NATIONAL ECONOMICS UNIVERSITY

Faculty of Mathematical Economics

Topic

COMPARISON OF CLAIMS RESERVES USING CHAIN LADDER METHOD VS. OVER – DISPERSED POISSON MODEL (GLMs)

A Graduation's Thesis

Author : Ngo Anh Thu
Student ID : 11194961
Class : Actuary 61
Supervisor : MsC. Nguyen Manh Duc

Hanoi, June 2023

ABTRACT

Conventionally, actuaries utilize deterministic approaches to establish the claims reserves, that it consequently leads to the point estimates of it. When applying stochastic reserving techniques, the accessibility of reserve estimation precision measures is provided, and in this point, attention is concentrated on the prediction error, known as the root mean squared error of prediction (rMSEP).

For loss reserving using the overdispersion Poisson model, it is presumed that the incremental claims stay independently and are Poisson distributed with the expectations being the product of two factor, depending on the accident year and the development year, respectively. Moreover, this overdispersion model is formed with a basis of generalized linear model, the utility of quasi – likelihood techniques. The actuary is allowed to generate point estimates and dispersed measures, as well as the entire distribution for delinquent claims that may be used to calculate reserves.

Keywords: *Vietnam Non-life Insurance market, Claims Reserving, Chain Ladder approach, Generalized Linear Models (GLMs), Over – Dispersed Poisson model.*

DEDICATION

To my supervisor, MsC. Nguyen Manh Duc, my family, and my friends, I would like to express my sincerest thanks for the enthusiastic support and encouragement from everyone during the implementation of my project.

TABLE OF CONTENTS

LIST OF TABLES

Figure

Table

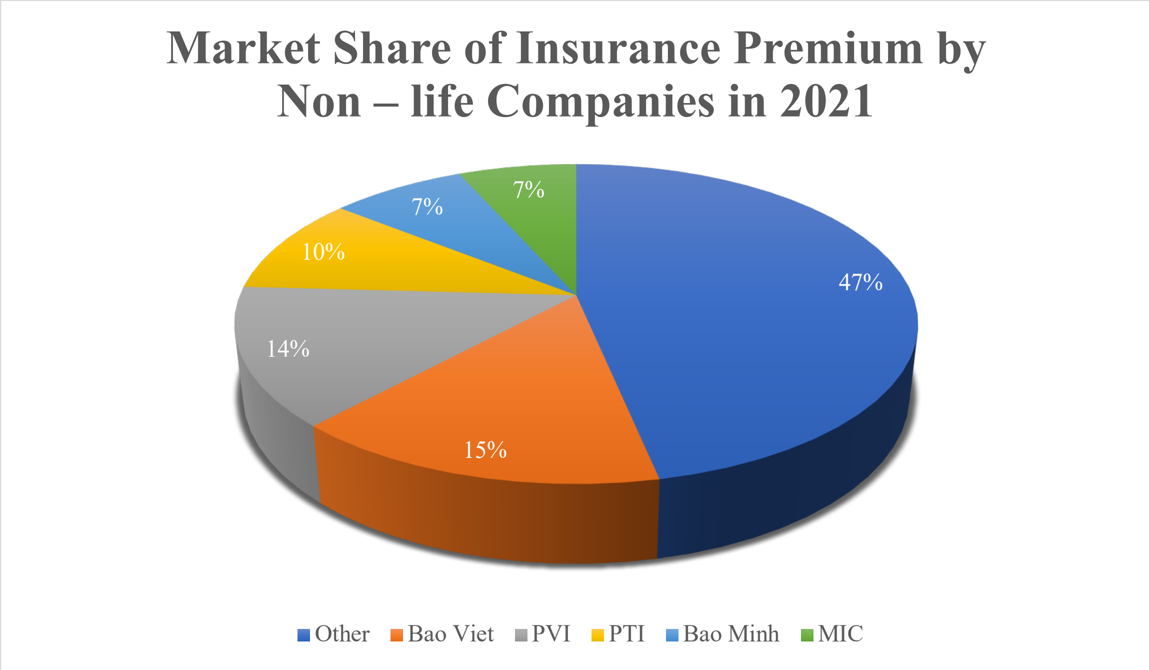
INTRODUCTION

1. Background of the Study

It is documented that the first social insurance in Vietnam was established in 1961, after the North was liberated and built socialism, the law on social insurance was developed and expanded rapidly. However, it stipulated that the insured was state employees. Before 1990, this segment comprised only non – life companies with a monopoly by the state – owned insurance group Bao Viet. Prior to 1995, the target customers of social insurance expanded, that the State budget insured for employees, employers and the State. Moreover, it was time for the Vietnamese to be more concerned with healthcare and insurance products of all types, as well as an aging population and an increase in disposable income; thus, life insurance is established. After twenty years, both life and non-life insurance sectors have experienced a significant increase in insurance premium revenues.

According to the Vietnam's Insurance Report 2021, the gross written premiums of non – life insurance was 59,135 (in billion VND) with 4.34% increase, which is compared to its in 2020. The market unceasingly paid most of the attention to the big 5 of non – life insurance companies, such as Bao Viet (at 15.13%), PVI Insurance (at 14.02%), PTI (at 9.87%), Bao Minh (at 7.60%), and MIC (at 6.65%); moreover, 27 other non – life insurers, as well as branches of foreign organizations in Vietnam contributed to the premium revenues as of 46.72%.

Figure 3.1. Market Share of Premium by Non – life Companies in 2021



(Source: *The Annual Report of Vietnam Insurance Market 2021*)

Particularly, in 2021, the gross claim payment and net retained claim payments accounted for 19.881 (in billion VND) and 12,625 (in billion VND), respectively. The non-life sector had fulfilled its responsibility of preventing and mitigating risks exposed to the insured, thereby reducing the burden on the State budget.

Table 3.1. The Non-life Claim Payments for the period 2017 – 2021

Unit: VND billion

Claims payments	2017	2018	2019	2020	2021
Gross Claim Payment	15,942	19.805	20,752	20,731	19,881
Net claims payments	12,147	13,054	13,887	13,314	12,625

(Source: "*The Annual Report of Vietnam Insurance Market 2021*")

By the end of 2021, total non-life technical reserves had increased by 3.74 percent over the previous year and amount to approximately VND 28,255 billion (compared to VND 27,237 billion in 2020).

Table 3.2. Technical Reserves for the period 2017 – 2021 in Non-life Insurance

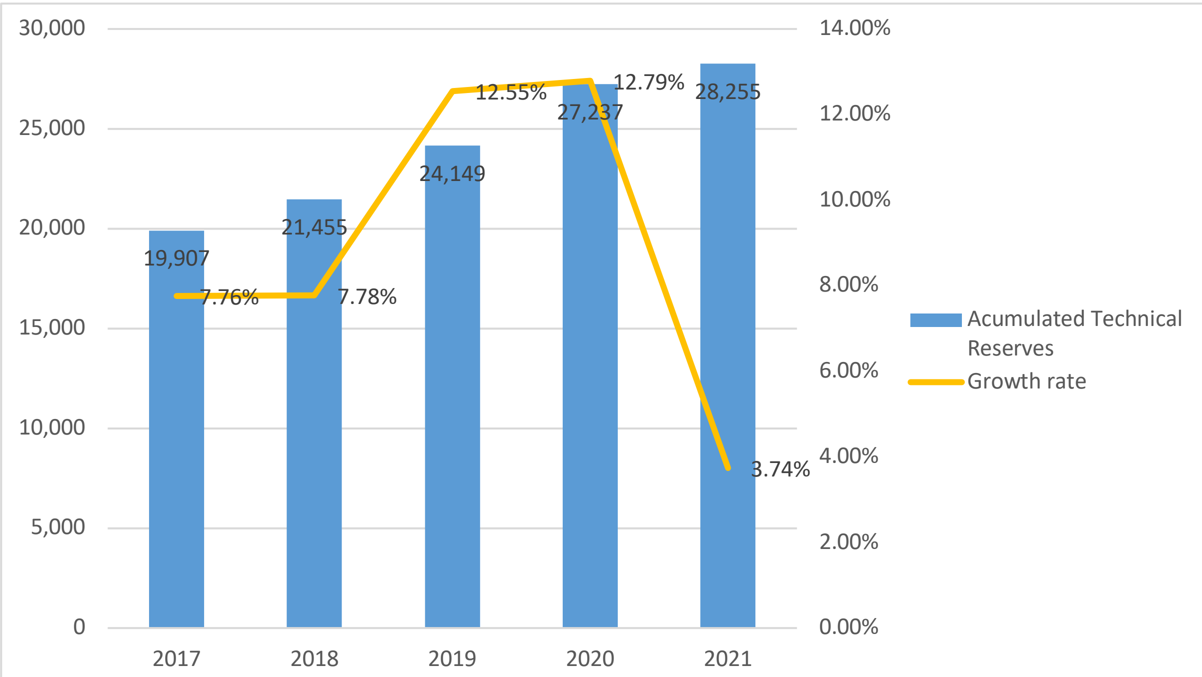
Unit: VND billion

Technical Reserves	2017	2018	2019	2020	2021
Unearned premium reserves	13,320	14,123	16,227	18,736	19,002
Outstanding claim reserves	5,072	5,611	5,962	6,271	6,597
Contingency Reserves	1,515	1,722	1.959	2,230	2,565
Total accumulated technical reserves	19,907	21,455	24,149	27,237	28,255

(Source: "*The Annual Report of Vietnam Insurance Market 2021*")

Figure 3.2. Accumulated Technical Reserves 2017 – 2021

Unit: VND billion



(Source: "*The Annual Report of Vietnam Insurance Market 2021*")

According to both chart and table, it is clearly seen that unearned premium reserves accounted for most of accumulated technical reserves; hence, why it happens? Theoretically, insurers collect premiums for coverage immediately that has not yet been rendered; this is known as written premium. Once insurance companies receive the premium, they typically divide it into two categories: earned and unearned premiums.

Unearned premium is the portion of a policy's premium that has been set aside for the remainder of the policy's term or the premium that must be earned. The unearned premium risk occurs when an unearned premium is insufficient to cover prospective losses; thus, the unearned premium reserve exists to cover losses if policies are terminated before the coverage period begins.

Moreover, according to the figure above, although the accumulated technical reserves increased over 5 years, the growth rate was unstable. In the first two years, 2017 and 2018, it mostly did not change significantly due to the cautious business targets of non – life insurers in 2017. The first quarter of 2017 documented major losses for non-life insurance businesses in Vietnam stemming from several serious incidents. For example, a $11 million loss was occurred due to a fire at manufacture of Truong Hai Auto Corporation (also known as Thaco) in Quang Nam province. Another case was pirate attacking Vinashin's Giang Hai cargo ship off the coast of Philippines. It leaded to the concern of non – life insurers about an abnormal increase in payouts.

In the same year, insurers anticipated that they could achieve an excessive premium earning by selling their insurance products on several sales platforms such as telesales, bancassurance and VNPost.

Until 2019, the growth rate of technical reserves in non – life insurance sector grew for 12.55% and continued rising in 2020, and dropped to 3.74% because of the pandemic, COVID – 19, had had a major effect on almost all economic segments since 2020. After two years of single-digit growth, the non-life insurance market experienced a significant improvement in the first half of 2022 across all lines, including health and auto insurance, which were severely impacted by the Covid – 19. Insurance distribution channels continue to diversify, primarily through e-commerce platforms such as Tiki, Lazada, and Shopee and service applications such as Grab, Be, and Gojek. The Covid-19 epidemic has accelerated and intensified the inevitable trend of digital transformation. Customers are becoming increasingly habituated to exploring for products, purchasing items, and claiming compensation online. Considering the intensive market competition in conventional insurance distribution channels, many organizations view digital transformation as a means to advance.

Therefore, with the development of non – life insurance market in Vietnam and the current volatile economic situation, careful and thorough calculation is required in determining the claims reserve. The actuary heretofore applies traditional method, deterministic approaches, in claims reserve estimation that the most prevalent is Chain Ladder. Despite its simple application, it only provides the result in the best estimate. If the actuaries want to accurately establish reserves for future losses, they might carry out more calculation of prediction error. However, in recent years, many actuaries have applied stochastic methods in claims reserving, which has precise merit over deterministic ones, such as provision for conducting diagnostic checks and the production of confidence intervals (C. I).

This study aims to review the Over – Dispersed Poisson model (ODP) under the Generalized Linear Structure; to analyze the characteristics of the model and compare it to the Chain Ladder method with an explanation of the results; and to visualize the practical application of the ODP model.

Stochastic claims reserving models consider the process of reserve estimation as a data analysis and construct a reserving model inside a statistical framework, which tends to produce measures of best estimates and variance. In this paper, the stochastic model is built based on the traditional one, Chain Ladder technique, with the purpose of providing the same reserve estimates. Besides, with the supplication of the link between two methods, it is beneficial the actuaries, who is familiar with traditional methods and need an inception to carrying out stochastic method research.

Sometimes, when the deterministic approaches fail, actuaries hope that stochastic techniques might provide the solution for claims reserving. In fact, stochastic models are sometimes assessed by their ability to succeed when basic deterministic models fail. There are some studies which were accomplished related to this topic, such as Chain

Ladder and maximum likelihood by Verrall, R. J (1989), Regression based models on log – incremental payments by Christofides, S (1990), A stochastic model underlying. the Chain Ladder technique by Renshaw, A. E & Verrall, R. J (1998), and Stochastic claims reserving in general insurance by England, P. D and Verrall, R. J (2002). And in this paper, we will concentrate on claims reserving in non – life insurance sector.

2. Objectives of the Study

This study aims to review the Over – Dispersed Poisson model (ODP) under the Generalized Linear Structure; to analyze the characteristics of the model and compare it to the Chain Ladder method with an explanation of the results; and to visualize the practical application of the ODP model.

The main target of this study is to develop a statistically valid reserving model and treat the reserving procedure as a data analysis in order to produce accurate estimates and measures of variability. The goals are:

- To review the Over – Dispersed Poisson model (ODP) under Generalized Linear Structure and the Chain Ladder method;
- To analyze the characteristics of the model and compare it to the Chain Ladder method with an explanation of the results;
- To visualize the practical application of two methods.

3. Significance of the Study

The significance of this research is to proving that both method give the actuary the same result (except for the special cases). In addition, it also compares the differences between the deterministic and stochastic applications.

LITERATURE REVIEW

Differ from traditional methods, stochastic approaches not only provide the best estimate claims reserve estimation, but also determine the uncertainty of measures. Previously, England and Verrall (2002) or Hess and Schmidt (2002) provided a comprehensive overview of the stochastic approaches presently used to estimate claims reserve.

The usage of generalized linear models (GLMs; Anderson et al., 2004) was first emphasized by England and Verrall (2002).

The framework for generalized linear models (GLMs; Anderson et al., 2004) is emphasized by England and Verrall (2002). Three examined models were able to accommodate negative values, according to England and Verrall (2002). They are overdispersion poisson model, negative binomial model, and normal approximation to the negative binomial model. When adverse incremental claims are made, the log-normal model, which Kremer (1982) and Verrall (1991) studied extensively, is also referenced. Researchers implied that the Poisson model does not only involve positive integer data. They, subsequently, offer a quasi-likelihood approach that can be used to circumvent this drawback and can be applied to both positive and negative non-integer data. The negative binomial is supported by an analogous argument. The purpose of this study is not to determine if they are utilizing these distributions or some continuous approximation. In addition, the normal distribution is unsuitable because long-tailed distributions are used to simulate claim data.

Verrall (2000), Mack and Venter (2000), Verrall and England (2000), Mack (1994), and Verrall (2000) all examine the stochastic foundation of Chain Ladder model. The link between the various models and whether they may legitimately be utilized to improve the deterministic Chain Ladder approach are at the core of the issue.

An Overdispersion Poisson distribution under Generalized Linear structure is the first stochastic variant of the Chain Ladder method (referred by Renshaw and Verrall,

1998) that can be utilized when the required values are expanding. Noted that the mean and variance of the over – dispersed Poisson model do not equal as in the Poisson distribution.

CLAIMS RESERVE ESTIMATION

1. Data Preparation

In non – life enterprises, the historical payments triangle, which was updated at the estimate date, is frequently used to compute the claims reserve for accidents yet to be paid generated by an insured risk portfolio at the end of the fiscal year.

In particular, we presume that the paid – payment observations are related to accidents that occurred within a certain past timeframe. It leads to the amounts paid for accidents that occurred or were created in prior years are available in this kind of diagram.

A variable that quantifies the claim payment year is created by dividing the data for each accident year into development years. Besides, both incremental claims or cumulative claims are used to model the portfolio.

When there are incremental claims, the loss of the accident year i is represented as a collection of random variables $\{S_{i,j}\}_{i,j \epsilon\{0,1,\ldots,n\}}$ which is settled with a delay of j years and hence in development year j and in calendar year $i + j$.

The incremental assertions $S_{i,j}$ are presumptively observable for calendar years $i + j \leq n$ and non-observable for calendar years $i + j \geq n + 1$. The observable incremental claims are represented by the following run-off triangle.

Table 5.1. The run-off triangle of incremental claims

Accident Year	Development Year				
	0	1	…	$n-1$	n
0	S_{00}	S_{01}	…	$S_{0,n-1}$	S_{0n}
1	S_{10}	S_{11}	…	$S_{1,n-1}$	
…	…	…	…		
$n-1$	$S_{n-1,0}$	$S_{n-1,1}$			
n	S_{n0}				

In case run-off triangle is in the form of cumulative claims, it is given C_{ij} as the total paid amount, with j as the delay in payment for accidents happened in the i-th year, usually called incremental payment. To model a portfolio, actuary gather a set of random $\{C_{i,j}\}_{i,j\,\epsilon\{0,1,\dots,n\}}$ and it is form from incremental losses by the equation

$$C_{i,j} := \sum_{l=0}^{j} S_{i,l} \qquad (5.1)$$

Thus, it is clear that modeling a portfolio by incremental claims is quite equivalent to modeling a portfolio by cumulative claims

Table 5.2. The run-off triangle in the form of cumulative claims

Accident Year	Development Year				
	0	1	…	$n-1$	n
0	$C_{0,0}$	$C_{0,1}$	…	$C_{0,n-1}$	$\boldsymbol{C_{0,n}}$
1	$C_{1,0}$	$C_{1,1}$	…	$\boldsymbol{C_{1,n-1}}$	$C_{1,n}$
…	…	…	**…**	…	…
$n-1$	$C_{n-1,0}$	$\boldsymbol{C_{n-1,1}}$	…	$C_{n-1,n-1}$	$C_{n-1,n}$
n	$\boldsymbol{C_{n,0}}$	$C_{n,1}$	…	$C_{n,n-1}$	$C_{n-,n}$

Sometimes, data is also earned premium and expected loss ratio to apply Bornhuetter – Ferguson Method, which is as common as Chain Ladder in claims reserve estimation. Notwithstanding, whilst the Chain Ladder method relies completely on the data contained in the run-off triangle, Bornhuetter – Ferguson method utilizes the earned premium and an expected loss ratio to estimate the expected ultimate loss and restricts the usage of the run-off triangle to the computation of the percentage of the outstanding loss.

$$L_i^{ult} = P^E * ELR \tag{5.2}$$

Where L_i^{ult} is expected loss ratio in year i

P^E is earned premium

ELR is expected loss ratio

When applying Bornhuetter – Ferguson method in reserve calculation, earned premium is one of the most significant factors, as policyholders pay premiums in advance, insurers do not immediately consider premiums paid for an insurance contract as earnings. Once the premium is paid, it is considered an unearned premium – not a profit.

According to the equation, the earned premium directly impacts on the expected loss ratio, then the reserves.

2. The Prediction and Prediction Error

The Claims Reserving is a process that actuary utilizes tools to estimate the future claims based on the given data. In this section, the Chain – Ladder as Over – dispersed Poisson model is applied to give a more diverse perspective in claims reserving using modern approach – Stochastic Claims Reserve Model; thus, the expected value by GLMs and the root mean squared error of prediction (rMSEP), also known as prediction error, are used for the prediction.

Consider a random variable y and a predicted value $\hat{y}$. The mean squared error of the prediction (MESP) equals:

$$E[(y - \hat{y})^2] = E[((y - E[y]) - (\hat{y} - E[\hat{y}]))^2] \tag{5.3}$$

Substituting $\hat{y}$ instead of y results in:

$$E[(y - \hat{y})^2] \approx E[(y - E[y])^2 - 2E[(y - E[y])] * E[(\hat{y} - E[\hat{y}])] + E[(\hat{y} - E[\hat{y}])^2] \tag{5.4}$$

Assuming the future observation are independent of the past observation gives:

$$E[(y - \hat{y})^2] \approx E[(y - E[y])^2 + E[(\hat{y} - E[\hat{y}])^2]$$

In other words, it is

Prediction Variance = Process Variance + Estimation Variance (5.5)

With the attempt to make an estimate of the prediction error of future payments and reserve estimation by using traditional approaches, there is only two components, process variance and estimation variance, that we need to be concerned about.

It is essential to distinguish the prediction error and the standard error. Strictly, whilst the square root of the estimation variance equals the standard error, the prediction error is concerned with a forecast's variability, accounting for both parameter estimate uncertainty and the inherent unpredictability in the data being anticipated. Unfortunately, nomenclature in the literature is imprecise; the rMSEP is sometimes referred to as the standard error of prediction or as the standard error.

3. The Chain Ladder Method

Chain Ladder method is a common method in reserve estimation. It was introduced by Mack in 1993 and has become the most popular method in reserve estimation in insurance industry. This actuarial method is based on the insured historical data to forecast the reserve in order to set up a coverage amount in the future.

This approach relies on run-off triangles of paid losses, incurred losses, and development factor, where run-off triangles, often referred to as delay triangles, are two-dimensional matrices created by gathering claim information over time. To create the run-off matrices, a stochastic process is used to the claim data after several degrees of freedom have been considered.

The initial data could be the incremental run-off triangle or the cumulative, it depends on the data collection. Table 2 demonstrates run-off triangle in cumulative claims, row represents accident year and column represents development year. In table 3 (in the previous section), there is a development triangle (for i + j ≤ n) and future triangle (for i + j > n). Mack assumes that the cumulative claims in Chain Ladder is distribution free and specify the first two moments for cumulative claims based on the following assumptions.

(1) Random variables $\{C_{i,0}, \dots, C_{i,j}\}$ and $\{C_{p,0}, \dots, C_{p,j}\}$ independent for accident year $i \neq p$.

(2) For $0 \leq i \leq n$ and $0 \leq j \leq n-1$ there is development factor $f_j > 0$ and $\sigma_j^2 > 0$ such that,

$$E\left(C_{i,j+1} \middle| C_{i,0}, C_{i,1}, \dots C_{i,j}\right) = C_{i,j} f_j \tag{5.6}$$

$$Var\left(C_{i,j+1} \middle| C_{i,0}, C_{i,1}, \dots C_{i,j}\right) = C_{i,j} \sigma_j^2 \tag{5.7}$$

with

$$\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}}; \text{ for } 0 \le j \le n-1 \quad (5.8)$$

$$\sigma_j^2 = \frac{1}{n-j-1} \sum_{i=0}^{n-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - f_i \right)^2 \text{ for } 0 \le j \le n-2 \quad (5.9)$$

In general, the prediction of ultimate claims for future triangle in every accident year i is

$$\hat{C}_{i,n}^{CL} = C_{i,n-i} * \prod_{n-i}^{n-1} \hat{f}_j, \ \ i + j > n \quad (5.10)$$

So that the claim reserve for every accident year i is

$$\hat{R}_{i,n}^{CL} = \hat{C}_{i,n}^{CL} - C_{i,n-i} \quad (5.11)$$

and the predicted value of total claim reserve is

$$\hat{R}^{CL} = \sum_{i=1}^{n} \hat{R}_i^{CL} \quad (5.12)$$

In claims reserving by using Chain Ladder method, the reason why average factor model, or n – average factor is not applied in claims reserving is that average factor model is used only when the number of claims is stable. However, the actuary apply volume weighted average factor in reserve estimation in case when data loss occurs abnormally, such as inflation, epidemic, ... using volume - weighted will reduce the impact of that abnormal data.

Thus, with the uncomplicated form, the Chain Ladder method comprises the ultimate claims only, where the ultimate is interpreted as the latest delay year so far observed and does not include any tail factors. From a statistical perspective, after arriving at a point estimate, the logical next step is to create estimates of the anticipated variability in the result. This allows evaluations to be made, such as whether additional reserves should be retained above and beyond the projected values out of caution. However, to calculate the variability of the reserve to make a more precise result, the next step is estimating the prediction error, known as the standard error of the distribution of possible reserve outcomes. We must figure out a model that can roughly represent the data in the triangle, including both past and future data, and, ideally, produce an estimate that is the same as or very comparable to that provided by the chain ladder model. The Over-Dispersed Poisson model enters the picture at this point.

THE PREDICTION ERROR OF CHAIN LADDER METHOD

Although the result from the Chain Ladder approach does not provide the prediction error (rMSEP), it can be estimated based on the equation (4.5), under the Murphy's Model (2007).

The Mack's Theory (1993) was the inspiration for the creation of the Murphy's Model. According to the Mack technique, the reserve risk is calculated using the mean square error, whereas the Murphy method calculates the reserve?risk using the total variance (Murphy, 2007). The mean square error may be divided down into three elements:

- Process risk;
- Parameter risk;
- Bias

The mean square error for the accident year i and the ultimate development year I for the estimator $\hat{C}$, can be defined as the expected squared deviation of $\hat{C}$ (which is a random variable) from the value of the random variable C which is estimated by

$$mse(\hat{C}) = E(\hat{C} - C)^2 \quad (5.13)$$

The mean square error can be decomposed into Variance and Bias as follows:

$$mse\ (\hat{C}) = Var(C) + Var(\hat{C}) + Bias^2(\hat{C}) \quad (5.14)$$

Therefore, the mean square error consists of the sum of the process risk, parameter risk, and the square of the bias of $\hat{C}$.

The Murphy's model predicts that process risk and parameter risk will both be calculated individually. The claim variance and variance of the estimated development factor must also be estimated prior to their calculation.

Estimated of the claims variance

$$\hat{\sigma}_k^2 = \frac{1}{I - k - 1} \sum_{i=1}^{I-k} C_{i,k} \left(\frac{C_{i,k+1}}{C_{i,k}} - \hat{f}_k \right)^2, \quad 1 \le k \le I - 2 \qquad (5.15)$$

$\hat{\sigma}_k^2$ is an unbiased estimator for σ_k^2. However, in the case of $1 \le k \le I - 2$, the estimator for σ_{I-1}^2 is acquired as well.

$$\sigma_{I-1}^2 = min(\hat{\sigma}_{I-2}^4 / \hat{\sigma}_{I-3}^2), min(\hat{\sigma}_{I-3}^2;\ \hat{\sigma}_{I-2}^2) \qquad (5.16)$$

Estimated of the variance of development factor

$$\widehat{Var}(\hat{f}_k) = \hat{\sigma}_{\hat{f}_k}^2 = \frac{\hat{\sigma}_k^2}{\sum_{r=1}^{I-k} C_{r,k}} \qquad (5.17)$$

$$\widehat{Var}(\hat{f}_{k-1}) = \hat{\sigma}_{\hat{f}_{k-1}}^2 = \frac{\hat{\sigma}_{k-1}^2}{\sum_{r=1}^{I-k-1} C_{r,k-1}} \qquad (5.18)$$

The Process & Parameter Risk

Equation (5.19)

$$\widehat{Process\ Risk}_{i,k} = \begin{cases} \hat{f}_{k-1}^2\ \widehat{Process\ Risk}_{i,k-1} + \hat{C}_{i,k-1}\ \hat{\sigma}_{k-1}^2 & for \quad k > I + 2 - i \\ C_{i,I+1-i}\ \hat{\sigma}_{k-1}^2 & for \quad k = I + 2 - i \end{cases}$$

Equation (5.20)

$$\widehat{Parameter\ Risk}_{i,k} = \begin{cases} \hat{f}_{k-1}^2\ \widehat{Parameter\ Risk}_{i,k-1} + \hat{C}_{i,k-1} \widehat{Var}(\hat{f}_k) + \widehat{Var}(\hat{f}_k) \widehat{Parameter\ Risk}_{i,k-1} & for \quad k > I + 2 - i \\ C_{i,I+1-i}\ \hat{\sigma}_{k-1}^2 & for \quad k = I + 2 - i \end{cases}$$

4. The Overdispersion Poisson Model under the GLMs

In this context, we will concentrate only on the overdispersion Poisson model (ODP) with the structure of the Generalised Linear Models (GLMs).

Generalised Linear Models (GLMs) are known as effective tools in regression and data prediction, which is based on the idea of classical linear models. The term of GLMs was first published by McCullagh (1982) and Nelder (2nd edition 1989), believing it to be a substantial class of non-linear regression models. Moreover, all special cases such as Poisson regression, Logistic, and Linear regression are included in Generalised Linear Models. The most significant inference qualities when it comes to estimate are driven by the variance function selection, not the distribution. In addition, the generality of GLMs will allow us to introduce new applications, such as Gamma regressions for fat-tailed distributions and Tweedie distributions for two-part data.

For count data, the Poisson Model with a log link is the fundamental GLM. When the response variable is a count, however, its conditional variance increases faster than its mean, resulting in a condition known as overdispersion and invalidating the use of the Poisson distribution.

The Over – Dispersed Poisson distribution has the same concept as the Poisson distribution; however, the mean and variance of the over-dispersed Poisson are proportional. Assume that

$$C_{ij} \sim \ iid\ ODP(\mu_{ij})$$

that iid is the abbreviation of independent, with

$$\begin{gathered} E[C_{ij}] = \mu_{ij} \ \text{ and } \ Var[C_{ij}] = \phi E[C_{ij}] \ = \phi\mu_{ij} \\ ln(\mu_{ij}) = \mu + \ \alpha_i \ + \ \beta_j \end{gathered} \tag{5.21}$$

which is known as a generalized linear model in which the responses C are modeled using a logarithmic link function and a linear predictor, η_{ij}. In this model, the parameters are linear, so that it is appropriate for fitting the chain ladder model as we have one parameter for each row i and each column j. Thus, we apply the corner constraints as we have many overdispersion parameters as values to

$$\begin{cases} \eta_{ij} = \mu + \alpha_i + \beta_j \\ \alpha_i = 0 \\ \beta_j = 0 \\ \log \mu_{ij} = \eta_{ij} \end{cases} \tag{5.22}$$

i.e. the first two parameters are zeroized.

For $\boldsymbol{\beta_j}$, the column parameter, generates the run-off data, we conclude the general hypotheses that the run-off pattern lacks a particular form since it comprises only one parameter for each column, conforming to the fundamental assumptions underlying the classic Chain Ladder model. Moreover, the overdispersion Poisson model is resilient in the context of a negligible number of negative incremental claims. This is a crucial trait, particularly for lines of products whose case reserves are determined by claims administrators who frequently make errors.

Besides, it is clearly seen that ϕ is a parameter to accommodate the Over - Dispersion. As the variance is proportional to mean by ϕ; thus, the overdispersion Poisson also relies on ϕ, which is unknown, and could be estimated from historical data.

THE ESTIMATION OF PARAMETERS

The difference between Poisson distribution and overdispersion Poisson is that the dispersion variance is proportional to the mean by ϕ; hence, the over-dispersed Poisson distribution relies on ϕ to estimate other parameters. Given that

$$\sum \left(\frac{\frac{C_{ij}}{\phi} - \frac{\hat{\mu}_{ij}}{\phi}}{\sqrt{\frac{\mu_{ij}}{\phi}}} \right) \sim \mathcal{X}^2_{n-p} \tag{5.23}$$

The above distribution is a result of the Pearson residual of Poisson distribution. i.e. the quotient which is squared in the expression above is N(0, 1) distributed and the sum of squared N(0, 1) variables are $\mathcal{X}^2$ distributed with the degree of freedom qual to the number of observations minus the number of parameters.

The moments method provides

$$\sum \left(\frac{C_{ij} - \hat{\mu}_{ij}}{\sqrt{\phi}\sqrt{\mu_{ij}}} \right)^2 = n - p \tag{5.24}$$

Since ϕ equals constant it is readily available from (function above)

$$\hat{\phi} = \frac{1}{n - p} \sum \left(\frac{C_{ij} - \hat{\mu}_{ij}}{\sqrt{\mu_{ij}}} \right)^2 \tag{5.25}$$

While C_{ij} represents the total number of claims, one may understand ϕ as the average claim size.

Subsequently, the estimation of other GLM parameters is carried out with the log – likelihood calculation function, I, and maximizing it:

$$I = \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \frac{1}{\phi} \left(C_{ij} * \log \mu_{ij} - \mu_{ij} \right) + \ldots \tag{5.26}$$

$$= \frac{1}{\phi} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \left[\left(C_{ij}\left(\mu + \alpha_i + \beta_j\right) - e^{\mu + \alpha_i + \beta_j} \right) \right] + \ldots \tag{5.27}$$

THE ESTIMATION OF PREDICTION VARIANCE

The estimation variance for each estimated value is

$$Var\left(\hat{\mu}_{ij}\right) = Var\left(e^{\hat{\eta}_{ij}}\right) = \left(\frac{\partial e^{\hat{\eta}_{ij}}}{\partial \hat{\eta}_{ij}} \right)^2 Var\left(\hat{\eta}_{ij}\right) \tag{5.28}$$

Taylor estimation of (function above) provides

$$Var\left(\hat{\mu}_{ij}\right) = \left(e^{\hat{\eta}_{ij}}\right)^2 Var\left(\hat{\eta}_{ij}\right) = \hat{\mu}_{ij}{}^2\, Var\left(\hat{\eta}_{ij}\right) \tag{5.29}$$

According to (5.5), the prediction variance for each values in cells equals

$$PE = \hat{\phi}\hat{\mu}_{ij} + \hat{\mu}_{ij}^{\,2}\, Var(\hat{\eta}_{ij}) \tag{5.30}$$

Even the covariance should be included when calculating the total prediction variance:

$$PE = \sum \hat{\phi}\hat{\mu}_{ij} + \sum \hat{\mu}_{ij}^{\,2}\, Var(\hat{\eta}_{ij}) + 2\sum Cov(\hat{\eta}_{ij}, \hat{\eta}_{ik})\, \hat{\mu}_{ij}\hat{\mu}_{ik} \tag{5.31}$$

Standard software programs like R, SPPS, and the likes, can be used to implement the over-dispersed Poisson model under the generalized linear structure.

DATA ANALYSIS

In this section, estimation of claims reserve by using the Chain Ladder method versus the overdispersion Poisson model is implemented with practical data collected from Bao Viet insurance company. Excel will be used to carry out the claims reserving using the Chain Ladder method; whereas, the over-dispersed Poisson model will be implemented in R program.

1. Data Description

Table 6.1. Data Description

Data	Practical data from Bao Viet General Insurance Corporation
Type of Data	Run – off Triangle form of Incremental Claims
Time range	2011 – 2018
Unit	VND million

About the data, it is collected from Bao Viet insurance company, which is an analysis of claims development of automobile line in 2011. It includes automobile claims during 8 years, from 2011 to 2018. We have the incremental claims under the run-off triangle form below

Table 6.2. Run – off Triangle of Incremental Claims

Unit: VND million

AY	**DY**							
	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
2011	387.441	1,724.078	5,620.458	3,024.839	5,726.136	4,237.830	1,578.160	2,282.757
2012	2,326.188	4,205.659	908.9385	10,289.794	4,070.344	2,991.686	3,283.593	
2013	3,597.041	6,019.537	4,090.429	2,701.838	2,972.990	1,569.183		
2014	1,280.728	8,855.254	2,767.459	3,081.241	2,108.870			
2015	2,030.654	5,391.969	4,679.294	5,118.204				
2016	3,783.948	5,770.583	7,226.111					
2017	3,490.599	9,128.425						
2018	3,981.962							

2. Application of the Chain Ladder Method using Excel

As the historical is under the incremental claims data form, we have to change it into the cumulative claims under run-off triangle by applying the below equation

$$C_{i,j} := \sum_{l=0}^{j} S_{i,l}$$

Table 6.3. Run – off Triangle of Cumulative Claims

Unit: VND million

AY	DY 2011	2012	2013	2014	2015	2016	2017	2018
2011	387.44	2,111.52	7,731.98	10,756.82	16,482.95	20,720.78	22,298.94	24,581.70
2012	2,326.19	6,531.85	7,440.79	17,730.58	21.800.92	24.792.61	**28,076.20**	
2013	3,597.04	9,616.58	13,707.01	16,408.85	19,381.84	**20,951.02**		
2014	1,280.73	10,135.98	12,903.44	15,984.68	**18,093.55**			
2015	2,030.65	7,422.62	12,101.92	**17,220.12**				
2016	3,783.95	9,554.53	**16,780.64**					
2017	3,490.60	**12,619.02**						
2018	**3,981.96**							

Afterwards, development factor is calculated by implementing volume-weighted factor, as the historical data loss was affected by economic factors; hence we can reduce the impact of abnormal data. The development factor is shown in table 3 below:

Table 6.4. Development factor of the Chain Ladder method

Development factor	
$\mathbf{f_1}$	3.4321757
$\mathbf{f_2}$	1.5574382
$\mathbf{f_3}$	1.4493989
$\mathbf{f_4}$	1.2443843
$\mathbf{f_5}$	1.1525811
$\mathbf{f_6}$	1.1068203
$\mathbf{f_7}$	1.1023707

For each year, the estimated reserve will be multiplied with corresponding development factor until the result of the ultimate year.

Table 6.5. Estimated reserves for each year

Unit: VND million

AY	DY							
	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
2011	387.44	2,111.52	7,731.98	10,756.82	16,482.95	20,720.78	22,298.94	**24,581.70**
2012	2,326.19	6,531.85	7,440.79	17,730.58	21,800.92	24,792.61	**28,076.20**	**30,950.38**
2013	3,597.04	9,616.58	13,707.01	16,408.85	19,381.84	**20,951.02**	23,189.01	**25,562.89**
2014	1,280.73	10,135.98	12,903.44	15,984.68	**18,093.55**	20,854.29	23,081.95	**25,444.86**
2015	2,030.65	7,422.62	12,101.92	**17,220.12**	21,428.45	24,698.03	27,336.28	**30,134.71**
2016	3,783.95	9,554.53	**16,780.64**	24,321.84	30,265.72	34,883.70	38,609.98	**42,562.51**
2017	3,490.60	**12,619.02**	19,653.35	28,485.54	35,446.96	40,855.50	45,219.70	**49,848.87**
2018	**3,981.96**	13,666.79	21,285.19	30,850.73	38,390.16	44,247.77	48,974.33	**53,987.86**

The total reserve $= (30{,}950.38 - 28{,}076.20) + (25{,}562.89 - 20{,}951.02) + (25{,}444.86 - 18{,}093.55) + (30{,}134.71 - 17{,}220.122) + (42{,}562.51 - 16{,}780.64) + (49{,}848.87 - 12{,}619.02) + (53{,}987.86 - 39{,}81.96) = \mathbf{140{,}769.56}$ million VND

Table 6.6. Reserves for each year of origin

Unit: VND million

AY	Reserve
2011	0
2012	2,874.18
2013	4,611.87
2014	7,351.31
2015	12,914.59
2016	25,781.87
2017	37,229.84
2018	50,005.90
Total Reserve	**140,769.56**

By applying the equation (5.15), the result of claims variance is shown in the table below,

Table 6.7. Claims Variance Estimation Result

k	1	2	3	4	5	6	7
$\hat{\sigma}_k^2$	5,602.427	2,379.685	1,990.843	387.811	142.316	35.751	80.656

To estimate the ultimate value of estimated claims variance, k = 7, then

$$\sigma_7^2 = min\left(\frac{\sigma_6^4}{\sigma_5^2}\right), min(\sigma_5^2, \sigma_6^2)$$

Then the variance of the estimated development factor $\widehat{Var}(\hat{f}_k)$ is calculated by using equation (5.17).

Table 6.8. Variance of Estimated Development Factor

k	$\widehat{Var}(\hat{f}_k)$
1	0.33157124
2	0.05244707
3	0.03694605
4	0.00636994
5	0.00246795
6	0.0007855
7	0.003617

The process risk is calculated by applying equation (5.20). In case that $I + 2 - i < k$, we use the upper part, otherwise, the lower part is used.

In the case of the second row, start with the case $k = I + 2 - i$ where $i = 2$ and $k = 8$, then

$\widehat{Process\ Risk}_{i,k} = C_{i,I+1-i}\hat{\sigma}_{k-1}^2$

$\widehat{Process\ Risk}_{2,8} = C_{2,7}\hat{\sigma}_7^2 = 28{,}076.202 \times 80.656 = 2{,}264{,}519$

In the case of the second row, start with the case $k = I + 2 - i$ where, $i = 3$

- In case of $k = 7$

$\widehat{Process\ Risk}_{3,7} = C_{3,6}\hat{\sigma}_6^2 = 20{,}951.018 \times 35.75833 = 749{,}016.36$

- In case of $k = 8$

$\widehat{Process\ Risk}_{i,k} = \hat{f}_7^2 \widehat{Process\ Risk}_{3,7} + C_{3,7}\hat{\sigma}_7^2$

$= 1.1024 \times 749{,}016.36 + 23{,}189.012 \times 80.656 = 2{,}780{,}557.4$

As a result, it is seen that the process risk is assessed for the third row, i = 3, going from left to right and starting with $I + 2 - i = k$, and so on until the eighth row, i = 8, is completed.

Table 6.9. The Process Risk Estimation

i/k	1	2	3	4	5	6	7	8
1								
2								**2,264,519**
3							**749,016**	2,780,557
4						**2,575,001**	3,900,066	6,601,144
5					**6,678,149**	8,871,595	10,868,176	13,207,335
6				**33,407,618**	61,163,709	85,559,812	106,062,262	132,003,229
7			**30,029,305**	63,085,003	97,687,687	129,773,861	158,981,118	193,198,875
8		**22,308,654**	54,112,893	113,678,831	176,030,979	233,847,264	286,474,879	348,130,486

Do the same as the process risk estimation, we have the parameter risk estimation table below,

Table 6.10. The Parameter Risk Estimation

i/k	1	2	3	4	5	6	7	8
1								
2								**2,851,216**
3							**344,792**	2,365,238
4						**807,948**	1,332,028	1,623,610
5					**1,888,909**	3,647,199	4,950,019	8,736,185
6				**10,403,637**	607,728,339	1,724,846,297	3,331,252,108	5,550,987,890
7			**8,351,659**	17,853,369	27,759,528	36,945,467	45,289,105	55,200,088
8		**5,257,401**	22,824,289	65,530,301	107,953,332	147,313,572	182,120,272	230,650,545

The process risk and parameter risk for accident year i are estimated by taking the square root of the values for column k = 8 (development year 8) as shown in Tables 8 and 9, respectively. In the chain ladder method, the sum of the process risk column, which represents the aggregate process risk, is estimated by taking the square root of the sum of the squares for this column's values. The sum of the values in the parameter risk column, which represents the parameter risk, is approximated by taking the square root of these values. The Chain Ladder method estimates the total standard error of the reserves by calculating the square root of the sum of the squares representing the total process risk and total parameter risk. Similar estimations are made for the total process risk and total parameter risk.

$$\sum \boldsymbol{Process\ risk} = \sqrt{\widehat{Process}\ Risk_2^2 + \widehat{Process}\ Risk_3^2 + \ldots + \widehat{Process}\ Risk_8^2}$$

$$= \sqrt{1{,}504.83188^2 + 1{,}667.500333^2 + \ldots + 13{,}899.5998^2 + 18{,}658.25516^2}$$

$$= \mathbf{26{,}423.21222}$$

$$\sum \boldsymbol{Parameter\ risk}$$

$$= \sqrt{Param\widehat{eter}\ R\imath sk_2^2 + Param\widehat{eter}\ R\imath sk_3^2 + \ldots + Param\widehat{eter}\ R\imath sk_8^2}$$

$$\sum Parameter\ risk = \sqrt{1{,}688.554338^2 + 1{,}537.932951^2 + \ldots + 15{,}187.18357^2}$$

$$= \mathbf{76{,}501.07693}$$

$$\sum Prediction\ Error = \sqrt{Proc\widehat{ess}\ R\imath sk^2 + Param\widehat{eter}\ R\imath sk^2}$$

$$= \sqrt{26{,}423.21222^2 + 76{,}501.07693^2} = \mathbf{80{,}935.7827}$$

Reserve Risk Estimates (process risk, parameter risk, and total standard error based on the accident year) are shown in Table 10:

Table 6.11. Risk Estimate for Reserves

i	Process Risk	Parameter Risk	Total
2	1,504.83188	1,688.554338	2,261.79901
3	1,667.500333	1,537.932951	2,268.43451
4	2,569.269063	1,274.209631	2,867.88314
5	3,634.189755	2,955.703803	4,684.39112
6	11,489.26579	74,504.95212	75,385.6161
7	13,899.59981	7,429.676148	15,760.6777
8	18,658.25516	15,187.18357	24,057.8684
Total	**26,423.21222**	**76,501.07693**	**80,935.7827**

The percentage of prediction error of chain ladder approach equals $\frac{s.e\,(\hat{R})}{\hat{R}} \times 100$

Thus, $\%rMSEP = \frac{80{,}935.7827}{140{,}769.56} \times 100 = 54.67\%$

3. Application of the Overdispersion Poisson model using R program

Total reserve can be calculated using R under Over – Dispersed Poisson Model.

Total Reserve = 104,769.555

It is equaled

$$PE = \sum \hat{\phi}\hat{\mu}_{ij} + \sum \hat{\mu}_{ij}{}^{2}\, Var(\hat{\eta}_{ij}) + 2\sum Cov(\hat{\eta}_{ij}, \hat{\eta}_{ik})\, \hat{\mu}_{ij}\hat{\mu}_{ik}$$

By using R, the total rMSEP = 46259.87, and total prediction error is approximated 33%.

CONCLUSION

Table 7.1. Total Reserves Comparison between using Chain Ladder vs GLMs

	Chain Ladder	GLMs
Total Reserves	140,769.56	140,769.555
rMSEP	80,935.7827	46,259.87
%rMSEP	55%	33%

Chain Ladder is one of the most common tools that actuaries use to estimate the claims reserves in insurance internationally. With the simple utility, it does not require various inputs to carry out the calculation. However, the output of the chain ladder method provides only the point estimates, if the actuaries want to determine the prediction error of it, it demands an additional calculation which is complicated in estimation. Besides, these formulas for calculating prediction error were initially developed experimentally, as shown in the article *"Afaf Antar Zohry & Mostafa Abdelghany Ahmed (2020)// The Prediction Error of the Chain Ladder Method (2020)"*, so it might not be relied upon when executing an estimate.

On the other hand, in recent decades, insurance companies begin to apply stochastic approaches more in estimating the claims reserves. The usage of stochastic methods brings to the actuaries many advantages such as more predictive information, about goodness-to-fit test statistic and the estimator error (which is included in R package). In addition, although the dispersion parameter was not mentioned in the preceding data analysis, the software R calculates it and uses it to compute the prediction error. However, compared to the chain ladder method, the ODP is much more complicated to use that we have to assume a model for the data to execute the estimation. Typically, to reach out the best result of modelling, the data is divided into two parts, training set and testing set, to measure the accuracy if the model is good enough. In other words, the ODP might provide us a more precise result, but it is more complicated than using the chain ladder method and that is why chain ladder method is more common in claims reserving than the ODP.

In this study, the total reserves of two methods are mostly the same; however, the result of prediction error using Chain Ladder is as 1.5 times as it using the ODP, about 80,935.7827 and 46,259.87, respectively. The rMSEP of chain ladder method account for approximately 55% of the total claims reserve, while it is 33% using the ODP.

In practical, when using the modern approaches, although the support of programming software allows us to generate extra information for forecasting, it is challenging for actuaries to construct or employ model which fits datasets. Therefore, actuaries tend to use the deterministic approaches in claims reserving due to its convenience and simple calculation, that we are able to calculate by hand or using Excel, although it does not provide the prediction error for the best reserve adjustment in the future. The stochastic approaches may not be as accurate as the deterministic ones, despite providing almost identical reserves and a reduced prediction error, but it offers a foundation that could become a new approach in the future.

REFERENCES

Vietnam Government (2021)// The Annual Report of Vietnam Insurance Market 2021 mof.gov.vn//theannualreportofvietnaminsurancemarket2021

Vietnam Government (2022)// Vietnam Insurance Business Law luatvietnam.vn/vietnaminsurancebusinesslaw

Vietnam National Reinsurance Corporation (2022)// Vietnam Insurance Market in the first 6 months: Recovering from the pandemic http://vinare.com.vn/vietnam_insurance_market_in_the_first_6_months

Ogutu Julie Amolo (2011)// Claims Reserving Using Overdispersed Poisson Model http://erepository.uonbi.ac.ke/OgutuJulieAmolo_ClaimsReservingusingOverdispersedPoissonModel.pdf

Afaf Antar Zohry & Mostafa Abdelghany Ahmed (2020)// The Prediction Error of the Chain Ladder Method https://www.researchgate/The_Prediction_Error_of_the_Chain_Ladder_Method

The Unearned Premium Reserve https://www.frontiersin.org/the_unearned_premium_reserves//

APPANDIX

R SCRIPTS

title: “Run Data” output: html_document date: “2023-04-08” —

Install package

```
install.packages("ChainLadder")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'

## (as 'lib' is unspecified)

library(ChainLadder)

##

## Welcome to ChainLadder version 0.2.17

## To cite package 'ChainLadder' in publications use:

##

##   Gesmann M, Murphy D, Zhang Y, Carrato A, Wuthrich M, Concina F, Dal

##   Moro E (2023). _ChainLadder: Statistical Methods and Models for

##   Claims Reserving in General Insurance_. R package version 0.2.17,

##   <https://mages.github.io/ChainLadder/>.

##

## To suppress this message use:

## suppressPackageStartupMessages(library(ChainLadder))
```

Importing data

```
library(readr)

Data <- read_csv("Data (demo 3).csv")

## Rows: 8 Columns: 8

##          ──          Column          specification
________________________________________________________

## Delimiter: ","
```

```
## dbl (8): 0, 1, 2, 3, 4, 5, 6, 7
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
View(Data)
## Warning in View(Data): unable to open display
## Error in .External2(C_dataviewer, x, title): unable to start data viewer
```

Cumulative data

```
Data.cum = incr2cum(Data)
Data.cum
##          0         1         2        3        4        5        6       7
## 1  387.4411  2111.519  7731.977 10756.82 16482.95 20720.78 22298.94 24581.7
## 2 2326.1882  6531.847  7440.785 17730.58 21800.92 24792.61 28076.20      NA
## 3 3597.0417  9616.578 13707.007 16408.85 19381.84 20951.02       NA      NA
## 4 1280.7287 10135.983 12903.442 15984.68 18093.55       NA       NA      NA
## 5 2030.6546  7422.624 12101.918 17220.12       NA       NA       NA      NA
## 6 3783.9487  9554.532 16780.642       NA       NA       NA       NA      NA
## 7 3490.5990 12619.024        NA       NA       NA       NA       NA      NA
## 8 3981.9621        NA        NA       NA       NA       NA       NA      NA

Data.mat = as.matrix(Data)
claims = as.vector(Data.mat)
n.origin = nrow(Data.mat)
n.dev = ncol(Data.mat)
origin = factor(rep(1:n.origin, n.dev))
dev = factor(rep(1:n.dev, each = n.origin))
w = data.frame(claims = claims, origin = origin, dev = dev)

model = glm(claims ~ origin + dev, family = quasipoisson(link = "log"), subset = !is.na(claims), data = w)
```

```
model.summary = summary(model)
model.summary
##
## Call:
## glm(formula = claims ~ origin + dev, family = quasipoisson(link = "log"),
##     data = w, subset = !is.na(claims))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -63.747  -19.501   -3.071   10.786   57.792
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.50277    0.35404  21.192 1.18e-15 ***
## origin2      0.23038    0.30549   0.754  0.45914
## origin3      0.03914    0.33141   0.118  0.90711
## origin4      0.03451    0.34819   0.099  0.92199
## origin5      0.20368    0.35810   0.569  0.57554
## origin6      0.54897    0.36969   1.485  0.15241
## origin7      0.70699    0.41063   1.722  0.09982 .
## origin8      0.78676    0.64546   1.219  0.23640
## dev2         0.88879    0.31124   2.856  0.00947 **
## dev3         0.64879    0.34650   1.872  0.07514 .
## dev4         0.87639    0.35667   2.457  0.02279 *
## dev5         0.63837    0.40184   1.589  0.12709
## dev6         0.38597    0.46919   0.823  0.41997
## dev7         0.17142    0.57872   0.296  0.76998
## dev8         0.23037    0.79589   0.289  0.77508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for quasipoisson family taken to be 1159.855)
##
##     Null deviance: 44874  on 35  degrees of freedom
## Residual deviance: 24513  on 21  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

Extract useful info from the model

```
coef = model$coefficients
disp = summary(model)$dispersion
cov.para = disp * model.summary$cov.unscaled

n.fut.points = length(claims[is.na(claims)])
fut.design = matrix(0, nrow = n.fut.points, ncol = length(coef))
fut.points = claims
fut.points[!is.na(claims)] = 0
fut.points[is.na(claims)] = 1:n.fut.points
for(p in 1:n.fut.points){
  fut.design[p, 1] = 1
  fut.design[p, 1 + as.numeric(origin[match(p, fut.points)]) - 1] = 1
  fut.design[p, 1 + (n.origin - 1) + as.numeric(dev[match(p, fut.points)]) - 1] =1
}

fitted.values = diag(as.vector(exp(fut.design %*% coef)))
total.reserve = sum(fitted.values)
total.reserve
## [1] 140769.6
```

Determine covariance matrix of linear predictors

```
cov.pred = fut.design %*% cov.para %*% t(fut.design)

cov.fitted = fitted.values %*% cov.pred %*% fitted.values

total.rmse = sqrt(disp * total.reserve + sum(cov.fitted))

total.predictionerror = round(100 * total.rmse/total.reserve)

total.rmse

## [1] 46259.87
```

TURNITIN CHECKING RESULT

Turnitin Originality Report

Processed on: 06-Jun-2023 19:42 +07
ID: 2108106426
Word Count: 7331
Submitted: 4

Demo 3 By Anh Thu Ngo

Similarity Index	Similarity by Source	
11%	Internet Sources:	8%
	Publications:	7%
	Student Papers:	3%

2% match (Internet from 15-Dec-2022)
http://erepository.uonbi.ac.ke/bitstream/handle/11295/14228/ABSTRACT.pdf?sequence=3

2% match (Afaf Antar Zohry, Mostafa Abdelghany Ahmed. "The Prediction Error of the Chain Ladder Method (With Application to Real Data)", International Journal of Economics and Finance, 2020)
Afaf Antar Zohry, Mostafa Abdelghany Ahmed. "The Prediction Error of the Chain Ladder Method (With Application to Real Data)", International Journal of Economics and Finance, 2020

1% match (S Karmila, S Nurrohmah, S F Sari. "Claim reserve prediction using the credibility theory for the Chain Ladder method", Journal of Physics: Conference Series, 2020)
S Karmila, S Nurrohmah, S F Sari. "Claim reserve prediction using the credibility theory for the Chain Ladder method", Journal of Physics: Conference Series, 2020

1% match (Internet from 16-Nov-2022)
https://irt.mof.gov.vn/webcenter/contentattachfile/idcplg?dDocName=MOFUCM245139&dID=247043&filename=Vietnam+Insurance+Market+2021.pdf

1% match (student papers from 03-Apr-2020)
Submitted to Oxford Brookes University on 2020-04-03

1% match (Internet from 11-Aug-2020)
http://docplayer.net/1237118-S-t-o-c-h-a-s-t-i-c-c-l-a-i-m-s-r-e-s-e-r-v-i-n-g-o-f-a-s-w-e-d-i-s-h-l-i-f-e-i-n-s-u-r-a-n-c-e-p-o-r-t-f-o-l-i-o.html

1% match (Internet from 19-Jul-2020)
https://www.actuaries.org.uk/system/files/documents/pdf/sm0201.pdf

1% match (student papers from 11-Feb-2017)
Submitted to Kwame Nkrumah University of Science and Technology on 2017-02-11

< 1% match ()
Tee, Liivika. "Stochastic reserving methods in non-life insurance", Tartu Ülikool, 2013

< 1% match ()
Boj del Val, Eva, Costa Cor, Teresa. "Provisions for claims outstanding, incurred but not reported, with generalized linear models: prediction error formulated according to calendar year", Cuadernos de Gestión, 2022

1% match (student papers from 11-Feb-2017)
Submitted to Kwame Nkrumah University of Science and Technology on 2017-02-11

< 1% match ()
Tee, Liivika. "Stochastic reserving methods in non-life insurance", Tartu Ülikool, 2013

< 1% match ()
Boj del Val, Eva, Costa Cor, Teresa. "Provisions for claims outstanding, incurred but not reported, with generalized linear models: prediction error formulated according to calendar year", Cuadernos de Gestión, 2022

< 1% match (Vajira Manathunga, Danlei Zhu. "Unearned premium risk and machine learning techniques", Frontiers in Applied Mathematics and Statistics, 2022)
Vajira Manathunga, Danlei Zhu. "Unearned premium risk and machine learning techniques", Frontiers in Applied Mathematics and Statistics, 2022

< 1% match (Internet from 17-Dec-2018)
https://res.mdpi.com/risks/risks-06-00139/article_deploy/risks-06-00139-v3.pdf?attachment=1&filename=

< 1% match (Internet from 25-Jan-2023)
https://signifiance.com/wireless-monitoring-of-water-level-using-radio-frequency-identification-6275.html

< 1% match (student papers from 07-Dec-2006)
Submitted to City University on 2006-12-07

< 1% match (Internet from 07-May-2010)
http://www.cass.city.ac.uk/facact/progression/actman/Elena_Pantzopoulou_MSc_dissertation.pdf

< 1% match (P.D. England, R.J. Verrall. "Stochastic Claims Reserving in General Insurance", British Actuarial Journal, 2011)
P.D. England, R.J. Verrall. "Stochastic Claims Reserving in General Insurance", British Actuarial Journal, 2011