{I-j}(C_{i,j} - \hat{\rho}C_{i,j-1})\hat{\mu} + \frac{\hat{\lambda}}{2}}{(I-j+1)\hat{\mu}^2} \right) = 1.$$

After regrouping the terms, we further obtain

$$\hat{\lambda} = \frac{\left(\sum_{j=0}^{I} \left(\frac{\sum_{i=0}^{I-j}(C_{i,j} - \hat{\rho}C_{i,j-1})}{(I-j+1)\hat{\mu}} \right) - 1 \right) \times 2}{-\left(\sum_{j=0}^{I} \frac{1}{(I-j+1)\hat{\mu}^2} \right)}. \tag{3.30}$$

Algorithm 2.

- First randomly select a starting value $\hat{\rho}_0$ between 0 and 1.
- Based on this known $\hat{\rho}_0$, the starting values of $\hat{\mu}$ and $\hat{\gamma}_j$'s can be estimated by Equations (3.23) and (3.24), respectively.
- After obtaining the starting values, the estimates of the parameters can be updated iteratively using Equations (3.27), (3.29) and (3.30). During each iteration, update $\hat{\rho}$ and $\hat{\mu}$ first, then calculate the Lagrange multiplier $\hat{\lambda}$ for this iteration, and finally update $\hat{\gamma}_j$'s with the updated $\hat{\rho}$, $\hat{\mu}$ and the calculated $\hat{\lambda}$.
- Repeat the iteration until the estimates coverages. The convergence criteria is the same as Algorithm 1.

Case II: All μ_i's are equal and all γ_j's are known

Under this situation, the estimation becomes quite straightforward. The objective function that needs to be minimized is just $Q(\theta) = \sum_{i=0}^{I}\sum_{j=0}^{I-i}(C_{i,j} - \rho C_{i,j-1} - \gamma_j\mu)^2$. In this case, ρ and μ can be estimated directly using (3.29) with $\hat{\gamma}$ replaced by the true γ. It is a one step calculation and no iterative algorithm is needed.

3.4.3 Iterative Weighted Conditional Least Squares Estimation

The iterative weighted conditional least squares (IWCLS) estimation is one step further from the CLS estimation. The IWCLS estimation of the parameters is obtained by minimizing the sum of the squared error between each observation and its conditional mean, weighted by the inverse of the conditional variance. This estimation becomes more reasonable because less weights are put on the observations with bigger uncertainty.

According to (3.8) and noting that $C_{i,-1} = 0$, for any $0 \le i \le I$ and $0 \le j \le I$, we have

$$\begin{aligned}\mathsf{E}[C_{i,j}|C_{i,j-1}] &= \gamma_j\mu_i + \rho \cdot C_{i,j-1}, \\ \mathsf{Var}[C_{i,j}|C_{i,j-1}] &= \gamma_j\mu_i + \rho \cdot (1-\rho) \cdot C_{i,j-1}.\end{aligned}$$

To get the estimation of ρ, μ_i's and γ_j's, one has to minimize the objective function

$$Q_W(\theta) = \sum_{i=0}^{I}\sum_{j=0}^{I-i} \frac{(C_{i,j} - \rho C_{i,j-1} - \gamma_j\mu_i)^2}{\gamma_j\mu_i + \rho \cdot (1-\rho) \cdot C_{i,j-1}}$$

under the constraint $\sum_{j=0}^{I} \gamma_j = 1$, where θ denotes the parameters that need to be estimated in the objective function. The Lagrange multiplier method could also be applied to solve the problem. The new objective function is

$$Q_W^*(\theta) = \sum_{i=0}^{I}\sum_{j=0}^{I-i} \frac{(C_{i,j} - \rho C_{i,j-1} - \gamma_j\mu_i)^2}{\gamma_j\mu_i + \rho \cdot (1-\rho) \cdot C_{i,j-1}} + \lambda\Big(1 - \sum_{j=0}^{I}\gamma_j\Big).$$

Letting $D_{i,j}(\theta) = \gamma_j\mu_i + \rho \cdot (1-\rho) \cdot C_{i,j-1}$, the objective function (3.31) can be rewritten as

$$Q_W^*(\theta) = \sum_{i=0}^{I}\sum_{j=0}^{I-i} \frac{(C_{i,j} - \rho C_{i,j-1} - \gamma_j\mu_i)^2}{D_{i,j}(\theta)} + \lambda\Big(1 - \sum_{j=0}^{I}\gamma_j\Big). \tag{3.31}$$

Taking derivatives with respect to each parameter could be troublesome because of the existence of the denominator $D_{i,j}(\theta)$. The solution would be to treat $D_{i,j}(\theta)$ as known and let it be estimated using a set of estimates of the parameters in θ, for instance, the CLS estimation. Now taking derivatives of (3.31) with respect to ρ, μ_i's and γ_j's and letting them equal to zero, and by solving the system of equations, the following estimation results are obtained:

$$\begin{aligned}
\hat{\rho} &= \frac{\sum_{i=0}^{I}\sum_{j=0}^{I-i}\frac{1}{D_{i,j}(\hat{\theta})}(C_{i,j}-\hat{\gamma}_j\hat{\mu}_i)C_{i,j-1}}{\sum_{i=0}^{I}\sum_{j=0}^{I-i}\frac{1}{D_{i,j}(\hat{\theta})}C_{i,j-1}^2}, \\
\hat{\mu}_i &= \frac{\sum_{j=0}^{I-i}\frac{1}{D_{i,j}(\hat{\theta})}(C_{i,j}-\hat{\rho}C_{i,j-1})\hat{\gamma}_j}{\sum_{j=0}^{I-i}\frac{1}{D_{i,j}(\hat{\theta})}\hat{\gamma}_j^2}, \qquad i=0,1,\ldots,I, \\
\hat{\gamma}_j &= \frac{\sum_{i=0}^{I-j}\frac{1}{D_{i,j}(\hat{\theta})}(C_{i,j}-\hat{\rho}C_{i,j-1})\hat{\mu}_i+\frac{\hat{\lambda}}{2}}{\sum_{i=0}^{I-j}\frac{1}{D_{i,j}(\hat{\theta})}\hat{\mu}_i^2}, \qquad j=0,1,\ldots,I.
\end{aligned} \tag{3.32}$$

Summing over j for $j = 0, 1, \ldots, I$ of the expression of $\hat{\gamma}_j$ and solving for $\hat{\lambda}$, we get

$$\hat{\lambda} = \frac{\left(\sum_{j=0}^{I}\frac{\sum_{i=0}^{I-j}\frac{1}{D_{i,j}(\hat{\theta})}(C_{i,j}-\hat{\rho}C_{i,j-1})\hat{\mu}_i}{\sum_{i=0}^{I-j}\frac{1}{D_{i,j}(\hat{\theta})}\hat{\mu}_i^2}-1\right)\times 2}{-\left(\sum_{j=0}^{I}\frac{1}{\sum_{i=0}^{I-j}\frac{1}{D_{i,j}(\hat{\theta})}\hat{\mu}_i^2}\right)}.$$

The algorithm for obtaining the estimation of all the parameters is presented below.

Algorithm 3.

- First we need to have a set of starting values for the parameters. We can either choose their corresponding CLS estimations as the starting values or use the method suggested by Algorithm 1. (We choose the estimates in Algorithm 1 in this project to guarantee fair comparison between the CLS and IWCLS estimations.)

- After obtaining the starting values, the estimates of the parameters can be updated iteratively using Equation (3.32). During each iteration, first calculate $D_{i,j}(\hat{\theta})$ using the previous estimates for the parameters, then update $\hat{\rho}$ and $\hat{\mu}_i$'s, next calculate the Lagrange multiplier $\hat{\lambda}$ for this iteration, and finally update $\hat{\gamma}_j$'s with the updated $\hat{\rho}$, $\hat{\mu}_i$'s and the calculated $\hat{\lambda}$.

- Repeat the iteration until the estimates coverages. The convergence criterion is the same as in Algorithm 1.

In the following, we consider two special cases the same as in the CLS estimation.

Case I: All $\boldsymbol{\mu_i}$'s are equal and all $\boldsymbol{\gamma_j}$'s are unknown

When $\mu_i = \mu$, $i = 0, 1, \ldots, I$, the objective function becomes

$$\tilde{Q}_W(\theta) = \sum_{i=0}^{I} \sum_{j=0}^{I-i} \frac{(C_{i,j} - \rho C_{i,j-1} - \gamma_j \mu)^2}{D_{i,j}(\hat{\theta})} + \lambda \left(1 - \sum_{j=0}^{I} \gamma_j \right),$$

where $D_{i,j}(\hat{\theta}) = \hat{\gamma}_j \hat{\mu} + \hat{\rho} \cdot (1 - \hat{\rho}) \cdot C_{i,j-1}$. The derivation of the estimation formulae under this case is very similar to the CLS estimation when all μ_i's are equal, so only the results are shown in the following.

The estimators of ρ and μ are given by

$$\begin{aligned} \hat{\rho} &= \frac{f_W(\hat{\gamma}) \cdot g_W(\hat{\gamma}) - b_W \cdot h_W(\hat{\gamma})}{f_W(\hat{\gamma})^2 - a_W \cdot h_W(\hat{\gamma})}, \\ \hat{\mu} &= \frac{b_W \cdot f_W(\hat{\gamma}) - a_W \cdot g_W(\hat{\gamma})}{f_W(\hat{\gamma})^2 - a_W \cdot h_W(\hat{\gamma})}, \end{aligned} \tag{3.33}$$

where $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_I)$, and

$$\begin{aligned} a_W &= \sum_{i=0}^{I} \sum_{j=0}^{I-i} \frac{C_{i,j-1}^2}{D_{i,j}(\hat{\theta})}, \\ b_W &= \sum_{i=0}^{I} \sum_{j=0}^{I-i} \frac{C_{i,j} C_{i,j-1}}{D_{i,j}(\hat{\theta})}, \\ f_W(\hat{\gamma}) &= \sum_{i=0}^{I} \sum_{j=0}^{I-i} \frac{\hat{\gamma}_j C_{i,j-1}}{D_{i,j}(\hat{\theta})}, \\ g_W(\hat{\gamma}) &= \sum_{i=0}^{I} \sum_{j=0}^{I-i} \frac{\hat{\gamma}_j C_{i,j}}{D_{i,j}(\hat{\theta})}, \\ h_W(\hat{\gamma}) &= \sum_{i=0}^{I} \sum_{j=0}^{I-i} \frac{\hat{\gamma}_j^2}{D_{i,j}(\hat{\theta})}. \end{aligned}$$

The estimators of γ_j's and λ is given by

$$\begin{aligned} \hat{\gamma}_j &= \frac{\left(\sum_{i=0}^{I-j} \frac{1}{D_{i,j}(\hat{\theta})}(C_{i,j} - \hat{\rho} C_{i,j-1})\right)\hat{\mu} + \frac{\hat{\lambda}}{2}}{\left(\sum_{i=0}^{I-j} \frac{1}{D_{i,j}(\hat{\theta})}\right)\hat{\mu}^2}, \qquad j = 0, 1, \ldots, I, \\ \hat{\lambda} &= \frac{\left(\sum_{j=0}^{I} \frac{\sum_{i=0}^{I-j} \frac{1}{D_{i,j}(\hat{\theta})}(C_{i,j} - \hat{\rho} C_{i,j-1})}{\sum_{i=0}^{I-j} \frac{1}{D_{i,j}(\hat{\theta})}\hat{\mu}} - 1\right) \times 2}{-\left(\sum_{j=0}^{I} \frac{1}{\sum_{i=0}^{I-j} \frac{1}{D_{i,j}(\hat{\theta})}\hat{\mu}}\right)}. \end{aligned} \tag{3.34}$$

Algorithm 4.

- The choice of the starting values is the same as shown in Algorithm 2.
- After obtaining the starting values, the estimates of the parameters can be updated iteratively using Equations (3.33) and (3.34). During each iteration, first calculate

$D_{i,j}(\hat{\theta})$ using the previous estimates for the parameters, then update $\hat{\rho}$ and $\hat{\mu}$, next calculate the Lagrange multiplier $\hat{\lambda}$ for this iteration, and finally update $\hat{\gamma}_j$'s with the updated $\hat{\rho}$, $\hat{\mu}$ and the calculated $\hat{\lambda}$.

- Repeat the iteration until the estimates coverages. The convergence criterion is the same as in Algorithm 1.

Case II: All $\boldsymbol{\mu_i}$'s are equal and all γ_j's are known

Under this situation, ρ and μ can still be estimated using (3.33) but with $\hat{\gamma}$ replaced by the true γ where $D_{i,j}(\hat{\theta})$ is estimated using the previous estimates for ρ and μ. The algorithm in this case is quite similar to Algorithm 4 but with only two parameters needed to be updated. The starting value of μ for a given $\hat{\rho}_0$ can be obtained by using (3.25).

Chapter 4

Numerical Illustration

Chapter 3 gives a thorough introduction to the Poisson INAR model that can be used to model the unclosed claims triangle. In Section 3.4, three different parameter estimation methods are considered and we also present the algorithms when the model has different numbers of the model parameters needed to be estimated. In this chapter, we conduct a simulation study to investigate the effectiveness of different estimation methods. We simulate data under different triangle sizes to see the influence of the relative size of the observed data compared to the number of parameters to be estimated. Section 3.3 has discussed the prediction of the Poisson INAR model. In this chapter, we generate a random sample when $I = 14$ to illustrate the accuracy of the prediction results. The chapter is divided into two sections. Section 4.1 presents the estimation of the parameters based on the results of the simulation study, and Section 4.2 shows the prediction results using the simulated sample and summarizes the accuracy of various prediction figures.

4.1 Estimation of the Parameters

In order to know the efficiency of the three estimation methods and the influence of the relative size of the observed data compared to the number of unknown parameters, a simulation study is conducted in this section in order to evaluate the estimation performance.

We choose three different sizes of the development triangle for our study with the final settlement happening at $I = 6$, $I = 10$ and $I = 14$, respectively. For all different sizes of development triangles, in the simulation we set $\mu_i = \mu = 2000$, $i = 0, 1, \ldots, I$ and $\rho = 0.5$. The choice of γ_j's for each size of the development triangle is quite random but with some rules. They need to be added up to 1 and also in a non-increasing order because usually the latter development years have less proportion of IBNyR claims. The following γ_j's are selected:

$I = 6$: $\gamma = (0.5, 0.2, 0.1, 0.1, 0.05, 0.03, 0.02)$;

$I = 10$: $\gamma = (0.3, 0.2, 0.1, 0.1, 0.1, 0.08, 0.07, 0.03, 0.01, 0.005, 0.005)$;

I	The number of unknown parameters			The observed data size
	Case 1	Case 2	General Case	
6	2	9	15	28
10	2	13	23	66
14	2	17	31	120

Table 4.1: The number of parameters that need to be estimated under different I and different cases compared to the observed data size

$I = 14$: $\gamma = (0.4, 0.2, 0.1, 0.1, 0.06, 0.04, 0.02, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01)$.

A total number of 1000 random samples are generated for each set of the parameters (i.e., for $I = 6$, $I = 10$ and $I = 14$, respectively), so there are 3000 random samples in total for all three sizes of triangles.

Consider three cases in our simulation study as mentioned in Section 3.4. Case 1 is that all $\mu_i = \mu$ and γ_j's are known so that we are to estimate ρ and μ only. Case 2 is the case when the information $\mu_i = \mu$ is available but γ_j's are unknown, so we are to estimate ρ, μ and γ_j's. The General Case is the case without knowing the information that μ_i's are all the same and the γ_j's are also unknown. In this case, we have to use the algorithms developed for the general model assumptions. Each set of 1000 simulated samples are used repeatedly to do the estimation under Case 1, Case 2 and the General Case, also under different estimation methods. Using the same simulated data under different cases and different estimation methods eliminates the possible influence for the estimation accuracy due to the randomness of the samples. Table 4.1 summarizes the number of parameters that need to be estimated under different triangle sizes and different cases and the observed data size under different I's. From the table we can see that the observed data size grows quicker than the number of unknown parameters as I increases.

To get a thorough understanding of the estimation results under different cases, different estimation methods and different sizes of the triangle, we consider the bias and the mean square error of the estimators for various parameters, and finally the distribution of the estimators for ρ and μ. Under the General Case, the distribution of the average of the estimators of all μ_i's are drawn to compare with the one under the other two cases where only one μ is estimated.

4.1.1 Bias and Mean Square Error

First of all, we calculate the relative bias and the square root of the mean square error of the estimators under each triangle size, each case and each estimation method. The relative bias is given by the distance between the estimated and the true value of the parameters divided by the true value. We present the relative bias instead of bias here so that the accuracy of the estimates for different model parameters are comparable to each other. The mean square error of estimators is the variance of the estimators plus the square of the bias. We present the square root of the mean square error here so that it has the same scale as the model parameters. The results are organized and shown in Tables 4.2 to 4.4 in this section.

$I = 6$		YW		CLS		IWCLS	
		Bias	Root MSE	Bias	Root MSE	Bias	Root MSE
Case 1	$\hat{\rho}$	-0.09886	0.13785	0.00012	0.00734	0.00013	0.00707
	$\hat{\mu}$	0.06653	369.69462	0.00015	23.71062	0.00019	23.58873
Case 2	$\hat{\rho}$	-0.09886	0.13785	-0.10082	0.13302	-0.09031	0.12229
	$\hat{\mu}$	0.09143	506.13061	0.09325	488.40579	0.08358	449.56267
	$\hat{\gamma}_j$	-0.04096	0.38362	-0.04704	0.37890	-0.04314	0.37874
		0.01525	0.16070	0.01742	0.15982	0.01607	0.16051
		0.05674	0.15624	0.06553	0.15463	0.05980	0.15495
		0.02881	0.15587	0.03357	0.15443	0.03055	0.15484
		0.07480	0.17471	0.08609	0.17331	0.07913	0.17381
		0.09812	0.18570	0.11197	0.18448	0.10291	0.18497
		0.10959	0.19113	0.12304	0.19010	0.11394	0.19091
General Case	$\hat{\rho}$			-0.53641	0.32307	-0.56474	0.31173
	$\hat{\mu}_i$			0.49350	1189.16190	0.51845	1147.21756
				0.49344	1190.87668	0.51957	1149.22785
				0.49477	1195.87579	0.52096	1153.62683
				0.49264	1191.98240	0.51879	1148.78634
				0.49617	1197.97751	0.52171	1154.67218
				0.49411	1194.36324	0.51999	1152.12680
				0.49774	1202.61042	0.52321	1157.49310
	$\hat{\gamma}_j$			-0.29092	0.29298	-0.32249	0.28115
				0.10461	0.13274	0.11611	0.12999
				0.41869	0.11527	0.46492	0.11153
				0.22152	0.11041	0.24619	0.10512
				0.53953	0.12735	0.59818	0.12219
				0.66796	0.14160	0.73728	0.13670
				0.67519	0.15180	0.74432	0.14728

Table 4.2: Summary of relative bias and square root of mean square error of the estimators under $I = 6$. The true values of the parameters are: $\rho = 0.5$, $\mu_i = \mu = 2000$ and $\gamma = (0.5, 0.2, 0.1, 0.1, 0.05, 0.03, 0.02)$.

$I = 10$		YW		CLS		IWCLS	
		Bias	Root MSE	Bias	Root MSE	Bias	Root MSE
Case 1	$\hat{\rho}$	-0.12939	0.10556	-0.00093	0.00667	-0.00062	0.00624
	$\hat{\mu}$	0.09948	325.74874	0.00020	23.36263	0.00004	22.93440
Case 2	$\hat{\rho}$	-0.12939	0.10556	-0.12859	0.10432	-0.11889	0.09551
	$\hat{\mu}$	0.12600	411.44294	0.12522	406.59919	0.11580	372.91806
	$\hat{\gamma}_j$	-0.09420	0.20318	-0.09419	0.20301	-0.08909	0.20336
		-0.02122	0.13119	-0.02122	0.13118	-0.02003	0.13150
		0.07502	0.08091	0.07502	0.08088	0.07090	0.08106
		0.03712	0.07975	0.03711	0.07975	0.03491	0.08012
		0.01834	0.07930	0.01834	0.07930	0.01719	0.07961
		0.03486	0.07910	0.03487	0.07909	0.03287	0.07944
		0.03450	0.08097	0.03450	0.08096	0.03262	0.08129
		0.16822	0.09700	0.16823	0.09699	0.15918	0.09746
		0.45939	0.10994	0.45931	0.10993	0.43504	0.11032
		0.60156	0.11419	0.60126	0.11419	0.57224	0.11450
		0.32049	0.11497	0.32049	0.11497	0.30677	0.11530
General Case	$\hat{\rho}$			-0.66433	0.35031	-0.54401	0.29120
	$\hat{\mu}_i$			0.64673	1366.79175	0.52910	1134.86239
				0.64702	1366.37727	0.52978	1136.03474
				0.64746	1367.26176	0.53055	1138.26104
				0.64846	1369.59326	0.53128	1139.01081
				0.64669	1367.33203	0.52950	1138.17723
				0.64845	1370.16621	0.53114	1139.95040
				0.64977	1374.10748	0.53217	1142.78759
				0.64853	1371.82981	0.53130	1142.23022
				0.64821	1370.93267	0.53091	1140.58644
				0.64883	1371.59685	0.53201	1143.99267
				0.64559	1368.42685	0.52883	1140.21470
	$\hat{\gamma}_j$			-0.38082	0.16240	-0.33398	0.16280
				-0.08773	0.11003	-0.07687	0.11300
				0.30262	0.07699	0.26521	0.07588
				0.15504	0.06548	0.13576	0.06680
				0.08097	0.06150	0.07082	0.06374
				0.15052	0.05983	0.13173	0.06226
				0.14469	0.06048	0.12682	0.06321

– continued on next page

– continued from previous page

$I = 10$		YW		CLS		IWCLS	
		Bias	Root MSE	Bias	Root MSE	Bias	Root MSE
				0.68794	0.07423	0.60303	0.07727
				1.81471	0.08782	1.59319	0.09085
				2.24310	0.09549	1.97313	0.09817
				1.15178	0.09929	1.01764	0.10147

Table 4.3: Summary of relative bias and square root of mean square error of the estimators under $I = 10$. The true values of the parameters are: $\rho = 0.5$, $\mu_i = \mu = 2000$ and $\gamma = (0.3, 0.2, 0.1, 0.1, 0.1, 0.08, 0.07, 0.03, 0.01, 0.005, 0.005)$.

$I = 14$		YW		CLS		IWCLS	
		Bias	Root MSE	Bias	Root MSE	Bias	Root MSE
Case 1	$\hat{\rho}$	-0.03956	0.08254	-0.00061	0.00487	-0.00054	0.00432
	$\hat{\mu}$	0.03314	278.93873	0.00014	18.45920	0.00013	18.00123
Case 2	$\hat{\rho}$	-0.03956	0.08254	-0.04084	0.08042	-0.03017	0.06724
	$\hat{\mu}$	0.03820	323.11770	0.03947	314.74369	0.02902	263.69561
	$\hat{\gamma}_j$	-0.01511	0.34678	-0.01771	0.34544	-0.01339	0.34627
		0.00029	0.16781	0.00033	0.16768	0.00028	0.16791
		0.01623	0.10856	0.01891	0.10835	0.01455	0.10821
		0.00834	0.10808	0.00970	0.10788	0.00763	0.10795
		0.01704	0.10259	0.01990	0.10233	0.01487	0.10258
		0.02071	0.10484	0.02422	0.10456	0.01838	0.10485
		0.03828	0.11265	0.04441	0.11237	0.03375	0.11262
		0.05781	0.11669	0.06652	0.11645	0.05172	0.11681
		0.02927	0.11599	0.03378	0.11577	0.02630	0.11611
		0.00856	0.11701	0.01093	0.11680	0.00675	0.11713
		0.01068	0.11688	0.01184	0.11667	0.00963	0.11702
		-0.00341	0.11609	-0.00262	0.11589	-0.00317	0.11621
		0.00850	0.11692	0.00896	0.11672	0.00864	0.11706
		-0.00290	0.11689	-0.00246	0.11669	-0.00385	0.11702
		-0.01716	0.11625	-0.01647	0.11605	-0.01805	0.11639
	$\hat{\rho}$			-0.57655	0.31133	-0.31039	0.18164
				0.56367	1219.72536	0.30345	714.01897

– continued on next page

– continued from previous page

$I = 14$		YW		CLS		IWCLS	
		Bias	Root MSE	Bias	Root MSE	Bias	Root MSE
General Case	$\hat{\mu}_i$			0.56529	1223.49418	0.30497	717.63585
				0.56519	1224.10913	0.30449	716.39022
				0.56375	1221.82708	0.30267	712.11029
				0.56138	1215.70909	0.30146	710.13724
				0.56428	1221.59427	0.30319	713.17936
				0.56374	1221.34126	0.30287	712.11347
				0.56471	1223.02642	0.30368	715.51060
				0.56386	1221.80602	0.30320	715.12868
				0.56570	1225.00806	0.30432	714.76026
				0.56289	1218.99465	0.30233	712.13633
				0.56451	1222.27252	0.30388	715.39006
				0.56343	1222.16816	0.30311	715.94596
				0.56456	1222.23501	0.30384	714.13453
				0.56459	1223.60645	0.30375	715.30190
	$\hat{\gamma}_j$			-0.34445	0.25503	-0.21694	0.27884
				0.00734	0.15511	0.00479	0.15920
				0.35918	0.11015	0.22670	0.10529
				0.18302	0.09438	0.11574	0.09729
				0.38780	0.08231	0.24421	0.08773
				0.46810	0.07926	0.29507	0.08731
				0.82048	0.08454	0.51734	0.09384
				1.17422	0.08998	0.74050	0.09936
				0.59223	0.09180	0.37259	0.10021
				0.29635	0.09401	0.18337	0.10192
				0.15496	0.09476	0.09831	0.10236
				0.07570	0.09454	0.04485	0.10189
				0.04680	0.09544	0.03106	0.10275
				0.02285	0.09549	0.01067	0.10277
				0.00583	0.09500	-0.00397	0.10222

Table 4.4: Summary of relative bias and square root of mean square error of the estimators under $I = 14$. The true values of the parameters are: $\rho = 0.5$, $\mu_i = \mu = 2000$ and $\gamma = (0.4, 0.2, 0.1, 0.1, 0.06, 0.04, 0.02, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01)$.

Based on the tables, we have the following observations. First we compare the estimation results under different cases. It can seen that the YW estimates of ρ are the same under Case 1 and Case 2. The reason is that ρ is estimated by (3.20) and the formula relates only to the observed data. If looking at the relative bias of $\hat{\rho}$ and $\hat{\mu}$ under the two least squares estimation methods, under Case 1, the estimates can be accurate up to around 0.001% of the true value, while under Case 2, the estimates are around 10% of the true value. Under the General Case, the relative bias of $\hat{\rho}$ and $\hat{\mu_i}$'s are around 50%. Therefore, we can conclude that the least squares estimation methods do not perform well under the General Case and because of the high bias, the estimation results under the General Case is not that reliable. The variability of the estimators for the YW estimation does not change much from Case 2 to Case 1, i.e., reducing the model parameters does not improve the YW estimation a lot. However, under Case 1, the mean square errors of the estimators with the least squares estimations are much better than those under Case 2, which means that the accuracy of the CLS and IWCLS estimations depends highly on the number of parameters needed to be estimated. Moreover, we can see that the estimation results for the least squares estimations are much better under Case 2 compared to the General Case, which again confirms that less model parameters do lead to better estimates. However, we can see from the tables that the square root of MSE of the $\hat{\gamma}_j$'s are mostly bigger under Case 2 than that under the General Case. The reason is that although the estimations have much bigger bias under the General Case, the variance of the estimators are actually smaller than that under Case 2. Because the square root of MSE depends mainly on the variance of the estimators, the square root of MSE under Case 2 is bigger as a result. It means that the estimates under the General Case is biased but clustered quite together for γ_j's.

As has mentioned before, the estimation results under the General Case is not that reliable, and therefore we compare different estimation methods under Case 1 and Case 2 only. Under Case 1, the least squares estimation methods perform much better than the YW estimation and there are not much difference between the CLS and the IWCLS estimations although that the IWCLS estimation tends to have smaller MSE values. Under Case 2, the three methods do not show much difference especially when looking at the estimation results of the γ_j's. However, in terms of the MSE, the IWCLS estimators for ρ and μ are always smaller than those under the other two methods.

When comparing the estimation results of different sizes of the development triangle, there is a big improvement from $I = 10$ to $I = 14$ compared to from $I = 6$ to $I = 10$. Although the estimation under Case 1 is always very good, the estimation is improved from $I = 10$ to $I = 14$ while there is not much improvement from $I = 6$ to $I = 10$ in terms of the estimation error. Under Case 2, when $I = 6$ or $I = 10$, the estimation of γ_j's is quite deviated from the true value at the tail, while for $I = 14$, the relative biases of $\hat{\gamma}_j$'s are always small. Moreover, if looking at the relative bias of $\hat{\rho}$ and $\hat{\mu}$, the biases are around 10% when $I = 6$ and $I = 10$ while they are in the range of 3% to 4% when $I = 14$. We also observe that, as the dimension of the triangle becomes bigger, the IWCLS estimation is more efficient compared to the CLS estimation.

To conclude, under simple cases (when there are not many parameters needed to be estimated), the least squares estimations are quit accurate compared to the YW estimation, but as the unknown parameters get increased, all three methods becomes less reliable. The IWCLS estimation method always shows the smallest estimation error among the three, and as the size of development triangle grows, it performs better than the other two methods. Increasing the size of development triangle could improve the accuracy of the estimation, but the improvement is not significant at the early stage.

4.1.2 Distribution of $\hat{\rho}$

Instead of the relative bias of $\hat{\rho}$, we directly draw the box plots for $\hat{\rho}$ to guarantee that the scale of the plots does not change under different cases or triangle sizes, so the comparison becomes more straightforward. However, the histograms are free scaled under different cases and triangle sizes; by this way we could see the shape of the distribution more clearly. If the same scale is used, the distribution under Case 2 and the General Case would be flat compared to Case 1. In addition to the observations presented in Section 4.1.1, there are several new findings by looking at the distribution of the estimators. Figure 4.1 show the box plots of $\hat{\rho}$ and Figure 4.2 shows the histograms of $\hat{\rho}$.

First of all, we can clearly see the improvement of the YW estimation under Case 1 from $I = 6$ to $I = 14$ by the box plots and the histograms. By looking at Tables 4.2 and 4.3, we can see that there is not much improvement of the estimation from $I = 6$ to $I = 10$. Now by looking at Figure 4.1, we observe that although the interquartile range is smaller, the number of outliers actually gets bigger as I changes from 6 to 10. For the figures under Case 2, the distribution seems more and more centred around the true value as the size of the triangles grow bigger. If looking at the histograms under the General Case in Figure 4.2, we could see that there are two modes among the estimates of $\hat{\rho}$, so it seems a bimodal distribution, but as the triangle size becomes larger, this phenomenon gets weaker: when I equals to 10 or 14, this phenomenon does not exist under the IWCLS estimation methods . This again demonstrates that a larger size of development triangle, which has more observed data, could make the estimation algorithms more reliable and the IWCLS performs better than the CLS estimation in the sense that it converges effectively to its true value.

Secondly, we discuss the skewness of the estimators. From Figures 4.1 and 4.2, we observe that the distributions of $\hat{\rho}$ under Case 1 and Case 2 are skewed to left while under the General Case, its distribution is skewed to the right. The large negative bias obtained from all three estimation methods under Case 2 could be well explained now. From the box plots, we see that all the medians are below the true value, and further more, because the distribution is left-skewed, the mean is even smaller than the median; as a result, the bias of the estimators is negative and relatively large under all three estimation methods. For the General Case, the interquartile ranges are all below the true values, which again shows that the estimation results under the General Case are not that reliable.

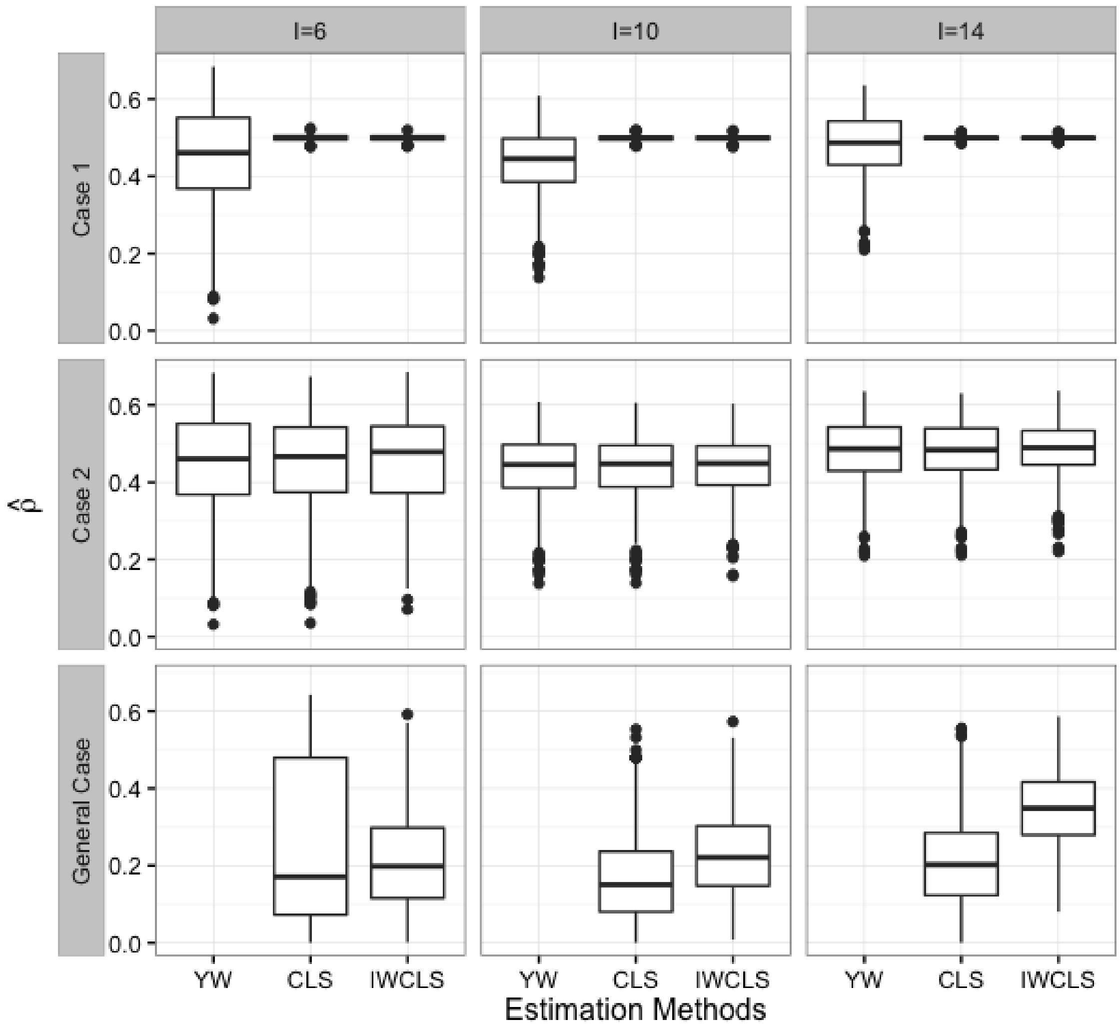

Figure 4.1: Box plots of $\hat{\rho}$. The true value of ρ is 0.5.

To conclude, by looking at the distribution of the estimators of ρ, the improvement of the estimations can be easily observed when I becomes large. However, since all the estimation methods give skewed distributions for the estimators of ρ and bimodal distributions under the CLS estimation method, other methods should be studied to get more accurate estimations. But fortunately when $I = 14$ and under Case 2, the estimation results are quite good and could be used to perform further studies.

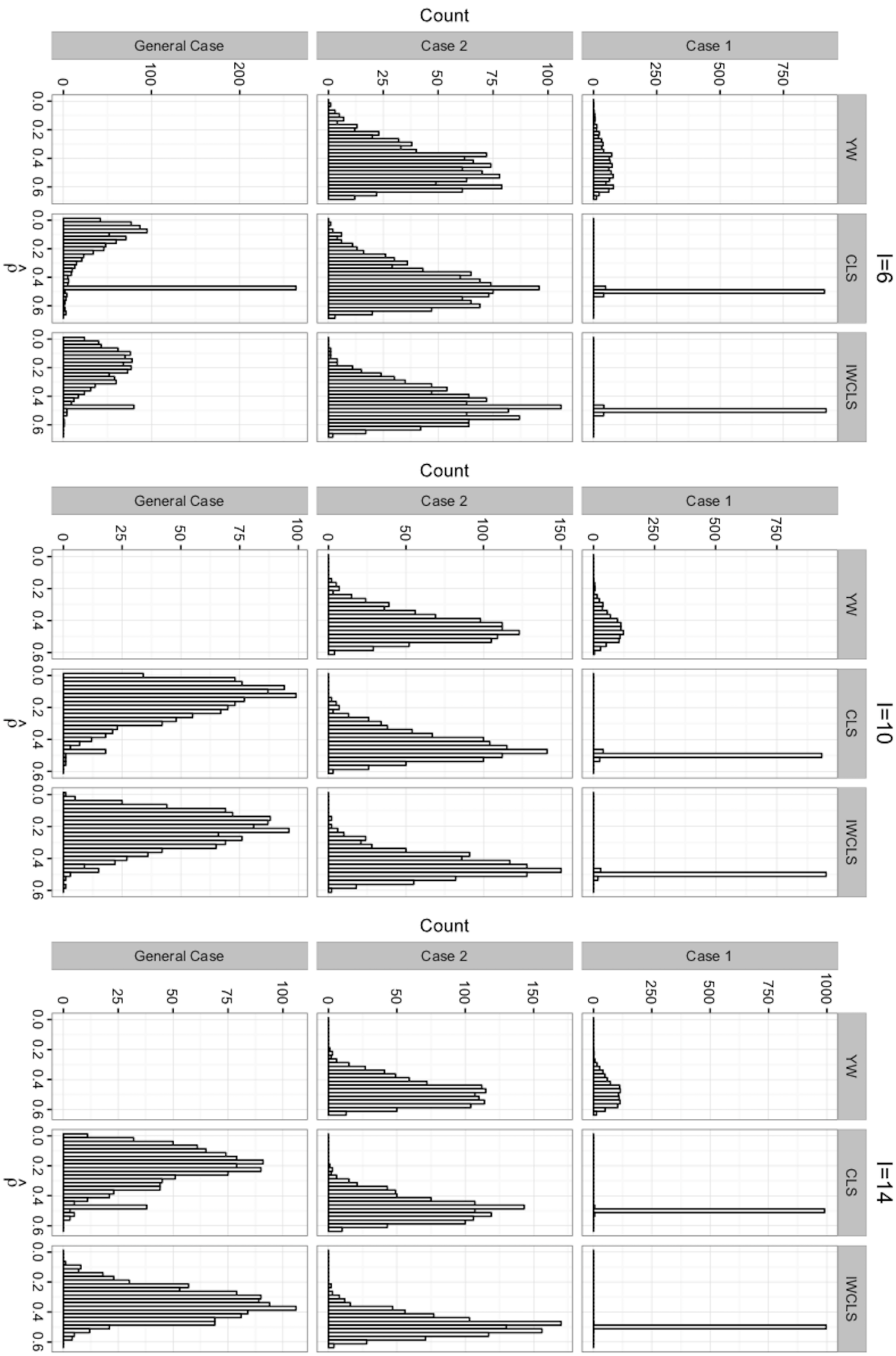

Figure 4.2: Histograms of $\hat{\rho}$. The true value of ρ is 0.5.

4.1.3 Distribution of $\hat{\mu}$

Similar to Section 4.1.2, the distributions of $\hat{\mu}$ under Case 1 and Case 2 and the distribution of the average of $\hat{\mu}_i$'s under the General Case are presented in Figures 4.3 and 4.4 in terms of the box plots and the histograms, respectively, under different methods and triangle sizes.

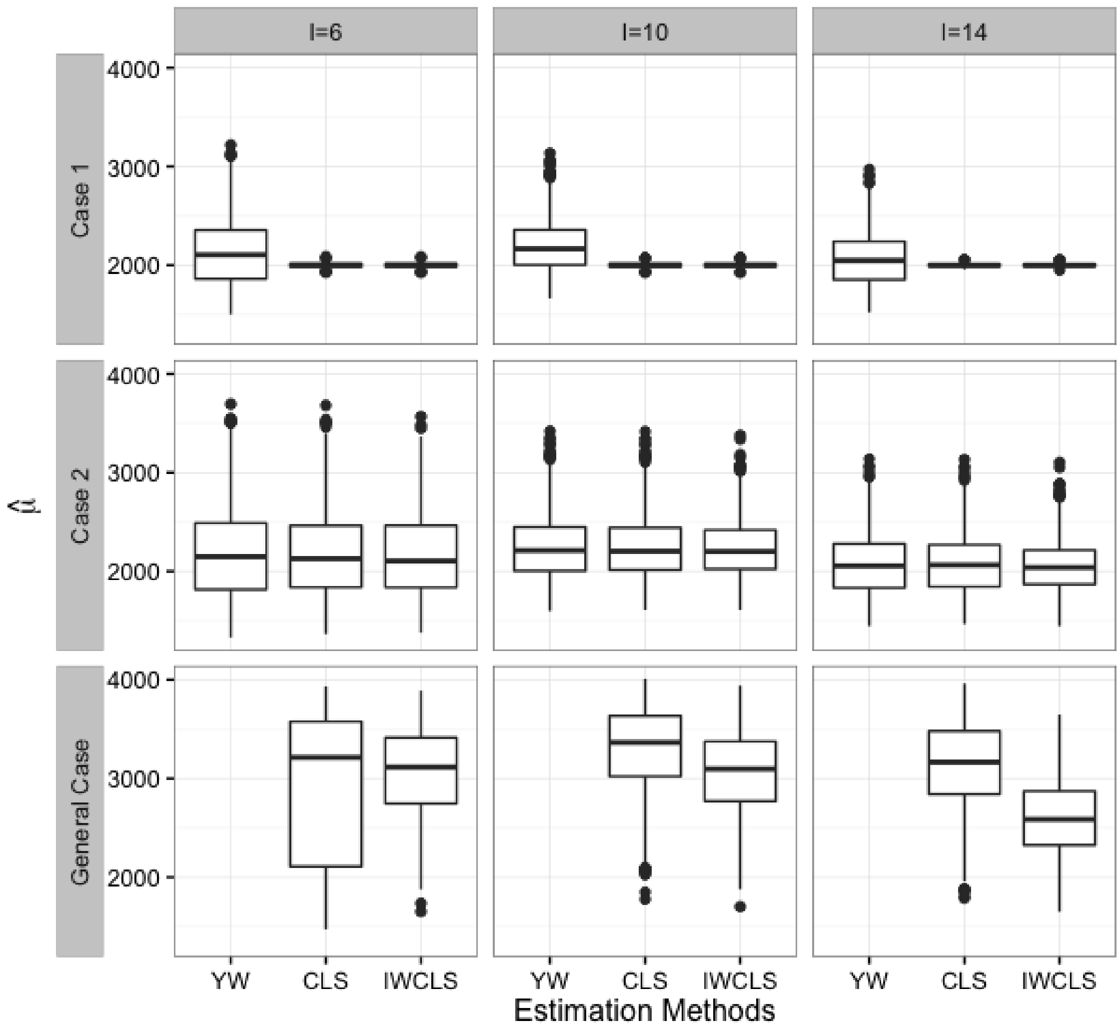

Figure 4.3: Box plots of $\hat{\mu}$. The true value of μ is 2000. Under the General Case, $\mu_i = \mu$ and the box plots of the averaged estimated μ_i's are shown.

The improvement of the YW estimation under Case 1 can be seen clearly from Figure 4.3 that the median of the estimates gets closer and closer to the true value and the interquartile range gets narrowed down. Under Case 2, there are not much difference between different estimation methods, but the distributions are skewed to the right. The skewness is bigger for $I = 10$ and $I = 14$ but the distribution has a smaller interquartile range compared to $I = 6$. Therefore, the estimation results do not improve much from $I = 6$ to $I = 10$, but get more accurate from $I = 10$ to $I = 14$. Because of the right skewness of the distribution,

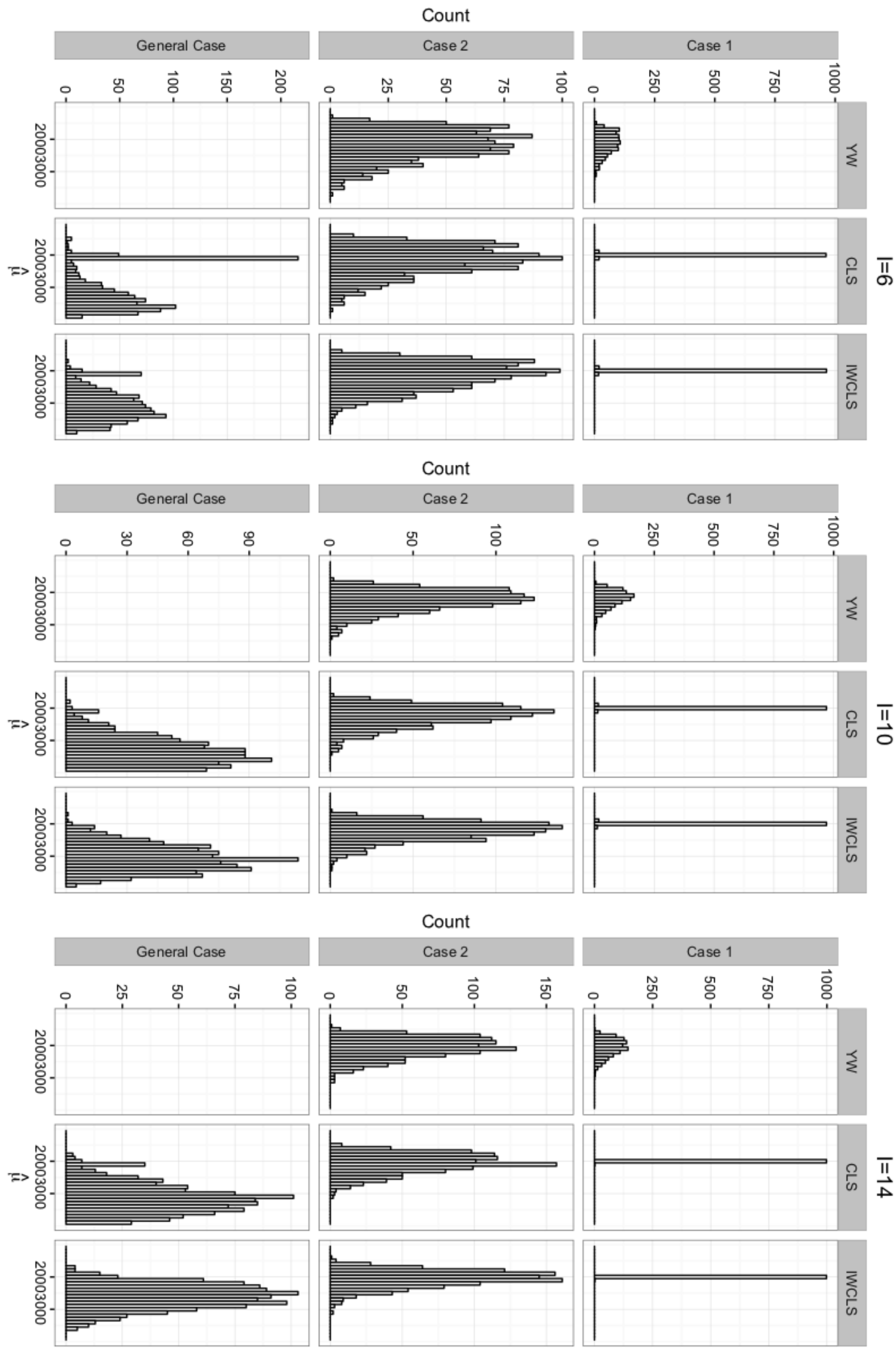

Figure 4.4: Histograms of $\hat{\mu}$. The true value of μ is 2000. Under the General Case, $\mu_i = \mu$, and the distributions of the averaged estimated μ_i's are shown.

the mean of the estimators is bigger than the median, and furthermore, the median is above the true value of the parameter μ, which implies the positive bias of $\hat{\mu}$ under Case 2. Under the General Case, the IWCLS estimation does show better results compared to the CLS estimation. However, both of them present a distribution far from the true value and the bi-modality is also observed for the CLS estimation. Therefore, to better estimate the parameters under the General Case, more advanced estimation methods should be used.

4.2 Prediction of Loss Reserve

According to the discussion in Section 4.1, we have concluded that larger triangle size leads to superior estimates of the model parameters and under Case 2 when assuming all $\mu_i = \mu$, the estimation results are good enough to do the prediction. Among the three estimation methods, the IWCLS estimation performs the best and it tends to perform much better than the other two as the triangle size grows. Therefore, in this section, we choose Case 2 with $I = 14$ to do the prediction. We choice the same parameters as in the simulation study ($\mu = 2000$, $\rho = 0.5$ and $\gamma = (0.4, 0.2, 0.1, 0.1, 0.06, 0.04, 0.02, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01)$) to generate a random sample first. Then, the IWCLS estimation is used to get the estimates for the model parameters. Based on the estimated model parameters and the prediction formulae in Section 3.3, we could get the prediction results for various numbers, and the MSEP is also calculated. The results are shown in Tables 4.5 to 4.7. Table 4.5 presents the estimated unclosed claims $\hat{C}_{i,j}^{Poi\_INAR}$, the outstanding claims $\hat{R}_{i,I-i}^{Poi\_INAR}$ and the estimated ultimate claim numbers $\hat{\mu}$ with the corresponding credibility type estimator $\hat{\mu}_i^{Poi\_INAR}$. Table 4.6 and Table 4.7 show the MSEP of the prediction results displayed in Table 4.5, and the percentage of the error of prediction, which is defiend as the square root of the MSEP over the estimated values; the latter is presented to show effectively the prediction error.

From Table 4.5, it can be seen that as the development year becomes bigger, the estimated $\hat{C}_{i,j}^{Poi\_INAR}$'s for different accident years become more close to each other; in other words, $\hat{C}_{i,j}^{Poi\_INAR}$ tends to have the same value for all $i \geq I - j$ as j increases. The reason is that for later development years, the unclosed claims depend mainly on the newly reported claims in that development year (the IBNyR claims) and merely on the observed data. Since the mechanism of the reporting process of different accident years are the same, i.e., different accident years share the same development pattern of the unclosed claims, their estimation results are thus similar. The empirical estimates $C_{i,I-i}/\hat{\beta}_{I-i}$ for μ varies a lot because it depends mainly on the diagonal values of the observed data. When the observed unclosed claims $C_{i,I-i}$'s are larger than usual, the empirical estimates for μ become larger than the true value, for example when $i = 1, 7, 9$ in our data. The credibility type estimator $\hat{\mu}_i^{Poi\_INAR}$ incorporates both the observed information and the prior information of μ,

	Estimated Unclosed Claims															$\hat{R}_{i,I-i}^{Poi\_INAR}$	$\hat{\mu}$	$C_{i,I-i}/\hat{\beta}_{I-i}$	$\hat{\mu}_i^{Poi\_INAR}$
Acci- -dent year i	Development year j																		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14				
0	817	799	599	495	364	277	180	99	75	60	45	38	42	45	41	41	2091	2120	2120
1	741	774	605	519	372	270	179	101	63	41	45	43	52	49	**44**	68	2091	2510	2293
2	778	830	607	501	370	265	171	127	81	43	50	55	47	**43**	**42**	91	2091	2362	2155
3	762	810	603	490	377	264	164	94	70	49	48	43	**41**	**41**	**40**	112	2091	2080	2089
4	765	782	604	528	405	277	173	111	81	66	51	**45**	**43**	**41**	**41**	143	2091	2286	2103
5	794	810	590	466	353	254	162	116	76	55	**47**	**44**	**42**	**41**	**40**	170	2091	2138	2092
6	735	782	578	505	369	274	162	111	70	**54**	**47**	**43**	**42**	**41**	**40**	202	2091	2130	2092
7	814	799	588	498	376	279	197	136	**86**	**62**	**51**	**45**	**43**	**41**	**41**	289	2091	2847	2102
8	833	835	631	479	368	242	167	**101**	**69**	**54**	**47**	**43**	**42**	**41**	**40**	349	2091	2116	2091
9	782	815	627	534	431	299	**185**	**109**	**73**	**56**	**48**	**44**	**42**	**41**	**40**	526	2091	2429	2094
10	832	827	629	493	353	**253**	**163**	**99**	**68**	**54**	**47**	**43**	**42**	**41**	**40**	669	2091	2033	2090
11	753	773	594	470	**350**	**251**	**162**	**98**	**68**	**53**	**46**	**43**	**42**	**41**	**40**	923	2091	1979	2090
12	813	795	602	**497**	**363**	**258**	**165**	**100**	**69**	**54**	**47**	**43**	**42**	**41**	**40**	1264	2091	2095	2091
13	775	757	**571**	**483**	**356**	**254**	**163**	**99**	**68**	**54**	**47**	**43**	**42**	**41**	**40**	1639	2091	1934	2091
14	789	**796**	**590**	**491**	**361**	**256**	**164**	**100**	**69**	**54**	**47**	**43**	**42**	**41**	**40**	2094	2091	1972	2091

Table 4.5: The estimated unclosed claims, outstanding claims and the ultimate claim numbers under the Poisson INAR model. The true value for μ_i is 2000.

Accident year i	Development year j														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	$\widehat{\mathsf{MSEP}}[\hat{C}_{i,j}^{Poi\_INAR}\|\mathcal{D}_I]$														
0															
1															33.14
2														32.64	39.21
3													31.64	38.51	39.90
4												33.63	39.92	40.68	40.53
5											34.63	40.63	41.07	40.73	40.44
6										38.38	43.27	42.53	41.47	40.80	40.45
7									54.85	54.93	48.97	44.75	42.42	41.23	40.64
8								62.58	60.40	52.00	46.29	43.18	41.60	40.82	40.44
9							116.43	93.71	69.67	55.15	47.50	43.69	41.83	40.93	40.49
10						171.72	144.17	94.52	67.20	53.31	46.48	43.16	41.57	40.80	40.43
11					242.73	226.66	156.44	97.18	67.74	53.40	46.48	43.16	41.56	40.80	40.43
12				359.30	331.80	250.30	163.41	99.52	68.63	53.77	46.65	43.23	41.60	40.82	40.44
13			397.98	442.80	347.27	252.09	162.98	99.02	68.32	53.61	46.57	43.19	41.58	40.81	40.44
14		615.04	548.51	481.90	358.43	255.71	164.31	99.57	68.57	53.72	46.62	43.22	41.59	40.81	40.44
	Percentage of Root $\widehat{\mathsf{MSEP}}[\hat{C}_{i,j}^{Poi\_INAR}\|\mathcal{D}_I]$														
0															
1															0.13
2														0.13	0.15
3													0.14	0.15	0.16
4												0.13	0.15	0.15	0.16
5											0.12	0.15	0.15	0.16	0.16
6										0.11	0.14	0.15	0.15	0.16	0.16
7									0.09	0.12	0.14	0.15	0.15	0.16	0.16
8								0.08	0.11	0.13	0.15	0.15	0.15	0.16	0.16
9							0.06	0.09	0.11	0.13	0.14	0.15	0.15	0.16	0.16
10						0.05	0.07	0.10	0.12	0.14	0.15	0.15	0.16	0.16	0.16
11					0.04	0.06	0.08	0.10	0.12	0.14	0.15	0.15	0.16	0.16	0.16
12				0.04	0.05	0.06	0.08	0.10	0.12	0.14	0.15	0.15	0.16	0.16	0.16
13			0.03	0.04	0.05	0.06	0.08	0.10	0.12	0.14	0.15	0.15	0.16	0.16	0.16
14		0.03	0.04	0.04	0.05	0.06	0.08	0.10	0.12	0.14	0.15	0.15	0.16	0.16	0.16

Table 4.6: The estimated MSEP of the estimated unclosed claims

which produces a reasonable estimation. For recent accident years, $\hat{\mu}_i^{Poi\_INAR}$ converges to the estimated prior information $\hat{\mu}$ because less weights are put on the empirical estimates.

As for the prediction error of the estimated unclosed claims $\hat{C}_{i,j}^{Poi\_INAR}$, it can be seen from Table 4.6 that for a particular accident year i, the prediction error increases as j increases and for a particular development year j, the prediction error gets slightly bigger for recent accident year i. The reason could be that, the later the accident year and the development year are, the less observed data we have, and therefore, the more uncertainty is observed in the prediction results. The development year j seems to have more impact on the MSEP of the estimators compared to the accident year i, partially because we ignore the parameter estimation errors so that the prediction errors for recent accident years are possibly underestimated. For the MSEP of the outstanding claims $\hat{R}_{i,I-i}^{Poi\_INAR}$, clearly, as the accident year i increases, the MSEP increases but the percentage of prediction error is nearly around 0.01. In contrary, the MSEP of the ultimate claims $\hat{\mu}_i^{Poi\_INAR}$ decreases dramatically as the accident year i increases, and the percentage of prediction error follows the same pattern. As mentioned earlier, $\hat{\mu}_i^{Poi\_INAR}$ converges to the estimated prior information μ_i because less weights are put on the empirical estimates as the accident year i increases. Hence, since we ignore the parameter estimation error, the MSEP for the latter accident years converges to 0.

Acci- -dent year i	$\hat{R}_{i,I-i}^{Poi\_INAR}$	$\widehat{\mathsf{MSEP}}[\hat{R}_{i,I-i}^{Poi\_INAR}]$	% Root MSEP	$\hat{\mu}_i^{Poi\_INAR}$	$\widehat{\mathsf{MSEP}}[\hat{\mu}_i^{Poi\_INAR}]$	% Root MSEP
0	41	0.0000	0.0000	2120	108094.0024	0.1551
1	68	0.1812	0.0062	2293	24985.5069	0.0689
2	91	0.9253	0.0106	2155	5832.0119	0.0354
3	112	2.3030	0.0135	2089	1388.1740	0.0178
4	143	4.0114	0.0140	2103	343.0491	0.0088
5	170	6.3091	0.0148	2092	90.5894	0.0045
6	202	8.2739	0.0143	2092	26.5008	0.0025
7	289	11.2566	0.0116	2102	8.8228	0.0014
8	349	15.8855	0.0114	2091	3.3383	0.0009
9	526	24.7088	0.0094	2094	1.1927	0.0005
10	669	47.8334	0.0103	2090	0.3853	0.0003
11	923	97.9947	0.0107	2090	0.1207	0.0002
12	1264	209.3019	0.0115	2091	0.0334	0.0001
13	1639	371.9909	0.0118	2091	0.0104	0.0000
14	2094	814.8506	0.0136	2091	0.0024	0.0000

Table 4.7: The estimated MSEP of the outstanding and ultimate claim numbers

To conclude, we can see that when $I = 14$ and under Case 2, we get pretty good prediction results for all the figures that we are interested in, and their prediction error is within a reasonable range. However, because we ignore the parameter estimation error, the MSEP is a bit underestimated for the later accident years and for the ultimate claim numbers.

Chapter 5

Conclusion and Further Discussion

The IBNR claims reserving problem deals with the major risk carried by the non-life insurers. Traditionally, the non-parametric models such as the CL and the BF methods are used widely in industry because of their nice interpretations and simple estimation procedures. However, more research has been done to study the stochastic frameworks underlying the non-parametric models with parametric assumptions. The accuracy of the loss reserves could be measured precisely when using the parametric models. As presented in Chapter 2, different parametric models are designed to be applied to different kinds of claim data, either claim counts or claim amount. Since the traditional time series are not suitable for modelling claim counts data, Kremer (1995) brought up the idea of using the INAR processes to model the IBNR claims. The idea is extended by this project to model the unclosed claims and both theoretical and numerical results are presented in this project.

The proposed Poisson INAR model has nice interpretations of the model parameters and the unclosed claims can be written as a summation of not yet settled claims from all previous and current development years. The Poisson INAR model also has the property of non-dispersion similar to the Poisson model. An unclosed development pattern is introduced and better describes the development of IBNR claims. The predictions can be done under both known and unknown model parameters, and by ignoring the parameter estimation error, we are able to successfully estimate the mean square error of prediction for various figures. As for the estimation of the parameters of the Poisson INAR model, the Yuller-Walker estimation is an extension from the Yuller-Walker estimates of the AR(1) processes while the least squares estimations are the commonly used statistical methods to estimate the model parameters, which require an iterative algorithm.

Among the three estimation methods discussed in this project, we conclude from the simulation study that the IWCLS estimation method always shows the smallest estimation error among the three, and as the size of development triangle grows, it also performs better than the other two methods. If assuming that all the accident years have the same ultimate claim numbers, which means that the average claim frequency does not change over years for this kind of insurance coverage, the estimation is quite accurate as there are less model parameters to be estimated. If further assumes that the pattern of the newly reported claims in each development year is known, the estimation is even better especially using the least square estimation methods. Another major finding is that increasing the size of the development triangle, which means that we are dealing with claims of long settlement pattern and more data is observed and needed, the accuracy of the estimation is improved. However, all of the three estimation methods do not provide good estimation results under the general model assumption. More advanced estimation methods such as the maximum likelihood estimation or the Bayesian methods could be studied.

For the prediction results under the case when the average claim frequency is the same for all the accident years, the credibility type of estimator for the ultimate claim numbers is related to the situation of different accident years by incorporating the empirical estimates and converges to the estimated prior estimate for the recent accident years when there are few observed information available. The uncertainty of the predicted unclosed claims for later accident years and development years are usually bigger so as to the uncertainty of the outstanding claims. In addition, because we ignore the parameter estimation error, the mean square error of prediction is a bit underestimated.

The proposed Poisson INAR model could be extended under the Bayesian framework by assuming a prior distribution for the expected number of claims of each accident year. In Appendix B, we present the properties of this extended model and provided the prediction and the MSEP formulas. However, the estimation procedure for this model could be more complicated.

Bibliography

[1] Alosh, M.A. and Alzaid, A.A. First-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis*, 8(3):261–275, 1987.

[2] Benktander, G. An approach to credibility in calculating IBNR for casualty excess insurance. *The Actuarial Review*, 1976.

[3] Bornhuetter, R.L. and Ferguson, R.E. The actuary and IBNR. *In Proceedings of the Casualty Actuarial Society*, 59:181-195, 1972.

[4] Buchwalder, M., Bühlmann, H., Merz, M. and Wüthrich, M.V. The mean square error of prediction in the chain ladder reserving method (mack and murphy revisited). *ASTIN Bulletin*, 36(2):521–542, 2006.

[5] Bühlmann, H. and Straub, E. Estimation of IBNR reserves by the methods chain ladder, Cape Cod and complementary loss ratio. In *International Summer School*, 1983.

[6] Friedland, J. Estimating unpaid claims using basic techniques. *Casualty Actuarial Society*, 2010.

[7] Gourieroux, C. and Jasiak, J. Heterogeneous INAR (1) model with application to car insurance. *Insurance: Mathematics and Economics*, 34(2):177–192, 2004.

[8] Hovinen, E. Additive and continuous IBNR. *Proceedings of the XV ASTIN Coll.*, 1981.

[9] Kremer, E. INAR and IBNR. *Blätter der DGVFM*, 22(2):249–253, 1995.

[10] Mack, T. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, 23(2):213–225, 1993.

[11] Peters, G.W., Shevchenko, P.V. and Wüthrich, M.V. Model uncertainty in claims reserving within Tweedie's compound poisson models. *ASTIN Bulletin*, 39(1):1–33, 2009.

[12] Silva, I., Silva, M.E., Pereira, I. and Silva, N. Replicated INAR (1) processes. *Methodology and Computing in applied Probability*, 7(4):517–542, 2005.

[13] Wüthrich, M.V. Claims reserving using Tweedie's compound poisson model. *ASTIN Bulletin*, 33(2):331–346, 2003.

[14] Wüthrich, M.V. and Merz, M. *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons, 2008.

[15] Zhang, T. Integer-valued autoregressive processes with dynamic heterogeneity and their applications in automobile insurance. Master's thesis, Dept. of Statistics and Actuarial Science-Simon Fraser University, 2009.

Appendix A

The Auto-correlation of the Poisson INAR Model

An alternative proof of the covariance formula in (3.7):

Since for each i, $X_{i,j}$'s are independent, using the variance formula given in (3.7), we have

$$\begin{aligned}
\mathsf{Cov}[C_{i,j}, C_{i,j-h}] &= \mathsf{Cov}\left[\sum_{k=0}^{j} \rho^k \circ X_{i,j-k}, \sum_{k=0}^{j-h} \rho^k \circ X_{i,j-h-k}\right] \\
&= \mathsf{Cov}\left[\sum_{k=h}^{j} \rho^k \circ X_{i,j-k}, \sum_{k=0}^{j-h} \rho^k \circ X_{i,j-h-k}\right] \\
&= \sum_{k=0}^{j-h} \mathsf{Cov}[\rho^{k+h} \circ X_{i,j-h-k}, \rho^k \circ X_{i,j-h-k}] \\
&= \sum_{k=0}^{j-h} \Big(\mathsf{E}[\mathsf{Cov}[\rho^{k+h} \circ X_{i,j-h-k}, \rho^k \circ X_{i,j-h-k} | \rho^k \circ X_{i,j-h-k}]] \\
&\quad + \mathsf{Cov}[\mathsf{E}[\rho^{k+h} \circ X_{i,j-h-k} | \rho^k \circ X_{i,j-h-k}], \mathsf{E}[\rho^k \circ X_{i,j-h-k} | \rho^k \circ X_{i,j-h-k}]] \Big) \\
&= \sum_{k=0}^{j-h} \mathsf{Cov}[\rho^h \cdot \rho^k \circ X_{i,j-h-k}, \rho^k \circ X_{i,j-h-k}] \\
&= \rho^h \sum_{k=0}^{j-h} \mathsf{Var}[\rho^k \circ X_{i,j-h-k}] \\
&= \rho^h \mathsf{Var}[C_{i,j-h}] \\
&= \rho^h \left(\sum_{k=0}^{j-h} \rho^k \gamma_{j-h-k} \right) \mu_i \\
&= \left(\sum_{k=h}^{j} \rho^k \gamma_{j-k} \right) \mu_i .
\end{aligned}$$

Appendix B

Model Unclosed Claims with Bayesian Method

The idea of this model comes from the Poisson-Gamma model mentioned in Section 2.4. It is also an extension of the Poisson INAR model discussed in this project.

Assumption B1. There exist random variables Θ_i, $X_{i,j}$, $C_{i,j}$ as well as constants $\gamma_0, \ldots, \gamma_I > 0$ with $\sum_{j=0}^{I} \gamma_j = 1$ such that for all $0 \leq i, j \leq I$,

- Conditionally, given Θ_i, the newly reported claims $X_{i,j}$, which is generated in accident year i but reported with delay j, are independently Poisson distributed with $\mathsf{E}[X_{i,j}|\Theta_i] = \mathsf{Var}[X_{i,j}|\Theta_i] = \Theta_i \gamma_j$.

- $(\Theta_i, (X_{i,0}, \ldots, X_{i,I}))$, $(i = 0, \ldots, I)$ are independent and Θ_i is Gamma distributed with shape parameter a_i and scale parameter b_i.

- The unclosed claims $C_{i,j}$ of different accident years are independent, and follow an INAR process such that

$$C_{i,j} = \rho \circ C_{i,j-1} + X_{i,j}$$

  with $\rho \circ C_{i,j-1} = \sum_{k=1}^{C_{i,j-1}} Y_k$, where $Y_k \sim \mathit{Bermoulli}(1, \rho)$ and $0 \leq \rho \leq 1$. Define that $C_{i,-1} = 0$.

Remark B1. Model Assumption B1 has a similar interpretation as for Assumption 1. The only difference between the two is that Assumption B1 assumes a prior distribution for the expected claim numbers for each accident year, which can be useful when the prior information for the expected claim numbers is available. The model presented here also resolves the non-dispersion property of the Poisson INAR model.

The properties of the Bayesian Poisson INAR model are presented in the following.

Proposition B1. The unclosed claims $C_{i,j}$ can be written as a summation of the not yet settled claims from all the past and current development years $j - k$, $0 \leq k \leq j$, i.e.,

$$C_{i,j} = \sum_{k=0}^{j} \rho^k \circ X_{i,j-k}, \qquad 0 \leq i, j \leq I. \tag{B.1}$$

Conditioning on $C_{i,j-h}$, $C_{i,j}$ can be written as

$$C_{i,j} = \sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} + \rho^h \circ C_{i,j-h}, \quad 0 \le i \le I, 1 \le j \le I,\ 1 \le h \le j. \tag{B.2}$$

Based on the observed data $\mathcal{D}_I$,

$$C_{i,j} = \sum_{k=0}^{i+j-I-1} \rho^k \circ X_{i,j-k} + \rho^{i+j-I} \circ C_{i,I-i}, \qquad 1 \le i \le I,\ j > I-i. \tag{B.3}$$

The proof is the same as Proposition 1 and are quite straightforward.

Proposition B2. The mean, variance and the auto-correlation of $C_{i,j}$ can be obtained as

$$\begin{aligned}
\mathsf{E}[C_{i,j}] &= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \frac{a_i}{b_i}, \\
\mathsf{Var}[C_{i,j}] &= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \frac{a_i}{b_i} + \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right)^2 \frac{a_i}{b_i^2}, \\
\mathsf{Cov}[C_{i,j}, C_{i,j-h}] &= \left(\sum_{k=h}^{j} \rho^k \gamma_{j-k}\right) \frac{a_i}{b_i} + \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \left(\sum_{k=0}^{j-h} \rho^k \gamma_{j-h-k}\right) \frac{a_i}{b_i^2}.
\end{aligned} \tag{B.4}$$

Proof. Note that Θ_i follows a Gamma distribution with shape parameter a_i and scale parameter b_i, and hence $\mathsf{E}[\Theta_i] = a_i/b_i$ and $\mathsf{Var}[\Theta_i] = a_i/b_i^2$. Conditioning on Θ_i, the Bayesian Poisson INAR model reduces simply to the Poisson INAR model with $\mu_i = \Theta_i$, and according to Proposition 2,

$$\begin{aligned}
\mathsf{E}[C_{i,j}|\Theta_i] &= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \Theta_i, \\
\mathsf{Var}[C_{i,j}|\Theta_i] &= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \Theta_i, \\
\mathsf{Cov}[C_{i,j}, C_{i,j-h}|\Theta_i] &= \left(\sum_{k=h}^{j} \rho^k \gamma_{j-k}\right) \Theta_i.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathsf{E}[C_{i,j}] &= \mathsf{E}[\mathsf{E}[C_{i,j}|\Theta_i]] \\
&= \mathsf{E}\left[\left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \Theta_i\right] \\
&= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \frac{a_i}{b_i},
\end{aligned}$$

$$\mathsf{Var}[C_{i,j}] = \mathsf{E}[\mathsf{Var}[C_{i,j}|\Theta_i]] + \mathsf{Var}[\mathsf{E}[C_{i,j}|\Theta_i]]$$

$$
\begin{aligned}
&= \mathsf{E}\left[\left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \Theta_i\right] + \mathsf{Var}\left[\left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \Theta_i\right] \\
&= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \mathsf{E}[\Theta_i] + \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right)^2 \mathsf{Var}[\Theta_i] \\
&= \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \frac{a_i}{b_i} + \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right)^2 \frac{a_i}{b_i^2},
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{Cov}[C_{i,j}, C_{i,j-h}] &= \mathsf{E}[\mathsf{Cov}[C_{i,j}, C_{i,j-h}|\Theta_i]] + \mathsf{Cov}[\mathsf{E}[C_{i,j}|\Theta_i], \mathsf{E}[C_{i,j-h}|\Theta_i]] \\
&= \mathsf{E}\left[\left(\sum_{k=h}^{j} \rho^k \gamma_{j-k}\right) \Theta_i\right] + \mathsf{Cov}\left[\left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \Theta_i, \left(\sum_{k=0}^{j-h} \rho^k \gamma_{j-h-k}\right) \Theta_i\right] \\
&= \left(\sum_{k=h}^{j} \rho^k \gamma_{j-k}\right) \mathsf{E}[\Theta_i] + \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \left(\sum_{k=0}^{j-h} \rho^k \gamma_{j-h-k}\right) \mathsf{Var}[\Theta_i] \\
&= \left(\sum_{k=h}^{j} \rho^k \gamma_{j-k}\right) \frac{a_i}{b_i} + \left(\sum_{k=0}^{j} \rho^k \gamma_{j-k}\right) \left(\sum_{k=0}^{j-h} \rho^k \gamma_{j-h-k}\right) \frac{a_i}{b_i^2}.
\end{aligned}
$$

□

Proposition B3. The conditional mean and variance of $C_{i,j}$ given $C_{i,j-h}$ can be obtained as

$$
\begin{aligned}
\mathsf{E}[C_{i,j}|C_{i,j-h}] &= \left(\sum_{k=0}^{h-1} \rho^k \gamma_{j-k}\right) \mathsf{E}[\Theta_i|C_{i,j-h}] + \rho^h \cdot C_{i,j-h}, \\
\mathsf{Var}[C_{i,j}|C_{i,j-h}] &= \left(\sum_{k=0}^{h-1} \rho^k \gamma_{j-k}\right)^2 \mathsf{Var}[\Theta_i|C_{i,j-h}] \\
&\quad + \left(\sum_{k=0}^{h-1} \rho^k \gamma_{j-k}\right) \mathsf{E}[\Theta_i|C_{i,j-h}] + \rho^h \cdot (1-\rho^h) \cdot C_{i,j-h}, \\
&\qquad 0 \le i \le I, 1 \le j \le I,\ 1 \le h \le j.
\end{aligned}
\tag{B.5}
$$

Based on the observed data $\mathcal{D}_I$, for any $1 \le i \le I$ and $j \ge I - i$,

$$
\begin{aligned}
\mathsf{E}[C_{i,j}|\mathcal{D}_I] &= \left(\sum_{k=0}^{i+j-I-1} \rho^k \gamma_{j-k}\right) \mathsf{E}[\Theta_i|\mathcal{D}_I] + \rho^{i+j-I} \cdot C_{i,j-h}, \\
\mathsf{Var}[C_{i,j}|\mathcal{D}_I] &= \left(\sum_{k=0}^{i+j-I-1} \rho^k \gamma_{j-k}\right)^2 \mathsf{Var}[\Theta_i|\mathcal{D}_I] \\
&\quad + \left(\sum_{k=0}^{i+j-I-1} \rho^k \gamma_{j-k}\right) \mathsf{E}[\Theta_i|\mathcal{D}_I] + \rho^{i+j-I} \cdot (1-\rho^{i+j-I}) \cdot C_{i,I-i}.
\end{aligned}
\tag{B.6}
$$

Proof. Taking conditional expectation of both sides of (B.2), we get

$$\begin{aligned}
\mathsf{E}[C_{i,j}|C_{i,j-h}] &= \mathsf{E}\left[\sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} + \rho^h \circ C_{i,j-h} \middle| C_{i,j-h}\right] \\
&= \mathsf{E}\left[\sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} \middle| C_{i,j-h}\right] + \rho^h \cdot C_{i,j-h} \\
&= \mathsf{E}\left[\mathsf{E}\left[\sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} \middle| \Theta_i\right] \middle| C_{i,j-h}\right] + \rho^h \cdot C_{i,j-h} \\
&= \mathsf{E}\left[\sum_{k=0}^{h-1} \rho^k \mathsf{E}[X_{i,j-k}|\Theta_i] \middle| C_{i,j-h}\right] + \rho^h \cdot C_{i,j-h} \\
&= \mathsf{E}\left[\sum_{k=0}^{h-1} \rho^k \gamma_{j-k} \Theta_i \middle| C_{i,j-h}\right] + \rho^h \cdot C_{i,j-h} \\
&= \left(\sum_{k=0}^{h-1} \rho^k \gamma_{j-k}\right) \mathsf{E}[\Theta_i|C_{i,j-h}] + \rho^h \cdot C_{i,j-h}.
\end{aligned}$$

To get the conditional variance, simply take variance of both sides of (B.2),

$$\begin{aligned}
\mathsf{Var}[C_{i,j}|C_{i,j-h}] &= \mathsf{Var}\left[\sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} + \rho^h \circ C_{i,j-h} \middle| C_{i,j-h}\right] \\
&= \mathsf{Var}\left[\sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} \middle| C_{i,j-h}\right] + \rho^h \cdot (1-\rho^h) \cdot C_{i,j-h} \\
&= \mathsf{Var}\left[\mathsf{E}\left[\sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} \middle| \Theta_i\right] \middle| C_{i,j-h}\right] \\
&\quad + \mathsf{E}\left[\mathsf{Var}\left[\sum_{k=0}^{h-1} \rho^k \circ X_{i,j-k} \middle| \Theta_i\right] \middle| C_{i,j-h}\right] + \rho^h \cdot (1-\rho^h) \cdot C_{i,j-h} \\
&= \mathsf{Var}\left[\sum_{k=0}^{h-1} \rho^k \mathsf{E}[X_{i,j-k}|\Theta_i] \middle| C_{i,j-h}\right] \\
&\quad + \mathsf{E}\left[\sum_{k=0}^{h-1} \mathsf{Var}[\rho^k \circ X_{i,j-k}|\Theta_i] \middle| C_{i,j-h}\right] + \rho^h \cdot (1-\rho^h) \cdot C_{i,j-h} \\
&= \mathsf{Var}\left[\sum_{k=0}^{h-1} \rho^k \mathsf{E}[X_{i,j-k}|\Theta_i] \middle| C_{i,j-h}\right] \\
&\quad + \mathsf{E}\left[\sum_{k=0}^{h-1} \left(\rho^k \cdot (1-\rho^k)\mathsf{E}[X_{i,j-k}|\Theta_i] + \rho^{2k}\mathsf{Var}[X_{i,j-k}|\Theta_i]\right) \middle| C_{i,j-h}\right] \\
&\quad + \rho^h \cdot (1-\rho^h) \cdot C_{i,j-h} \\
&= \mathsf{Var}\left[\sum_{k=0}^{h-1} \rho^k \gamma_{j-k} \Theta_i \middle| C_{i,j-h}\right]
\end{aligned}$$

$$
+ \mathsf{E}\left[\sum_{k=0}^{h-1}\left(\rho^k(1-\rho^k)\gamma_{j-k}\Theta_i + \rho^{2k}\gamma_{j-k}\Theta_i\right)\middle| C_{i,j-h}\right] + \rho^h\cdot(1-\rho^h)\cdot C_{i,j-h}
$$

$$
= \left(\sum_{k=0}^{h-1}\rho^k\gamma_{j-k}\right)^2 \mathsf{Var}[\Theta_i|C_{i,j-h}] + \left(\sum_{k=0}^{h-1}\rho^k\gamma_{j-k}\right)\mathsf{E}[\Theta_i|C_{i,j-h}]
$$

$$
+ \rho^h\cdot(1-\rho^h)\cdot C_{i,j-h}.
$$

□

According to Propositions B2 and B3, the non-parametric assumptions for this Bayesian Poisson INAR model can be introduced assuming an unclosed claims development pattern $\{\beta_j\}_{j=0}^I$ with $\beta_j = \sum_{k=0}^j \rho^k\gamma_{j-k}$. For any $0 \le i \le I$, $0 \le j \le I-1$ and $1 \le h \le I-j$,

$$
\begin{aligned}
\mathsf{E}[C_{i,0}] &= \beta_j \mathsf{E}[\Theta_i], \\
\mathsf{E}[C_{i,j+h}|C_{i,0},\ldots,C_{i,j}] &= \rho^h\cdot C_{i,j} + (\beta^{j+h} - \rho^h\cdot\beta_j)\cdot\mathsf{E}[\Theta_i|C_{i,0},\ldots,C_{i,j}].
\end{aligned} \tag{B.7}
$$

The prediction formula for $C_{i,j}$, $1 \le i \le I$ and $j \ge I-i$ with unknown parameters is given by

$$
\begin{aligned}
\hat{C}_{i,j}^{PoiGa\_INAR} &= \hat{\mathsf{E}}[C_{i,j}|\mathcal{D}_I] \\
&= \hat{\rho}^{i+j-I}\cdot\hat{\beta}_{I-i}\cdot\frac{C_{i,I-i}}{\hat{\beta}_{I-i}} + (\hat{\beta}_j - \hat{\rho}^{i+j-I}\cdot\hat{\beta}_{I-i})\cdot\hat{\mathsf{E}}[\Theta_i|\mathcal{D}_I].
\end{aligned} \tag{B.8}
$$

The prediction combines the empirical estimates with the posterior mean of the expected ultimate claim numbers.

To obtain the prediction results, one has to estimate the posterior mean of Θ_i's. The posterior distribution of Θ_i, $i = 0, 1, \ldots, I$, are stated in the following lemma and the corresponding posterior mean is straightforward from the lemma.

Lemma B1. Under Assumption B1, we have the following results for the posterior distribution of Θ_i, $i = 0, 1, \ldots, I$.

(1) Given $c_{i,0}$, $\Theta_i \sim Gamma\left(a_{i,0}^{Post}, b_{i,0}^{Post}\right)$, where

$$
a_{i,0}^{Post} = a_i + c_{i,0}, \qquad b_{i,0}^{Post} = b_i + \gamma_0.
$$

(2) Given $c_{i,0}, c_{i,1}$, Θ_i follows a weighted sum of gamma distributions with its probability density function (pdf) given by

$$
\sum_{y_1=0}^{\min(c_{i,0},c_{i,1})} \frac{w(y_1, c_{i,1}, c_{i,0})}{\sum_{y_1=0}^{\min(c_{i,0},c_{i,1})} w(y_1, c_{i,1}, c_{i,0})}\cdot g\left(\theta_i; a_{i,1}^{Post}(y_1), b_{i,1}^{Post}\right)
$$

where $g\left(\theta_i; a_{i,1}^{Post}(y_1), b_{i,1}^{Post}\right)$ is the pdf of a gamma distribution with parameters

$$
a_{i,1}^{Post}(y_1) = a_i + c_{i,0} + c_{i,1} - y_1, \qquad b_{i,1}^{Post} = b_i + \gamma_0 + \gamma_1,
$$

and

$$w(y_1, c_{i,1}, c_{i,0}) = \binom{c_{i,0}}{y_1} \cdot \left(\frac{\rho}{1-\rho}\right)^{y_1} \cdot \frac{1}{\gamma_1^{y_1}(c_{i,1}-y_1)!} \cdot \frac{\Gamma(a_{i,1}^{Post}(y_1))}{(b_{i,1}^{Post})^{a_{i,1}^{Post}(y_1)}}.$$

(3) In general, given $c_{i,0}, c_{i,1}, \ldots, c_{i,j}$, the pdf of Θ_i is given by

$$\sum_{y_j=0}^{\min(c_{i,j},c_{i,j-1})} \cdots \sum_{y_1=0}^{\min(c_{i,1},c_{i,0})} \frac{w(y_1,\ldots,y_j,c_{i,0},\ldots,c_{i,j})}{\bar{w}(c_{i,0},\ldots,c_{i,j})} g\left(\theta_i; a_{i,j}^{Post}(\bar{y}), b_{i,j}^{Post}\right),$$

where $g\left(\theta_i; a_{i,j}^{Post}(\bar{y}), b_{i,j}^{Post}\right)$ is the pdf of a gamma distribution with parameters

$$a_{i,j}^{Post}(\bar{y}) = a_i + \sum_{k=0}^{j} c_{i,k} - j \cdot \bar{y}, \qquad b_{i,j}^{Post} = b_i + \sum_{k=1}^{j} \gamma_k$$

with $\bar{y} = \sum_{k=1}^{j} y_k / j$. The weight function can be written as

$$\begin{aligned} w(y_1,\ldots,y_j,c_{i,0},\ldots,c_{i,j}) = &\binom{c_{i,j-1}}{y_j} \cdots \binom{c_{i,0}}{y_1} \cdot \left(\frac{\rho}{1-\rho}\right)^{j\cdot\bar{y}} \\ &\times \prod_{k=1}^{j} \left(\frac{1}{\gamma_k^{y_k}(C_{i,k}-y_k)!}\right) \cdot \frac{\Gamma(a_{i,j}^{Post}(\bar{y}))}{(b_{i,j}^{Post})^{a_{i,j}^{Post}(\bar{y})}}, \end{aligned}$$

and

$$\bar{w}(c_{i,0},\ldots,c_{i,j}) = \sum_{y_j=0}^{\min(c_{i,j},c_{i,j-1})} \cdots \sum_{y_1=0}^{\min(c_{i,1},c_{i,0})} w(y_1,\ldots,y_j,c_{i,0},\ldots,c_{i,j}).$$

Proof. The proof is very similar to that of Proposition 1 in Gourieroux and Jasiak (2004).

(1) According to Assumption B1, Θ_i has a prior distribution with Gamma pdf, that is,

$$\pi(\theta_i) = \frac{b_i^{a_i}}{\Gamma(a_i)} \theta_i^{a_i-1} e^{-b_i\theta_i}.$$

Given $c_{i,0}$, the posterior distribution of Θ_i can be written as

$$\begin{aligned} \pi(\theta_i|c_{i,0}) &= \frac{\mathsf{Pr}[C_{i,0}=c_{i,0}|\theta_i]\pi(\theta_i)}{\int \mathsf{Pr}[C_{i,0}=c_{i,0}|\theta_i]\pi(\theta_i)d\theta_i} \\ &= \frac{\frac{e^{-\theta_i\gamma_0}(\theta_i\gamma_0)^{c_{i,0}}}{c_{i,0}!} \frac{b_i^{a_i}}{\Gamma(a_i)} \theta_i^{a_i-1} e^{-b_i\theta_i}}{\int \frac{e^{-\theta_i\gamma_0}(\theta_i\gamma_0)^{c_{i,0}}}{c_{i,0}!} \frac{b_i^{a_i}}{\Gamma(a_i)} \theta_i^{a_i-1} e^{-b_i\theta_i} d\theta_i} \\ &= \frac{\theta_i^{a_i+c_{i,0}-1} e^{-(b_i+\gamma_0)\theta_i}}{\int \theta_i^{a_i+c_{i,0}-1} e^{-(b_i+\gamma_0)\theta_i} d\theta_i} \\ &= \frac{\frac{(b_i+\gamma_0)^{(a_i+c_{i,0})}}{\Gamma(a_i+c_{i,0})} \theta_i^{a_i+c_{i,0}-1} e^{-(b_i+\gamma_0)\theta_i}}{\int \frac{(b_i+\gamma_0)^{(a_i+c_{i,0})}}{\Gamma(a_i+c_{i,0})} \theta_i^{a_i+c_{i,0}-1} e^{-(b_i+\gamma_0)\theta_i} d\theta_i} \end{aligned}$$

$$= \frac{(b_i+\gamma_0)^{(a_i+c_{i,0})}}{\Gamma(a_i+c_{i,0})}\theta_i^{a_i+c_{i,0}-1}e^{-(b_i+\gamma_0)\theta_i}.$$

Therefore, given $c_{i,0}$, $\Theta_i \sim Gamma\left(a_{i,0}^{Post}, b_{i,0}^{Post}\right)$, with $a_{i,0}^{Post} = a_i + c_{i,0}$ and $b_{i,0}^{Post} = b_i + \gamma_0$.

(2) Given $c_{i,0}$, $c_{i,1}$, the posterior distribution of Θ_i is given by

$$\pi(\theta_i|c_{i,0}, c_{i,1}) = \frac{\mathsf{Pr}[C_{i,1} = c_{i,1}, C_{i,0} = c_{i,0}|\theta_i]\pi(\theta_i)}{\int \mathsf{Pr}[C_{i,1} = c_{i,1}, C_{i,0} = c_{i,0}|\theta_i]\pi(\theta_i)d\theta_i}$$
$$= \frac{\mathsf{Pr}[C_{i,1} = c_{i,1}|C_{i,0} = c_{i,0}, \theta_i]\mathsf{Pr}[C_{i,0} = c_{i,0}|\theta_i]\pi(\theta_i)}{\int \mathsf{Pr}[C_{i,1} = c_{i,1}|C_{i,0} = c_{i,0}, \theta_i]\mathsf{Pr}[C_{i,0} = c_{i,0}|\theta_i]\pi(\theta_i)d\theta_i}.$$

where

$$\mathsf{Pr}[C_{i,1} = c_{i,1}|C_{i,0} = c_{i,0}, \theta_i]\mathsf{Pr}[C_{i,0} = c_{i,0}|\theta_i]\pi(\theta_i)$$
$$= \sum_{y_1=0}^{\min(c_{i,0},c_{i,1})} \binom{c_{i,0}}{y_1}\rho^{y_1}(1-\rho)^{c_{i,0}-y_1} \cdot e^{(-\theta_i\gamma_1)}\frac{(\theta_i\gamma_1)^{c_{i,1}-y_1}}{(c_{i,1}-y_1)!}\frac{e^{(-\theta_i\gamma_0)}(\theta_i\gamma_0)^{c_{i,0}}}{c_{i,0}!}\frac{b_i^{a_i}}{\Gamma(a_i)}\theta_i^{a_i-1}e^{(-b_i\theta_i)}.$$

By omitting the terms that are not related to θ_i and y_1's,

$$\pi(\theta_i|c_{i,0}, c_{i,1}) \propto \sum_{y_1=0}^{\min(c_{i,0},c_{i,1})} \binom{c_{i,0}}{y_1}\cdot\left(\frac{\rho}{1-\rho}\right)^{y_1}\cdot\frac{1}{\gamma_1^{y_1}(c_{i,1}-y_1)!}\theta_i^{a_i+c_{i,0}+c_{i,1}-y_1-1}e^{-(b_i+\gamma_0+\gamma_1)\theta_i}$$

Let $a_{i,1}^{Post}(y_1) = a_i + c_{i,0} + c_{i,1} - y_1$, $b_{i,1}^{Post} = b_i + \gamma_0 + \gamma_1$ and

$$w(y_1, c_{i,1}, c_{i,0}) = \binom{c_{i,0}}{y_1}\cdot\left(\frac{\rho}{1-\rho}\right)^{y_1}\cdot\frac{1}{\gamma_1^{y_1}(c_{i,1}-y_1)!}\cdot\frac{\Gamma(a_{i,1}^{Post}(y_1))}{(b_{i,1}^{Post})^{a_{i,1}^{Post}(y_1)}}.$$

Noting that $g(\theta_i; a_{i,1}^{Post}(y_1), b_{i,1}^{Post})$ is the pdf of the gamma distribution with parameter $a_{i,1}^{Post}(y_1)$ and $b_{i,1}^{Post}$, and therefore,

$$\pi(\theta_i|c_{i,0}, c_{i,1}) \propto \sum_{y_1=0}^{\min(c_{i,0},c_{i,1})} w(y_1, c_{i,1}, c_{i,0})g(\theta_i; a_{i,1}^{Post}(y_1), b_{i,1}^{Post}).$$

Thus,

$$\pi(\theta_i|c_{i,0}, c_{i,1}) = \frac{\sum_{y_1=0}^{\min(c_{i,0},c_{i,1})} w(y_1, c_{i,1}, c_{i,0})g(\theta_i; a_{i,1}^{Post}(y_1), b_{i,1}^{Post})}{\sum_{y_1=0}^{\min(c_{i,0},c_{i,1})} w(y_1, c_{i,1}, c_{i,0})}.$$

Now the prove completes.

(3) The proof for this general case is similar as (2), so we omit it here.

□

The accuracy of the estimates can be measured by the MSEP (mean square error of prediction). We present a similar lemma below as Lemma 1.

Lemma B2. Under Assumption B1, if the model parameters are known, then $\hat{C}_{i,j}^{PoiGa\_INAR} = \mathsf{E}[C_{i,j}|\mathcal{D}_I]$, denoted as $\hat{C}_{i,j}$ for simplicity below.

(1) The conditional MSEP is given by

$$\begin{aligned}\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I] = &(\beta_j - \rho^{i+j-I} \cdot \beta_{I-i})^2 \mathsf{Var}[\Theta_i|\mathcal{D}_I] + (\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}) \mathsf{E}[\Theta_i|\mathcal{D}_I] \\ &+ \rho^{i+j-I} \cdot (1 - \rho^{i+j-I}) \cdot C_{i,I-i}.\end{aligned}$$

(2) The unconditional MSEP is

$$\begin{aligned}\mathsf{MSEP}[\hat{C}_{i,j}] = &\left(\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}\right)^2 \mathsf{E}[\mathsf{Var}[\Theta_i|\mathcal{D}_I]] \\ &+ (\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}) \frac{a_i}{b_i} + \rho^{i+j-I} \cdot (1 - \rho^{i+j-I}) \cdot C_{i,I-i}.\end{aligned}$$

Proof. By the definition of the conditional MSEP, we can easily get

$$\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I] = \mathsf{E}[(C_{i,j} - \hat{C}_{i,j})^2|\mathcal{D}_I] = \mathrm{Var}[C_{i,j}|\mathcal{D}_I].$$

According to (B.6) given in Proposition B3,

$$\begin{aligned}\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I] = &\left(\sum_{k=0}^{i+j-I-1} \rho^k \gamma_{j-k}\right)^2 \mathsf{Var}[\Theta_i|\mathcal{D}_I] \\ &+ \left(\sum_{k=0}^{i+j-I-1} \rho^k \gamma_{j-k}\right) \mathsf{E}[\Theta_i|\mathcal{D}_I] + \rho^{i+j-I} \cdot (1 - \rho^{i+j-I}) \cdot C_{i,I-i}.\end{aligned}$$

If written using the development pattern β_j given by (B.7),

$$\begin{aligned}\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I] = &(\beta_j - \rho^{i+j-I} \cdot \beta_{I-i})^2 \mathsf{Var}[\Theta_i|\mathcal{D}_I] + (\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}) \mathsf{E}[\Theta_i|\mathcal{D}_I] \\ &+ \rho^{i+j-I} \cdot (1 - \rho^{i+j-I}) \cdot C_{i,I-i},\end{aligned}$$

where $\mathsf{E}[\Theta_i|\mathcal{D}_I]$ and $\mathsf{Var}[\Theta_i|\mathcal{D}_I]$ can be calculated using the results presented in Lemma B1.

The unconditional MSEP is the expectation of the conditional MSEP, which gives

$$\begin{aligned}\mathsf{MSEP}[\hat{C}_{i,j}] &= \mathsf{E}[\mathsf{MSEP}[\hat{C}_{i,j}|\mathcal{D}_I]] \\ &= \mathsf{E}\left[(\beta_j - \rho^{i+j-I} \cdot \beta_{I-i})^2 \mathsf{Var}[\Theta_i|\mathcal{D}_I]\right] + \mathsf{E}\left[(\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}) \mathsf{E}[\Theta_i|\mathcal{D}_I]\right] \\ &\quad + \rho^{i+j-I} \cdot (1 - \rho^{i+j-I}) \cdot C_{i,I-i} \\ &= (\beta_j - \rho^{i+j-I} \cdot \beta_{I-i})^2 \mathsf{E}[\mathsf{Var}[\Theta_i|\mathcal{D}_I]] + (\beta_j - \rho^{i+j-I} \cdot \beta_{I-i}) \mathsf{E}[\Theta_i] \\ &\quad + \rho^{i+j-I} \cdot (1 - \rho^{i+j-I}) \cdot C_{i,I-i},\end{aligned}$$

in which $\mathsf{E}[\Theta_i] = a_i/b_i$ and $\mathsf{E}[\mathsf{Var}[\Theta_i|\mathcal{D}_I]]$ can be calculated using the results presented in Lemma B1. □