Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS

Tongtong Gao

Actuarial Reserving Methods for Non-Life Insurance

Department of Probability and Mathematical Statistics

Supervisor of the master thesis:	RNDr. Martin Branda, Ph.D.
Study programme:	Financial and Insurance Mathematics
Specialization:	Financial and Insurance Mathematics

Prague 2023

I would like to express my deepest gratitude to RNDr. Martin Branda, Ph.D. for all the guidance, patience and support he provided me during the writing process of my thesis. I am deeply thankful to my friends for their unwavering support and encouragement throughout this challenging period. Last but not least, I am profoundly grateful to my family for their boundless love and continued support. Their understanding and patience provided me with the strength and confidence needed to complete my studies.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague 12.07.2024 Author signature

Název práce: Aktuárské rezervovací metody v neživotním pojištění

Autor: Tongtong Gao

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Martin Branda, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato práce zkoumá běžné metody výpočtu pojistných rezerv v neživotním pojištění, včetně tradiční metody řetězového žebříku (CL) a metody Bornhuetter-Ferguson (BF), stejně jako stochastické metody, jako jsou generalizované lineární modely (GLM) a generalizované lineární smíšené modely (GLMM). Porovnáním těchto metod zjistíme, že stochastické metody jsou účinnější při zachycování náhodných výkyvů v datech. Abychom tyto závěry ověřili, aplikujeme tyto modely na skutečná pojistná data a analyzujeme výsledky. Zjištění této studie poskytují důležité vodítko a reference pro odhad rezerv v pojišťovnictví.

Klíčová slova: rezerva na škody, řetězový žebřík, generalizované lineární modely, generalizované lineární smíšené modely

Title: Actuarial Reserving Methods for Non-Life Insurance

Author: Tongtong Gao

Department: Department of Probability and Mathematical Statistics

Supervisor of the master thesis: RNDr. Martin Branda, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This thesis examines common methods for calculating insurance reserves in non-life insurance, including the traditional chain ladder method (CL) and the Bornhuetter-Ferguson method (BF), as well as stochastic methods such as Generalized Linear Models (GLM) and Generalized Linear Mixed Models (GLMM). Through comparisons of these methods, we find that stochastic methods are more effective in capturing random fluctuations in the data. To validate these conclusions, we apply these models to real insurance data and analyze the results. The findings of this study provide important guidance and references for reserve estimation in the insurance industry.

Keywords: loss reserve, chain ladder, generalized linear models, generalized linear mixed models

Contents

Introduction

Reserving is crucial for insurance companies as it typically represents a significant portion of the non-life insurers' balance sheet and plays a vital role in ensuring the company's financial strength. However, it involves great uncertainty, making it challenging to establish a precise mathematical model for estimating the amount of loss reserves. Various approaches have been developed to provide reasonable estimations, which can be categorized into two groups: deterministic models and stochastic models.

Deterministic models, such as the chain ladder and Bornhuetter-Ferguson methods, are simple and easy to operate but have certain limitations. At times, they may overlook important information, such as random fluctuations. In contrast, stochastic models perform better in addressing these aspects. Generalized linear models (GLM) are emerging as a popular statistical analysis method. Generalized linear mixed models (GLMM) extend GLM by incorporating random effects into the linear predictor, challenging the assumption of independence. This paper focuses on using both GLM and GLMM to estimate loss reserves and compares its performance with other traditional methods.

This paper is divided into four chapters. In the first chapter, we discuss the basic notation and the existing loss reserving estimation methods. First of all, it introduces the basic notation of loss reserving and development triangle. Then it introduces the methods and principles of the chain ladder method and B-F method.

The second chapter first discusses the theoretical basis of generalized linear models and generalized linear mixed models. It focuses on the exponential family of distributions, link functions, parameter estimation methods, and model evaluation techniques.

In the third chapter, we establish the generalized linear models (GLM) and generalized linear mixed models (GLMM) for loss reserving. We specify several GLM and GLMM models separately for claim counts and average claim amounts, each characterized by distributions that align with the theoretical framework associated with actuarial practice.

Finally, in the fourth chapter, we apply both the current methods and generalized linear mixed models to real-life data. We use R software to fit the considered models to the data and subsequently obtain the final calculated IBNR reserves. The comparison of the fitted models, along with a discussion of their performance and adequacy, is included.

1. Introduction of claims reserving

There are two main categories of claims reserves. The first one is a reserve on claims that have incurred but have not been reported, called IBNR reserve, and the second is a reserve on claims that have been reported but have not yet been settled, called RBNS reserve. Incurred but not reported reserves estimate the liability for claim-generating events that have already taken place but have not yet been reported to the insurer or self-insurer, as described by Tarbell (1934). The IBNR reserve is the primary area of our concern. The definition and notation of the claims reserving problem, as well as the common methods stated in this section, refer to Wüthrich and Merz (2008).

1.1 Notation of loss triangle

Before evaluating and calculating the outstanding claims reserve, it is essential to organize and analyze the raw data, transforming it into a form so-called *claims development triangles* or a *run-off triangle*. This is an important tool for assessing outstanding claims reserves. We categorize loss data based on the accident year, denoted as i, representing the year of occurrence, and the development year, denoted as j. We assume $i \in \{0, 1, \ldots, I\}$ and $j \in \{0, 1, \ldots, J\}$, where I represents the most recent accident year, and J denotes the last development year. Note that we further assume I = J. The incremental data is denoted as X_{ij}.

Table 1.1: Claims Development Triangle

	Development Year						
Accident Year	1	2	...	j	...	J-1	J
1	$X_{1,1}$	$X_{1,2}$	...	$X_{1,j}$	...	$X_{1,J-1}$	$X_{1,J}$
2	$X_{2,1}$	$X_{2,2}$	...	$X_{2,j}$	....	$X_{2,J-1}$	
...	...	...	...	...	...		
i	$X_{i,1}$	$X_{i,2}$	...	$X_{i,j}$			
...	...	...	...				
I-1	$X_{I-1,1}$	$X_{I-1,2}$					
I	$X_{I,1}$						

And then the data should be structured into a format known as the loss development triangle. The loss triangle is illustrated in the Table 1.1. It is represented as an upper triangular matrix, where each row corresponds to different accident years, and each column represents different development years. The observations available at time I are represented by the set

$$D_I^U = \{X_{i,j} : i + j \leq I, 0 \leq j \leq J\} \tag{1.1}$$

which corresponds to an upper triangle in table 1.1, and the outstanding payments for the lower triangle can be denoted as

$$D_I^L = \{X_{i,j} : i + j > I, 0 \leq j \leq J\} \tag{1.2}$$

The cumulative amount $C_{i,j}$ for accident year i after j development years is then given by

$$C_{i,j} = \sum_{k=0}^{j} X_{ik} \tag{1.3}$$

In this thesis, we assume no further developments after development year J and consider the equivalence between I and J refer to Wüthrich and Merz (2008). Thus, C_{iJ} represents the aggregate loss from all claims that occurred in accident year i, which is called the *ultimate claims amount.* For the cumulative claims amounts, we can also represent the claims amounts as components of a cumulative claims development triangle.

The *outstanding loss liabilities* for accident year i at time I are subsequently determined by:

$$R_i = \sum_{k=j+1}^{J} X_{ik} = C_{i,J} - C_{i,I-i} \tag{1.4}$$

And the estimation of IBNR reserves depends on the estimate of ultimate loss liabilities.

1.2 Basic methods of claims reserving

This section primarily introduces current methods for estimating outstanding claim reserves. Commonly used estimation methods include chain ladder method, Bornhuetter-Ferguson method, and others.

1.2.1 Chain-ladder method

The chain ladder method is one of the most widely used evaluation methods. There are several stochastic models that justify the CL method, and one particularly well-known model is introduced by Mack (1993). Mack's stochastic model is specified by three assumptions. We will initiate our discussion by outlining the assumptions first.

Model assumptions (CL)

CL1 Cumulative claims C_{ij} of different accident years i are independent.

$$\{C_{i,0}, ..., C_{i,I}\}, \{C_{j,0}, ..., C_{j,I}\}\, for\, i \neq j\, are\, independent. \tag{1.5}$$

CL2 The basic chain ladder assumption is that there exists development factors or link ratios $f_1, ..., f_{J-1} > 0$ such that for all $0 \leq i \leq I$ and all $1 \leq j \leq J$ we have

$$E(C_{i,j}|C_{i,0}, ..., C_{i,j-1}) = E(C_{i,j+1}|C_{i,j}) = C_{i,j-1}f_{j-1}. \tag{1.6}$$

CL3 There exist parameters $\sigma_0^2, ..., \sigma_{J-1}^2 > 0$ such that for all $0 \leq i \leq I$ and all $1 \leq j \leq J$ we have

$$Var(C_{i,j}|C_{i,0}, ..., C_{i,j-1}) = \sigma_{j-1}^2 C_{i,j-1}. \tag{1.7}$$

Under the model assumptions we can derive that the ultimate claims amount C_i for accident year i is then given by

$$E(C_{i,J}|D_I) = E(C_{i,J}|C_{i,I-i}) = C_{i,I-i} f_{I-i} ... f_{I-1} \tag{1.8}$$

for all $1 \leq i \leq I$. The chain ladder estimator for $E(C_{i,j}|D_I)$ motivated by this equation is

$$\hat{C_{i,j}^{CL}} = E\left[C_{i,j} \hat{|} D_I\right] = C_{i,I-i} \hat{f_{I-i}} ... \hat{f_{I-1}} \tag{1.9}$$

for $i + j > I$. The development factor can be estimated using arithmetic average method, volume-weighted method and so on. More specifically, actuaries often use data from only the most recent accident years to calculate in practice. This thesis uniformly adopts the simple average method, where the development factors are determined as the average values for each accident year.

$$\hat{f}_i = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} \tag{1.10}$$

which shows that f_i is a weighted average of the individual development factors.

Based on equations 1.4 and 1.9, the estimated outstanding claims reserves for accident year i in the chain-ladder method can be obtained by

$$\begin{aligned} \hat{R}_i^{CL} &= E(C_{i,J} \hat{|} D_I) - C_{i,I-i} = \hat{C_{i,I}^{CL}} - C_{i,I-i} \\ &= (\hat{f_{I-i}} ... \hat{f_{I-1}} - 1) C_{i,I-i} \end{aligned} \tag{1.11}$$

Mack (1993)shows that under model assumptions CL1 (1.5) and CL2 (1.6), the estimated development factors $\hat{f}_i$ are unbiased and uncorrelated estimators of the true parameters f_i. This implies that the Mack chain-ladder model provides unbiased estimators $\hat{R}_i^{CL}$ of the outstanding loss liabilities R_i.

1.2.2 Bornhuetter-Ferguson method

The Bornhuetter-Ferguson method was first described in the paper by Bornhuetter and Ferguson (1972). This method estimates the ultimate loss by considering both incurred losses and their expected development in the future. The fundamental steps of evaluating the outstanding claims reserve using the Bornhuetter-Ferguson method begin with the computation of the ultimate loss. We will start by presenting model assumptions.

Model assumptions (BFI)

BFI1 Cumulative claims C_{ij} of different accident years i are independent.

$$\{C_{i,0}, ..., C_{i,I}\}, \{C_{j,0}, ..., C_{j,I}\} \; for \, i \neq j \; are \, independent. \tag{1.12}$$

BFI2 There exists parameters $\mu_0, ..., \mu_I > 0$ and a pattern $\beta_0, ..., \beta_J > 0$ with $\beta_J = 1$ such that

$$EC_{i0} = \beta_0 \mu_i, \tag{1.13}$$

$$E(C_{i,j+k}|C_{i,0}, ..., C_{i,j}) = C_{i,j} + (\beta_{j+k} - \beta_j)\mu_i \tag{1.14}$$

holds for all $0 \leq i \leq I$, $0 \leq j \leq J - 1$ and $1 \leq k \leq J - j$.

Under the assumptions (BFI) it holds $EC_{i,j} = \beta_j \mu_i$ and $EC_{i,J} = \mu_i$ for all $0 \leq i \leq I$ and $0 \leq j \leq J$, while the sequence $\beta_0, ..., \beta_J$ denotes the *claims development pattern*.

We summarize the weaker set of assumptions BFII as follows.

Model assumptions (BFII)

BFII1 Cumulative claims $C_i j$ of different accident years are independent.

$$\{C_{i,0}, ..., C_{i,I}\}, \{C_{j,0}, ..., C_{j,I}\} \, for \, i \neq j \, are \, independent. \tag{1.15}$$

BFII2 There exists parameters $\mu_0, ..., \mu_I > 0$ and a pattern $\beta_0, ..., \beta_J > 0$ with $\beta_J = 1$ such that

$$EC_{i0} = \beta_0 \mu_i, \tag{1.16}$$

holds for all $0 \leq i \leq I$, $0 \leq j \leq J - 1$ and $1 \leq k \leq J - j$.

Under the assumption BFI, the conditional expected value of aggregate loss for a given accident year based on known historical data at the end of year I is derived by

$$\begin{aligned} E(C_{i,J}|D_I) &= E(C_i|C_{i,0}, ..., C_{i,I-i}) \\ &= C_{i,I-i} + E(C_i - C_{i,I-i}) \\ &= C_{i,I-i} + (1 - \beta_{I-i})\mu_i \end{aligned} \tag{1.17}$$

Under the weaker assumption BFII, we obtain the same result when assuming the independence of incremental claims amount $C_{i,J} - C_{i,I-i}$ of $C_{i,0}, ..., C_{i,I-i}$.

The BF estimator for $E(C_{i,J}|D_I)$ based on (1.16) is given by the following formula.

$$\hat{C_{i,J}^{BF}} = E(\hat{C_{i,J}}|D_I) = C_{i,I-i} + (1 - \hat{\beta_{I-i}})\hat{\mu_i} \tag{1.18}$$

for $0 \leq i \leq I$, where $\hat{\beta_{I-i}}$ is an appropriate estimate for β_{I-i} and $\hat{\mu_i}$ is a given prior estimate for the expected ultimate claim $E(C_{i,J})$.

Based on equations (1.4) and (1.18), the estimated outstanding claims reserves for accident year i in the BF method can be obtained by

$$\hat{R}_i^{BF} = \hat{E}(C_{i,J}|D_I) - C_{i,I-i} = \hat{C_{i,I}} - C_{i,I-i} = (1 - \hat{\beta_{I-i}})\hat{\mu_i}$$

.

Both BF and chain-ladder method are used for estimating IBNR reserves, and they can produce different results based on the assumptions made. But in some cases, the BF method can be viewed as a generalization of the CL method when certain parameters are set accordingly. Under the model assumption CL, it follows for all $1 \leq j \leq J$

$$E(C_{i,j}) = E(E(C_{i,j}|C_{i,j-1})) = f_{j-1}EC_{i,j-1} = ... = EC_{i,0}\prod_{k=0}^{j-1} f_k \tag{1.19}$$

$$E(C_{i,J}) = EC_{i,0}\prod_{k=0}^{J-1} f_k \tag{1.20}$$

Combining (1.19) and (1.20), we obtain

$$E(C_{i,j}) = EC_{i,J}\prod_{k=0}^{J-1} f_k^{-1} \tag{1.21}$$

It corresponds to the model assumption BFII with

$$\mu_i = E(C_{i,J}), \beta_j = \prod_{k=0}^{J-1} f_k^{-1}, for\, j = 0, ..., J-1\, and\, \beta_J = 1. \tag{1.22}$$

When we derive the claim development pattern parameters $\beta_0, ..., \beta_{J-1}$ using the estimated CL development factors $\hat{f}_j$, we obtain

$$\hat{\beta_j^{CL}} = \prod_{k=j}^{J-1} \frac{1}{\hat{f}_k} \tag{1.23}$$

And consequently we can rewrite the BF estimator for $E(\hat{C_{i,J}})|D_I)$ based on the following formula.

$$\begin{aligned} \hat{C_{i,J}^{BF}} &= C_{i,I-i} + (1 - \hat{\beta_{I-i}^{CL}})\hat{\mu_i} \\ &= C_{i,I-i} + (1 - \frac{1}{\prod_{k=I-i}^{J-1} \hat{f}_k})\hat{\mu_i} \end{aligned} \tag{1.24}$$

And the Chain-Ladder estimator can be interpreted by the formula

$$\begin{aligned} \hat{C_{i,J}^{CL}} &= C_{i,I-i} \prod_{j=I-i}^{J-1} \hat{f}_j \\ &= C_{i,I-i} + C_{i,I-i}(\prod_{j=I-i}^{J-1} \hat{f}_j - 1) \\ &= C_{i,I-i} + \frac{\hat{C_{i,J}^{CL}}}{\prod_{j=I-i}^{J-1} \hat{f}_j}(\prod_{j=I-i}^{J-1} \hat{f}_j - 1) \\ &= C_{i,I-i} + (1 - \hat{\beta_{I-i}^{CL}})\hat{C_{i,J}^{CL}} \end{aligned} \tag{1.25}$$

We can see that the difference between the Chain-Ladder and BF estimators is that in the Chain-Ladder method, the estimate is based only on observations, while in the BF method, we use a prior estimate $\hat{\mu_i}$. In summary, Chain Ladder relies on historical patterns with a deterministic approach, whereas BF introduces stochastic elements and is more flexible in accommodating different assumptions.

2. Gneralized linear mixed models

The concept of Generalized Linear Models (GLM) was developed by Nelder and Wedderburn (1972) in an extension of the classical linear model. GLM extended the normal distribution of response variables to exponential family distributions, which significantly advancing the application of statistical methods in the actuarial field. Subsequently, Generalized Linear Mixed Models (GLMM) not only introduced random effects on the basis of GLM but also transformed the mean of response variables into a linear form by linking with a link function. This transformation effectively expanded the linear form to nonlinear forms, demonstrating excellent adaptability for non-independent data. In this section, we introduce the basic assumptions and model structures of GLM and GLMM.

2.1 Generalized Linear Models

Gneralized linear models (GLM) is a generalization of linear regression to response types other than Gaussian, as long as the distribution of that response is a member of the exponential family. The definition and basic assumptions of the generalized linear models and the common forms of the distribution functions stated in this section, refer to McCullagh and Nelder (1989).

2.1.1 Basic assumptions

Before discussing the structure of Generalized Linear Models (GLM), let us first discuss the distribution of the response variable, specifically the exponential distribution family. The exponential family implies that the density of a single random variable y can be represented in the following form:

$$f(y) = exp\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\} \tag{2.1}$$

In (2.1), it typically takes $a(\phi) = \frac{\phi}{\omega}$, where $\omega > 0$ is a given weight and ϕ is called the *dispersion parameter*, representing the scale. θ represents the *natural parameter*. The functions $a(\cdot)$, $b(\cdot)$, and $c(\cdot, \cdot)$ are some real-valued functions, and function $b(\cdot)$ (*cumulant function*) is supposed to be twice-differentiable. The exponential distribution family, as represented in (2.1), includes many common distributions by specifying the functions$a(\cdot)$, $b(\cdot)$, and $c(\cdot, \cdot)$. These incorporate various continuous distributions such as the normal distribution, exponential distribution, gamma distribution, and discrete distributions like the Poisson distribution, binomial distribution, etc.

Assuming independent random variables $y_1, ..., y_n$ have the form of (2.1), the log-likelihood function for each random variable y_i is:

$$l_i(\theta_i, \phi; y_i) = \frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i), i = 1, 2, ..., n \tag{2.2}$$

Then the log-likelihood function is typically takes the form:

$$\boldsymbol{l(\theta, \phi; y)} = \sum_{i=1}^{n} l(\theta_i, \phi; y_i) \tag{2.3}$$

Integrating both sides of the probability density function (2.1), the following equation holds.

$$\int exp\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i)\}dy_i = 1, i = 1, 2, ..., n \tag{2.4}$$

Based on (2.4), taking the first and second derivatives with respect to θ_i on both sides, we obtain:

$$\int (y_i - b^{'}(\theta_i))exp\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i)\}dy_i = 0, i = 1, 2, ..., n$$

the expectation can be expressed as

$$\int y_i exp\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i)\}dy_i = EY_i, i = 1, 2, ..., n$$

and then the expectation have the form:

$$E[y_i] = \mu_i = b^{'}(\theta_i) \tag{2.5}$$

in the same way it can be proved that the variance have the form:

$$Var[y_i] = a(\phi_i)b^{''}(\theta_i) \tag{2.6}$$

Let $b^{''}(\theta_i) = V(\mu_i)$, where $V(\mu_i)$ is referred to as the variance function. Typically, taking $a(\phi_i) = \frac{\phi_i}{\omega_i}$ where ω_i represents a given weight, then $Var(y_i) = \frac{\phi_i}{\omega_i}b^{''}(\theta_i)$. The variance function $V(\mu_i)$ describes the relationship between variance and mean. For generalized linear models, having knowledge of the variance function $V(\mu_i)$ allows the determination of the distribution type. Conversely, if the distribution type is known, its variance structure is consequently defined. Taking Poisson distribution as an example, its density function is given by

$$f(y) = \frac{1}{y!}e^{-\lambda}\lambda^y = exp(ylog(\lambda) - \lambda - log(y!))$$

corresponding to the density function of the form of exponential family (2.1): $\theta = log(\lambda), b(\theta) = e^\theta, \phi = 1, c(y) = -log(y!)$. And also the equation $b^{'}(\theta) = b^{''}(\theta) = e^\theta = \lambda$ holds. This shows that Poisson distribution satisfies the definition and propeety of the exponential family of distributions and that the variance function is determined if the dictribution is determined and vice versa. Table 2.1 illustrates the common forms of distribution functions and their relationships with the variance functions.

Table 2.1: Generalized Linear Models and Their Variance Functions

Distribution	Variance function
Normal distribution $N(\mu, \sigma^2)$	1
Poisson distribution $Poi(\lambda)$	μ
Binomial distribution $B(n, p)$	$\mu(1-\mu)$
Gamma distribution $Gamma(v, \frac{\mu}{v})$	μ^2
Inverse Gaussian distribution	μ^3

Next we introduce the model components of generalized linear models. Generalized Linear Models (GLM) consist of three components: the random component, the systematic component, and the link function, which are shown below:

1. The random component: this refers to the probability distribution of the response variable or the random error component. Generalized linear models assume that the response variables $y_1, ..., y_n$ are mutually independent, and their distribution belong to the same exponential distribution family.

2. The systematic component: the systematic component is also known as the linear predictor part and the linear predictor $\boldsymbol{\eta}$ is given by

$$\boldsymbol{\eta} = \sum_{j=1}^{p} \boldsymbol{x_j} \beta_j$$

   where $\boldsymbol{x_j}, j = 1, ..., p$ are column vectors of the matrix $\boldsymbol{X}$. The linear prediction for the ith object can be represented as: $\eta_i = x_{i1}\beta_1 + x_{i2}\beta_2 + ... + x_{ip}\beta_p = X_i^T \cdot \beta$, i=1,2,...,n, where $\beta = (\beta_1, \beta_2, ..., \beta_p)^T, X_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$.

3. Link functions: link function $g(\mu_i)$ transforms the expected value of the response variable to the linear predictor scale, allowing for a wider range of distributions. The relationship is given by $g(\mu_i) = \eta_i = X_i^T \cdot \beta$. It is assumed to be a sufficiently smooth and strictly monotonic function. Therefore, $g(\cdot)$ has an inverse function $g^{-1}(\cdot)$, $\mu_i = g^{-1}(\eta_i)$

The link function plays a crucial role in Generalized Linear Models. Commonly used link functions include:

1. Identity link: $\eta_i = g(\mu_i) = \mu_i$

2. Log link: $\eta_i = g(\mu_i) = \log(\mu_i)$

3. Logit link:$\eta_i = g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$

4. Probit link: $\eta_i = g(\mu_i) = \Phi^{-1}(\mu_i)$

5. Complementory log-log link: $\eta_i = g(\mu_i) = \log(-\log(1-\mu_i))$

The choice of link function depends on the nature of the response variable and the characteristics of the data being modeled. For a given exponential family, there exists a special link function called the *canonical link function.* As stated in McCullagh and Nelder (1989), the canonical link function results from setting the linear component of the model equal to the natural parameter θ.

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \boldsymbol{\theta}$$

The canonical link function offers a key advantage by rendering the canonical parameter of the joint distribution equal to $\boldsymbol{\beta}$. This simplifies the interpretation of model parameters and facilitates maximum likelihood estimation. Additionally, it often leads to more stable and reliable model estimation. However, despite the mentioned statistical property, it is not always justified to apply the canonical link to specific datasets. The use of the canonical link is entirely optional; in this thesis, we consider both canonical link functions and other link functions.

The general form of generalized linear models and their canonical link functions are shown in the table 2.2:

Table 2.2: Generalized Linear Models and Their Canonical Link Functions

Distribution	Link function
Normal distribution $N(\mu, \sigma^2)$	$g(\mu) = \mu = \eta$
Poisson distribution $Poi(\lambda)$	$g(\lambda) = log(\lambda) = \eta$
Binomial distribution B(n,p)	$g(\mu) = log(\frac{p}{1-p}) = \eta$
Gamma distribution $Gamma(v, \frac{\mu}{v})$	$g(\mu) = \frac{1}{\mu} = \eta$
Inverse Gaussian distribution	$g(\mu) = \frac{1}{\mu^2} = \eta$

2.1.2 Parameter estimation

Parameter estimation in generalized linear models (GLM) is based on the Maximum Likelihood Theory. The log-likelihood function for each component of the random vector $\mathbf{Y}$ takes the form as in 2.2, then the log-likelihood for all components of $\mathbf{Y}$ is given by

$$\boldsymbol{l(\theta, \phi; y)} = \sum_{i=1}^{n} l(\theta_i, \phi; y_i) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i), i = 1, 2, ..., n \qquad (2.7)$$

The maximum likelihood estimators of model parameters β, which maximize the log-likelihood, are the roots of the score function. From the chain rule, the derivative of log-likelihood with respect β follows

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}(\frac{\partial l_i}{\partial \boldsymbol{\beta}}) = \sum_{i=1}^{n}(\frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}})$$

Using the derivative of the inverse function we have

$$\frac{\partial \theta_i}{\partial \mu_i} = (\frac{\partial \mu_i}{\partial \theta_i})^{-1} = \frac{1}{b''(\theta_i)} = \frac{1}{V(\mu_i)}$$
$$\frac{\partial \mu_i}{\partial \eta_i} = (\frac{\partial \eta_i}{\partial \mu_i})^{-1} = \frac{1}{g'(\mu_i)}$$

Then the maximum likelihood estimators of $\beta_0, ..., \beta_k$ are obtained from the solution of the equation

$$\begin{aligned} 0 = \frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{1}{a(\phi)} \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{V(\mu_i)} \frac{1}{g'(\mu_i)} \boldsymbol{x_i} \right] \\ &= \frac{1}{a(\phi)} \sum_{i=1}^{n} \left[\omega(\mu_i) g'(\mu_i)(y_i - \mu_i) \boldsymbol{x_i} \right] \end{aligned} \tag{2.8}$$

where

$$\omega(\mu_i) = \frac{1}{V(\mu_i)(g'(\mu_i))^2} > 0 \, for \, i = 1, ..., n$$

.

Adding $(\boldsymbol{X^T \hat{W} X})\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} w(\hat{\mu_i}) \boldsymbol{x_i x_i^T}\right] \hat{\boldsymbol{\beta}}$ to both sides of (2.8), the equation can be rewritten in a matrix form

$$(\boldsymbol{X^T \hat{W} X})\hat{\boldsymbol{\beta}} = \boldsymbol{X^T \hat{W} X \hat{z}} \tag{2.9}$$

hence

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X^T \hat{W} X})^{-1} \boldsymbol{X^T \hat{W} \hat{z}}$$

where$\hat{\boldsymbol{z}} = (\hat{z_1}, ..., \hat{z_n})^T$ denotes an adjusted rsponse with

$$\hat{\boldsymbol{z_i}} = \boldsymbol{x_i^T} \hat{\boldsymbol{\beta}} + (y_i - \hat{\mu_i})(\frac{\partial \hat{\eta_i}}{\partial \hat{\mu_i}}) \, for \, i = 1, ..., n.$$

and the matrix of weights $\hat{\boldsymbol{W}} = diag\{w(\hat{\mu_1}), ..., w(\hat{\mu_n})\}$.

In practice, obtaining an analytical solution for the above equation is challenging, and therefore numerical methods must be employed for solving. McCullagh and Nelder (1989) demonstrated that by using the Newton-Raphson method in conjunction with the Fisher scoring algorithm, the maximization of the logarithmic likelihood function can be transformed into an *iterative weighted least squares procedure* (IWLS). The IWLS algorithm for generalized linear models is as follows.

1. Start with $\hat{\boldsymbol{\beta}}^{\boldsymbol{(0)}}$ and the initial linear predictor $\hat{\boldsymbol{\eta}^{(0)}}$ can be obtained based on $\hat{\boldsymbol{\eta}^{(0)}} = \mathbf{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{(0)}}$. The corresponding initial estimates for the mean of the response $\hat{\boldsymbol{\mu}^{(0)}}$ based on the chosen link function $\hat{\boldsymbol{\mu}^{(0)}} = g^{-1}(\hat{\boldsymbol{\eta}^{(0)}})$.

2. Construct an adjusted response $\hat{\boldsymbol{z}^{(0)}}$ and a matrix of weights $\hat{\boldsymbol{W}^{(0)}}$; The vector $\hat{\boldsymbol{z}}^{\boldsymbol{(0)}} = (\hat{z_1^{(0)}}, ..., \hat{z_n^{(0)}})^T$ denotes the adjusted dependent variable, where

$$\hat{z_i^{(0)}} = \boldsymbol{x_i^T} \hat{\boldsymbol{\beta}^{(0)}} + (y_i - \hat{\mu_i^{(0)}})(\frac{\partial \hat{\eta_i^{(0)}}}{\partial \hat{\mu_i^{(0)}}}) \, for \, i = 1, ..., n. \tag{2.10}$$

   And the matrix of weights $\hat{\boldsymbol{W}^{(0)}} = diag\{w(\hat{\mu_1^{(0)}}), ..., w(\hat{\mu_n^{(0)}})\}$.

3. Calculate $\hat{\boldsymbol{\beta}^{(1)}}$ by weighted lease squares (WLS) method based on equation (2.9); Use the constructed $\hat{\boldsymbol{z}^{(0)}}$ and weights $\hat{\boldsymbol{W}^{(0)}}$ from step 2 to fit an ordinary linear model, predicting new means $\hat{\boldsymbol{\mu}^{(1)}}$ and new linear predictors $\hat{\boldsymbol{\eta}^{(1)}}$.

4. Repeat steps 2 and 3 till $\hat{\boldsymbol{\beta}}$ converges.

We obtained the iterative formula used for the GLM model

$$\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{W}}^{(m)}\boldsymbol{X}\boldsymbol{\beta}^{(\hat{m}+1)} = \boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{W}}^{(m)}\hat{\boldsymbol{z}}^{(m)}$$

hence

$$\boldsymbol{\beta}^{(\hat{m}+1)} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}^{\hat{(m)}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}^{\hat{(m)}}\hat{\boldsymbol{z}}^{(m)}$$

2.1.3 Statistical Tests for Parameter Significance

For a given set of data in generalized linear regression analysis, it's common to establish more than one model. We typically use statistical tests to validate the appropriateness of the built models and assess the relative merits among them. Next, we will elaborate on the calculation principles of the two most commonly used statistical testing methods in generalized linear regression.

Likelihood ratio test

The Likelihood Ratio Test (LRT) is a generic hypothesis test that can also be specifically applied to GLM. LRT based on the comparison of the maximized loglikelihood functions under the full model and its respective reduced model.

$H_0 : \beta_1 = ... = \beta_r = 0$, which means the reduced model is adequate.
H_1 : There are no constraints on β_j, which means the full model is better.

The Likelihood Ratio Test (LRT) is employed to compare these two nested models since the model under H_0 represents a special or simplified case of the model under H_1. The test statistic is given by:

$$LRT = -2(l_1 - l_0)$$

where l_0 and l_1 are the maximum loglikelihoods under the two models corresponding to H_0 and H_1. The LR statistic follows a chi-square distribution, with degrees of freedom equal to the number of predictor variables removed in the reduced model. If the p-value from the LRT is less than the chosen significance level (typically 0.05), the null hypothesis is rejected, suggesting that the model improvement is statistically significant.

Wald test

The Wald test is a statistical method to assess the significance parameters. The test can be used to test whether the linear constraint $\boldsymbol{C\beta} = \boldsymbol{r}$ holds:

$$H_0 : \boldsymbol{C\beta} = \boldsymbol{r}$$

$$H_1 : \boldsymbol{C\beta} \neq \boldsymbol{r}$$

where $\boldsymbol{C}$ is a $k \times p$ matrix and $\boldsymbol{\beta}$ is a column vector with p elements. The test statistic is given in the following form:

$$W_n = \frac{1}{a(\phi)}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^T \left[\boldsymbol{C}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-\boldsymbol{1}}\right]^{-1} (\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{r})$$

The Wald test statistic under the null hypothesis follows a chi-square distribution with k degrees of freedom. A large W_n indicates a small p-value, leading to the rejection of the null hypothesis. This suggests there is evidence of a significant effect for the coefficient of interest.

2.1.4 Model Fitting and Evaluation

Goodness-of-fit test is a statistical method used to assess how well a statistical model fits the observed data. It is commonly applied to compare the distribution expected by the model with the actual distribution of the observed data. The basic steps include: (1) Formulating the null hypothesis, assuming the model fits the data well. (2) Classifying the data into categories and generating observed frequencies. (3) Calculating expected frequencies based on the model. (4) Computing the goodness-of-fit test statistic, such as chi-square or Deviance. (5) Using the statistic for hypothesis testing. (6) Drawing a conclusion based on the p-value or other criteria to decide whether to reject the null hypothesis. Goodness of fit relies on choosing suitable statistic. In this thesis we introduce two common ones: Pearson statistic and Deviance statistic.

Pearson statistic

This method compares the differences between observed and expected frequencies. In a goodness-of-fit test, data is first classified, and the observed frequencies in each category are compared to the expected frequencies. The chi-square statistic is used to measure the degree of fit between observed and expected frequencies, and this statistic approximately follows a chi-square distribution under the null hypothesis.

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu_i})^2}{V(\hat{\mu_i})}$$

Deviance statistic

For models like generalized linear models (GLM), the Deviance goodness-of-fit test is more widely used. Deviance measures the difference between the fit of the observed data and a perfect fit. In a Deviance goodness-of-fit test, the model's fit is compared to a special model (often the saturated model), and the test determines whether this difference is significant.

$$d_i = sign(y_i - \hat{\mu_i})\sqrt{2\left[l(\hat{\boldsymbol{\theta}}^{*}, \phi; \mathbf{y}) - l(\hat{\boldsymbol{\theta}}, \phi; \mathbf{y})\right]}$$

2.2 Generalized Linear Mixed Models

A Generalized Linear Mixed Model (GLMM) is a combination of a Generalized Linear Model (GLM) and a Linear Mixed Model (LMM), which can be used to fit non-independent data, including discrete and other types of data. Refer to McCulloch et al. (2001), the model structure is as follows.

Let Y_{ij} be the response variable, i=1,...,n represent n observations, and j=1,...,n_i denote the j-th observation of the i-th cluster. Assuming given the random effects

$\upsilon_i, i = 1..., n$ for each observation, the response variables $Y_{i,1}, ..., Y_{i,n_i}$ are assumed to be independently distributed according to the density from the exponential family.

$$f(y_{ij}|\upsilon_i, \beta, \theta) = exp\{\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi)\} \tag{2.11}$$

In Generalized Linear Mixed Models, the expected value and variance of the response variable are respectively:

$$\mu_{ij} = E\left[y_{ij}|\upsilon_i\right] = b^{'}(\theta_{ij})$$

$$Var\left[y_{ij}|\upsilon_i\right] = a(\phi)b^{''}(\theta_{ij}) = a(\phi)V(\mu_{ij})$$

The random and the fixed effects are combined to form the linear predictor

$$\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Z\upsilon}$$

where **X** and **Z** are the design matrices for fixed effects and random effects, respectively. $\boldsymbol{\beta}$ $(p \times 1)$ denotes the vector of the fixed effects parameter and $\boldsymbol{\upsilon}$ $(q \times 1)$ denotes the vector of random effects. The model requires the random effects υ_i to be mutually independent and identically distributed with the density function $f(\upsilon_i|\alpha)$, where α is an unknown parameter. Generally, in GLMM model one assumes the random effects υ_i follow a normal distribution $\boldsymbol{N(0, \Sigma_i)}$, where the covariance matrix $\boldsymbol{\Sigma_i}$ is a positive definite matrix defined by the parameter α, with α representing the parameter in the density function of the random effects. The random effect reflects the homogeneity within the same group of variables and the heterogeneity between different groups of variables, a relationship that cannot be captured by fixed effects.

The expectation μ_{ij} of the response variable y_{ij} is linked to the systematic component through a monotonically differentiable function g(·). In this context, "monotonically differentiable" indicates that g(·) changes smoothly and consistently across its domain, ensuring a clear and well-defined relationship between η_{ij} and μ_{ij}:

$$g(\mu_{ij}) = \eta_{ij} = x_{ij}^{'}\beta + z_{ij}^{'}\upsilon$$

GLMMs allow for the modeling of both fixed effects and random effects, which capture the variability associated with groups or clusters within the data. By incorporating random effects, GLMMs provide more flexibility in modeling correlated data and hierarchical structures, thus offering a more robust and flexible framework for modeling a wide range of real-world datasets.

2.2.1 Parameter estimation

In GLMM, the construction of likelihood function is based on the marginal density $f(y_{ij})$ and has the following form:

$$L(y_{ij}; \phi, \beta, \alpha) = \prod_{i=1}^{n} f(y_i; \phi, \beta, \alpha) = \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} f(y_{ij}|\phi, \beta, \upsilon_i) f(\upsilon_i|\alpha) d\upsilon_i \tag{2.12}$$

In order to derive the likelihood solution, integration over the random-effects distribution is essential. However, this task often involves complex numerical computations, and closed-form expressions for the integral are typically not available. The maximum likelihood estimation method frequently requires handling high-dimensional integrals, which can be challenging, particularly when the response variable is not normally distributed. In general cases, direct computation of the integral becomes difficult, leading to the necessity of employing numerical approximation methods. Commonly used techniques include Laplace approximation, penalized quasi-likelihood, and Gaussian Hermite methods.

Laplace approximation

Laplace Approximation employs a second-order Taylor expansion to approximate intricate integrals as Gaussian integrals, thereby streamlining computational processes. Specifically, in maximum likelihood estimation, Laplace approximation approximates the function near its global maximum in parameter space using a Taylor expansion, facilitating more efficient calculations. The Laplace approximation can be employed to approximate the integral by rewriting the integral term in (2.12) as:

$$\begin{aligned}
&\int \prod_{j=1}^{n_i} f(y_{ij}|\phi, \beta, v_i) f(v_i|\alpha) dv_i \\
&= \int f(\boldsymbol{y}|\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{v}) f(\boldsymbol{v}|\boldsymbol{\alpha}) d\boldsymbol{v} \\
&= \int e^{logf(\boldsymbol{y}|\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{v}) f(\boldsymbol{v}|\boldsymbol{\alpha})} d\boldsymbol{v} \\
&= \int e^{g(v)} dv
\end{aligned} \tag{2.13}$$

where $g(v) = logf(\boldsymbol{y}|\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{v}) f(\boldsymbol{v}|\boldsymbol{\alpha})$. Assuming $g(v)$ is differentiable, we aim to select $\hat{v}$ in a way that maximizes $g(v)$, ensuring that the necessary and sufficient conditions $g^{'}(v) = 0$ and $g^{''}(v) < 0$ for achieving the global maximum of $g(v)$. The second order Taylor expansion around $\hat{v}$ for $g(v)$ is given by:

$$\begin{aligned}
g(v) \approx \tilde{g}(v) &= g(\hat{v}) + (v - \hat{v}) g^{'}(\hat{v}) + \frac{1}{2}(v - \hat{v})^2 g^{''}(\hat{v}) \\
&= g(\hat{v}) - \frac{1}{2}(v - \hat{v})^2 (-g^{''}(\hat{v}))
\end{aligned} \tag{2.14}$$

It can be seen that $e^{\tilde{g}(v)}$ is proportional to the normal distribution (μ_L, σ_L^2) where $\mu_L = \hat{v}$ and $\sigma_L^2) = -\frac{1}{g^{''}(\hat{v})}$. Then, from (2.12) to (2.14), the Laplace approximation for the likelihood $L(y_{ij}; \phi, \beta, \alpha)$ can also be written as:

$$
\begin{aligned}
L(y_{ij};\phi,\beta,\alpha) &= \prod_{i=1}^{n} f(y_i;\phi,\beta,\alpha) = \prod_{i=1}^{n}\int\prod_{j=1}^{n_i} f(y_{ij}|\phi,\beta,\upsilon_i)f(\upsilon_i|\alpha)d\upsilon_i \\
&= \int e^{g(\upsilon)}d\upsilon \approx \int e^{\tilde{g}(\upsilon)}d\upsilon \\
&= \int e^{g(\hat{\upsilon})-\frac{1}{2}(\upsilon-\hat{\upsilon})^2(-g''(\hat{\upsilon}))}d\upsilon \\
&= \int e^{g(\hat{\upsilon})}e^{-\frac{1}{2\sigma_L^2}(\upsilon-\mu_L)^2}d\upsilon \\
&= e^{g(\hat{\upsilon})}\int e^{-\frac{1}{2\sigma_L^2}(\upsilon-\mu_L)^2}d\upsilon \\
&= e^{g(\hat{\upsilon})}\sqrt{2\pi\sigma_L^2}
\end{aligned}
\tag{2.15}
$$

When using the *glmer* function in the *lme4* package in R software, Laplace approximation is employed as the default method for regression analysis. Moreover, the accuracy of this approximation increases if higher order of Taylor expansion is used. A more detailed explanation of Laplace approximation of generalized linear mixed model could be found in Handayani et al. (2017).

Gauss-Hermite

Adaptive Gauss-Hermite quadrature is an extension of traditional Gaussian quadrature that provides a method to numerically compute integrals involving Gaussian weight functions. This method is particularly effective for integrals of the form: $\int_{-\infty}^{\infty} e^{-z^2}h(z)\,dz$, where $h(z)$ is a smooth function and e^{-z^2} is the weight function. The formula for Gauss-Hermite quadrature can be expressed as:

$$
\int_{-\infty}^{\infty} e^{-z^2}h(z)\,dz \approx \sum_{q=1}^{Q} w_q \cdot h(z_q) \tag{2.16}
$$

The z_q denote the roots of the Qth order Hermite polynomial $H_Q(z)$, where Q indicates the order of the approximation. And w_q represents the corresponding weights, the weights w_q have a particular functional form involving lower-order Hermite polynomials $H_{Q-1}(z)$ and the number of specified nodes Q. Both the nodes (or quadrature points) z_q and the weights w_q are tabulated in Abramowitz and Stegun (1968).

Assuming a single random effect υ_i follows a normal distribution N(0,σ_i^2), the contribution of the i^{th} random effect to the marginal likelihood is represented as:

$$
l_i(y_i;\phi,\beta,\alpha) = \int\prod_{j=1}^{n_i} f(y_{ij}|\phi,\beta,\upsilon_i)f(\upsilon_i|\alpha)d\upsilon_i \tag{2.17}
$$

In practice, maximum likelihood estimates are obtained from the log of the quadrature approximation. After a reparameterization to $\delta_i = \sigma^{-1}\upsilon_i$, we have $\upsilon_i = \delta_i\sigma$, and the marginal likelihood function can be written as

$$
\begin{aligned}
l_i(y_i;\phi,\beta,\alpha) &= \int\prod_{j=1}^{n_i} f(y_{ij}|\phi,\beta,\upsilon_i)f(\upsilon_i|\alpha)d\upsilon_i \\
&= \int\prod_{j=1}^{n_i} f(y_{ij}|\phi,\beta,\delta_i\sigma)\phi(\delta_i;0,1)d\delta_i
\end{aligned}
\tag{2.18}
$$

In order to calculate (2.18), we consider integrating $\int_{-\infty}^{\infty} h(z) \cdot \phi(z; 0, 1)dz$, where $\phi(z; 0, 1)$ denotes the standard normal density. Let $\hat{\mu}$ and $\hat{\sigma}$ be

$$\hat{\mu} = mode\,[h(z) \cdot \phi(z; 0, 1)]$$

$$\hat{\sigma^2} = \left[-\frac{\partial^2}{\partial z^2} ln(h(z) \cdot \phi(z; 0, 1))|_{z=\hat{\mu}} \right]^{-1}$$

then $\hat{\mu}$ and $\hat{\sigma^2}$ would be the mean and variance of the distribution with density $h(z) \cdot \phi(z; 0, 1)$. Let $V(z) = \frac{h(z)\cdot\phi(z;0,1)}{\phi(z;\mu,\sigma)}$, then

$$\begin{aligned} &\int_{-\infty}^{\infty} h(z) \cdot \phi(z; 0, 1)dz \\ &= \int_{-\infty}^{\infty} V(z)\phi(z; \mu, \sigma)dz \\ &= \int_{-\infty}^{\infty} V(z)\frac{1}{\sqrt{2\pi}\sigma} exp\{-\frac{1}{2}(\frac{z-\mu}{\sigma})^2\}dz \end{aligned} \tag{2.19}$$

We introduce a new variable $\tilde{z}$ for reparameterization, where $\tilde{z} = \frac{z-\mu}{\sqrt{2}\sigma}$. We then substitute z with $z = \mu + \sqrt{2}\sigma\tilde{z}$ to continue.

$$\begin{aligned} &\int_{-\infty}^{\infty} V(z)\frac{1}{\sqrt{2\pi}\sigma} exp\{-\frac{1}{2}(\frac{z-\mu}{\sigma})^2\}dz \\ &= \int_{-\infty}^{\infty} V(\mu + \sqrt{2}\sigma\tilde{z})\frac{1}{\sqrt{2\pi}\sigma} exp\{-\tilde{z}^2\}\sqrt{2}\sigma d\tilde{z} \\ &= \int_{-\infty}^{\infty} \frac{V(\mu + \sqrt{2}\sigma\tilde{z})}{\sqrt{\pi}} exp\{-\tilde{z}^2\}d\tilde{z} \end{aligned} \tag{2.20}$$

Based on Gauss-Hermite quadrature formula (2.16) and (2.20) we obtain

$$\int_{-\infty}^{\infty} \frac{V(\mu + \sqrt{2}\sigma\tilde{z})}{\sqrt{\pi}} exp\{-\tilde{z}^2\}d\tilde{z} \approx \sum_{q=1}^{Q} \frac{V(\mu + \sqrt{2}\sigma\tilde{z}_q)}{\sqrt{\pi}} w_q \tag{2.21}$$

replacing μ and σ^2 by $\hat{\mu}$ and $\hat{\sigma^2}$, we have

$$\int_{-\infty}^{\infty} V(z)\phi(z; \mu, \sigma)dz \approx \sum_{q=1}^{Q} \frac{V(\hat{\mu} + \sqrt{2}\hat{\sigma}\tilde{z}_q)}{\sqrt{\pi}} w_q \tag{2.22}$$

Based on the calculations above, it is sufficient to integrate $\int_{-\infty}^{\infty} h(z)\cdot\phi(z; 0, 1)dz$. The process is shown as follows.

$$\begin{aligned} &\int_{-\infty}^{\infty} h(z) \cdot \phi(z; 0, 1)dz \\ &= \int_{-\infty}^{\infty} \frac{h(z) \cdot \phi(z; 0, 1)}{\phi(z; \mu, \sigma)} \phi(z; \mu, \sigma)dz \\ &= \int_{-\infty}^{\infty} V(z)\phi(z; \mu, \sigma)dz \\ &\approx \sum_{q=1}^{Q} \frac{V(\hat{\mu} + \sqrt{2}\hat{\sigma}\tilde{z}_q)}{\sqrt{\pi}} w_q \end{aligned} \tag{2.23}$$

By substituting $V(z) = \frac{h(z)\cdot\phi(z;0,1)}{\phi(z;\mu,\sigma)}$ and $z = \hat{\mu} + \sqrt{2}\hat{\sigma}\tilde{z}$ into the above equation

$$\begin{aligned}
&\sum_{q=1}^{Q} \frac{V(\hat{\mu} + \sqrt{2}\hat{\sigma}\tilde{z}_q)}{\sqrt{\pi}} w_q \\
&= \sum_{q=1}^{Q} \frac{\sqrt{2\pi}\hat{\sigma}}{\sqrt{\pi}} exp\{\frac{(\sqrt{2}\hat{\sigma}\tilde{z}_q)^2}{2\hat{\sigma}^2}\}\phi(\hat{\mu} + \sqrt{2}\hat{\sigma}\tilde{z}_q; 0, 1) w_q \cdot h(\hat{\mu} + \sqrt{2}\hat{\sigma}\tilde{z}_q) \\
&= \sum_{q=1}^{Q} \sqrt{2}\hat{\sigma} exp\{\tilde{z}_q^2\}\phi(z_q; 0, 1) w_q \cdot h(z_q) \\
&= \sum_{q=1}^{Q} w_q^* h(z_q)
\end{aligned} \tag{2.24}$$

where $w_q^* = \sqrt{2}\hat{\sigma} exp\{\tilde{z}_q^2\}\phi(z_q; 0, 1) w_q$ and $z = \hat{\mu} + \sqrt{2}\hat{\sigma}\tilde{z}$. (2.24) is called an *adaptive Gauss–Hermite quadrature formula* and can be used to approximate the integration $\int_{-\infty}^{\infty} h(z) \cdot \phi(z; 0, 1) dz$. Full details and references for Gauss–Hermite quadrature for generalized linear mixed models are in Antonio and Beirlant (2007).

In Gauss-Hermite quadrature, the integration of the polynomial function $h(z)$ achieves high precision. Increasing the number of quadrature points can improve the accuracy of integral approximations. Results from other methods are often compared to adaptive Gauss-Hermite quadrature due to its high accuracy when Q is large, refer to Kim et al. (2013). However,there are limitations to its usage: the computational burden increases with more quadrature points and grows exponentially as the number of random effects increases. Complex models with multiple, especially nested or crossed, random effects can become computationally unmanageable, refer to Fitzmaurice et al. (2011).

The Laplace method approximates the objective function around its mode, providing asymptotically unbiased estimates with lower computational burden compared to adaptive Gauss-Hermite quadrature. As a result, Laplace approximation may yield less accurate estimates when dealing with datasets that have small cluster sizes. Laplace approximation is more flexible and slightly less computationally intensive than adaptive Gauss-Hermite quadrature (GHQ). For appropriately structured data, adaptive GHQ is expected to produce more accurate estimates. Additionally, Laplace approximation is numerically equivalent to Adaptive Gauss-Hermite quadrature with a single quadrature node, refer to Tuerlinckx et al. (2006). It might seem that adaptive Gauss-Hermite quadrature (GHQ) should always be employed to obtain a more accurate approximation of the likelihood integral.

In addition, PQL (Penalized Quasi-Likelihood) is a method proposed by Breslow and Clayton (1993) for estimating penalized quasi-likelihood, which describes a linearization strategy for approximate inference in GLMM. The technique is also known as 'pseudolikelihood' approximation, utilizing Laplace approximation based on quasi-likelihood. It aims to maximize the likelihood function (2.12). While widely used, likelihood inference can be inaccurate, leading to biased estimates in cases with large variances or small means. Further descriptions of the pseudolikelihood approach, including derivations for the general case of GLMM, the random intercept logistic regression, and other multilevel logistic regression models, can be found in Handayani et al. (2017).

3. Generalized linear mixed models in claims reserving

In the previous chapter, we provided a detailed overview of the structure and parameter estimation of generalized linear mixed models. In this chapter, we will delve into a comprehensive discussion on the application of generalized linear mixed models in the estimation of outstanding claims reserves.

In the estimation model for outstanding claim reserves, the frequency of claims constitutes a pivotal factor, yet it is often disregarded in favor of directly assuming and estimating claim amounts. The two-stage Generalized Linear Model (GLM) delineated below presents a preferred methodology to address this issue. This two-stage model facilitates a more comprehensive evaluation of various factors within the claims process, encompassing the interplay between claim frequency and claim amounts.

In this approach, we initially employ a GLM or GLMM model to model the claim counts, followed by modeling the average claim amounts using another GLM or GLMM model. In this way, we yield the benefit of better considering the influence of claim frequency on claim amounts, thereby augmenting the precision of reserve estimation.

3.1 Generalized Linear Model

Assuming that the total number of accident years and the total number of development years are related as $I = J$. The increments of number of claims for the jth development year of the ith accident year are denoted as $\{n_{ij}, i, j = 1, ..., I\}$. Here, the data differs from the aggregate data $\{N_{ij}, i, j = 1, ..., I\}$ where $N_{ij} = N_{i,j-1} + n_{i,j}$. Similarly, the settled claim amounts for the jth development year of the ith accident year are denoted as $\{c_{ij}, i, j = 1, ..., I\}$, while the cumulative claim amounts are denoted as $\{C_{ij}, i, j = 1, ..., I\}$ where $C_{ij} = C_{i,j-1} + c_{i,j}$. Let $\{y_{ij}, i, j = 1, ..., I\}$ be the increments of average claim amounts for the jth development year of the ith accident year and $\{Y_{ij}, i, j = 1, ..., I\}$ be the cumulative average claim amounts.

The basic idea of the model is as follows:

1. Based on the triangle of incremental claim counts $\{n_{ij}, i = 1, ..., I, j \leq I + i - 1\}$, establish a generalized linear model to estimate incremental claim counts $\{\hat{n_{ij}}, i = 1, ..., I, j > I + i - 1\}$ within the lower right triangle, and fill in the triangle.

2. Based on the triangle of incremental claim amounts $\{c_{ij}, i = 1, ..., I, j \leq I + i - 1\}$ and the triangle of incremental claim counts $\{n_{ij}, i = 1, ..., I, j \leq I + i - 1\}$, calculate the incremental average settled claim amounts $\{y_{ij}, i = 1, ..., I, j \leq I + i - 1\}$ according to $y_{ij} = \frac{c_{ij}}{n_{ij}}$. Then establish a generalized linear model to estimate the incremental average settled claim amounts $\{\hat{y_{ij}}, i = 1, ..., I, j > I + i - 1\}$ within the lower right triangle, and fill in the triangle.

3. According to $\hat{c_{ij}} = \hat{n_{ij}} \cdot \hat{y_{ij}}$, obtain the total incremental claim amounts for each cell in the incremental claim amounts triangle $\{\hat{c_{ij}}, i = 1, ..., I, j > I + i - 1\}$. The cumulative claim amounts triangle $\{\hat{C_{ij}}, i = 1, ..., I, j > I + i - 1\}$ could be obtained subsequently.

4. From $\hat{C_{i,j}} = \hat{C_{i,j-1}} + \hat{c_{i,j}}$ and (1.4), the cumulative outstanding claim reserves can be obtained.

3.1.1 Modelling of claims counts

In the claims frequency model, we assume the incremental claims counts $\{n_{ij}, i, j = 1, ..., I\}$ is independent and follows a distribution from the exponential family. Claim frequency is a type of discrete count data, commonly modeled using distributions such as the Poisson distribution and negative binomial distribution.

Poisson model

The Poisson probability distribution function has the form

$$f(y) = \frac{\lambda^y}{y!} e^{-\mu}, y = 0, 1, 2, ...$$

with parameter $\lambda > 0$. The positive real number λ_{ij} is equal to the expected value of n_{ij} and also to its variance.

$$Var(n_{ij}) = E(n_{ij}) = \lambda_{ij} \tag{3.1}$$

The linear predictor η_{ij} is related to the mean λ_{ij} by a logarithmic link

$$log(\lambda_{ij}) = \eta_{ij} = c + a_i + b_j \tag{3.2}$$

The parameter a_i represents the effects of the accident year i, b_j represents the effects of development year j on the expected value of the incremental claim counts $n_{i,j}$, and c is a constant term for all i and j, where $a_1 = b_1 = 0$.

In order to measure the deviation between observed and expected values, Pearson chi-square statistic for Poisson distribution is shown below:

$$\begin{aligned} \hat{\varphi} &= \frac{1}{n-p} \sum_{i=1}^{I} \sum_{j=1}^{I-i} \frac{(n_{ij} - \hat{\mu_{ij}})^2}{V(\hat{\mu_{ij}})} \\ &= \frac{1}{n-p} \sum_{i=1}^{I} \sum_{j=1}^{I-i} \frac{(n_{ij} - \hat{\lambda_{ij}})^2}{\hat{\lambda_{ij}}} \end{aligned}$$

where n is the number of observations with $n = I(I+1)/2$ and p is the number of parameters with $p = 2I + 1$. n_{ij} is the observed increment count for the accident year i and development year j and $\hat{\lambda_{ij}}$ is the predicted count for the n_{ij} observation.

In order to measure the goodness of fit for Poisson distribution, the deviance is computed by the formula:

$$D = \frac{2}{\hat{\varphi}} \sum_{i=1}^{I} \sum_{j=1}^{I-i} \left[n_{ij} \log \left(\frac{n_{ij}}{\hat{\lambda}_{ij}} \right) - (n_{ij} - \hat{\lambda}_{ij}) \right]$$

The deviance statistic follows a χ^2 distribution with n-p degrees of freedom. If the deviance statistic is large compared to this chi-squared distribution, it suggests that the model does not fit the data well.

Negative binomial model

The negative binomial distribution is closely related to the Poisson distribution, serving as a generalization that allows for overdispersion. While the Poisson distribution assumes constant variance equal to its mean, the negative binomial distribution relaxes this assumption by introducing an additional parameter that allows for variance greater than the mean.

The probability mass function of the negative binomial distribution is given by

$$f(y) = \binom{y+r-1}{y} \left(\frac{\mu}{\mu+r}\right)^y \left(\frac{r}{\mu+r}\right)^r, y = 0, 1, 2, ...$$

with parameters $r, \mu > 0$. μ_{ij} is the expected value of n_{ij} and the variance can be expressed in terms of the mean:

$$E(n_{ij}) = \mu_{ij}$$

$$Var(n_{ij}) = \frac{\mu_{ij} + r}{r} \mu_{ij} \tag{3.3}$$

In negative binomial distribution for the generalized linear model, the logarithmic link function is also typically used.

$$log(\mu_{ij}) = \eta_{ij} = c + a_i + b_j \tag{3.4}$$

where a_i represents the effects of the accident year i, b_j represents the effects of development year j on the expected value of the incremental claim counts $n_{i,j}$ with $a_1 = b_1 = 0$. Pearson chi-square statistic for negative binomial distribution is shown in the following

$$\begin{aligned} \hat{\varphi} &= \frac{1}{n-p} \sum_{i=1}^{I} \sum_{j=1}^{I-i} \frac{(n_{ij} - \hat{\mu_{ij}})^2}{V(\hat{\mu_{ij}})} \\ &= \frac{1}{n-p} \sum_{i=1}^{I} \sum_{j=1}^{I-i} \frac{(n_{ij} - \hat{\mu_{ij}})^2}{\frac{\hat{\mu_{ij}}+r}{r} \hat{\mu_{ij}}} \end{aligned}$$

where n is the number of observations with n = I(I + 1)/2 and p is the number of parameters with p = 2I + 1. μ_{ij} is the fitted expected value of n_{ij}. The deviance measures how well the negative binomial regression model fits the observed data.

$$D = \frac{2}{\hat{\varphi}} \sum_{i=1}^{I} \sum_{j=1}^{I-i} \left[n_{ij} \log\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right) - (n_{ij} + r) log(\frac{n_{ij} + r}{\hat{\mu_{ij}} + r}) \right]$$

In this section, we introduced two different models to model the claim counts. We can use these two models in R to predict the final claim counts and conduct model testing and comparison. And in the following section, we continued to model the average claim amounts.

3.1.2 Modelling of claims amounts

As we metioned at the begining of this chapter, we use the symbol y_{ij} to represent the incremental average claim amounts, and Y_{ij} to denote the cumulative average claim amounts. In this section, we use the average claim amounts data from the development triangle of incremental average claim amounts. When modeling average claim amounts, continuous distributions should be considered. The most commonly used models for average claim amounts are the gamma distribution and the inverse Gaussian distribution. Next, we will describe these two distributions.

Gamma models

The probability density function of gamma distribution is defined as

$$f(y) = \frac{1}{\Gamma(k)\theta^k} y^{k-1} e^{-\frac{y}{\theta}}, y > 0$$

where k and θ are parameters, with $k > 0$ being the shape parameter and $\theta > 0$ being the scale parameter. The random variable y with the probability density function $f(y)$ is used to represent the average claim amount.

The mean and variance functions of the gamma distribution are represented as

$$E(y_{ij}) = \mu_{ij} = k\theta$$
$$Var(y_{ij}) = \varphi\mu_{ij}^2 = k\theta^2 \tag{3.5}$$

where φ is defined as the dispersion parameter with $\varphi = \frac{1}{k}$. In the gamma distribution, the inverse reciprocal link function, which is the canonical link function, is used. The inverse link function is defined as:

$$\eta_{ij} = \frac{1}{\mu_{ij}} = c + a_i + b_j \tag{3.6}$$

for all i and j with $a_1 = b_1 = 0$. In addition, the logarithmic link function, although not the canonical link for the gamma distribution, is commonly used because it models the multiplicative relationship between predictors and the mean, which simplifies the interpretation of the model in terms of relative changes in the mean. The logarithmic link for the gamma distribution is defined as:

$$log(\mu_{ij}) = \eta_{ij} = c + a_i + b_j \tag{3.7}$$

for all i and j with $a_1 = b_1 = 0$. The parameters of the given models are estimated using the maximum likelihood method, as described previously.

After the model is fitted, to examine the goodness of fit, the deviance for the fitted gamma distribution model is computed as follows:

$$D = \frac{2}{\hat{\varphi}} \sum_{i=1}^{I} \sum_{j=1}^{I-i} \left[\frac{y_{ij} - \hat{\mu_{ij}}}{\hat{\mu}_{ij}} - log(\frac{y_{ij}}{\hat{\mu_{ij}}}) \right]$$

where y_{ij} is the observed increment average claim amount for the accident year i and development year j and $\hat{\mu_{ij}}$ is the predicted average claim amount for the y_{ij} observation.

Inverse Gaussian models

The inverse gaussian distribution is appropriate for modeling right-skewed positive data and is applicable in situations where all observations are positive and unbounded. The probability density function of the inverse gaussian distribution is defined as

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}}\exp\left(-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right), y > 0$$

with parameters $\lambda > 0$ and $\mu > 0$. The mean and variance functions of the inverse gaussian distribution are represented as

$$E(y_{ij}) = \mu_{ij}$$

$$Var(y_{ij}) = \varphi\mu_{ij}^3 \tag{3.8}$$

where the dispersion parameter here is defined as $\varphi = \frac{1}{\lambda}$.

The canonical link function for the inverse Gaussian distribution is the reciprocal of the square of its mean parameter.

$$\eta_{ij} = \frac{1}{\mu_{ij}^2} = c + a_i + b_j \tag{3.9}$$

for all i and j with $a_1 = b_1 = 0$. Same as before, in practice, we also consider the logarithmic link for the inverse Gaussian distributed response variables due to its convenient properties.

$$log(\mu_{ij}) = \eta_{ij} = c + a_i + b_j \tag{3.10}$$

for all i and j with $a_1 = b_1 = 0$. The deviance for the fitted model of inverse gaussian distribution is computed as follows:

$$D = \frac{1}{\hat{\varphi}}\sum_{i=1}^{I}\sum_{j=1}^{I-i}\left[\frac{(y_{ij}-\hat{\mu_{ij}})^2}{\hat{\mu}_{ij}^2 y_{ij}})\right]$$

3.2 Generalized Linear Mixed Model

Generalized Linear Mixed Models (GLMMs) extend the capabilities of Generalized Linear Models (GLMs) by incorporating random effects. The most significant improvement of GLMMs over GLMs is their ability to account for correlated data and hierarchical structures. As we methioned in the previous chapter, glmm has the form:

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \boldsymbol{X\beta} + \boldsymbol{Zv}$$

In this formulation, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are typically constructed based on the structure of the data and the specified model. The random effects $\boldsymbol{v}$ are often assumed to be normally distributed.The residual errors are assumed to follow a distribution from the exponential family, such as normal, Poisson, or binomial.

To build a more specific GLMM model for predicting outstanding reserves, it is necessary to align the elements of the GLMM with the corresponding claims notation. It is evident that the incremental claim counts $\{n_{ij}, i, j = 1, ..., I\}$ and the incremental average claim amounts $\{y_{ij}, i, j = 1, ..., I\}$ with accident year i and development j will serve as the dependent variables Y_{ij} in the GLMM framework, respectively. The claims counts or average claim amounts originating from the same accident year i are expected to be clustered together.

In the triangular dataset, observations within the same accident year may be correlated. To model these correlations, an appropriate random effects structure must be chosen. Properly identifying and incorporating random effects helps the model account for these correlations by capturing variations across different accident years or other grouping factors.

In loss reserving calculations, both accident year and development year can be considered as factors susceptible to random effects. Accident year may be influenced by external factors such as unpredictable natural disasters or fluctuations in the macroeconomic environment. Development year may also be influenced by various factors such as the efficiency of claims processing, adjustments in compensation amounts, etc. Therefore, in our GLMM model, we consider two random effect variables: the first from accident year i and the second from development year j. The vector form of random effects can be expressed as

$$\boldsymbol{v} = (d, u_1, ..., u_I, v_1, ..., v_J)^T$$

where u_i represents the random effect of the accident year i, v_j represents the random effect of the development year j, and d is the constant term. The corresponding design vector can be expressed as

$$\boldsymbol{z_{ij}} = (1, \delta_{1i}, ..., \delta_{Ii}, \delta_{1j}, ..., \delta_{Jj})^T$$

where the Kronecker's delta δ_{ij} represents the dummy variables. It equals 1 when the indices i and j in δ_{ij} are equal to the indices i and j in $\boldsymbol{z_{ij}}$, and it equals 0 when the indices are different.

However, in practical applications, we should also consider the issue of overparameterization. Including random effects for each level of a grouping variable in GLMMs can lead to overparametrization if there are too many levels of the grouping variable or too few observations per level. When we are considering the random effect for both row specific and column specific effects, the number of random effect parameters in the fitted model equals to $2I - 1$. Therefore, to address this issue, a simpler mixed model called *the random intercept model* is often utilized in loss reserving analysis.

In random intercept model, we omit the column specific effect v_j and consider only the row specific effect u_i. The random effects of the random intercept model can be expressed as

$$\boldsymbol{v} = (d, u_1, ..., u_I)^T$$

where u_i represents the random effect of the accident year i and d is a constant term. The corresponding design vector can be expressed as

$$\boldsymbol{z_i} = (1, \delta_1, ..., \delta_I)^T$$

where the Kronecker's delta δ_i represents the dummy variables. It equals 1 when the indices i in δ_i and in $\boldsymbol{z_i}$ are equal, and it equals 0 when the indices are different. Then the number of random effect parameters in the fitted random intercept model reduced to I.

The mean structure for claims associated with accident year i and development year j is represented by the following form:

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^{T}\boldsymbol{\beta} + \boldsymbol{z}_{i}^{T}\boldsymbol{\upsilon} \tag{3.11}$$

where $\boldsymbol{\beta}$ is the fixed coefficient vector for the predictor variable and $\boldsymbol{\upsilon}$ is the random intercept term for the ith accident year. These random effects are typically assumed to follow a normal distribution with mean zero and a variance parameter that needs to be estimated from the data. This model allows for the intercept term to vary across different accident year i, which is captured by the random effect.

We assume claim development patterns follow historical trends. Changes are primarily due to inflation, making column-specific effects unnecessary in the model. Thus, we can simplify the model by excluding these effects and focusing solely on capturing inflation-induced trends. In addition, we have also made adjustments to the fixed effects of the model, and then (3.11) could be expressed as

$$g(\mu_{ij}) = c + b_j + u_i \tag{3.12}$$

where c is a constant term reflects the baseline or average level of claims, the fixed effect b_j represents the influence of the development year j with $b_1 = 0$, and the random effect u_i represents the deviation or fluctuation of the claim for accident year i from the expected level c. Assuming the random effect u_i follows a normal distribution.

In generalized linear mixed models, we adopt the same modeling approach as GLMs to predict both the frequency of losses $\{n_{ij}, i, j = 1, ..., I\}$ and the average loss amounts $\{y_{ij}, i, j = 1, ..., I\}$.

3.2.1 Modelling of claims counts

Poisson model

The probability distribution function of the Poisson distribution is the same as the pdf we mentioned in the previous section in GLMs. The difference is that GLMMs introduces a random effect $g(\boldsymbol{\mu}) = \boldsymbol{X\beta} + \boldsymbol{Z\upsilon}$. The logarithmic link function of the Poisson GLMM model is shown as follows.

$$log(\lambda_{ij}) = \eta_{ij} = c + b_j + u_i \tag{3.13}$$

where the parameter b_j represents the fixed effect of development year j on the expected value of the incremental claim counts $n_{i,j}$, u_i represents the random effect of accident year i, and c is a constant term for all accident year i and development year j, with $b_1 = 0$.

Negative binomial model

The probability distribution function of the negative binomial distribution is the same as the pdf we mentioned in the previous section in GLMs. In negative binomial distribution for the generalized linear mixed model, the logarithmic link function is also typically used.

$$log(\mu_{ij}) = \eta_{ij} = c + b_j + u_i \tag{3.14}$$

for all i and j with $b_1 = 0$. b_j represents the fixed effects of development year j and u_i represents the random effects of different accident year i.

3.2.2 Modelling of claims amounts

Gamma models

The probability distribution function of the gamma distribution is the same as the pdf we mentioned in the previous section in GLMs. Similarly, in gamma distribution for the generalized linear mixed model, the inverse reciprocal link function, which is the canonical link function, is used. The inverse link function is defined as:

$$g(\mu_{ij}) = \eta_{ij} = \frac{1}{\mu_{ij}} = c + b_j + u_i \tag{3.15}$$

for all i and j with $b_1 = 0$. The difference between the link functions of GLMs and GLMMs is that in GLMMs there exists random effect for accident year i, represented by the parameter u_i.

Inverse Gaussian models

The probability distribution function of the inverse Gaussian distribution is the same as the pdf we mentioned in the previous section in GLMs. The same link function for the inverse Gaussian distribution, which is the reciprocal of the square of its mean parameter, has been used.

$$g(\mu_{ij}) = \eta_{ij} = \frac{1}{\mu_{ij}^2} = c + b_j + u_i \tag{3.16}$$

for all i and j with $b_1 = 0$. Similarly, the difference between the link functions of GLMs and GLMMs is that the parameter u_i represents the random effect of accident year i.

3.2.3 Model Fitting and Evaluation

In addition to the statistical methods mentioned in section 2.1.4, information criteria such as AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are often used in practice to compare a finite set of models. These criteria help evaluate the relative quality of statistical models for a given claims dataset.

Akaike's Information Criterion (AIC) for a statistical model is given by:

$$AIC = 2p - 2\ln(L)$$

where p represents the number of parameters in the model, and L denotes the maximized value of the likelihood function. Another criterion, known as the *Bayesian Information Criterion* (BIC), is computed as follows:

$$BIC = p\ln(n) - 2\ln(L)$$

where P is the number of parameters in the model, L is the maximized value of the likelihood function and n is the number of observations in the dataset.

These criterias allow for the comparison of GLMMs with GLMs or among different GLMMs, providing a way to select the model that best balances fit and complexity.

4. Real data analyze

In this chapter, we will calculated the INBR reserves based on the real data. First we use the Chain Ladder method as traditional deterministic model. Then we use double generalized linear models to model the claim counts and average claim amounts separately. The computational part of the analysis was constructed in a software R, and a complete source code is attached in the Appendix A.

4.1 Data set

The dataset utilized in this chapter in the analysis is sourced from Chen (2009). It illustrates seven years (1999-2005) of observations of paid losses with seven development lags. The data is presented in two tables. They all have been organized into the format of incremental triangles. Table 4.1 shows the incremental claim counts and Table 4.2 displays the incremental total claim amounts.

Table 4.1: Development triangle of incremental claim counts

Accident year i	Development year i						
	1	2	3	4	5	6	7
1	23355	8585	1348	572	231	156	47
2	22662	7632	1294	541	194	110	
3	18951	6246	1017	368	195		
4	16631	6263	912	423			
5	17381	7200	1184				
6	12666	4003					
7	10592						

Table 4.2: Development triangle of incremental claim amounts

Accident year i	Development year i						
	1	2	3	4	5	6	7
1	22607640	2455310	508196	150436	45276	19968	2961
2	22050126	3648096	527952	203957	34726	10010	
3	20163864	2885652	410868	115920	40365		
4	19291960	2956136	442320	192888			
5	20561723	4046400	821696				
6	16997772	2690016					
7	11344032						

The incremental triangle of total claim amounts is calculated as the product of the incremental triangle of the number of claims and the incremental triangle of the average claim amounts. Dividing Table 4.2 by Table 4.1, we obtain the incremental triangle of average claim amounts, as shown in Table below.

Table 4.3: Incremental average claim amounts triangle

Accident year i	Development year i						
	1	2	3	4	5	6	7
1	968	286	377	263	196	128	63
2	973	478	408	377	179	91	
3	1064	462	404	315	207		
4	1160	472	485	456			
5	1183	562	694				
6	1342	672					
7	1071						

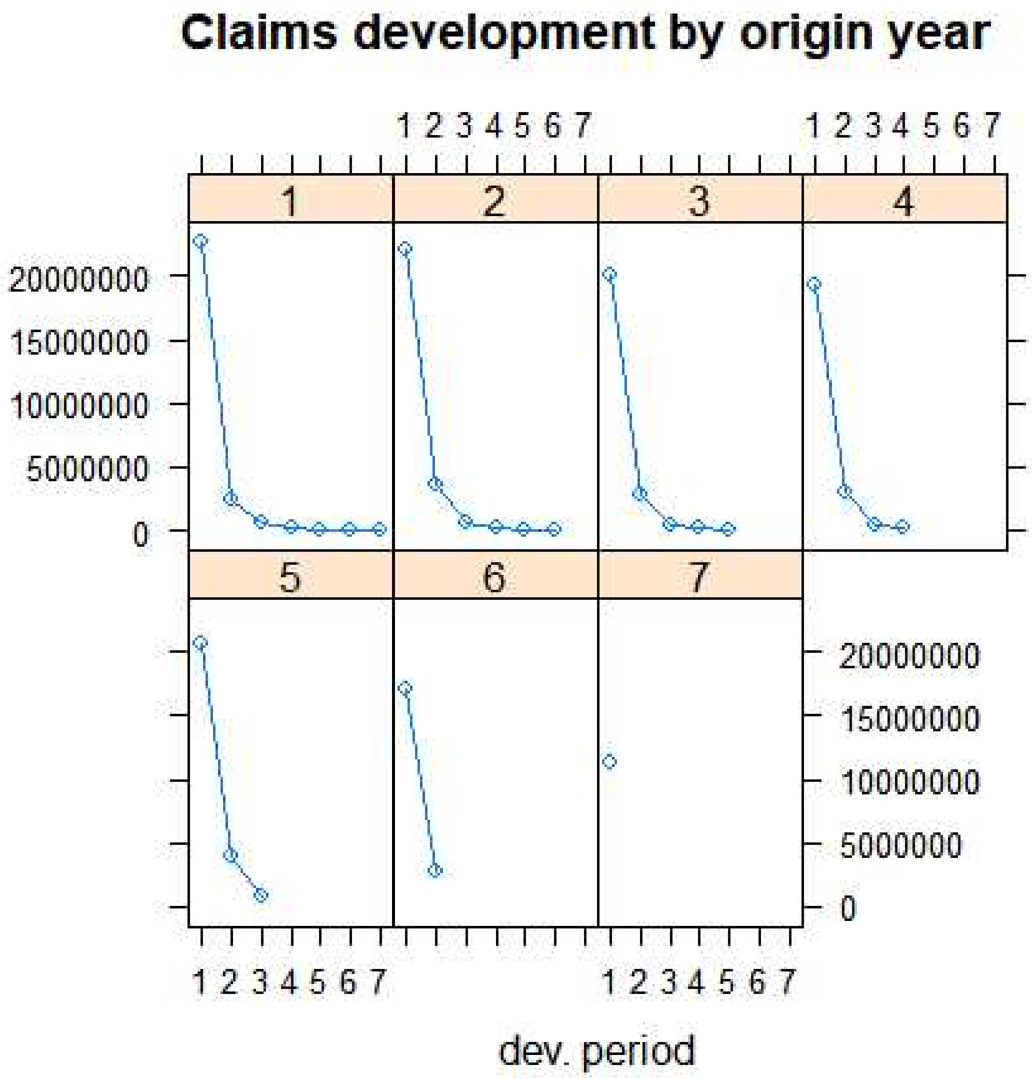

Figure 4.1: Development of incremental claims amounts by year of origin

In order to better illustrate the data, figure 4.1 displays the development of incremental claims for each year individually, and the cumulative claims development

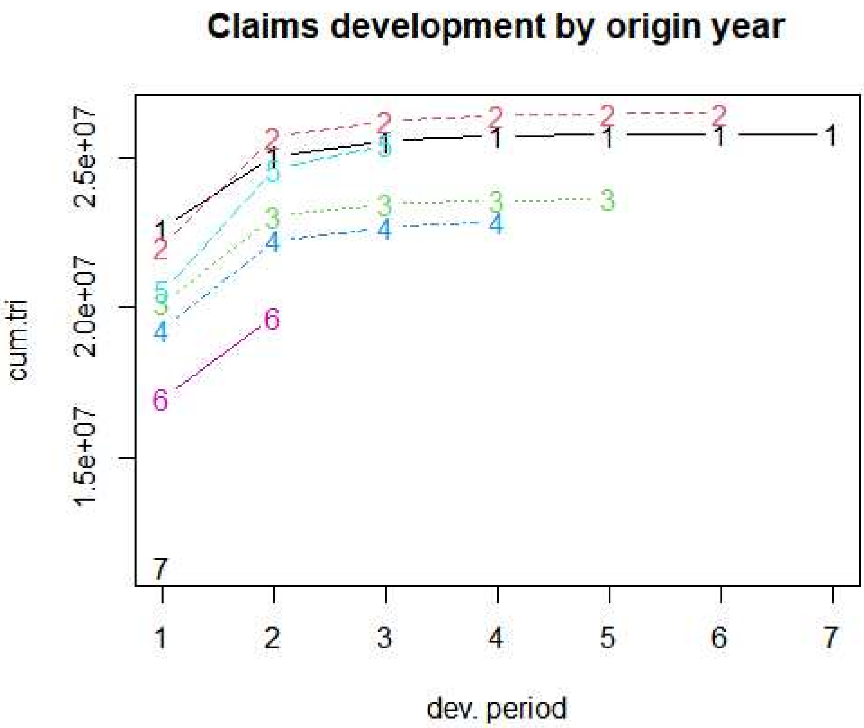

Figure 4.2: Development of cumulative claims amounts

is visualised in figure 4.2. From these figures, we can observe that as the development years increase, the incremental claims exhibit a trend of gradually decreasing to zero, indicating that the cumulative claims increase at a slower rate. This trend is attributed to the fact that the majority of insurance incidents can be resolved relatively quickly, with only a few cases requiring more time to settle.
In the next section, we will establish different models for the upper triangles of both claim counts and average claim amounts separately based on real data, estimating the values to fill the lower triangles. And then we will calculate the IBNR values accordingly.

4.2 Chain Ladder method

The chain ladder method is based on cumulative claim data, so the first step is to convert all the incremental claim data provided in the previous section into cumulative claim data, as shown in Table 4.4.

The development factors are calculated as the average cumulative claim values $\{C_{i,j}, i = 1, ..., I, j \leq I + i - 1\}$ for each accident year based on equation (1.10). The result for each year are shown in Table 4.5.

By multiplying the reported claim amounts for each accident year by the corresponding claim development factors, we can obtain the future claims for each accident year. This process fills in the total claim amount triangle. The claims for the 7th development year, which represent the ultimate losses (UL),

Table 4.4: Development triangle of cumulative claim amounts

Accident year i	Development year i						
	1	2	3	4	5	6	7
1	22607640	25062950	25571146	25721582	25766858	25786826	25789787
2	22050126	25698222	26226174	26430131	26464857	26474867	
3	20163864	23049516	23460384	23576304	23616669		
4	19291960	22248096	22690416	22883304			
5	20561723	24608123	25429819				
6	16997772	19687788					
7	11344032						

Table 4.5: Development factor for CL method

Accident year i	Development factor
1999	1.153539
2000	1.022467
2001	1.006771
2002	1.001589
2003	1.000574
2004	1.000115
2005	1.000000

can be obtained by subtracting the reported losses (RL) for each accident year from the reported losses for the calendar year 2005. This process yields the IBNR reserves for each year, as shown in the following table.

Table 4.6: Observed values of claims reserves based on CL method

Accident year i	UL	RL	IBNR
1999	25789787	25789787	0
2000	26474867	26477907	3040
2001	23616669	23632937	16268
2002	22883304	22935464	52160
2003	25429819	25660360	230541
2004	19687788	20312610	624822
2005	11344032	13501085	2157053
Total IBNR			3083884

Therefore, according to Table4.6, The total IBNR reserve is obtained by summing the IBNR reserves for each year, resulting in 3,083,884.

4.3 Generalized Linear Models

The evaluation of IBNR using a double generalized linear model is a stochastic model that not only provides point estimates of IBNR reserves with higher accuracy than deterministic methods but also yields confidence intervals for IBNR reserves. In this section, we employ generalized linear models to predict both the claim counts and the average claim amounts.

4.3.1 GLM for claim counts

First, we focus on forecasting future claim counts using Poisson and negative binomial distributions for fitting. We will first illustrate claim counts model with Poisson distribution.

Poisson Models

The model results for predicting future claim counts using the Poisson distribution function in the generalized linear model are shown in the Table 4.7.

Table 4.7: Poisson Model Parameter Estimation Results

Parameters	Estimate	Std.Error	z value	P value
Intercept	10.06672	0.00567	1775.787	0.0000
a2	-1.02823	0.00583	-176.337	0.0000
a3	-2.84876	0.01355	-210.280	0.0000
a4	-3.75255	0.02316	-161.999	0.0000
a5	-4.64205	0.04033	-115.102	0.0000
a6	-5.14953	0.06146	-83.789	0.0000
a7	-6.21658	0.14598	-42.587	0.0000
b2	-0.05442	0.00775	-7.024	0.0000
b3	-0.24206	0.00816	-29.660	0.0000
b4	-0.33538	0.00840	-39.913	0.0000
b5	-0.25747	0.00826	-31.158	0.0000
b6	-0.65117	0.00948	-68.727	0.0000
b7	-0.79887	0.01125	-71.015	0.0000

Table 4.7 provides key information about the model: the maximum likelihood estimation results for all parameters. The last column presents the associated probabilities of significance for each parameter, and observing that they are all less than 0.05 indicates that the estimates for these parameters are highly significant. Next, we are going to perform model diagnostics.

Model Diagnostics

Table 4.8: Model diagnostics for Poisson model

dependent variables	df	LR Test		Wald Test	
		Chi-Square	Pr>ChiSq	Chi-Square	Pr>ChiSq
Accident Years	6	-9878.40	0.00	9257.60	0.00
Development years	6	-228831.00	0.00	107418.10	0.00

Based on the table 4.8, both the LR test and the Wald test passing indicate that the parameter estimates in the model are significant, suggesting that both accident year and development year in the model have a statistically significant impact on the incremental claim counts. And then we show scaled deviance and Pearson residuals for the Poisson model in figures 4.3 and 4.4 respectively to evaluate their adequacy of fit. From the figures, the residuals have a mean of zero but exhibit a decreasing trend in variance, it may indicate that the error structure of the model varies across different predicted values. Specifically, the decreasing trend in variance may suggest that the predictive accuracy of the Poisson model is higher within certain ranges of predicted values and lower within others. This could be due to the model being overly conservative or aggressive in predicting certain data points, resulting in the variance changing. It is also possible that the trend in variance may be due to the small size of the dataset.

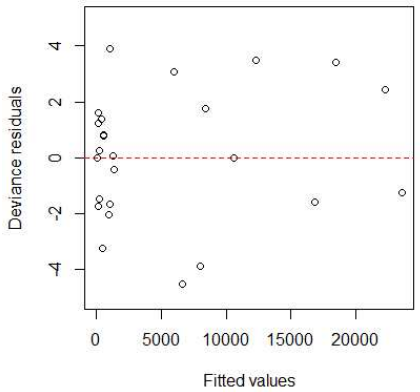

Figure 4.3: Deviance residuals

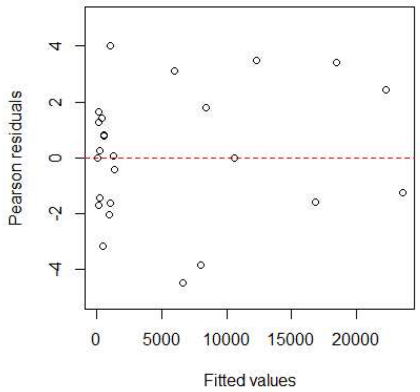

Figure 4.4: Pearson residuals

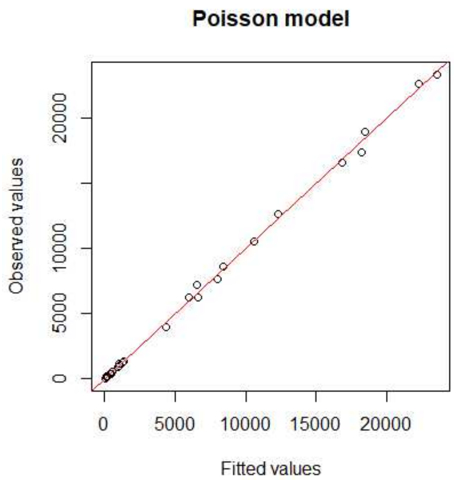

Figure 4.5: Observed and fitted values of Poisson model

In figure 4.14, another graphical tool is employed to evaluate the model fit. Points indicating observed values in comparison to fitted values are anticipated to exhibit a linear pattern along the main diagonal in the plot.

Table 4.9: Negative Binomial Model Parameter Estimation Results

Parameters	Estimate	Std. Error	z value	P value
Intercept	10.08424	0.02864	352.145	0.0000
a2	-1.03511	0.02622	-39.478	0.0000
a3	-2.85703	0.03060	-93.358	0.0000
a4	-3.74986	0.03779	-99.241	0.0000
a5	-4.63273	0.05225	-88.673	0.0000
a6	-5.15231	0.07279	-70.779	0.0000
a7	-6.23409	0.15509	-40.197	0.0000
b2	-0.08741	0.03268	-2.675	0.0075
b3	-0.28952	0.03381	-8.563	0.0000
b4	-0.34484	0.03493	-9.872	0.0000
b5	-0.21773	0.03678	-5.919	0.0000
b6	-0.69216	0.04137	-16.733	0.0000
b7	-0.81638	0.05357	-15.239	0.0000

Negative Binomial Models

The last column of the table 4.9 displays the associated probabilities of significance for each parameter. Notably, all probabilities are less than 0.05, indicating a high level of significance for the parameter estimates. Subsequently, model diagnostics will be conducted to evaluate the fit of the negative binomial model and identify any potential shortcomings or areas for refinement.

Model Diagnostics

Table 4.10: Model diagnostics for negative binomial model

dependent variables	df	LR Test		Wald Test	
		Chi-Square	Pr>ChiSq	Chi-Square	Pr>ChiSq
Accident Years	6	69.83	0.00	461.60	0.00
Development Years	6	191.93	0.00	21126.90	0.00

The results of the Wald test and likelihood ratio test are shown above in table 4.10. They primarily assess the significance of the statistical model. It can be observed that the p-values are very small, indicating that the model is significant.

Then we will illustrate deviance residuals and Pearson residuals. From figure 4.6 and 4.7, the residuals have a mean close to zero and a relatively constant variance. This indicates that the counts predictions of negative binomial model are generally accurate and consistent across the dataset. These conditions suggest that the model's assumptions regarding the mean and variance of the residuals are being satisfied, which is crucial for the reliability of its predictions.

The plot of observed values against predicted values forms a straight line, it indicates a linear relationship between the predicted and observed values. This suggests that the negative binomial model fits well, and the predicted values can effectively explain the variability in the observed values.

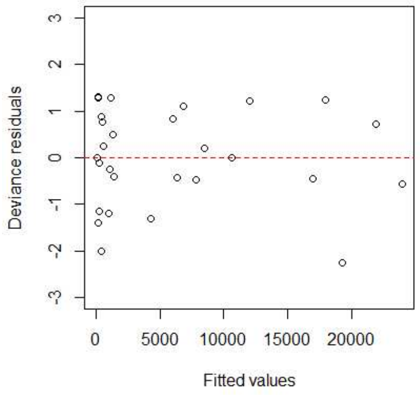

Figure 4.6: Deviance residuals

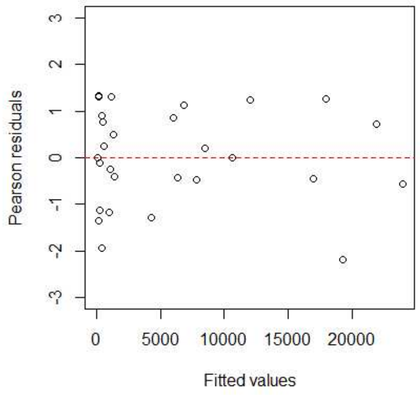

Figure 4.7: Pearson residuals

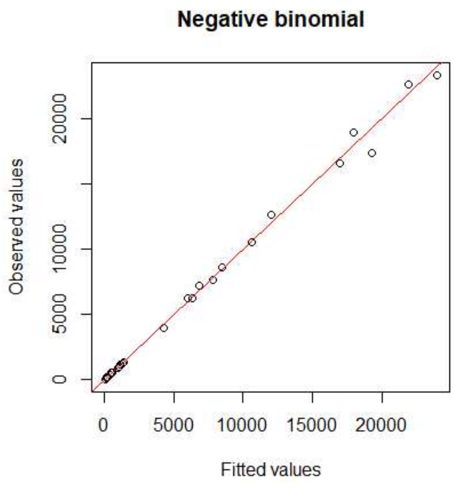

Figure 4.8: Observed and fitted values of negative binomial model

Claim Counts Model Evaluation

As we cannot know in advance whether the number of claims follows a Poisson distribution, a negative binomial distribution, or another distribution form, to choose the best distribution function for fitting the final claim count, we compared the model selection criteria statistics of the test models for these two distributions. The relevant data has been organized into the Table 4.11.

Table 4.11: Pearson χ^2 statistic, deviance and AIC for claim counts under different models

Model	Poisson	Negative Binomial
Deviance	271.2467	28.8649
Pearson Chi-square	271.8178	28.4318
AIC	558.8787	368.9500

It is clear that smaller values of model selection statistics in the table indicate better fit, as they measure the degree to which the model fits the observed data. Both the Poisson and negative binomial distributions provide good fits, but based on various statistics and indicators, it is evident that the negative binomial distribution performs better than the Poisson distribution. Therefore, we can conclude that among these three distribution functions, the negative binomial distribution is the most suitable model for predicting future claim counts.

Table 4.12: Estimation values and confidence intervals for claim count parameters under negative binomial model.

parameters	Estimate	95% CI LL	95% CI UL
Intercept	10.08424	10.0281233	10.1409147
a2	-0.08741	-0.1514556	-0.0233745
a3	-0.28952	-0.3558459	-0.2231825
a4	-0.34484	-0.4133994	-0.2762186
a5	-0.21773	-0.2899933	-0.1452813
a6	-0.69216	-0.7730621	-0.6107506
a7	-0.81638	-0.9207803	-0.7104518
b2	-1.03511	-1.0865424	-0.9836759
b3	-2.85703	-2.9171536	-2.7969114
b4	-3.74986	-3.8239609	-3.6759199
b5	-4.63273	-4.7356809	-4.5308662
b6	-5.15231	-5.2966049	-5.0112432
b7	-6.23409	-6.5507923	-5.9413962

Table 4.13: Incremental claim counts fitted with negative binomial distribution

Accident year i	Development year i						
	1	2	3	4	5	6	7
1	23355	8585	1348	572	231	156	47
2	22662	7632	1294	541	194	110	43
3	18951	6246	1017	368	195	104	35
4	16631	6263	912	423	165	98	33
5	17381	7200	1184	453	188	112	38
6	12666	4003	689	282	117	69	24
7	10592	3762	608	249	103	61	21

The estimated values and confidence intervals for the model parameters of the negative binomial model are shown in the Table 4.12. Using the parameters in the table 4.12, we can predict the ultimate claim counts for each accident year. For example, the ultimate claim counts for accident year 2 is calculated as exp(Intercept + a2 + b7) = exp(10.08424 - 0.08741 - 6.23409) = 43.07.

The estimated values can be used to fill the claim counts triangle, as shown in the Table 4.13. Table 4.13 represents the incremental claim counts fitted with a negative binomial distribution. Adding up the claim counts for each accident year in every row yields the ultimate claim counts for each accident year.

4.3.2 GLM for average claim amounts

For the GLM model for average claim amounts, we followed the same approach as the previous section for the GLMs for claim counts. We fitted the average claim amounts using gamma distribution and inverse Gaussian distribution.

Gamma Models

The model results for predicting average claim amounts using the gamma distribution function in the generalized linear model are shown in the table below.

Table 4.14: Gamma Model Parameter Estimation Results

Parameter	Estimate	Std.Error	z value	P value
Intercept	0.0012947	0.0001928	6.716	0.0000
a2	-0.0002281	0.0002411	-0.946	0.3591
a3	-0.0002661	0.0002391	-1.113	0.2832
a4	-0.0004481	0.0002275	-1.969	0.0677
a5	-0.0005622	0.0002232	-2.519	0.0236
a6	-0.0006104	0.0002257	-2.704	0.0163
a7	-0.0003610	0.0002623	-1.376	0.1890
b2	0.0011251	0.0001730	6.504	0.0000
b3	0.0011351	0.0001961	5.789	0.0000
b4	0.0017846	0.0002852	6.258	0.0000
b5	0.0040274	0.0005774	6.975	0.0000
b6	0.0079532	0.0012375	6.427	0.0000
b7	0.0145784	0.0030305	4.811	0.0002

From table 4.14, some of the accident years are not statistically significant predictors based on the p-values. It indicates that certain levels of the accident year variable may not have a substantial impact on the dependent variable in the current model. This could be attributed to the small size of the dataset. In small datasets, statistical tests may lack the power to detect significant effects, resulting in larger p-values and potentially inconclusive results.

However, this does not directly imply that the accident year has no effect on the average claim amounts. We will conduct model diagnostics to determine whether this variable should be removed.

Table 4.15: Model diagnostics for gamma model

dependent variables	df	LR Test		Wald Test	
		Chi-Square	Pr>ChiSq	Chi-Square	Pr>ChiSq
Accident Years	6	-0.41	0.08	10.80	0.09
Development years	6	-9.68	0.00	185.40	0.00

Based on the Wald and LR tests in table 4.15, the accident year is not considered as a predictor variable for the average claim amounts, possibly because, in a specific dataset or analysis, the accident year does not demonstrate a significant impact on the average claim amounts. This may be due to a lack of direct causality between the accident year and the average claim amounts, or the accident year may no longer be a significant predictor after considering other variables. Therefore, based on model evaluation and testing results, it is decided to exclude the accident year and retain the development year variable to construct a new model. The model diagnostic of the new model are showing in the following.

Model Diagnostics

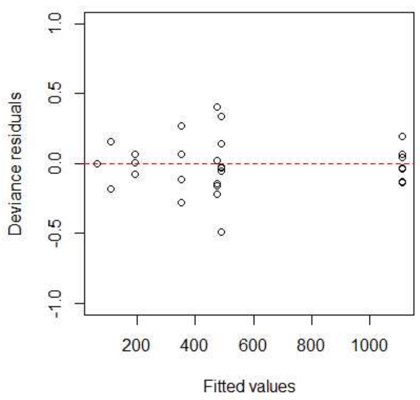

Figure 4.9: Deviance residuals

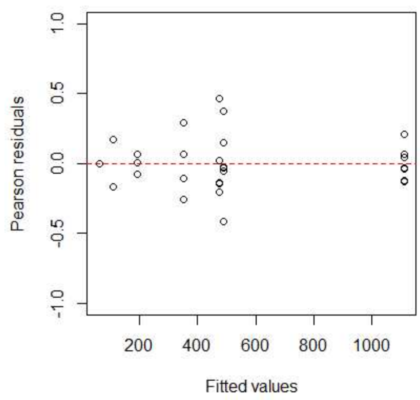

Figure 4.10: Pearson residuals

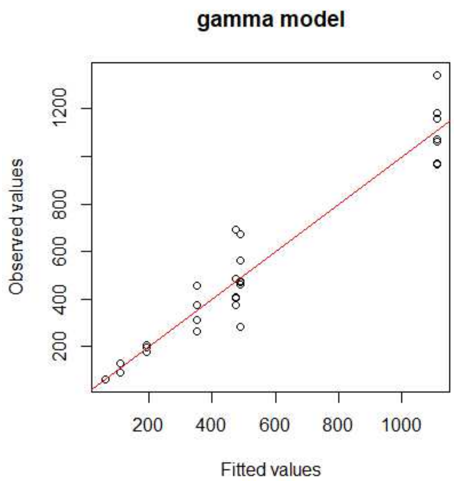

Figure 4.11: Observed and fitted values of gamma model

The linear relationship between predicted values and observed values is apparent but not entirely clear, it may indicate that the model partially explains the variation in the data, but there are still other factors or complexities not fully captured. This could be because we removed one independent variable, and now the average claim amounts depend solely on the development year.

Inverse Gaussian Model

The model results for predicting average claim amounts using the inverse gaussian distribution function in the generalized linear model are shown in the following table. From table 4.16, based on the results of the inverse Gaussian

Table 4.16: Model Parameter Estimation Results

parameters	Estimate	Std. Error	z value	P value
Intercept	0.0000020	0.0000010	1.960	0.0688
a2	-0.0000008	0.0000012	-0.629	0.5388
a3	-0.0000009	0.0000012	-0.727	0.4783
a4	-0.0000013	0.0000011	-1.162	0.2635
a5	-0.0000015	0.0000011	-1.368	0.1914
a6	-0.0000015	0.0000011	-1.418	0.1767
a7	-0.0000011	0.0000012	-0.916	0.3743
b2	0.0000032	0.0000009	3.545	0.0029
b3	0.0000034	0.0000011	3.227	0.0056
b4	0.0000068	0.0000018	3.784	0.0018
b5	0.0000251	0.0000050	5.028	0.0002
b6	0.0000818	0.0000144	5.688	0.0000
b7	0.0002500	0.0000466	5.367	0.0001

model, it can be observed that some of the accident years are not statistically significant predictors of the average claim amounts based on the p-values in the table. Similarly, we will conduct model diagnostics to determine whether the accident year should be removed.

Model Diagnostics

Table 4.17: Model diagnostics for inverse Gaussian model

dependent variables	df	LR Test		Wald Test	
		Chi-Square	Pr>ChiSq	Chi-Square	Pr>ChiSq
Accident Years	6	-0.000468	0.75	3.00	0.81
Development years	6	-0.031687	0.00	115.60	0.00

The p-values of both the LR and Wald tests for the accident year variable are greater than 0.05. The same conclusion is drawn from the average claim amount models considered. Hence, we proceed to exclude the accident year as an independent variable and conduct model diagnostics based on the revised model.

Figures 4.12 and 4.13 shown the deviance residuals and Pearson residuals plots which are generated based on the reduced submodel. Figure 4.14 comparing fitted values with observed values. The residual plot shows residuals close to zero, indicating good model fit with random differences between predicted and observed values. There is also no noticeable variance trend, suggesting stable prediction errors across observed value ranges, ensuring consistent predictive accuracy throughout the dataset.

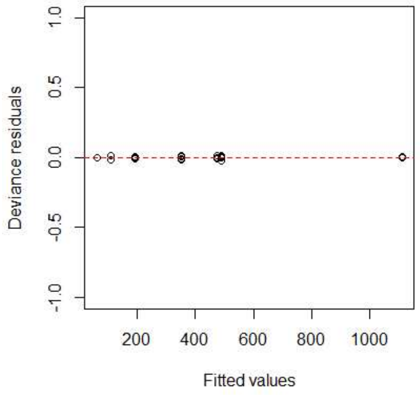

Figure 4.12: Deviance residuals

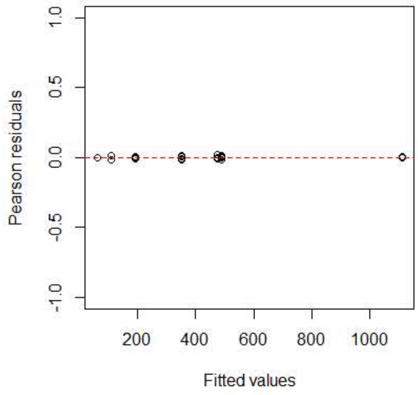

Figure 4.13: Pearson residuals

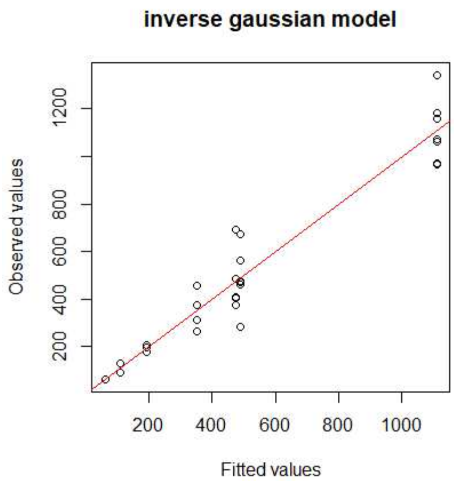

Figure 4.14: Observed and fitted values of inverse gaussian model

Claim Amounts Model Evaluation

Similarly, to select the best function distribution for fitting the ultimate average claim amounts, we compared the model evaluation and selection criteria for these two distributions to test the model fit, and organized the relevant data into the following table:

Table 4.18: Pearson χ^2 statistic, deviance and AIC for claim amounts under different models

Model	Gamma	Inverse Gaussian
Deviance	0.95874	0.00253
Pearson Chi-square	0.95846	0.00245
AIC	339.84050	342.8222

As seen from the Table, both the gamma distribution and the inverse Gaussian distribution show good fit, with the inverse Gaussian distribution slightly outperforming the gamma distribution. The estimated values for the model parameters and confidence intervals for the average claim amounts based on inverse gaussian model are provided below.

Table 4.19: Estimation values and confidence intervals for average claim amounts under inverse Gaussian model

parameters	Estimate	95% CI LL	95% CI UL
Intercept	0.0000008	0.0000004	0.0000013
b2	0.0000034	0.0000018	0.0000052
b3	0.0000036	0.0000019	0.0000057
b4	0.0000072	0.0000043	0.0000108
b5	0.0000258	0.0000175	0.0000356
b6	0.0000826	0.0000585	0.0001108
b7	0.0002511	0.0001736	0.0003430

Using the parameters in the table 4.19, we can predict the average claim amounts for each accident year based on the link function. The estimated values can be used to fill the average claim amounts triangle, as shown in the Table 4.20.

Table 4.20: incremental average claim amounts fitted with inverse Gaussian distribution

Accident year i	Development year i						
	1	2	3	4	5	6	7
1	968.00	286.00	377.00	263.00	196.00	128.00	63.00
2	973.00	478.00	408.00	377.00	179.00	91.00	63.00
3	1064.00	462.00	404.00	315.00	207.00	109.50	63.00
4	1160.00	472.00	485.00	456.00	194.00	109.50	63.00
5	1183.00	562.00	694.00	352.75	194.00	109.50	63.00
6	1342.00	672.00	473.60	352.75	194.00	109.50	63.00
7	1071.00	488.67	473.60	352.75	194.00	109.50	63.00

Table 4.13 presents the fulfilled incremental claim counts triangle fitted with a negative binomial distribution, while Table 4.20 represents the fulfilled incremental average claim amounts triangle fitted with an inverse Gaussian distribution. Therefore, by combining the data from Tables 4.13 and 4.20 and multiplying the corresponding values, we can estimate the incremental claim amounts for each accident year in its development year, as shown in the table 4.21 below.

Subsequently, we may proceed to convert the incremental claim amounts into cumulative totals. Following this, deduction of the claims reported in the calendar year 2005 from the total claims is required. While this process involves summing the incremental claims in the lower triangle, which have just been computed, row by row. By undertaking this, we derive the IBNR reserves for each accident year. The estimated IBNR reserves based on GLMs presented in Table 4.22.

At this point, the entire process of assessing IBNR reserves using two-stage generalized linear models has been completed.

Table 4.21: Estimated triangle of incremental claim amounts based on GLMs

Accident year i	Development year i						
	1	2	3	4	5	6	7
1	22607640	2455310	508196	150436	45276	19968	2961
2	22050126	3648096	527952	203957	34726	10010	2713
3	20163864	2885652	410868	115920	40365	11366	2217
4	19291960	2956136	442320	192888	32033	10754	2097
5	20561723	4046400	821696	159916	36375	12211	2382
6	16997772	2690016	326247	99506	22634	7598	1482
7	11344032	1838438	288135	87882	19990	6711	1309

Table 4.22: Estimated IBNR reserves based on GLMs

Accident year	1999	2000	2001	2002	2003	2004	2005	Total reserves
IBNR	0	2713	13582	44885	210884	457468	2242465	2971996

4.4 Generalized Linear Mixed Models

In this section on GLMMs, we continue to model both the frequency of losses and the average claim amounts.

4.4.1 GLMM for claim counts

Poisson Mixed Effects Model

The fixed effect parameters of the Poisson model are presented in the following table:

Table 4.23: Poisson Model Fixed Effects Estimation

Parameter	Estimate	Std.Error	z value	P value
Intercept	9.732560	0.103078	94.420	0.000
b2	0.103078	94.420	0.000	0.000
b3	0.005831	-176.330	0.000	0.000
b4	0.013547	-210.270	0.000	0.000
b5	0.023164	-162.000	0.000	0.000
b6	0.040330	-115.100	0.000	0.000
b7	0.061460	-83.780	0.000	0.000

The diagnostic results of the Poisson model are as follows. Figures 4.15 and 4.16 illustrate the model residuals for both deviance residuals and Pearson residuals, both of which have a mean of zero and exhibit quite constant variance. However there are some outlying observations. Figure 4.17 illustrates the observed and fitted values of poisson mixed model, the fitted values closely align with the observed values along a straight line, it typically indicates a strong alignment between model predictions and actual observations.

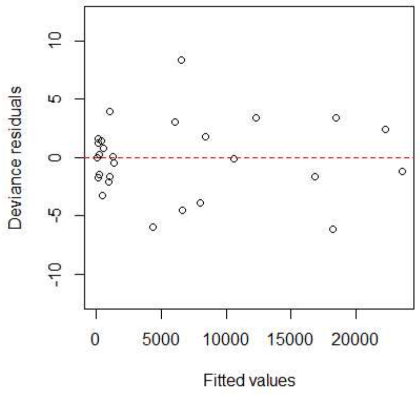

Figure 4.15: Deviance residuals

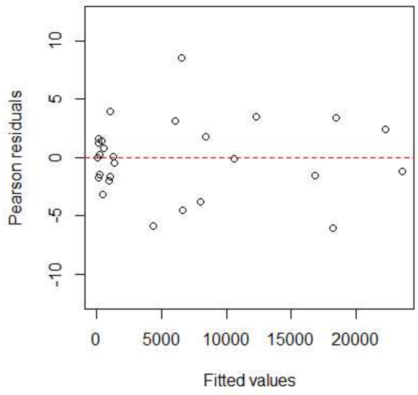

Figure 4.16: Pearson residuals

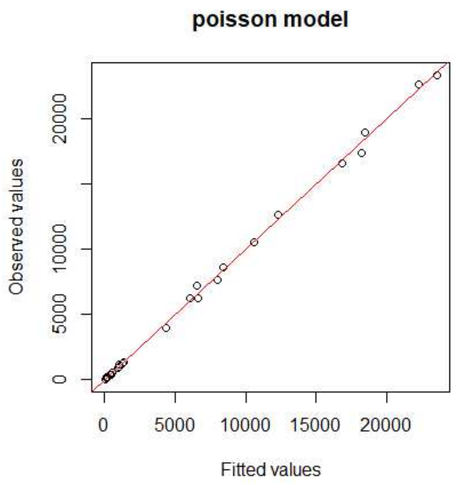

Figure 4.17: Observed and fitted values of Poisson mixed model

Negative Binomial Mixed Effects Model

The fixed effect parameters of the negative binomial model are presented in the following table:

Table 4.24: Negative Binomial Model Fixed Effects Estimation

Parameter	Estimate	Std.Error	z value	P value
Intercept	9.734630	0.105780	92.030	0.000
b2	-1.032090	0.032730	-31.540	0.000
b3	-2.852500	0.037220	-76.640	0.000
b4	-3.744890	0.044100	-84.910	0.000
b5	-4.625880	0.058050	-79.690	0.000
b6	-5.146200	0.078570	-65.500	0.000
b7	-6.229750	0.160080	-38.920	0.000

From Table 4.24, all p-values for the estimated model parameters being 0.00 indicate that these estimated parameters are reliable and significant, with significant implications for model interpretation. Next, we will proceed to illustrate the model diagnostics.

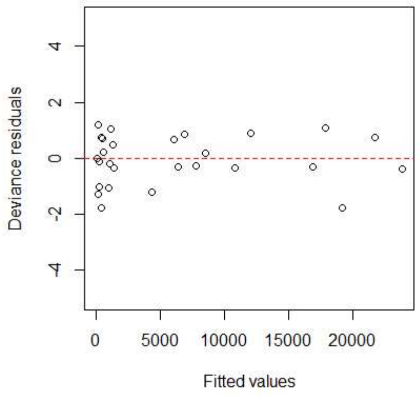

Figure 4.18: Deviance residuals

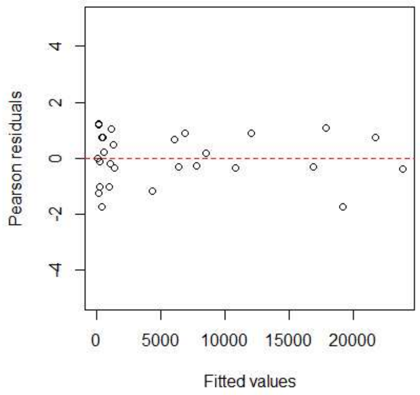

Figure 4.19: Pearson residuals

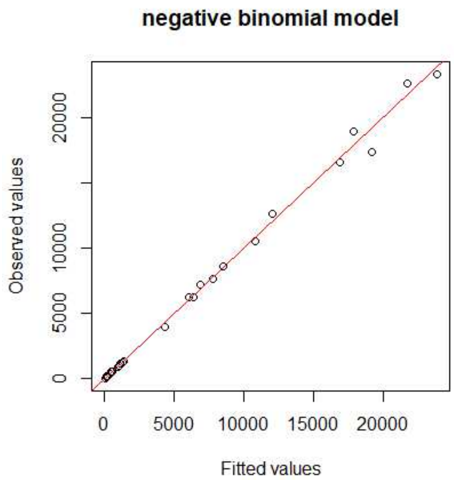

Figure 4.20: Observed and fitted values of negative binomial mixed model

The diagnostic results for the negative binomial model are as follows. Figures 4.18 and 4.19 depict the model residuals, showing a mean of zero and constant variance, indicating a well-fitted model. Additionally, Figure 4.20 demonstrates the alignment between observed and fitted values, suggesting strong correspondence between model predictions and actual observations.

Claim Counts Model Evaluation

This section involves comparing the model selection criteria statistics of the test models for these two distributions. The relevant data has been organized into Table 4.25.

Table 4.25: Pearson χ^2 statistic, deviance and AIC for claim counts under different mixed models

Model	Poisson	Negative Binomial
Deviance	271.2527	21.3197
Pearson Chi-square	271.8148	20.9958
AIC	607.9548	396.8314

Based on various criteria statistics from Table 4.25, it can be observed that the negative binomial distribution yields smaller values. Therefore, the negative binomial distribution is more suitable for predicting claim counts. The predicted random intercepts of negative binomial model are shown in the Table 4.26.

Table 4.26: Predicted random intercepts of negative binomial model

Accident year i	Prediction $\hat{u}_i$
1	0.345040325
2	0.252900046
3	0.055087141
4	0.002412919
5	0.129866553
6	-0.337330978
7	-0.447541735

The parameter b_j in Table 4.24 captures the effect of different development year j and the intercept term reflects the baseline or average level of the number of claims, representing the expected level of claims before accounting for other factors. As the development year increases, the estimated coefficients tend to decrease, suggesting a decreasing trend in claim development over time.

The random intercepts u_i in Table 4.26 captures the additional random variation specific to each accident year i from the expected level. It accounts for variations in claims that are not explained by the average level and the influence of the development year j. Based on the assumptions of normality, the estimated intercepts are randomly scattered around zero. Such an arrangement implies no significant deviations among the claims within one column of the incremental triangle.

Using the fixed effect parameters from negative binomial mixed model in Table 4.24 and the random effect parameters in Table 4.26, we can predict the claim counts for each accident year based on the log link function and formula (3.14). The estimated values can be used to fill the claim counts triangle, as shown in Table 4.27.

4.4.2 GLMM for average claim amounts

Gamma Mixed Effects Model

The fixed effect parameters of the Gamma model are presented in Table 4.28.

Table 4.27: Incremental claim counts fitted with negative binomial mixed model

Accident year i	Development year i						
	1	2	3	4	5	6	7
1999	23355	8585	1348	572	231	156	47
2000	22662	7632	1294	541	194	110	43
2001	18951	6246	1017	368	195	104	35
2002	16631	6263	912	423	166	99	33
2003	17381	7200	1184	455	188	112	38
2004	12666	4003	696	285	118	70	24
2005	10592	3847	623	255	106	63	21

Table 4.28: Gamma Model Fixed Effects Estimation

Parameter	Prediction b	Std.Error	z value	P value
Intercept	7.03234	0.09614	73.15	0.00
b2	-0.83928	0.06822	-12.30	0.00
b3	-0.82419	0.07295	-11.30	0.00
b4	-1.05167	0.07905	-13.30	0.00
b5	-1.58419	0.08796	-18.01	0.00
b6	-2.12219	0.10325	-20.55	0.00
b7	-2.64224	0.13772	-19.18	0.00

From the model parameters, it appears that the model fitting is good. Next, we will proceed with model diagnostics.

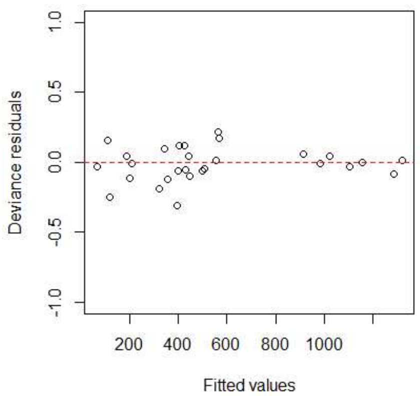

Figure 4.21: Deviance residuals

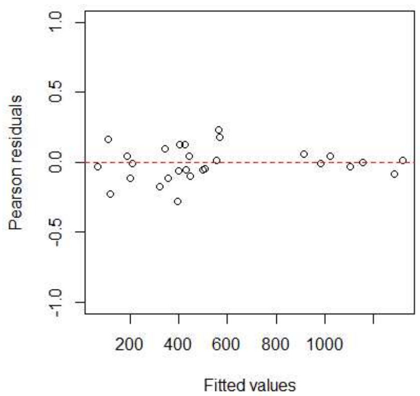

Figure 4.22: Pearson residuals

From the residual plots, it can be observed that the variance of the residuals exhibits a trend of increasing followed by decreasing. This suggests that the predictions of gamma mixed model may be relatively inaccurate within certain ranges of the independent variables, while being more accurate within other ranges. From figure 4.23, apparent linear relationship between predicted values and observed values may indicate that the model explains some of the variation in the data.

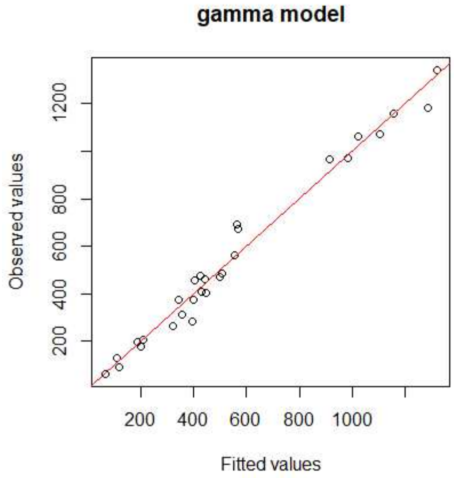

Figure 4.23: Observed and fitted values of gamma mixed model

Inverse Gaussian Mixed Effects Model

The fixed effect parameters of the Inverse Gaussian model are presented in Table 4.29.

Table 4.29: Inverse Gaussian Model Fixed Effects Estimation

Parameter	Prediction b	Std.Error	z value	P value
Intercept	7.0676	0.1291	54.757	0.00
b2	-0.8375	0.1131	-7.403	0.00
b3	-0.8397	0.1171	-7.169	0.00
b4	-1.0563	0.1177	-8.974	0.00
b5	-1.5927	0.1146	-13.893	0.00
b6	-2.1435	0.1160	-18.473	0.00
b7	-2.7118	0.1238	-21.899	0.00

The residual plots in Figure 4.24 and 4.25 demonstrate a good model fit, with residuals exhibiting a random distribution and no apparent trend or pattern. Figure 4.26 demonstrates that fitted values closely align with observed values, indicating a strong fit of the model. Next, we select one of the two average claim amounts mixed models to forecast the average claim amounts.

Average Claim Amounts Model Evaluation

From Table 4.30, the inverse gaussian mixed model shows lower deviance and Pearson Chi-square statistics. Since the aim of the model is to accurately capture the variability in the data and predict, model with lower deviance and Pearson statistics might be more appropriate. We choose the inverse gaussian mixed model as the estimation model for the average claim amounts. The random intercept of the model are shown Table 4.31.

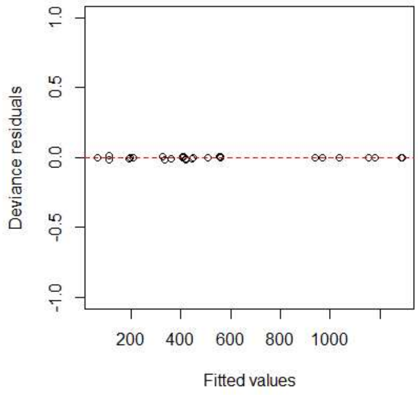

Figure 4.24: Deviance residuals

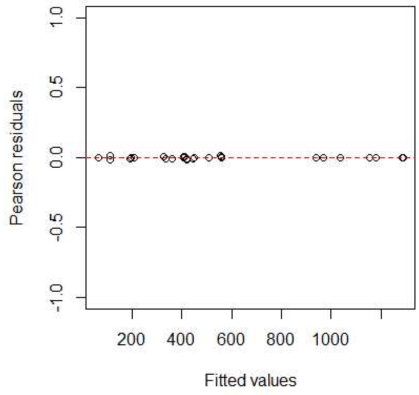

Figure 4.25: Pearson residuals

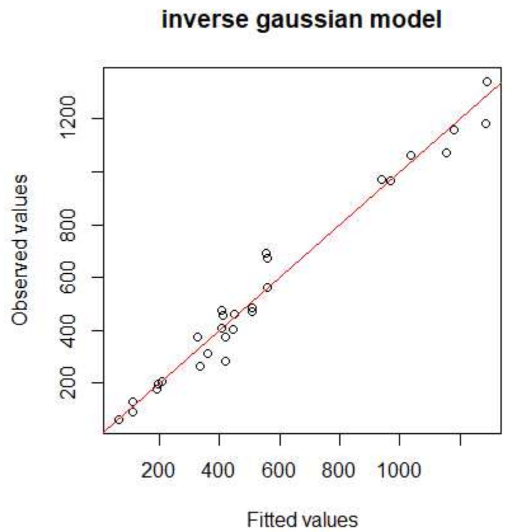

Figure 4.26: Observed and fitted values of inverse gaussian mixed model

Table 4.30: Pearson χ^2 statistic, deviance and AIC for claim amounts under different mixed models

Model	Gamma	Inverse Gaussian
Deviance	0.3993	0.0015
Pearson Chi-square	0.3806	0.0013
AIC	328.6740	339.2791

Based on the fixed effects in Table 4.29 from the inverse gaussian mixed model and the random effects parameters in Table 4.31, the predicted incremental average claim amounts are shown in Table 4.32.

Table 4.31: Predicted random intercepts of inverse gaussian model

Accident year i	Prediction $\hat{u}_i$
1	-0.192882683
2	-0.222058683
3	-0.12556235
4	0.004730318
5	0.092594644
6	0.093018508
7	-0.014755184

Table 4.32: Estimated development triangle of incremental average claim amounts

Accident year i	Development year i						
	1	2	3	4	5	6	7
1999	968.00	286.00	377.00	263.00	196.00	128.00	63.00
2000	973.00	478.00	408.00	377.00	179.00	91.00	62.41
2001	1064.00	462.00	404.00	315.00	207.00	121.34	68.74
2002	1160.00	472.00	485.00	456.00	239.76	138.23	78.30
2003	1183.00	562.00	694.00	447.60	261.77	150.92	85.49
2004	1342.00	672.00	556.09	447.79	261.89	150.98	85.53
2005	1071.00	500.41	499.28	402.04	235.13	135.56	76.79

As we can see from the table, different development and accident years generate varying values, which differ from those of the inverse Gaussian GLM model in the previous section. In the GLMM, we introduced the random intercept, leading to slight variations between different accident years. This suggests that the GLMM predictions may be more reliable for the prediction.

Combining the incremental claim counts triangle fitted using the negative binomial mixed model from Table 4.27 and the incremental average claim amounts triangle fitted using the inverse Gaussian mixed model from Table 4.32, we obtain the incremental total claim amounts triangle, which is shown in Table 4.33.

Table 4.33: Estimated triangle of incremental claim amounts based on GLMMs

Accident year i	Development year i						
	1	2	3	4	5	6	7
1999	22607640	2455310	508196	150436	45276	19968	2961
2000	22050126	3648096	527952	203957	34726	10010	2675
2001	20163864	2885652	410868	115920	40365	12608	2417
2002	19291960	2956136	442320	192888	39767	13626	2612
2003	20561723	4046400	821696	203520	49321	16900	3240
2004	16997772	2690016	386829	127611	30925	10596	2031
2005	11344032	1924973	311064	102617	24868	8521	1633

The estimated IBNR reserves based on GLMMs presented in Table 4.34.

Table 4.34: Estimated IBNR reserves based on GLMMs

Accident year	1999	2000	2001	2002	2003	2004	2005	Total reserves
IBNR	0	2675	15025	56005	272980	557993	2373677	3278355

By comparing the estimation results from Table 4.6 with those from Tables 4.21 and 4.33, it is observed that the estimated IBNR reserves using the chain ladder method are 3,083,884, while the calculated result from the generalized linear model is 2,971,996, lower than that of the chain ladder method. The estimated IBNR using the generalized linear mixed model is 3,278,355, which is higher than the estimates obtained using the chain ladder method and the generalized linear models.

Conclusion

The main goal of the thesis was to apply stochastic models, including GLM and GLMM models, to estimate claims reserves and compare them with traditional deterministic methods. In this thesis, the claim amounts were divided into two sets of data: claim counts and average claim amounts. The thesis focused on modeling and predicting each of them separately. Finally, a practical application of the developed models to a real dataset is performed.

In the first chapter, a brief summary of non-life insurance was provided, including basic notation and techniques for outstanding claim reserves. The concept of the loss triangle was introduced first, followed by a description of the most commonly used methods for calculating insurance reserves, including the chain ladder method and the Bornhuetter-Ferguson method. However, these traditional deterministic methods may overlook important information, such as random fluctuations. Therefore, it is advisable to consider stochastic models, which perform better in addressing these aspects.

When calculating claim reserves, the most commonly used stochastic methods are GLM and GLMM. Therefore, it was important to include the theory of generalized linear models. In the second chapter, we first presented the exponential family distribution and the most commonly used link functions. Then we introduced the basic concepts of these two models, including model formulas, parameter estimation methods, and model diagnostic methods.

The third chapter presents the application of GLM and GLMM for claims reserving. It outlines how the classical structure of GLM and GLMM was transformed and adjusted to create a suitable framework for both claim counts and average claim amounts, utilizing the notation from the previous chapter. Additionally, it provides a list of the most commonly used distributions, mean structures, and link functions for application to the claims reserving datasets.

Finally, the application of different models to a real dataset was presented. We select the most appropriate GLM and GLMM models to estimate reserve amounts respectively, and compare them with the chain ladder method. The obtained results were summarized and compared in the final section of the this chapter.

Bibliography

Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1968.

Katrien Antonio and Jan Beirlant. Actuarial statistics with generalized linear mixed models. In *Insurance: Mathematics and Economics*, volume 40, pages 58–76. Elsevier, 2007.

Ronald L Bornhuetter and Ronald E Ferguson. The actuary and ibnr. In *Proceedings of the casualty actuarial society*, volume 59, pages 181–195, 1972.

Norman E Breslow and David G Clayton. Approximate inference in generalized linear mixed models. volume 88, pages 9–25. Taylor & Francis, 1993.

Rui Chen. Application of kalman filter in estimating no-life outstanding claims reserving. Jan 2009.

G.M. Fitzmaurice, N.M. Laird, and J.H. Ware. *Applied Longitudinal Analysis*. Wiley Series in Probability and Statistics. Wiley, 2011. ISBN 9780470380277. URL `https://books.google.com/books?id=qOmxRtdNJpEC`.

Dian Handayani, Khairil Anwar Notodiputro, Kusman Sadik, and Anang Kurnia. A comparative study of approximation methods for maximum likelihood estimation in generalized linear mixed models (glmm). In *AIP Conference Proceedings*, volume 1827. AIP Publishing, 2017.

Yoonsang Kim, Young-Ku Choi, and Sherry Emery. Logistic regression with multiple random effects: a simulation study of estimation methods and statistical packages. In *The American Statistician*, volume 67, pages 171–182. Taylor & Francis, 2013.

Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. In *ASTIN Bulletin*, volume 23, page 213–225, 1993. doi: 10.2143/AST.23.2.2005092.

P. McCullagh and J.A. Nelder. *Generalized Linear Models, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1989. ISBN 9780412317606. URL `https://books.google.cz/books?id=h9kFH2_FfBkC`.

Charles E McCulloch, Shayle R Searle, and John M Neuhaus. *Generalized, linear, and mixed models*, volume 325. Wiley Online Library, 2001.

J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384, 1972. ISSN 00359238. URL `http://www.jstor.org/stable/2344614`.

Thomas F Tarbell. Incurred but not reported claim reserves. In *Proceedings of the Casualty Actuarial Society*, volume 20, pages 275–280, 1934.

Francis Tuerlinckx, Frank Rijmen, Geert Verbeke, and Paul De Boeck. Statistical inference in generalized linear mixed models: A review. In *British Journal of Mathematical and Statistical Psychology*, volume 59, pages 225–255. Wiley Online Library, 2006.

M.V. Wüthrich and M. Merz. *Stochastic Claims Reserving Methods in Insurance*. The Wiley Finance Series. Wiley, 2008. ISBN 9780470772720. URL `https://books.google.com/books?id=Wo5DIyzqaB4C`.

List of Figures

List of Tables

Source code

The source code used to perform the analysis for the last chapter is included in this appendix.

```
library(ChainLadder)
library(lmtest)
library(MASS)
library(readxl)
library(lme4)
library(Matrix)
library(aod)

## prepare data
a=read.csv(choose.files(),header=TRUE)
count.tri<-as.triangle(as.matrix(count),origin="ay",
   dev="dev",value="count")
amount.tri<-as.triangle(as.matrix(amount),origin="ay",
   dev="dev",value="amount")
cum.tri<-as.triangle(as.matrix(cum),origin="ay",
   dev="dev")
incre.tri<-cum2incr(cum.tri)

#plot the data
plot(cum.tri,  main = "Claims development by origin
   year")
plot(cum.tri, lattice=TRUE, main = "Claims development
   by origin year")
plot(incre.tri,  main = "Claims development by origin
   year")
plot(incre.tri, lattice=TRUE, main = "Claims
   development by origin year")

##chainLadder method
mack<-MackChainLadder(cum.tri,est.sigma = "Mack")
fullprediction<-mack$FullTriangle
developmentfactors<-mack$f
##Generalized Linear Models
#for claim counts
m1 <- glm(value ~ as.factor(dev)+as.factor(ay), data =
   count.tri, family = poisson("log"))
m2 <- glm.nb(value ~ as.factor(dev)+as.factor(ay),
   data = count.tri)

#for claim amounts
m3 <- glm(value ~ as.factor(dev)+as.factor(ay), data =
   amount.tri,
family = Gamma("inverse"))
```

```
m4 <- glm(value ~ as.factor(dev)+as.factor(ay), data =
   amount.tri,
family =inverse.gaussian("1/mu^2"))
m=m1,m2,m3,m4
summary(m)
confint(m)

##Genralized Linear Mixed Models
#for claim counts
gm1 <- glmer(Freq ~ as.factor(dev) + (1 | ay), data =
   flat_count,family=poisson("log"))
gm2<-glmer.nb(Freq ~ as.factor(dev) +(1 | ay),
   data=flat_count)
#for average claim amounts
gm3<-glmer(Freq ~ as.factor(dev) +(1 | ay),
   data=flat_amount,family = Gamma("log"))
gm4<-glmer(Freq ~ as.factor(dev)+(1 | ay) ,
   data=flat_amount,
family = inverse.gaussian("log"))
m=gm1,gm2,gm3,gm4
summary(m)
ranef(m)

##Model Diagnostics
#LR test
m11 <- glm(value ~ as.factor(dev), data = count.tri,
   family = poisson("log"))
m12 <- glm(value ~ as.factor(ay), data = count.tri,
   family = poisson("log"))
m21 <- glm.nb(value ~ as.factor(dev), data = count.tri)
m22 <- glm.nb(value ~ as.factor(ay), data = count.tri)
M31 <- glm(value ~ as.factor(dev), data = amount.tri,
   family = Gamma("inverse"))
M32 <- glm(value ~as.factor(ay), data = amount.tri,
   family = Gamma("inverse"))
M41 <- glm(value ~ as.factor(dev), data = amount.tri,
   family = inverse.gaussian("1/mu^2"))
M42 <- glm(value ~ as.factor(ay), data = amount.tri,
   family = inverse.gaussian("1/mu^2"))
anova(m1, m11, test = "Chisq")
anova(m1, m12, test = "Chisq")
anova(m2, m21, test = "Chisq")
anova(m2, m22, test = "Chisq")
anova(m3, m31, test = "Chisq")
anova(m3, m32, test = "Chisq")
anova(m4, m41, test = "Chisq")
anova(m4, m42, test = "Chisq")
#Wald test
```

```
m=m1,m2,m3,m4
wald.test(Sigma = vcov(m), b = coef(m), Terms = 2:7)
wald.test(Sigma = vcov(m), b = coef(m), Terms = 8:13)

# Model deviance
m=m1,m2,m3,m4,gm1,gm2,gm3,gm4
mDeviance <- sum(residuals(m, "deviance")^2)
# Scaled Pearson Chisq statistic
mPearson <- sum(residuals(m, "pearson")^2)
#AIC and BIC
mAIC<-AIC(m)
mBIC<-BIC(m)
#residuals plot
resid(m)
plot(resid(m),xlab="",ylab="residuals")
abline(h=0,lty=2,col="red")
#scaled deviance and Pearson residuals plots
mderesid <- residuals(m, "deviance")
plot(mderesid ~ fitted(m),
     ylab="Deviance residuals", xlab="Fitted
        values",ylim=c(-1,1))
abline(h=0, "lty"=2, col="red")
mperesid <- residuals(m, "pearson")
plot(mperesid ~ fitted(m),
     ylab="Pearson residuals", xlab="Fitted
        values",ylim=c(-1, 1))
abline(h=0, "lty"=2, col="red")
#Fitted value vs observed value plot
flat_count <- as.data.frame(as.table(count.tri))
observed_counts <- na.omit(flat_count$Freq)
flat_amount <- as.data.frame(as.table(amount.tri))
observed_amounts <- na.omit(flat_amount$Freq)
plot(observed_counts~fitted(m),xlab="Fitted
   values",ylab="Observed values",
main="negative binomial model")
abline(0, 1, col = "red", lwd = 1)
plot(observed_amounts~fitted(m),xlab="Fitted
   values",ylab="Observed values",
main="gamma model")
abline(0, 1, col = "red", lwd = 1)


##Model predictions
newdata <- expand.grid(dev = 1:7, ay = 1:7)

#predicted claim counts
predictions2 <- predict(m2, newdata, type = "response")
predictions2
```

```
matrixm2 <- matrix(predictions2, nrow = 7, byrow =
   TRUE)
incre_count.pre<-as.triangle(matrixm2)
#predicted claim amounts
predictions4 <- predict(m4, newdata, type = "response")
predictions4
matrixm4 <- matrix(predictions4, nrow = 7, byrow =
   TRUE)
incre_amount.pre<-as.triangle(matrixm4)
cum_count.pre<-incr2cum(incre_count.pre)

# calculating total claim amounts development triangle
cum_matrix <- matrixm2 * matrixm4
cum.tri<-as.triangle(cum_matrix)
reserves<- sum(cum.tri[lower.tri(cum.tri, diag =FALSE
   )])
```