University of Nairobi

College of Biological and Physical Sciences

School of Mathematics

Analysis of Variability in IBNR Estimates with Run off Triangles

By Antony Gitonga

MASTERS OF SCIENCE IN ACTUARIAL SCIENCE

A dissertation submitted to the graduate school in partial fulfillment of the requirements for the award of Master of Science in Actuarial Science.

Year, 2015

Declaration of Authorship

This dissertation, as presented in this document, is my original work and has not been replicated, extracted or copied from any other published or unpublished sources.

Antony Gitonga Wambugu

Signature .... Date ....

I56/70005/2013

This project was supervised and approved as original work by:

Prof. Patrick G.O. Weke

Signature .... Date ....

Dr Joseph Ivivi Mwaniki

Signature .... Date ....

Abstract

The many uncertainties involved in the payment of losses makes the estimation of the required reserves more difficult. Yet, some of the existing methods, such as the popular chain-ladder, are simple to apply. However, it has become evident that there is a need for better ways not only to estimate the reserves, but also to obtain some measures of their variability. The chain-ladder is used as a benchmark in this study due to its generalized use and ease of application. Although this facilitates comparisons between different methods, studies have shown that different classes of insurance present different development patterns hence the need to apply a variety of methods.

This dissertation aims at presenting different loss reserving models both deterministic and stochastic and compare the variability in the models. The objective will be to develop and implement a loss reserving model that combines both deterministic and stochastic methods to estimate reserve provisions for different classes in general insurance. In this study we also present Bayesian method to model both claim frequencies and severity using some well defined assumptions and to use the resulting predictive distributions to estimate loss reserves, allowing for negative values. In this study we assume that the expected loss payments depends upon unknown parameters that determine the expected loss ratio for each accident year and the expected payment lag. The distribution of outcomes is given by a collective risk model in which the expected claim severity increases with the settlement lag. The claim count distribution is given by a Poisson distribution with its mean determined by dividing the expected loss by the expected claim severity. The parameter that describe the posterior distribution are calculated using a Monte Carlo simulation algorithm. Models back testing with real life data have shown that in some classes of insurance the actual and expected estimates vary significantly there by discrediting the models and hence the need to compare different models and study the variability presented.

Acknowledgements

Foremost, I am a highly grateful to the almighty Lord for seeing me through my academic years in the university. I swept along in the turbulence of great events and storms of life to finally come up with this dissertation but He never abandoned me. I have taken efforts in this dissertation. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere appreciation to all of them. Leading on this list is the intellectual debt that I owe to my lecturers at the School of Mathematics.

It is inevitable that when one embarks on a long-term dissertation whose primary objective is to put forward an alternative to the traditional approach, the road would often be arduous and the going would get very toilsome and solitary at times. During periods of forlorn hope, a pat on the back or a vote of confidence from colleagues whose views one values, is of paramount importance in keeping one's enthusiasm going.

It gives me great pleasure to thank Professor P.G.O Weke and Dr Joseph Ivivi Mwaniki for their guidance and constant supervision as well as providing me with necessary information regarding my dissertation and also for their support in completing this dissertation. Your experience and knowledge in the field of loss reserving was evident in my work. I should certainly say I had a great team of supervisors.

I would like to express my gratitude towards my parents and members of the University of Nairobi for their kind co-operation and encouragement which helped me in completion of this dissertation. I would like to express my special gratitude and thanks to industry persons for giving me such attention and time during data collection.

Special thanks are due to David Muigai Wanjohi for providing important and essential input at crucial points especially in the R models, your kindness and support is highly appreciated.

My thanks and appreciations also go to my colleagues in class and people who have willingly helped me out with their abilities. My sincerest apologies to any colleagues and students whose help to me I have inadvertently failed to acknowledge. I hope to have an opportunity to correct the omission in another work.

Contents

List of Figures

List of Tables

Abbreviations

ODP	**O**ver **D**espersed **P**oisson
CL	**C**hain **L**adder
CCL	**C**lassical **C**hain **L**adder
CLM	**C**hain **L**adder **M**odel
ANOVA	**A**nnalysis **O**f **V**ariance
IBNyR	**I**ccured **B**ut **N**ot yet **R**eported
IBNeR	**I**ccured **B**ut **N**ot enough **R**eported
MSEP	**M**ean **S**quare **E**rror of **P**rediction
MCMC	**M**arkov **C**hain **M**onte **C**arlo
FFT	**F**ast **F**ourier **T**ransform
DFT	**D**iscrete **F**ourier **T**ransform
PDF	**P**robability **D**ensity **F**unction
ELR	**E**xpected **L**oss **R**atio

Dedicated to my Dear Mum and Dad

Chapter 1

Introduction

1.1 Background Information

While a certain degree of variation in the reserves estimate is to be expected, it's the duty of the financial actuary to calculate as accurately as possible the amount of variation to be expected. The achievement of this goal necessitates an understanding of the difference between the process variance, measured by the standard deviation of the underlying claim case estimates and payment process, and the method variance, which is a characteristic of the measurement method. Due to the heuristic nature of most of the calculation methods used by actuaries, a certain amount of method variance is to be expected. However, a critical evaluation of the most common methods used shows that the techniques used yield for the most part a much higher error due to methodology than is necessary. One of the basic properties of variance reveals why some common reserve calculation methods result in a high variance.

This is the key property why variances are additive under additions but increase polynomially under multiplication. The variance of the sum of a collection of random variables is, in general, the sum of the variances of the individual variables, while any multiplication process applied to a random variable increases variance in proportion to the square of the multiplier assuming the covariances are negative. To keep the error variance to a minimum, one should seek to use methods that rely on the summation of data, and avoid methods that use or result in multiplicative factors.

Many studies have shown that there has been an increasing recognition in consideration of the random nature of the insurance loss processes that leads to better predictions of ultimate losses. Some of the studies that led to this recognition include Stanard (1985) and Barnett and Zehnwirth (2000). The other approach that has gained popularity is that of recognizing outside information in the formulas that predict ultimate losses, Bornhuetter and Ferguson (1972) represents one of the early papers exemplifying this approach. More recently, papers by Meyers (2007) and Verrall (2007) have combined these two approaches with a Bayesian methodology. In this study we advance the development of the approach started by Meyers(2007) and draw from the methodology described by Verrall(2007).

In his work Meyer(2007) accomplished to make predictions of the distribution of future losses and validated the predictions on subsequent reported losses. The analysis of loss reserving data is based on a run-off triangle. The runoff triangle is a matrix which contains claims data, where each row corresponds to the year of an accident (the so-called 'Accident Year'), and each column corresponds to the delay, or the number of years between the accident year and the year in which the claim was made, the claim was paid or outstanding (the so-called "Development Year"). The matrix contains data in the upper left triangle, and the aim of this study is to estimate future claims, to be filled in the empty lower right triangle. The entries of each diagonal therefore correspond to one single calendar period.

As indicated the chain-ladder method is the most popular method of loss reserving, however extensions to the chain-ladder method based on the stochastic nature of the losses have started gaining popularity in the recent past. To cite but a few broadly used models are the log-linear and the log-normal ANOVA-type models (Verrall, 2000). Others include the dynamic state space models (Ntzoufras and Dellaportas, 2002), as an extension to the log-normal models and Bayesian models using Markov chain Monte Carlo techniques (Makov, 2001; Scollnik, 2001). Methods based on simulation include, Stanard (1985) and Pentikainen and Rantala (1995) that describe methods of simulating random loss triangles in which data of the upper triangle are simulated and used to estimate the simulated losses, that is, the missing values of the lower triangle.

Settlement of claims is one of the primary objective of an insurance company. Many people take up insurance so that in return for the premiums paid, the insurance company can accept the liability to make a monetary payment to the insured or reinstate them to the position they were before the occurrence of a specified event within a specified period of time. The insurer's liability to pay a claim materializes on occurrence of the insured risk. However, there are many factors that can lead to considerable delays between occurrence, notification and payment. The insured risk itself may not occur in a single phase and may not even be recognized as claimable events until after many years. In other cases the legal liability of the insured to third parties may not always be clear-cut, and there may be considerable delays before the insurer or the court decides that liability exists. The quantum of damages may be impossible to determine until some period of time has elapsed since occurrence of the event leading to further delays. Estimating of outstanding liability in non-life insurance policies is highly speculative, details of methodologies for making such predictions are contained in a comprehensive and highly detailed survey conducted by Taylor (1986). The research sponsored by Society of actuaries on comparison of incurred but not reported methods by (Cabe Chadick et al, 2009) found out that seasonality of the claims can have a material impact on the mean error. For example, seasonality exhibited by claim concentration in early calendar years resulted in typical material positive errors. Seasonality exhibited by claims concentration in later calendar year depicted large deductibles as a result of recoveries. With this in mind, development methods can be employed with completion factors and projection factors that result in varying degrees of conservatism. It is therefore essential to estimate variability in the estimates provided by the various method and consider averaging and smoothing techniques that will typically create a series of potentially usable completion factors. The use of projection factors or other techniques to estimate the most recent years incurred also involves selection of various trends. Implicit margin can be created by conservatism in the choice of completion and projection factors.

A significant accomplishment of the Meyers (2007) paper cited above was that it made predictions of the distribution of future losses of real insurers, and successfully validated these predictions on subsequent reported losses. To do this, it was necessary to draw upon proprietary data that, while generally available, comes at a price. While this made a good case that the underlying model is realistic, it tended to inhibit future research on this methodology.

Recognizing lack of comparative information on techniques used for estimating incurred but not reported (IBNR) and reported but not enough reserves, presented an opportunity to assess the accuracy of commonly used IBNR estimation methods over a wide range of scenarios. To conduct the study, a stochastic model was constructed to compare and score estimates produced by the loss reserving methods that were selected for testing. The testing was done over a significant number of iterations and alternative business situations.

This dissertation follows Arjas (1989). It presents a mathematical framework for claims reserving and formulate the claims-reserving problem in the language of stochastic processes. All actions relating to a claim are listed in the order of their notification at the insurance company. From a statistical point of view, this makes perfect sense; however, from an accounting point of view, one should preferably list the claims according to their occurrence or accident date as suggested in Norberg (1993,1999)

The primary traditional loss reserving techniques used to determine loss reserve variability are based on the inherent assumption that this variability arises from a single factor. The Mack (1993), bootstrap (England and Verrall, 1999) and simulation approaches all measure the variability in historical loss development patterns and use the derived measures to determine a distribution of loss development factors and to calculate loss reserve ranges. The Mack method uses chain ladder link ratios to obtain a mean value of the loss reserve. The statistical features of chain ladder link ratios are used to derive formulas for the process and parameter risk of the loss reserve. These two components are combined to obtain a standard deviation measure. A log normal or normal distribution is then fit to the mean and standard deviation to obtain a distribution of loss reserves.

The bootstrap method calculates a triangle of cumulative fitted values by working backwards on the data triangle using chain ladder link ratios. The residuals between the actual and fitted values are randomly arranged to obtain a new triangle of data that has the same statistical characteristics as the actual data. New link ratios are obtained from the sampled triangle to calculate a point estimate of the loss reserve. The sampling of residuals is performed many times to obtain a distribution of the loss reserve. The statistical characteristics of the sampled data are then used to derive the parameter variance and the total standard deviation of the loss reserve. The simulation method calculates the mean and standard deviation of the link ratios, which are then fitted to a log normal or normal distribution. Link ratios are simulated based on those distributions to obtain a point estimate of the loss reserve. Then Monte Carlo simulation is performed to obtain a distribution of the loss reserve.

Although the Mack, bootstrap and simulation approaches are able to calculate loss reserve variability based on a single factor, there are many situations in which these methods would fail to produce an accurate and meaningful distribution. First, there are multiple factors that impact the variability of loss development. A change in inflation can affect loss payment patterns, impacting loss development. Factors within a company, such as claim department staffing, selected year end cut-off dates, loss settlement practices and the experience of claim department personnel, will all affect loss payment patterns and loss development factors.

In addition, random factors that would affect when claims occur during the year and, more importantly, when the outlier large claims occur, will impact the variability of loss development factors. These factors would impact loss development in different directions based on the calendar year, accident year and development year, and are often highly uncorrelated. For example, inflation is a calendar year effect, impacting all claims that are open in a calendar year. A high inflation rate in one calendar year would affect the loss development factors along the diagonal on an accident year basis. A small change in the inflation rate in either direction is likely to produce loss development patterns in aggregate that are significantly different from the historical patterns. The approaches aggregate all of these effects into a single variability measure.

Usually, different methods and differently aggregated data sets lead to very different results. It is only through experience that one is able to tell which is an accurate or good estimate for future liabilities for a specific data set, and which method applies to which data set. Often there are many phenomena in the data that first need to be understood before applying any claims reserving method. Especially in direct insurance, the understanding of the data can even go down to single claims and to the personal knowledge of the managing claims adjusters. With this in mind, one is able to describe different methods that can be used to estimate loss reserves, but only practical experience will tell which of the methods should be applied in any particular situation. That is, the focus of this dissertation is on the mathematical description of relevant models and we will derive various properties of these models and their variations.

The question of an appropriate model choice for a specific data set is only partially treated here. In fact, the model choice is probably one of the most difficult questions in any application in practice. Moreover, the claims reserving literature on the topic of choosing a model is fairly limited – for example, for the chain-ladder method certain aspects are considered in Barnett and Zehnwirth (2000) and Venter (1998). In classical claims reserving literature, claims reserving is often understood to be providing a best estimate to the outstanding loss liabilities. Providing a best estimate means that one applies an algorithm that gives a number or amount. In recent years, especially under new solvency regimes, one is also interested in the development of adverse claims reserves, and estimating potential losses that may occur in the future in these best estimate reserves. Such questions require stochastic claims reserving models that can justify the claims reserving algorithms and quantify the uncertainties in these algorithms. From this point of view one should be aware of the fact that stochastic claims reserving models do not provide solutions where deterministic algorithms fail, they rather quantify the uncertainties in deterministic claims reserving algorithms using appropriate stochastic models.

The basic chain-ladder is used as a benchmark, due to its generalized use and ease of application. This facilitates comparison between methods. However, in this dissertation our aim in not to develop methods that provide results close to those of the chain-ladder method. Rather, we aim at studying results of different methods from model that utilize both claim intensity and severity using some common assumptions and to use the resulting predictive variability to adjust estimated loss reserves, especially with data allowing for negative values. Particular one would be concerned with the situation when there are negative values in the development triangle of the incremental claim amounts or cumulative claim amounts.

The purpose of this study is not to add or make recommendations to any of the existing methodologies but rather to return to the grass roots of the subject and exploring more carefully the statistical variations in the estimates for the classical chain-ladder and compare with related techniques, both deterministic and stochastic.

1.2 Problem Statement

The chain ladder method is intuitively appealing and simple to calculate and which often give reasonable results these attributes makes the method popular. Because of these strengths, there may have been some reluctance to adopting alternative methods of estimating outstanding liabilities. However, the chain ladder method comes with its shortcomings:

1. It is a purely multiplicative approach, the estimates for each origin period is formed

by multiplying the most recent value in each value in each origin period by development factor. For the long-tailed business the most recent values could be zero or negative leading to underestimation of the ultimate claims. On the other hand if the recent are unusually large perhaps because of some large claims, the development factor may overstate the ultimate claims for the period.

2. The link ratios must be stable across the origin periods for the method to produce sensible results and such stability is rare. In particular, the method is vulnerable to changes in pace of claims settlement, especially when applied to claims paid.

3. Special adjustment to the data using information from the claims settled can help to deal with changes in claims settlement however this kind of information is not always available.

4. Mathematically, the chain ladder method works by calculating a series of linear regression of the form.

$$y = Ax + \varepsilon \tag{1.1}$$

   Where x represents the values in column N, y represents the value in column $N + 1$, A is an estimate parameter and ε is a random error. However there is no particular reason why the regression cannot take another form, hence motivation to explore other models and

1.2.1 Objectives

The general objective of this study is to formulate a collective risk model aimed at analyzing variability in loss reserves by incorporating information other than the claims amounts. This is achieved by smoothing loss development factors and adjusting variation in average claim reserves by accident year. This has addressed in two fold by looking at the process variations and parameter estimation variations.

The specific objectives are:

1. To estimate parameters of the posterior distribution given the prior and likelihood distribution using a Bayesian Markov Chain Monte Carlo Algorithm.

2. To compare different loss reserving method with an aim of bridge the gap between the stochastic underpinnings of the chain ladder method and its implementation in practice.

3. Restate the reserves for each accident year using the characteristics of the open claims for the accident year under evaluation, this eliminates inconsistency in the subsequent loss development factors.

4. Apply the Bayesian approach in estimation of parameters as well as evaluation of prediction uncertainty used in estimating reserves.

5. To bridge the gap between the stochastic underpinnings of the chain ladder method and its implementation in practice especially when link ratios are selected based on judgement.

6. Provide a model whose underlying assumptions and actuarial inputs are testable within a rigorously-defined statistical framework.

1.2.2 Significance of Study

Claims liability forms a significant proportion of insurance total liability, if claims reserves are understated this would lead to collapse of a company. It is therefore not only necessary to build premium reserves for future exposures, but on the other hand one would need also to build claims reserves for unsettled claims of past exposures. There are two different types of claims reserves for past exposures:

1. IBNyR reserves (Incurred But Not yet Reported): Claims reserves for claims which have occurred, but which have not been reported by the end of the valuation year (i.e. the reporting delay laps into the next accounting years).

2. IBNeR reserves (Incurred But Not enough Reported) or RBNS (reported but not settled): Claims reserves for claims which have been reported but have not been settled yet, i.e. still expect payments in the future, which need to be financed by the already earned premium.

This study is geared toward estimating the ultimate reserves and measuring variability by constructing upper limits at an adequate confidence level using both deterministic and stochastic approaches. The reserving methods used in practice are frequently deterministic. For instance, the claims reserve is often obtained by case estimation of individual claims. A popular statistical method is the chain-ladder method, which originally was deterministic. In most cases adjustments are applied, for example projection of payments into the future can sometimes be done by extrapolating by eye, hence the need to measure the standard error in the reserves estimates.

Regulatory authorities in different parts of the world especially developing countries have also set guidelines on non-life insurance reserving to safeguard policyholders from losing out on claims payments when insurance companies get ruined, in Kenya specifically this is an establishment by Act of Parliament Chapter 487, this was a major motivation to carry out this study.

Chapter 2

Literature Review

The linage of chain ladder can be traced to the mid - 60's and the name refers to the chaining of a sequence of age-to-age development factors into a ladder of factors by which one can climb from the observations to date to the predicted ultimate claim cost. The chain-ladder was originally deterministic, but in order to assess the variability of the estimate it has been developed into a stochastic method. Taylor (2000) presents different derivations of the chain-ladder procedure; one of them is deterministic while another one is based on the assumption that the incremental observations are Poisson distributed. Verrall (2000) provides several models which under maximum likelihood estimation reproduce the chain-ladder reserve estimate.

In the recent past a large variety of methods of loss reserving based on run-off triangles have been proposed. In each of these methods, it is assumed that all claims are settled within a fixed number of development years and that the development of incremental or cumulative losses from the same number of accident years is known up to the present calendar year such that the losses can be represented in a run-off triangle.

The most venerable and most famous of these methods are certainly the chain-ladder method and the Bornhuetter-Ferguson method. The basic idea of the chain-ladder method was already known to Tarbell (1934) while the Bornhuetter-Ferguson method was first described almost four decades later in the paper by Bornhuetter and Ferguson (1972) in his paper he recognized outside information into the formulas to better predict the ultimate losses.

Meyers and Verral(2007) combined the chain ladder and the Bornhuetter and Ferguson approach with a bayesian methodology this was a significant accomplishment in that it made predictions of the distribution of future losses, and successfully validated these predictions on subsequent reported losses.

In 2007 and 2008, the General Insurance Reserving Oversight Committee, under the Institute of Actuaries in the U.K commissioned a study to test the model proposed by England and Verrall (2002). The findings of the study were that even under ideal conditions the probabilities of extreme results could be under-stated using the Mack and the over dispersed poison bootstrap models. The detailed model proposed by Meyer (2007) was also tested with UK motor data they fitted the model on the data excluding that of the most recent diagonal, and then simulated distributions of the next diagonal to compared with the actual diagonal. The model allowed for the error in parameter selection to help overcome some of the underestimation of risk seen in the Mack(2000 and 2007) and Over Desperesed Poisson bootstrap models. However, this was not a guarantee of correctly predicting the underlying distribution and the variability in the reserves was still evident.

Meyer etal (2011) performed a retrospective testing of stochastic loss reserves on the over dispersed poison bootstrap model as well as a hierarchical Bayesian model, using commercial auto liability data from U.S. annual statements for reserves as of December 2007. The first was to test the modelled distribution of each projected incremental loss for a single insurer. The second was to test the modelled distribution of the total reserve for many insurers. The findings were that in case there are environmental changes that cannot be identified by the model under study then one cannot solely rely on stochastic loss reserve models to manage the reserve risk and it would be desirable to develop other risk management strategies to deal with the unforeseen environmental changes.

Stephen P. D'Arcy(2008) traditional loss reserving techniques measure variability based on a single factor on historical loss development factors and loss reserves ranges. This limits the calculated variability to what occurred during the experience period. However, there are multiple factors that impact the variability of loss development and they are not always stationary. Inflation is a key element in loss development. The traditional approach for determining loss reserve variability is reasonable as long as inflation is relatively constant. If inflation and its volatility, were to change, actual loss reserve variability would turn out to be higher or lower than expected based on the traditional approaches. Mack and bootstrap methods use only information from the historical loss development patterns and assume future development would follow those patterns. Simulation method allow for customized inputs for simulating link ratios, but an increase or decrease in the mean or the standard deviation compared to that obtained from the historical data is difficult to justify, or properly quantify, on a one-factor basis. The objective was to accurately estimate the inflation variability and the residual reserve variability using a two factor model. The model accommodated shifts in inflation as well as residual standard errors.

The greater predictive power in calculating loss reserve variability by using multiple uncorrelated factors has been recognized by the increasingly popular use of statistical modeling techniques in loss reserving. The statistical nature of the modeling framework also allows separation of parameter uncertainty and process variability Barnett and Zehnwirth (1999). These parameters are not easy to extract from the data and sometimes their introduction to the statistical model ends up distorting the framework hence producing an incorrect distribution of the reserves, hence statistical modeling techniques limit the calculated variability to what occurred during the experience period to a certain degree. To overcome such limitations one would need to accurately extract information from the data trends, one would also need to have some flexibility in introducing variability that is different from what occurred during the experience period, sound actuarial judgment and ability to produce a reasonable distribution of the reserves.

One of the major advantages of the classical IBNR claims reserving methods, like the chain-ladder, Cape Cod and Bornhuetter-Ferguson methods or credibility like methods by Mack(2000), Hürlimann(2005)), is their distribution-free validity. However, the insurance industry is slowly changing and the complexity in term of framework is increasing becoming evidnet, there is an accrued interest to know more about the standard deviation and the higher percentile values. Therefore, attempts to model adequately not only the mean of the IBNR claims reserves but also its full distribution have the potential to retain more attention from both a theoretical and practical viewpoint. Early developments in this area include work by Bühlmann et al.(1980), and Hertig(1985).

The present approach is inspired from Mack(1997), which proposes distribution dependent IBNR claims reserving methods, in particular a cross-classified parametric method of multiplicative type.

Studies on the statistical basis of the chain-ladder method, with a focus on the distributional assumptions of the aggregate data and the use of generalized linear models have been advanced and the recent works focused on, over-dispersed Poisson (ODP) model Renshaw and Verrall (1998), negative binomial model Verrall (2000), Mack's model

Mack (1993), and log-normal model Kremer (1982).

In recent years, the understanding of the chain-ladder technique has been further developed. Kuang et al. (2008, 2011) extends the chain-ladder model with a calendar effect and uses time-series analysis to forecast this effect. Verrall et al. (2010) and Martinez Miranda et al. (2011, 2012) proposes a double chain-ladder method that simultaneously uses a triangle of paid losses and a triangle of incurred claim counts. Martinez-Miranda et al. (2013) reformulates the triangular data as a histogram and proposes a continuous chain-ladder model through the use of a kernel smoother.

Some studies done have also addressed the limitations of the chain ladder method, notably, over-parametrization of the chain-ladder method Wright (1990), unstable predictions for recent accident years Bornhuetter and Ferguson (1972), problems with the presence of zero or negative cells in run-off triangles Kunkler (2004), difficulties in separating assessment of RBNS and IBNR claims Schnieper (1991), Liu and Verrall (2009), difficulties in the simultaneous use of incurred and paid claims Quarg and Mack (2008). At the heart of the limitations of such models is the small sample size and the inability to use any information about the individual claims. These issues are derived from the inherent nature of the use of aggregate data and thus generally cannot be addressed by any adjustments within the framework of chain ladder models. The observed data in a runoff triangle is typically small, leading to a prediction error that and could be very large England and Verrall (2002).

A run-off triangle is essentially a summary of the underlying individual ideally homogeneous claims data. If claims are believed to be heterogeneous, then they are often segmented by certain characteristics usually discrete and compiled into multiple triangles. In this respect, individual claim level information is used to segment the data before the modeling phase. Nevertheless, under circumstances when the heterogeneity of claims is due to many characteristics including continuous characteristics, or the characteristics that contribute to the heterogeneity is not clear, or the number of claims in the portfolio is big enough, the segmentation may not be practical and the incorporation of individual claim level information would be desirable in the reserving model.

England and Verrall (2002) questioned the continuing use of aggregate data when the underlying extensive micro-level information is available and the computation is feasible. Parodi (2012) points out the misalignment of rate-making and reserving: they both value the same risk but the former is based on individual data whereas the latter is based on aggregate data.

Run-off triangles are used in general insurance to forecast future claim numbers and amounts. Usually run-off triangles arise in non-life insurance where it may take some time to establish the full extent of the claim before the final payment can be made. Run off triangles attribute the claims to the year in which the accident occurred. The idea is to estimate how much of each class of business an insurance company is liable to pay in claims so that it can make adequate provisions. It is clear that although the exact figure for total claims is unknown because of delays in the claim settlements, provisions can be made for future claims settlements with as much confidence and accuracy as possible.

In any claim event there may be delays in between the occurrence of the claim event and the date on which the claim is reported to the insurer (reporting delay) and another delay between the reporting date and the date on which the claim loss is finally settled (settlement delay).

The first step in creating the claims loss settlement run-off triangle is to group the claims loss settlement amounts by the year in which the associated claims events occurred; this is called the claims occurrence year. Typically, claims losses settled for each claims occurrence year are not paid on one date but rather over a number of years (or time periods). This leads to development periods or delay lags measured from the accident date.

Chapter 3

Methodology

The approach of this project was to consider the claims incurred in an insurance company in two fold 1.) the claims paid during a given accounting period and 2.) the case estimates as at a given accounting period. In practice this data is presented based on transactions. Data preparation and structuring for use in a reserving exercise is assumed to be basic and done prior to beginning this study. We however assume that the accident date, notification dates and the pay dates are available in the data. The case estimates are assumed have been classified per accounting period and from this we are able to lag the data per the accident periods for triangulation.

3.1 Incremental and Cumulative Losses

The incremental triangle can be based on paid claims or incurred claims depending on data availability. The ultimate result should be a matrix of incremental aggregated payments or payments plus outstanding claims located in the upper triangle and the lower triangle consisting in missing or zero values. Depending on the model data requirements the triangles can also be constructed to contain incremental number of reported claims. For each claims occurrence year the incremental claims loss settled for a particular development year is the amount settled in that development year. The next step employed is to develop cumulative incurred or paid claims triangles by accident year, that is the total amount settled up to that development year. Cells are summed forward to obtain cumulative incurred and paid claims. The result is a progression of payments toward ultimate payout for a given accident year.

3.2 Basic Chain Ladder

The chain ladder method is the most popular method for estimating outstanding claims reserves. The main reason for this is its simplicity and the fact that it is distribution-free. In most cases the results from the straight forward basic chain ladder method are used as benchmarks. We will relax the assumptions to this impression because it is clear that the chain ladder algorithm has far-reaching implications. These implications also allow it to measure the variability of chain ladder reserve estimates and with the help of this measure it is possible to construct a confidence interval for the estimated ultimate claims amount and for the estimated reserves.

Traditional methodologies such as the chain ladder, though not necessarily stochastic based are robust and when used as intended tend to be a holistic approach to estimating reserves. Using such approaches may lead one to develop a gut feel for the uncertainty in his or her estimates, but may not necessarily be able to quantify that gut feel. Conversely, more modern stochastic methods bring with them quantification of the volatility of the forecasts, but usually conditioned on a specific set of assumptions.

The chain-ladder method assumes that factors such as inflation, changes in portfolio mix and changes in rate of settlement of claims can be ignored and assumes the form,

$$C_{w,d} = X_w R_d + e_{w,d} \tag{3.1}$$

and with known parameters taking the following observations,

$$C_{w,d} = X_w R_d + e_{w,d} \tag{3.2}$$

where X_w is the ultimate total cost of claims in the period of origin w and R_d is the proportion of total claims incurred by the end of development period d. $C_{w,d}$ is the cumulative amount of claims incurred to the end of the period d. The claims development pattern is assumed to be constant.

The chain-ladder method operates on cumulative observations, assume we have a set of incremental claims data.

$$\{X_{w,d} : w = 1, ...n; d....n - i + 1\} \tag{3.3}$$

Where w indicates the accident year or the notification year and in some cases the underwriting year.

d indicated the delay or the lag typically measured in years but could be adjusted to quarterly or monthly.

Cumulative claims are derived from the incremental claims:

$$C_{w,d} = \sum_{k=1}^{d} X_{w,k} \tag{3.4}$$

Incremental claims can be obtained from the cumulative claims by letting.

$$X_{w,d} = \begin{cases} C_{w,d} & \text{if } d = 0, \\ \\ C_{w,d} - C_{w,d-1} & \text{elsewhere.} \end{cases} \tag{3.5}$$

The development factor or link ratio method is the oldest tradition approach to estimating of technical liabilities. Assume that the development factors are denoted by:

$$\Lambda_d : d = 2, ...n \tag{3.6}$$

The chain-ladder technique estimates the development factors as:

$$\Lambda_d = \frac{\sum_{w=1}^{n-d+1} C_{w,d}}{\sum_{w=1}^{n-d+1} C_{w,d-1}} \tag{3.7}$$

and

$$r_j = \prod_{k=1}^{d} \Lambda_k \tag{3.8}$$

This method assumes that the claims development pattern is stable between each year of origin and that the future claims inflation follows the past trends. In cases of volatile claims inflation the method fails to provide reliable estimates. The purpose of loss reserving is to predict the ultimate losses $C_{w,n}$ and accident year reserves $C_{w,n} - C_{w,n-w}$. The development factors obtained above are applied to the latest cumulative claims in each accident year (row) $C_{w,n-w+1}$ to produce forecasts of future values of cumulative claims.

$$C_{w,n-w+2} = C_{w,n-w+1}\Lambda_{n-w+2},$$
$$C_{w,k} = C_{w,k-1}\lambda_k, \tag{3.9}$$
$$k = n-w+3, n-w+4, ..., n$$

A development pattern of quotas consists of parameters $\gamma_0, \gamma_1, \gamma_2, ..., \gamma_n$ with

$$\gamma_k = \frac{E\left[C_{w,k}\right]}{E\left[C_{w,n}\right]} \tag{3.10}$$

The quotas represent the percentage of claims reported. A development pattern of factors consists of parameters $\psi_1, \psi_2, ..., \psi_n$ with

$$\psi_k = \frac{E\left[C_{w,k}\right]}{E\left[C_{w,k-1}\right]} \tag{3.11}$$

These parameters are the development factors and simply represent the age-to-age factors.

The basic chain ladder is easy to explain and has been the subject of much literature. It was not originally grounded in mathematical or statistical theory; though in the recent past scholars have done work to set it into a statistical framework. In addition, it is known to be quite volatile, particularly for less mature exposure periods.

Thus, the chain-ladder technique, in its simplest form, consists of a way of obtaining forecasts of ultimate claims only. Ultimate is interpreted as the latest delay year so far observed, and does not include any tail factors. From a statistical viewpoint, given a point estimate, the natural next step is to develop estimates of the likely variability in the outcome so that assessments can be made, for example, of whether extra reserves should be held for prudence, over and above the predicted values. In this respect, the measure of variability commonly used is the prediction error is defined as the standard deviation of the distribution of possible reserve outcomes. It is also desirable to take account of other factors, such as the possibility of unforeseen events occurring which might increase the uncertainty, but which are difficult to model.

Mack (1993) also consider a distribution free chain ladder model with cumulative claims $C_{w,d}$ of different accident years w being assumed to be independent. He observed that there exists positive development factors $\Lambda_1, \Lambda_2, ..., \Lambda_{d-1}$ and positive parameters for the standard error $\sigma_1^2, \sigma_2^2..., \sigma_{d-1}$ such that for all $0 \leq w \leq n$ and all $1 \leq d \leq n$.

$$\begin{aligned} E\left[C_{w,d} \mid C_{w,0}, C_{w,1}.........C_{w,d-1}\right] &= \Lambda_{d-1} \times C_{w,d-1} \\ Var\left[C_{w,d} \mid C_{w,0}, C_{w,1}.........C_{w,d-1}\right] &= \sigma_{d-1}^2 \times C_{w,d-1} \end{aligned} \tag{3.12}$$

The ultimate claims amount can be estimated as:

$$C_{w,k}^{CL} = C_{w,k-1} \times \prod_{k=1}^{n-1} \Lambda_k \tag{3.13}$$

The Chain Ladder reserve for accident year w is given by:

$$\overline{R}_w = C_{w,n-w+1} \times \left[\prod_{k=1}^{n-1} \Lambda_k - 1\right] \tag{3.14}$$

The CL development pattern is estimated by

$$\overline{\beta}_d^{CL} = \left[\prod_{k=1}^{n-1} \Lambda_k\right] \tag{3.15}$$

Where $\overline{\beta}_d^{CL}$ is equal to 1.

That is, β_d^{CL} is an estimate of the proportion of the expected ultimate claim which emerges up to development year d. From the assumption of cumulative claims independence over the lags.

$$E\left[C_{w,n} \mid D_I\right] = E\left[C_{w,n} \mid C_{w,1}, C_{w,2}, ..., C_{w,n-w}\right] = C_{w,n-w} \times \left[\prod_{k=1}^{n-1} \Lambda_k\right] \tag{3.16}$$

Mack(1993) showed that under assumptions of independence of cumulative claims data the development factors estimators $\overline{\Lambda}_w$ are uncorrelated and unbiased for Λ_w. This is so even when $\overline{\Lambda}_{w-1}$ and $\overline{\Lambda}_w$ depend on the same data $C_{n-d-1,d}$.

The estimate for variance is:

$$\sigma_d^2 = \frac{1}{n-d-1} \sum_{w=0}^{n-d-1} C_{w,d} \left(\frac{C_{w,d+1}}{C_{w,d}} - \dot{\Lambda}_d \right)^2 \tag{3.17}$$

3.3 Bornhuetter-Ferguson

Another traditional approach is the Bornhuetter-Ferguson method. Rather than being multiplicative and leveraged for less mature exposure periods, this method is additive and tends to be more stable. However, the method needs both an estimate of the loss emergence or development (as does the development factor method) but as well as a prior estimate of ultimate losses for each exposure year. This latter requirement can be overcome using a variant approach known as the Cape Cod method. In the cape cod method one estimates the initial "seed" by using an approach equivalent to the development factor projection method.

Let $C_{w,d}$ denote the cumulative amount of claims incurred of accident year w after d years of development and that the following holds.

$1 \leq w$ and $d \leq n$.

Let ν_w denote the earned premiums for accident year w.

$C_{w,n+1-w}$ is the current known claim amount for accident year w, if U_w is the ultimate claim amount of accident year w, then the claims reserve for accident year w can be represented as:

$$R_w = U_w - C_{w,n+1-w} \tag{3.18}$$

The underlying ultimate loss ratio for accident year w is.

$$q_w = \frac{U_w}{\nu_w} \tag{3.19}$$

Let F_d denote the lag factor for lag d then we can estimate the reserve as:

$$R_w = \nu_w \times \left(1 - \frac{1}{F_w}\right) \times q_w \tag{3.20}$$

The chain ladders reserve strongly depends on the current amount $C_{w,n+1-w}$ which can lead to a negative reserve or a zero reserve for accident years where no claims are paid or reported which is not unusual in excess of loss reinsurance.

The BF reserve estimate avoid this discrepancy from the current claims amounts $C_{w,n+1-w}$ hence:

$$\overline{R}_w^{BF} = \nu_w \times q_w \times (1 - z_{n+1-w}) \tag{3.21}$$

Where $1 - z_{n+1-w}$ is therefore an estimate for the percentage of the expected claims outstanding of accident year w.

The Frequency/Severity method presented by Berquist and Sherman (1977) is similar to the Bornhuetter-Ferguson method. The focus of this method is on incremental average cost per claim with separate selections for claim counts and trends in the incremental averages. It exhibits some of the stability of the Bornhuetter-Ferguson method for less mature exposure periods, and does not require a prior estimate of ultimate losses. It does exhibit some volatility due to the forecasts of ultimate claim counts, and in the selection of trends for both current levelling and forecasting into the future.

The BF method is based on the following assumptions.

- All incremental claims $X_{w,d}$ are independent.

- There are parameters μ_w, β_d , $0 \leq w \leq n$ and $0 \leq d \leq n$, with $E\left[X_{w,d}\right] = \mu_w \beta_d$ and $\beta_0 + ... + \beta_n = 1$.

- There are proportionality constants σ_d^2 , $0 \leq d \leq n$, with $Var(X_{w,d}) = \mu_w \sigma_d^2$.

- There are given unbiased a prior estimates $\dot{\mu_w}$ for μ_i, $0 \leq w \leq n$.

Further assume that a development pattern of quotas consists of parameters $\gamma_0, \gamma_1, \gamma_2, ..., \gamma_n$ with

$$\gamma_k = E[C_{w,k}] \tag{3.22}$$

The quotas represent the percentage of claims reported. A development pattern of factors consists of parameters $\psi_1, \psi_2, ..., \psi_n$ with

$$\psi_k = \frac{E\left[C_{w,k}\right]}{E\left[C_{w,k-1}\right]} \tag{3.23}$$

If the parameters $\gamma_0, \gamma_1, \gamma_2, ..., \gamma_n$ form a development pattern for the quotas, then the parameters $\psi_1, \psi_2, ..., \psi_n$ with

$$\psi_k = \frac{\gamma_k}{\gamma_{k-1}} \tag{3.24}$$

form a development pattern for the factors.

If the parameters $\psi_1, \psi_2, ..., \psi_n$ form a development pattern for the quotas, then the parameters $\gamma_0, \gamma_1, \gamma_2, ..., \gamma_n$ with

$$\gamma_k = \prod_{i=k+1}^{n} \frac{1}{\psi_i} \tag{3.25}$$

form a development pattern for the quotas.

Estimation of the parameter γ_k of a development pattern for quotas provided in the run off triangle is an empirical individual quota.

$$\overline{\gamma}_{0,k} = \frac{C_{0,k}}{C_{0,n}} \tag{3.26}$$

Estimation of the parameter ψ_k of a development pattern for factors provided in the run off triangle is an empirical individual factors.

$$\overline{\psi}_{w,k} = \frac{C_{w,k}}{C_{w,k-1}} \tag{3.27}$$

Moreover any weighted mean of the estimators is an estimator as well.

If $\gamma_0, \gamma_1, \gamma_2, ..., \gamma_n$ represent development patterns for quotas, then the expected reserve satisfied the following model.

$$E\left[R_w\right] = \left(1 - \gamma_{n-w}\right) E\left[C_{w,n}\right] \tag{3.28}$$

Therefore the predictors for the reserves can be defined as:

$$\overline{R}_w = (1 - \gamma_{n-w})\pi_w \overline{K}_w \tag{3.29}$$

where π_w is a premiums volume measure while K_w is an estimator of the expected loss.

$$\overline{K}_i = E\left[\frac{C_{w,n}}{\pi_w}\right] \tag{3.30}$$

The future cumulative losses satisfy the model equation.

$$E\left[C_{w,k}\right] = E\left[C_{w,n-w}\right] + (\gamma_k - \gamma_{n-w})E\left[C_{w,n}\right] \tag{3.31}$$

Mack(2006; 2008) criticized the use of the Chain Ladder development pattern in the BF method due to the fundamental assumption of independence between past and future claims, which underlies the BF method. Due to the proportionality of the Chain Ladder reserve to the current claims amount, the Chain Ladder reserve for accident year w is smaller, the smaller the current claims amount $C_{w,n,w}$ is. Mack (2008) observed that the systematic use of the Chain Ladder link ratios assumes that the outstanding claims part is a direct multiple of the already known claims part at each point of the development. Therefore, the development pattern should be estimated differently from the Chain Ladder development pattern.

According to Roger M. Hayne (2002) reserve uncertainty is the distribution of the amount and timing of future payments for a particular book of policies. Timothy Peterson (1980) noted that the variability of loss reserves might be estimated using maximum and minimum link ratios, instead of just averages but he noted that this method was flawed.

3.4 Over-Dispersed Poisson Model

The over-dispersed Poisson distribution differs from the Poisson distribution in that the variance is not equal to the mean, but, instead, is proportional to the mean. In claims reserving, the over-dispersed Poisson model assumes that the incremental claims $X_{w,d}$ are distributed as independent over-dispersed Poisson random variables, with mean and variance:

The assumptions of this model are similar to those in the Bayesian claims reserving models presented in England-Verrall(2002,2006), the assumption is that the parameters are modeled through prior distributions and conditional on these parameters are the incremental claims $X_{w,d}$ have independent over dispersed Poisson distributions for accident years w and development years d. The final development year is given by I and the observations at time I are given in the upper run-off triangle.

$$D_I = \{X_{w,d} : w + d \leq I\} \tag{3.32}$$

The Bayesian over-dispersed poisson model is build on the following assumptions:

- $\mu_0, ..., \mu_I, \gamma_0, ..., \gamma_I, \psi$ are independent positive random variables with joint density u(.)
- Conditionally given $\theta = \mu_0, ..., \mu_I, \gamma_0, ..., \gamma_I, \psi$ are $X_{w,d}$ random variables with

$$E\left[X_{w,d}\right] = m_{w,d} = x_w y_d \tag{3.33}$$

$$Var\left[X_{w,d}\right] = \phi x_w y_d \tag{3.34}$$

Here, x_w is the expected ultimate claims (where ultimate means up to the latest development year observed in the triangle), and y_d is the proportion of ultimate claims to emerge in each development year. Over-dispersion is introduced through the parameter ϕ, which is unknown and estimated from the data. Allowing for over-dispersion does not affect estimation of the parameters, but does have the effect of increasing their standard errors as pointed out by Renshaw and Verrall (1998).

It should be noted that, since y_d appears in the variance, the restriction that y_d must be positive is automatically imposed. This implies that the sum of incremental claims in column d must also be positive, which was a limitation of the model. Note that some negative incremental are allowed, as long as any column sum is not negative. In this formulation, the mean has a multiplicative structure, that is, it is the product of the row effect and the column effect. Both the row effect and the column effect have specific interpretations(being the expected ultimate claims in each origin year and proportion of ultimate to emerge in each development year, respectively), and it is sometimes useful to preserve the model in this form. However, for estimation purposes, it is often better to re-parametrize the model so that the mean has a linear form.

$$\log [m_{w,d}] = c + \alpha_w + \beta_d \tag{3.35}$$

This predictor structure is still a chain-ladder type, in the sense that there is a parameter for each row w and a parameter for each column d. There are some advantages and some disadvantages to this form of the model.As a generalized linear model, it is easy to estimate, and standard software packages can be used; the estimates should be well behaved. However, the parameter values themselves will be harder to interpret, making it necessary to convert them back into more familiar quantities. Note that constraints have to be applied to the sets of parameters, which could take a number of different forms. For example, the corner constraints would put

$$\alpha_1 = \beta_1 = 0 \tag{3.36}$$

Although this kind of model is based on the Poisson distribution, this does not imply that it is only suitable for data consisting exclusively of positive integers. That constraint can be overcome using a quasi-likelihood approach described by McCullagh and Nelder (1989), which can be applied to non-integer data, positive and negative. With quasi-likelihood, in this context, the likelihood is the same as a Poisson likelihood up to a constant of proportionality. For data consisting entirely of positive integers, identical parameter estimates are obtained using the full or quasi-likelihood.

Many statistical packages fit GLMs using quasi-likelihood by default, the user being entirely unaware. In modelling terms, the crucial assumption is that the variance is proportional to the mean, and the data are not restricted to being positive integers.

Assume there are N open and unknown claims, all of which are statistically independent, and have the same probability distribution with mean μ and variance σ^2. The distribution of the reserves will have a mean and variance of:

$$E\left[R\right] = N\mu \tag{3.37}$$

$$Var\left[R\right] = N\sigma^2 \tag{3.38}$$

If the distribution of X_i is known then the resulting reserve distribution will only exhibit process uncertainty. In some cases the distribution of claim size will be known and exhibit a closed form, however if the number of claims N is random and independent from the claim size distribution and has a mean λ and a variance τ^2 then the reserve will have the following mean and variance.

$$E\left[R\right] = \lambda\mu \tag{3.39}$$

$$Var\left[R\right] = \lambda\sigma^2 + \mu^2\tau^2 \tag{3.40}$$

Consider a collective risk model, the random variable N is assumed to have a poison distribution, in which case τ^2 is equal to λ hence:

$$Var\left[R\right] = \lambda\left(\sigma^2 + \mu^2\right) \tag{3.41}$$

With a poison claim count distribution, the variance of the average reserve is:

$$Var\left[\frac{R}{\lambda}\right] = \frac{\lambda(\sigma^2 + \mu^2)}{\lambda^2} = \frac{(\sigma^2 + \mu^2)}{\lambda} \tag{3.42}$$

The variance approaches zero as λ becomes arbitrary large. Assuming the claim count have a poison distribution, the process uncertainty inherent in the average reserve will effectively disappear as the expected number of claims becomes large.

Consider a classical collective risk model and incorporate the following parameters α and β. To this end we will assume that α and β are two random variables. The essence of this is to solve the problem of calculating aggregate distribution for collective risk models with weak restrictions on the claims size distribution.

Therefore:

$$E\left[\alpha\right] = 1 \tag{3.43}$$

$$Var\left[\alpha\right] = c \tag{3.44}$$

and

$$E\left[\frac{1}{\beta}\right] = 1 \tag{3.45}$$

$$Var\left(\frac{1}{\beta}\right) = b \tag{3.46}$$

Next we formulate an algorithm to generate one observation of aggregate reserves.

- Randomly select a value of α from a gamma distribution.
- Randomly select the number of claims N from a poison distribution with parameters α^{λ}.
- Randomly select a value for β from a gamma distbution.
- Randomly select N claims from a claim size distribution.
- Add the value of the N claims and divide the results by β.

Here c is a contagion parameter while b is a mixing parameter.

Under the assumption that the claims count and claim size distribution are independent and claims size selection in step 4 are independent of each other and the random variable α and β. We can calculate the expected value and the variance of aggregate reserves.

$$E\left[R\right] = \lambda\mu \tag{3.47}$$

$$Var\left[R\right] = \lambda\left(\mu^2 + \sigma^2\right)(1+b) + \lambda^2\mu^2\left(b+c+bc\right) \tag{3.48}$$

The algorithm presented here allows for the combination of aggregate loss distributions for several lines of insurance, each with its own contagion parameter c but with a global mixing parameter b. We will take advantage of this feature and have different contagion parameters for each accident year as well as a single global mixing parameter reflecting uncertainty that affects reserves for all accident years at once. Global uncertainty would take forms of future inflation or court decisions.

3.5 Formulation of the Collective Risk Model

3.5.1 Mean Square Error of Prediction

The primary goal of this project is to estimate variability in reserves, in this section we address the Mean Square Error technique that estimates how good the mean or expected ultimate claims are ie the quality of the estimates.

Assume we have a random variable X (incremental claims) and a set of observations D. In this case X denotes the incremental entries in the triangle while D is the set of training data.

Assume that $\overline{X}$ is a D - measurable estimator for $E[X|D]$

The conditional mean square error for the prediction of the estimator $\overline{X}$ can be defined as:

$$msep_{X|D}\left(\overline{X}\right) = E\left[\left[\overline{X} - X\right]^2 |D\right] \tag{3.49}$$

For a D - measurable estimator $\overline{X}$ we have:

$$msep_{X|D}\left(\overline{X}\right) = Var\left[X|D\right] + \left[\overline{X} - E\left[X|D\right]\right]^2 \tag{3.50}$$

where:

$Var\,(X|D)$ is the process variance ie. the variance which is within the stochastic model (pure randomness which cannot be eliminated).

$\left(\overline{X} - E\,[X|D]\right)^2$ is the parameter or estimation variance. It reflects the uncertainty in the estimation parameter and of the expectation.

We assume that X is independent of D in a situation where we have iid experiments and we want to estimate the average outcome.

$$E\,[X|D] = E\,[X] \tag{3.51}$$

$$Var\,[X|D] = Var\,[X] \tag{3.52}$$

The unconditional mean square error of prediction for the estimator $\overline{X}$ is:

$$msep_X\left[\overline{X}\right] = E\left[msep_{X|D}\left[\overline{X}\right]\right] \tag{3.53}$$

$$msep_X\left[\overline{X}\right] = Var\,[X] + Var\left[\overline{X}\right] \tag{3.54}$$

Hence the parameter error is estimated by the variance of $\overline{X}$.

Assume that $X_1, ..., X_n$ are iid with mean μ and variance σ^2. We have the estimator.

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n} \tag{3.55}$$

and therefore:

$$msep_{X|D}\left(\overline{X}\right) = \sigma^2 + \left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2 \tag{3.56}$$

By law of large numbers the last term disappears as n approaches ∞. In order to determine this term for finite n, we explicitly calculate the distance between $\sum_{i=1}^{n} \frac{X_i}{n}$

and μ. However since in general μ is not known we can only give an estimate for that distance.

The unconditional mean square error of prediction is:

$$msep_X\left(\overline{X}\right) = \sigma^2 + \frac{\sigma^2}{n} \tag{3.57}$$

We can assume that the deviation of $\sum_{i=1}^{n} \frac{X_i}{n}$ around μ is on average of order $\frac{\sigma}{n}$.

In case X is not independent of the observation D.

$$msep_X\left(\overline{X}\right) = E\left[msep_{X|D}\overline{X}\right] \tag{3.58}$$

This can be expressed as:

$$\begin{aligned} msep_X\left[\overline{X}\right] &= E\left[VarX|D\right] + E\left[\left[\overline{X} - E\left[X|D\right]\right]^2\right] \\ &= Var\left[X\right] - Var\left[E\left[X|D\right]\right] + E\left[\left[\overline{X} - E\left[X|D\right]\right]^2\right] \\ &= Var\left[X\right] + E\left[\left[\overline{X} - E\left[X\right]\right]^2\right] - 2E\left[\left[\left(\overline{X} - E\left[X\right]\right)\right]\left[E\left[X|D\right] - E\left[X\right]\right]\right] \end{aligned} \tag{3.59}$$

If the estimator $\overline{X}$ is unbiased for $E\left[X\right]$ we obtain:

$$msep_X\left[\overline{X}\right] = Var\left[X\right] + Var\left[\overline{X}\right] - 2Cov\left[\overline{X}, E\left[X|D\right]\right] \tag{3.60}$$

To quantify the variability in the chain ladder reserve estimates we will assume that the claims amount do not have a specific distribution function and establish a formula for the standard error which is an estimate for the standard deviations of outstanding claims reserves.

Let:

- w = Accident year.
- d = Development year/ Lag period.
- $C_{w,d}$ = Cumulative losses (either paid or incurred)
- $X_{w,d}$ = Incremental (paid or incurred) losses

Assume that we have an $n \times n$ triangle of incremental paid or incurred losses organized by rows for accident years and by columns for development lags. We also have the premium associated with each accident year.

We estimate the age to age factors by:

$$\Lambda_d = \sum_{w=1}^{n-d} \times \frac{C_{w,d}}{\sum_{w=1}^{n-d} C_{w,d}} \times \frac{C_{w,d+1}}{C_{w,d}} \tag{3.61}$$

This can be shown to be equivalent to:

$$\begin{aligned} \Lambda_d &= \frac{\sum_{w=1}^{n-d} C_{w,d+1}}{\sum_{w=1}^{n-d} C_{w,d}} \\ &= \sum_{w=1}^{n-d} \frac{C_{w,d+1}}{C_{w,d}} \end{aligned} \tag{3.62}$$

For $d = (1, 2..., n)$

This is the weighted average of the observed individual development factors $\frac{C_{w,d+1}}{C_{w,d}}$ and the weights are proportional to $C_{w,d}$. Like Λ_d every individual development factor $\frac{C_{w,d+1}}{C_{w,d}}$ is also an unbiased estimator of Λ_d because:

$$\begin{aligned}
E\left[\frac{C_{w,d+1}}{C_{w,d}}\right] &= E\left[E\left[\frac{C_{w,d+1}}{C_{w,d}}|C_{w,1}.....C_{w,n}\right]\right] \\
&= E\left[\frac{E\left[C_{w,d+1}|C_{w,1}.....C_{w,n}\right]}{C_{w,n}}\right] \\
&= E\left[\frac{C_{w,d}\Lambda_d}{C_{w,d}}\right] \\
&= E\left[\Lambda_d\right] \\
&= \Lambda_d
\end{aligned} \tag{3.63}$$

This hold because by iterative rule $E[X] = E[E[X|Y]]$ and Λ_d is a scalar parameter and the chain ladder uses Λ_d as an estimator of $\overline{\Lambda}_d$.

Applying the principle of the theory of point estimation that states that among several unbiased estimators preference should be given to the one with the smallest variance.

Assume that we estimate the age to age factors of the observed development factors which is also an unbiased estimator as is in the case of weighted average.

$$\overline{\Lambda}_d = \sum_{w=1}^{n-d} W_{w,d}\frac{C_{w,d+1}}{C_{w,d}} \tag{3.64}$$

with

$$\sum_{w=1}^{n-d} W_{w,d} = 1 \tag{3.65}$$

$W_{w,d}$ is a scalar if $C_{w,1},...,C_{w,n}$ are unknown.

$W_{w,d}$ is inversely proportional to:

$$Var\left[\frac{C_w,d+1}{C_{w,d}}|C_{w,1},...,C_{w,n}\right]$$

The chain ladder estimator Λ_d uses weights which are proportional to $C_{w,d}$ and therefore $C_{w,d}$ is assumed to be inversely proportional to:

$$Var\left[\frac{C_w, d+1}{C_{w,d}}|C_{w,1}, ..., C_{w,n}\right] = \frac{\alpha_d^2}{C_{w,d}} \tag{3.66}$$

with a proportionality constant α_d^2 which may depend on d but not on w and which must be non negative because variance is always positive.

$C_{w,d}$ is a scalar and generally $Var\left[\frac{X}{c}\right] = \frac{Var[X]}{c^2}$ for any scalar c, we can state the above proportionality condition in the form.

$$Var\left[C_{w,d+1}|C_{w,1}, ..., C_{w,n}\right] = C_{w,d}\alpha_d^2 \tag{3.67}$$

Note that the aim of claims reserving methods is to estimate the ultimate claim amounts $C_{w,n}$

Where $w = 2, ..., n$

This is accomplished by applying.

$$C_{w,n} = C_{w,n+1-w} \times (\Lambda_{n+1-w}, ..., \Lambda_{n-1}) \tag{3.68}$$

The ultimate claims amounts are unbiased under the following assumptions.

- There are unknown constants $\Lambda_1, ..., \Lambda_{n-1}$ with

$$E\left[C_{w,d+1}|C_{w,1}, ..., C_{w,n}\right] = C_{w,d}\Lambda_d \tag{3.69}$$

- The variables $C_{w,1}, ..., C_{w,n}$ and $C_{d,1}, ..., C_{d,n}$ of different accident years $w \neq d$ are independent.

The expected value of the estimator

$\overline{C}_{w,n} = C_{w,n+1-w}\left[\overline{\Lambda}_{n+w-1}, ..., \overline{\Lambda}_{n-1}\right]$ for the ultimate claim amounts and the true claim amounts $C_{w,n}$ are equal that is:

$$E\left[\overline{C}_{w,n}\right] = E\left[C_{w,n}\right] \tag{3.70}$$

To estimate the mean square error we need to know the average distance between the forecast $\overline{C}_{w,n}$ and the future realizations $C_{w,n}$

$$mse\left[\overline{C}_{w,n}\right] = E\left[\left[C_{w,n} - \overline{C}_{w,n}\right]^2 |D\right] \tag{3.71}$$

Where $D = \{C_{w,d}|w + d \leq n + 1\}$ is the set of all data observed so far. The error we want to estimate is that due to future randomness.

The unconditional error

$$E\left[C_{w,n} - \overline{C}_{w,n}\right]^2 = E\left[E\left[C_{w,n} - \overline{C}_{w,n}\right]^2 |D\right] \tag{3.72}$$

This is the exact concept underlying

$$Var\left[X\right] = E\left[X - E\left[X\right]\right]^2 \tag{3.73}$$

Due to the general rule;

$$E\left[X-c\right]^2 = Var\left[X\right] + \left[E\left[X\right]-c\right]^2$$
$$mse\left[\overline{C}_{w,n}\right] = Var\left[C_{w,n}\right] + \left[E\left[C_{w,n}|D\right] - \overline{C}_{w,n}\right]^2 \tag{3.74}$$

$mse\left[C_{w,n}\right]$ depends on the unknown model parameters f_k and α_k^2. The standard error $s.e\left[\overline{C}_{w,n}\right]$ of $C_{w,n}$ is the standard error of the reserves $s.e\left[\overline{R}_{w,n}\right]$.

$\overline{R}_w = \overline{C}_{w,n} - C_{w,n+1-w}$ of the outstanding claim reserves. Therefore:

$$mse\left[\overline{R}_w\right] = E\left[\left[\overline{R}_w - R_w\right]^2 |D\right] = E\left[\left[\overline{C}_{w,n} - C_{w,n}\right]^2 |D)\right] = mse\left[\overline{C}_{w,n}\right] \tag{3.75}$$

Equality of the mean square error also implies the equality of the standard errors.

$$s.e\left[\overline{R}_w\right] = s.e\left[\overline{C}_w\right] \tag{3.76}$$

The standard error of claims reserves can be computed as:

$$\left[s.e\left[(\overline{C}_{w,n}\right]^2\right] = \overline{C}_{w,n}^2 \sum_{k=n+1-w}^{n-1} \frac{\alpha_k^2}{f_k} \left(\frac{1}{\overline{C}_{w,n}} + \frac{1}{\sum_{d=1}^{n-1} C_{d,k}} \right) \tag{3.77}$$

where

$$\overline{\alpha}_k^2 = \frac{1}{n-k-1} \sum_{d=1}^{n-k} C_{d,k} \left(\frac{C_{d,k+1}}{C_{d,k}} - \overline{f}_k \right) \tag{3.78}$$

This is an unbiased estimator of α_k^2.

We have estimated $\overline{R}_w$ and $s.e\left(\overline{R}_w\right)$ for the mean and standard deviation. The assumption was that the outstanding claims were large enough and due to the central limit theorem we had assumed that the distribution function was normal with the expected value equal to point estimate given by R_w and the standard deviation equal to the standard error $s.e\left(R_w\right)$.

A symmetric 95% confidence interval for the reserves under normal distribution would be:

$$\left(\overline{R}_w - 2s.e\left(\overline{R}_w\right), \overline{R}_w + 2s.e\left(\overline{R}_w\right)\right) \tag{3.79}$$

This is a weak assumption especially if the distribution of the reserves is skewed that is for the cases where $s.e(\overline{R}_w)$ is greater than 50% of $\overline{R}_w$. In this case we recommend to use an approach based on the log-normal distribution. For this purpose we approximate the unknown distribution of R_w by a log-normal distribution with parameters μ_w and σ_w. Such that the mean values as well as variance of both distributions are equal.

$$\begin{aligned} \exp\left(\mu_w + \frac{\sigma_w^2}{2}\right) &= \overline{R}_w \\ \exp\left(2\mu_w + \sigma_w^2\right)\left(\exp\left(\sigma_w^2\right) - 1\right) &= Var\left(\overline{R}_w\right) \end{aligned} \tag{3.80}$$

This leads to:

$$\begin{aligned} \mu_w &= \ln\left(R_w\right) - \frac{\sigma_w^2}{2} \\ \sigma_w^2 &= \ln\left(\frac{1 + Var\left(\overline{R_w}\right)}{R_w^2}\right) \end{aligned} \tag{3.81}$$

3.5.2 Bayesian Approach Models

Let $C_{w,d}$ be random variables of accumulated claim amounts for accident years w and development years d. We will denote all observed data in the triangle ie the training data by D. We will assume that the claim amounts in the first year has fully developed as so $C_{w,d}$ for $(2 \leq w \leq n)$ are considered as ultimate claims to be estimated.

Given D the development factor Λ_d are estimated by:

$$\Lambda_d|D = \frac{\sum_{w=1}^{n-d} C_{w,d+1}}{\sum_{w=1}^{n-d} C_{w,d}} \tag{3.82}$$

and the ultimate claim amounts for the w^{th} $(w \geq 2)$ year denoted as:

$$\overline{C}_{w,n}|D = C_{w,n-w+1} \prod_{d=n-w+1}^{n-1} \overline{\Lambda}_d \tag{3.83}$$

Model assumptions

- $E\left[C_{w,d+1}|C_{w,1},...,C_{w,d}\right] = \Lambda_d C_{w,d}$
- $C_{w,1},...,C_{w,n},\ C_{k,1},...,C_{k,I}$ are independent and
- $Var\left[C_{w,d+1}|C_{w,1},...,C_{w,d}\right] = \sigma_d^2 C_{w,d}$

To make simulations possible we assume that $C_{w,d+1}$ is normally distributed with mean $\Lambda_d C_{w,d}$ and variance $\sigma_d^2 C_{w,d}$ that is, this allows one to calculate t-tests, p-values, F tests, etc. and draw inferences about how well simulated data describe the distribution of the reserves.

$$C_{w,d+1}|\left(C_{w,1},...,C_{w,d}\right) \sim N\left(\Lambda_d C_{w,d}, \sigma_d^2 C_{w,d}\right) \tag{3.84}$$

Let $Y_{w,d} = \frac{C_{w,d+1}}{C_{w,d}}$ this assumption is equivalent to:

$$Y_{w,d}|\left(C_{w,1},...,C_{w,d}\right) \sim N\left(\Lambda_d, \frac{\sigma_d^2}{C_{w,d}}\right) \tag{3.85}$$

The distribution of $C_{w,d}$ is defined by parameters Λ_d, σ_d^2, the posterior distribution of Λ_d, σ_d^2 is calculated so that the posterior distribution of $C_{w,d}$ can be evaluated.

For simplicity assume that we have a vector:

$$\Lambda = (\Lambda_1, ..., \Lambda_{n-1})$$
$$\sigma^2 = \left(\sigma_1^2, ..., \sigma_{n-1}^2\right) \tag{3.86}$$

The first step of the Bayesian approach is to calculate the posterior distribution.

$$P\left(\Lambda, \sigma^2 | D\right) \propto P\left(C | \Lambda, \sigma^2\right) P\left(\Lambda, \sigma^2\right) \tag{3.87}$$

Where $P(\Lambda, \sigma^2)$ is the joint distribution of Λ and σ^2 and $P(\Lambda, \sigma^2|D)$ is the joint posterior distribution. $P(C|\Lambda, \sigma^2)$ is determined by the assumptions of the model.

$$\begin{aligned}
P\left(C|\Lambda, \sigma^2\right) &= \prod_{w=1}^{n} P\left(x_{w,1}, ..., x_{w,n+1-w} | \Lambda, \sigma^2\right) \\
&= \prod_{w=1}^{n} \left[P\left(x_{w,1} | \Lambda, \sigma^2\right) \prod_{d=2}^{n+1-w} P\left(x_{w,d} | x_{w,1}, ..., x_{w,d-1}\right), \Lambda, \sigma^2 \right] \\
&= \left[\prod_{w=1}^{n} \left[P\left(x_{w,1} | \Lambda, \sigma^2\right) \right] \prod_{d=2}^{n} \prod_{w=1}^{n-d+1} P\left(x_{w,d} | x_{w,1}, ..., x_{w,d-1}), \Lambda, \sigma^2\right) \right] \\
&\propto \prod_{d=2}^{n} \left[\prod_{w=1}^{n-d+1} \left[\frac{1}{\sqrt{2\Pi(\sigma_{d-1}^2 x_{w,d-1})}} \right] exp\left(-\frac{(x_{w,d} - \Lambda_{d-1} x_{w,d-1})^2}{2(\sigma_{d-1}^2)} \right) \right] \\
&\propto \prod_{d=1}^{n-1} \left[\prod_{w=1}^{n-d} \left(\frac{1}{\sqrt{\sigma_d^2}} exp\left(-\frac{(y_{w,d} - \Lambda_d)^2}{2\frac{\sigma_d^2}{x_{w,d}}} \right) \right) \right]
\end{aligned} \tag{3.88}$$

Note that $y_{w,d}$ is defined as $y_{w,d} = \frac{x_{w,d+1}}{x_{w,d}}$ if both $x_{w,d+1}$ and $x_{w,d}$ are known.

By definition $P(\Lambda, \sigma^2)$ is a multi-dimensional distribution and generally there is no guarantee of independency between pairs(Λ_d, σ_d^2). However any appropriate distribution can be chosen as prior distribution so it is reasonable to assume that the chosen prior distributions have the features of independency that is any pair of (Λ_d, σ_d^2) is independent of other pairs.

$$P\left(\Lambda, \sigma^2\right) = \prod_{d=1}^{n-1} P\left(\Lambda_d, \sigma_d^2\right) \tag{3.89}$$

Substituting accordingly:

$$P\left(\Lambda, \sigma^2 | D\right) \propto \prod_{d=1}^{n-1} \left[\prod_{w=1}^{n-d} \left[\frac{1}{\sqrt{\sigma_d^2}\ \exp\left(-\frac{(y_{w,d} - \Lambda_d}{\frac{2\sigma_d^2}{x_{w,d}}} \right) P(\Lambda_d, \sigma_d^2)} \right] \right] \tag{3.90}$$

which shows that the joint posterior distribution can be factorized. This gives an important conclusion that if the prior distribution is independent, the joint posterior distribution of the pair $\left(\Lambda_d, \sigma_d^2 | D\right)$ is also independent of other pairs.

$$\begin{aligned} P\left(\Lambda_d, \sigma_d^2 | D\right) &\propto \prod_{d=1}^{n-d} \frac{1}{\sigma_d^2} \exp\left[-\frac{(y_{w,d} - \Lambda_d)^2}{2\left(\frac{\sigma_d^2}{x_{w,d}}\right)} P(\Lambda_d, \sigma_d^2) \right] \\ &\propto \left(\sigma_d^2\right)^- \left(\frac{(n-d)}{2} \right) \exp\left[-\frac{1}{2\sigma_d^2} \sum_{w=1}^{n-d} x_{w,d} (y_{w,d} - \Lambda_d)^2 \right] P\left(\Lambda_d, \sigma_d^2\right) \end{aligned}$$

The second step is to compute the marginal posterior distributions $P(\Lambda_d|D)$ and $P(\sigma_d^2|D)$. We calculate this by integrating out the unwanted variables in the joint posterior distribution.

$$\begin{aligned} P\left(\Lambda_d | D\right) &\propto \int_{d=0}^{\infty} P\left(\Lambda_d | \sigma_d^2, D\right) P\left(\sigma_d^2 | D\right) d\sigma_d^2 \\ &\propto \int_{d=0}^{\infty} P\left(\Lambda_d | \sigma_d^2, D\right) d\sigma_d^2 \end{aligned} \tag{3.91}$$

and similarly

$$\begin{aligned} P\left(\sigma_d^2 | D\right) &\propto \int_{d=0}^{\infty} P\left(\sigma_d^2 | \Lambda_d, D\right) P\left(\Lambda_d | D\right) d\Lambda_d \\ &\propto \int_{d=0}^{\infty} P\left(\sigma_d^2 | \Lambda_d, D\right) d\Lambda_d \end{aligned} \tag{3.92}$$

However this cannot be computed analytically hence we need to use numerical techniques to compute the marginal posterior distributions. This makes the bayesian technique tricky and the alternative is to resort to simulation techniques. To compute the variance we assume independence in pairs $(\Lambda_d, \sigma_d^2)|D)$

$$E\left[_{w,d+1}|D\right] = E\left[Lambda_d C_{w,d}|D\right] = E\left[Lambda_d|D\right] E\left[C_{w,d}|D\right] \tag{3.93}$$

The second central moment is

$$\begin{aligned} E\left[C^2_{w,d+1}|D\right] &= E\left[\left[[\Lambda_d C_{w,d}]^2 + \sigma_d^2 C_{w,d}\right]|D\right] \\ &= E\left[\Lambda_d^2|D)\right] E\left[C^2_{w,d}|D\right] + E\left[\sigma_d^2|D\right] E\left[C_{w,d}|D\right] \end{aligned} \tag{3.94}$$

Hence the variance of $C_{w,d+1}|D$ is

$$\begin{aligned} Var(C_{w,d+1}|D) &= E(C^2_{w,d+1}|D) - (E(C_{w,d+1}|D))^2 \\ &= E(\Lambda_d^2|D)E(C^2_{w,d}|D) + E(\sigma_d^2|D)E(C_{w,d}|D) - (E(\Lambda_d|D))^2(E(C_{w,d}|D))^2 \\ &= Var(\Lambda_d|D)(E(C_{w,d}|D))^2 + E(\Lambda_d^2|D)Var(C_{w,d}|D) + E(\sigma_d^2|D)E(C_{w,d}|D) \end{aligned}$$

The values of $E\left[\Lambda_d|D\right]$, $Var\left[\Lambda_d|D\right]$ and $Var\left[\sigma_d^2|D\right]$ can be calculated from the posterior distribution. The boundary is needed for this recursive formula, because

$C_{w,n-w+1}|D = C_{w,n-w+1}$

and its mean is

$$E\left[C_{w,n-w+1}|D\right] = C_{w,n-w+1} \tag{3.95}$$

and its variance is

$$Var\left[C_{w,n-w+1}|D\right] = 0 \tag{3.96}$$

3.5.2.1 Parameter Estimation

The distribution of each pair of parameter (Λ_d, σ_d^2) can be calculated individually and the posterior distributions have similar forms for different lag periods.

$$P(\Lambda_d, \sigma_d^2|D) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{w=1}^{n} C_w(y_w - \Lambda)^2\right) P(\Lambda_d, \sigma_d^2) \tag{3.97}$$

3.5.2.2 Known σ^2

In order to make comparison we analyse σ^2 when it is known. This can be simplified to:

$$P\left(\Lambda_d|D\right) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{w=1}^{n} x_w\left(y_w - \Lambda\right)^2\right) P\left(\Lambda\right) \tag{3.98}$$

For a non-informative prior distribution.

$$P\left(\Lambda\right) = 1 \tag{3.99}$$

The equation simplifies to:

$$P\left(\Lambda_d|D\right) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{w=1}^{n} x_w\left(y_w - \Lambda\right)^2\right) \tag{3.100}$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\left(\Lambda - \overline{\Lambda}\right)^2\right) \tag{3.101}$$

$$\tag{3.102}$$

where

$$\overline{\Lambda} = \frac{\sum\limits_{w=1}^{n} x_w y_w}{\sum\limits_{w=1}^{n} x_w}$$

The posterior distribution of Λ is normally distributed.

$$\Lambda|D \sim N\left(\overline{\Lambda}, \frac{\sigma^2}{\sum\limits_{w=1}^{n} x_w}\right) \tag{3.103}$$

and the mean is

$$E\left[\Lambda|D\right] = \overline{\Lambda} \tag{3.104}$$

and the variance is

$$Var\left[\Lambda|D\right] = \frac{\sigma^2}{\sum\limits_{w=1}^{n} x_w} \tag{3.105}$$

With prior knowledge of Λ, it is useful to use an informative prior distribution.

$$P\left(\Lambda\right) = \frac{1}{\sqrt{2\Pi\sigma^2}} \exp\left(-\frac{(\Lambda-\mu_0)^2}{2\sigma_0^2}\right) \tag{3.106}$$

Where μ_0 is the prior knowledge of Λ and σ_0^2 indicates the confidence about the prior distribution. The larger the variance the lower the confidence. Using this prior the posterior distribution becomes.

$$P\left(\Lambda_d|D\right) \propto \exp -\frac{1}{2\sigma^2}\sum_{w=1}^{n} x_w(y_w-\Lambda)^2 - \frac{1}{2\sigma_0^2}(\Lambda-\mu_0)^2$$
$$P\left(\Lambda_d|D\right) \propto \exp -\frac{1}{2\sigma^2}\sum_{w=1}^{n} x_w + \frac{1}{2\sigma_0^2}\left(\Lambda-\oslash\right)^2$$

where

$$\oslash = \left\{\frac{\frac{\overline{\Lambda}}{\sigma^2}\sum\limits_{w=1}^{n} x_w + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma^2}\sum\limits_{w=1}^{n} x_w + \frac{1}{\sigma_0^2}}\right\} \tag{3.107}$$

Hence the posterior distribution is normally distributed.

$$\Lambda|D \sim N\left\{\frac{\frac{\overline{\Lambda}}{\sigma^2}\sum_{w=1}^{n}x_w + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma^2}\sum_{w=1}^{n}x_w + \frac{1}{\sigma_0^2}}, \frac{1}{\frac{1}{\sigma^2}\sum_{w=1}^{n}x_w + \frac{1}{\sigma_0^2}}\right\} \tag{3.108}$$

3.5.2.3 Unknown σ^2 Conjugate Prior

The idea is to choose a prior distribution that provides convenience in calculation. The conjugate distribution to be used is the distribution which makes the prior and the posterior belong to the same family of distribution. For the likelihood information the conjugate distribution is the Normal-Inverse-Gamma distribution defined as:

$$P = \left(\Lambda, \sigma^2\right) = P\left(\Lambda|\sigma^2\right) P\left(\sigma^2\right) \tag{3.109}$$

Where σ^2 has an Inverse-Gamma distribution.

$$\sigma^2 \sim IG(\frac{\omega_0}{2}, \sigma_0^2) \tag{3.110}$$

and Λ has a normal distribution with variance relate to σ^2.

$$\Lambda|\sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{\eta_0}\right)$$

$\omega_0, \sigma_0^2, \mu_0$ and η_0 are all parameters that can be chosen based on the prior knowledge. With these prior the posterior distribution becomes.

$$P\left(\Lambda, \sigma_0^2|D\right) \propto (\sigma^2)^{-\frac{n+\omega_0+3}{2}} \exp\frac{1}{2\sigma^2}\left\{\eta_0(\Lambda-\mu_0)^2 + \sum_{w=1}^{n}x_w(\Lambda-y_w)^2 + \sigma_0^2\right\}$$

The marginal distribution of Λ is:

$$
\begin{aligned}
P(\Lambda|D) &\propto \int_0^\infty \sigma^{2-\frac{n+\omega_0+3}{2}} exp -\frac{1}{2\sigma^2}\left\{\eta_0(\Lambda-\mu_0)^2+\sum_{w=1}^{n} x_w(\Lambda-y_w)^2+\sigma_0^2\right\} \text{ d } \sigma^2 \\
&\propto \left\{\eta_0\left(\Lambda_0-\mu_0\right)^2+\sum_{w=1}^{n} x_w\left(\Lambda-y_w\right)^2+\sigma_0^2\right\}^{-\frac{n+w+1}{2}} \\
&\propto \left\{1+\frac{\frac{(\Lambda-\mu_n)^2}{\frac{\sigma_n^2}{\omega_n\eta_n}}}{\omega_n}\right\}^{-\frac{\omega_n+1}{2}}
\end{aligned}
$$

Where

$$
\begin{aligned}
\mu_n &= \frac{\eta_0}{\eta_0+\sum_{w=1}^{n} x_w}\mu_0+\frac{\sum_{w=1}^{n} x_w}{\eta_0+\sum_{w=1}^{n} x_w}\overline{\Lambda} \\
\eta_n &= \eta_0+\sum_{w=1}^{n} x_w \\
\omega_n &= \omega_0+n \\
\sigma_n^2 &= \sigma_0^2+(n-1)s^2+\frac{\eta_0}{\eta_n}(\Lambda-\mu_0)^2\sum_{w=1}^{n} x_w
\end{aligned}
\tag{3.111}
$$

Hence $\frac{\Lambda-\mu_n}{\sqrt{\frac{\sigma_n^2}{\omega_n\eta_n}}}$ has a t distribution with ω_n degrees of freedom.

$$
\Lambda|D \sim t_{\omega_n}(\mu_n,\frac{\sigma_n^2}{\omega_n\eta_n}) \tag{3.112}
$$

This gives a mean of

$$
E\left[\Lambda|d\right]=\mu_k \tag{3.113}
$$

and a variance of

$$
Var\left[\Lambda|d\right]=\frac{\sigma_n^2}{\omega_n\eta_n}\frac{\omega_n}{\omega_n-2}=\frac{\sigma_n^2}{(\omega_n-2)\eta n} \tag{3.114}
$$

Likewise the marginal posterior distribution of σ^2 is

$$P\left(\sigma^2|D\right) \propto \left(\sigma^2\right)^{\left(-\frac{\omega_n-2}{2}\right)}\left(-\frac{\sigma_n^2}{2\sigma^2}\right) \tag{3.115}$$

σ^2 has an inverse gamma distribution with parameters $\frac{\omega_n}{2}$ and $\frac{\sigma_n}{2}$ that is $\sigma^2|D \sim IG\left(\omega_n^2, \sigma_n^2\right)$ with a mean $E\left[\sigma^2|D\right] = \frac{\sigma_n^2}{\omega_n-2}$. The choice of distributions to use require a thorough analysis of data especially where the data would result to negative or zero entries in the last diagonal.

3.5.3 Stochastic Framework

Having build on the Bayesian framework that derives a posterior distribution from a prior and a conjugate distribution, we can now extend this to a Bayesian Markov Chain Monte-Carlo (MCMC) models. This model provides an unprecedented flexibility in stochastic model development. In our model we wish to:

- Examine correlation accident years.
- Analyse a skewed distribution defined over the entire real line to deal with negative incremental paid data.
- Accommodate changes in the claims settlement rate and payment year trends for paid claims data set.

3.5.3.1 Lognormal Loss Development Model

We wish to examine two sources of variation. These effects are captured by the log-normal random variable estimated for the next year of development with respect to the accident year under review.

- Process variance and
- Estimation Variation

For the paid losses triangle the development lags are independent from one period to the next. The ultimate loss ratios are estimated based on the incurred losses.

We assume that the age to age development factors are log-normally distributed.

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\Pi}} \exp\left\{-\frac{1}{2}\left(\frac{ln(x)-\mu}{\sigma}\right)^2\right\} \tag{3.116}$$

The expected values and the variance of the distribution are given by;

$$E(x) = \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \tag{3.117}$$

and

$$Var(x) = \exp\left\{2\mu + \sigma^2\right\}\left(\exp\left\{\sigma^2\right\} - 1\right) \tag{3.118}$$

The product of independent log-normal random variables is also log-normal implying that the age to ultimate loss development factors are log-normal. The product of a constant and a log-normal random variable is log-normal, hence given the loss ratios and the estimate age to ultimate factors we can compute the ultimate loss ratios.

To estimate the parameters for the log-normal distribution we use the unbiased estimators.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.119}$$

and

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1} \tag{3.120}$$

where x_i's are the observed age to age factors and $\overline{x}_i$ is a maximum likelihood estimator. Using the estimated parameters that is μ_i and σ_i we can get the mean loss development factors. The parameters define the log-normal distribution that best fits the data.

We combine the parameter estimates for the prospective age to age factors using the multiplicative property of log-normal distribution to determine parameter estimates for the prospective age to ultimate factors.

The product of n log-normal random variables with respective parameters $(\mu_1, \sigma_1), (\mu_2, \sigma_2), ..., (\mu_n, \sigma_n)$ is a log normal random variable with parameters.

$$\mu = \sum_{i=1}^{n} \mu_i \tag{3.121}$$

and standard deviation

$$\sigma = \sqrt{\sum_{i=1}^{n} \sigma_i^2} \tag{3.122}$$

To achieve the age to ultimate factors we need to take the natural logarithms of the age to age factors. The ultimate development factor is the sum of the mean age to age factor natural logarithms. The corresponding σ estimate is the square root of the sum of variances of the natural logarithms for the age to age factors.

Measuring effect of log-normal loss development can be modelled for the next year on the existing mean age to age and age to ultimate factors.The effect of adding a data point arises from a log-normal distribution, the existing mean remains while the standard deviation gets revised.

Examining the product of a constant and two log-normal random variables.

$$\mu = ln(p) + \mu_1 + \mu_2$$

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$$

The chain ladder random variable can be expressed as a product of two log-normal random variables. The idea is to model revised loss ratio estimates for other time horizons and the confidence intervals for the same and those of the estimated reserves.

$$x_{cl} = x_w E\left[tail\right] \tag{3.123}$$

For the BF method.

$$x_{BF} = x_p - E\left[x_p\right] + E\left[x_p\right] E\left[tail\right] \tag{3.124}$$

If we assume that $\mu = \overline{x}$ and $\sigma = s$ the 95% confidence interval is

$$\exp(\overline{x} - N^{-1}(0.975 \times \sigma)) \tag{3.125}$$

and

$$\exp(\overline{x} + N^{-1}(0.975 \times \sigma)) \tag{3.126}$$

The assumption of log-normal distribution is a weak one since it does not take in to account the distribution of the claims count and the claims sizes. The fact that the log-normal estimation of development factors reverts back to the traditional chain ladder factors only gives a good frame work for studying the variability in the reserves.

3.5.3.2 Poisson Gamma Mixture

Assume that the claims frequencies are poisson distributed $N \sim Po(\lambda)$, the number of claims in a fixed period of time from an insured in a large pool of insured. λ represents the nature of risk.

- Large λ represents high risks while
- Small λ represents low risks

Hence λ is itself a random variable in our case we assume that λ has a gamma distribution with parameters α and β.

$\lambda \sim gamma(\alpha, \beta)$

$g(\lambda) = \frac{\alpha^\beta}{\Gamma\beta}\lambda^{\beta-1}e^{-\alpha\lambda}$

The joint density of N and λ is:

$$\begin{aligned}
P(N = n) &= \int_0^\infty P(N = n|\Lambda = \lambda)g\left(\lambda\right)d\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^n}{n!}\frac{\alpha^\beta}{\Gamma\beta}\alpha^{\beta-1}e^{-\alpha\lambda}d\lambda \\
&= \frac{\alpha^\beta}{\Gamma\beta n!}\int_0^\infty \lambda^{n+\beta-1}e^{-(\alpha+1)\lambda}d\lambda
\end{aligned}$$

We introduce a normalization parameter

$\frac{\Gamma(n+\beta)}{(\alpha+1)^{(n+\beta)}}$ This reduces to an integrand of a gamma distribution with parameters $n+\beta$ and $\alpha+1$ this is equal to one.

$$= \frac{\Gamma(n+\beta)}{(\alpha+1)^{(n+\beta)}}\frac{\alpha^\beta}{\Gamma\beta n!}\int_0^\infty \frac{\Gamma(n+\beta)}{(\alpha+1)^{(n+\beta)}}\lambda^{n+\beta-1}e^{-(\alpha+1)\lambda}d\lambda \tag{3.127}$$

This is a negative binomial distribution a mixture of a family of poisson distribution with gamma mixing weights.

$$= \frac{\Gamma(n+\beta)}{\Gamma(n+1)\Gamma\beta}\left(\frac{\alpha}{\alpha+1}\right)^{\beta}\left(\frac{1}{\alpha+1}\right)^{n}$$
$$= \begin{pmatrix} n+\beta-1 \\ n \end{pmatrix}\left(\frac{\alpha}{\alpha+1}\right)^{\beta}\left(\frac{1}{\alpha+1}\right)^{n}$$

This discretizes the claims severity distributions This model is sometimes problematic in practical applications. It assumes that we have no negative increments in the incremental triangle.However if $X_{i,j}$ denotes incremental payments, we can have negative values. The negative values in a triangle can be accommodated by preserving the sum of the incremental column. The second way to allow for negative increments in the triangle is to opt for a Mixed Log normal-Normal distribution by replacing the truncated normal distribution with another skewed distribution, such as the log normal distribution. Define $X \sim N(z, \sigma)$ where $z \sim LN(\mu, \sigma)$. The loss distribution has the desired features of skewness and a domain that includes negative numbers, This can help us describe a model for incremental paid losses with a calendar-year trend.

$$PredictionVariance = ProcessVariance + EstimationVariance \tag{3.128}$$

The negative binomial model can be fitted using incremental or cumulative data. Unlike the over dispersed poisson model which has an origin and development year parameter, the parameters of a negative binomial model relate to development year only.

$$\begin{aligned}
E\left[X_{w,d}\right] &= (\lambda_d - 1)C_{w,d-1} \\
Var\left[X_{w,d}\right] &= \Phi\lambda_d(\lambda_d - 1)C_{w,d-1} \\
E\left[X_{w,d}\right] &= m_{w,d} = (\lambda_d - 1)C_{w,d-1} \\
log(m_{w,d}) &= log(\lambda_d - 1) + log(C_{w,d-1}) \\
log(\lambda_d - 1) &= c + \alpha_{d-1} \\
log(m_{w,d}) = c + \alpha_{d-1} + log(C_{w,d-1}) &
\end{aligned} \tag{3.129}$$

This specifies a generalized linear model with logarithmic link function and a negative error structure.

$$\begin{aligned}
Var\left[\overline{\lambda}_d\right] &= Var\left[\overline{\lambda}_d - 1\right] \\
&= \exp(\overline{C} - \alpha_{d-1})Var\left[\overline{C} - \alpha_{d-1}\right]
\end{aligned} \tag{3.130}$$

3.5.3.3 Shifted Pareto Distribution

Let the number of claims N be exponentially distributed N $\sim \exp(\lambda)$ and the claims sizes be gamma distributes $\lambda \sim \gamma(\alpha, \beta)$. The joint probability is a shifted pareto distribution that allows us to model policies with deductibles.

The joint density of N and λ is:

$$\begin{aligned}
P(N=n) &= \int_0^\infty P(N=n|\Lambda=\lambda)g\left(\lambda\right)d\lambda \\
&= \int_0^\infty \lambda e^{-\lambda x}\frac{\alpha^\beta}{\Gamma\beta}\alpha^{\beta-1}e^{-\alpha\lambda}d\lambda \\
&= \frac{\alpha^\beta}{\Gamma\beta}\int_0^\infty \lambda^\beta e^{-(\alpha+x)\lambda}d\lambda
\end{aligned}$$

We introduce a normalization parameter

$$= \frac{\alpha^\beta\Gamma(\beta+1)}{\Gamma\beta(x+\alpha)^{\beta+1}}\int_0^\infty \lambda^\beta\frac{(x+\alpha)^\beta}{\Gamma(\beta+1)}e^{-(x+\beta)\lambda}d\lambda \tag{3.131}$$

$$\tag{3.132}$$

But $\Gamma\alpha = (\alpha-1)!$

and $\Gamma(\alpha+1) = \int_0^\infty e^{-\lambda}\lambda^\alpha dx$

Hence this is a shifted Pareto distribution of the form

$$\frac{\beta\alpha^\beta}{(x+\alpha)^{\beta+1}} \tag{3.133}$$

It can be viewed as a compound gamma distribution because an exponential distribution is a special case of a gamma distribution.

3.5.3.4 Metropolis Algorithim

The metropolis algorithm provides an elegant method for obtaining sequences of random samples from probability distributions. The main idea is to draw samples from a distribution which can be evaluated from any point not necessarily an integral.

$$F(\theta|D) = \frac{F(D|\theta)\, F(\theta)}{F(D)} \tag{3.134}$$

F(D) is the unconditional probability of observing the data. This is not dependant on the parameters of the model θ on which we wish to perform the inference, hence $F(D)$ is effectively a normalizing constant which makes $F(D|\theta)$ a proper probability density function. In case we have non-normalized probability density functions which we wish to estimate by taking random samples. The process of generating random samples may be very difficult for complex models, we therefore need to explore the distributions using Markov chain.

The chain we need is that if we run it long enough will constitutes as a whole of random sample from the distribution of interest. This property of the Markov chain is called ergodicity. The metropolis Hastings algorithm is a well structured model for constructing such a chain. This algorithm can be constructed as follows.

- Choose a vantage point (starting point) in the parameter space X.
- Choose a candidate point such that $Y \sim N(X, \sigma)$ this is defined as the proposal distribution.
- Move the candidate point with probability $\min\left(\frac{\Phi(Y)}{\Phi(X)}, 1\right)$
- Repeat this over the number of iterations

The limitation of this approach is the selection of σ. Efficient mixing (the rate at which the chain converges to the target distribution) occurs when σ approximates the standard deviation of the target distribution. When the value of σ is not known in advance we allow σ to adapt that of the history of the chain so far.

A Markov Chain is Stochastic process in which future states are independent of past states given the present state. A Stochastic Process is a consecutive set of the random quantities indexed at time (t) defined on some known state space Θ, where Θ is a parameter space.

$$P\left(\theta^{t+1}|\theta^{1},\theta^{2},...,\theta^{t}\right)=P\left(\theta^{t+1}|\theta^{t}\right) \tag{3.135}$$

The Markov chain is a bunch of draws of θ that are slightly dependant on the previous one. The chain wanders around the state space, remembering only where it has been in the last period. The jumping rule is governed by a transition kernel which is a mechanism that describes the probability of moving to some other state based on the current state. The rows sums to 1 we therefore have to define a conditional Probability Mass Function, conditional on the current a state. The column are the marginal probabilities of being in a certain state in the next period.

$$M=\begin{bmatrix} P\left(\theta_A^{t+1}|\theta_A^t\right) & P\left(\theta_B^{t+1}|\theta_A^t\right) & P\left(\theta_C^{t+1}|\theta_A^t\right) \\ P\left(\theta_A^{t+1}|\theta_B^t\right) & P\left(\theta_B^{t+1}|\theta_B^t\right) & P\left(\theta_C^{t+1}|\theta_B^t\right) \\ P\left(\theta_A^{t+1}|\theta_C^t\right) & P\left(\theta_B^{t+1}|\theta_C^t\right) & P\left(\theta_C^{t+1}|\theta_C^t\right) \end{bmatrix}$$

For a continuous state space infinite possible states the transition kernel is a bunch of conditional Probability Density Function $f\left(\theta_j^{t+1}|\theta_i^t\right)$. The steps are as follows.

- Define a starting distribution $\pi^{(\theta)}$ a $1 \times k$ vector of probability that sums to 1.

- The first iteration of our distribution $\pi^{(1)}$ from which θ^1 is drawn is

$$\pi^{(1)} = \pi^{(0)} \times P \tag{3.136}$$

  where P is a $k \times k$ matrix

- At iteration 2 the distribution of $\pi^{(2)}$ from which θ^2 is drawn is

$$\pi^{(2)} = \pi^{(1)} \times P \tag{3.137}$$

- At iteration t our distribution $\pi^{(t)}$ from which θ^t is drawn is

$$\begin{aligned}\pi^{(t)} &= \pi^{(t-1)} \times P \\ &= \pi^{(0)} \times P^t\end{aligned} \tag{3.138}$$

Define a stationary distribution π to be same distribution as Φ such that

$$\pi = \pi P \tag{3.139}$$

For all MCMC algorithm we use the Bayesian statistics, the Markov chain will typically converge to π regardless of our starting point. We can devise a Markov chain whose stationary distribution π is our desired posterior distribution $P(\theta|y)$, such that when we run this chain we get draws that are approximately from $P(\theta|y)$ once the chain converges.

Regardless the starting point the chain will always converge but the time taken to converge varies depending on the starting point. In practice we throw out a certain number of the first draws a process known as burn in. This makes our draws closer to stationary distribution and less dependant on the starting point. Once we obtain a Markov chain that converges to the stationary distribution then the draws from the Markov chain appear to be like those drawn from $P(\theta|y)$.

We can therefore use Monte Carlo simulation to find quantities of interest. The draw back is that the draws are not independent which is a requirement for Monte Carlo to work. Evoking the ergodic theorem, Let $\theta^1, \theta^2, ..., \theta^m$ be M values from a Markov chain that is aperiodic, irreducible and positive recurrent the chain is ergodic and $E[g(\theta)] \leq \infty$ with probability 1.

$$\frac{1}{m}\sum_{i=0}^{\infty} \rightarrow \int_{0}^{\infty} g(\theta)\,\pi(\theta)\,d\theta \tag{3.140}$$

This is the Markov chain analog that allows us to ignore the dependence between draws of the Markov chain when calculating quantities of interest. A markov chain is egordic if:

- The Markov chain is aperiodic - if the only length of time for which the chain repeats same cycle of values is the trivial case with cycle length equal to one. A
- The Markov chain is then irreducible - if it is possible to go from any state to any other state not necessarily in one step.
- The Markov chain is recurrent - if for any give state i (if the chain starts at i) it will eventually return to it with a probability of 1.
- And if it is positive recurrent - if the expected return time to state i is infinite otherwise it is null recurrent.

To break the dependence between draws in the Markov chain we only keep every d^{th} draw of the chain a process called thinning. This is however not necessary with ergodic theorem. It also tends to increase the variances. Therefore MCMC is a class of methods in which we can stimulate draws that are slightly dependent and are approximately from a posterior distribution. In Bayesian statistics we have two MCMC algorithms that we can use.

- Gibbs Sampler algorithm
- Metropolis-Hastings algorithm

In this project we take advantage of Metropolis-Hastings algorithm due to its ease in implementing. It is particularly useful especially when the posterior distribution doesn't look like any known distribution that is there is no conjugacy, when the posterior distribution has more than two parameters and when some or all of the full conditionals do not look like any known distribution we know and we cannot do Gibbs Sampling.

3.5.3.5 Fast Fourier Transform

The Fast Fourier Transform is an efficient method to calculate compound distributions via the inversion of the characteristic function. The method has been known for many decades and originates from the signal processing. The existence of the algorithm became generally known recently in the mid-1960s. It is now being used to model aggregate loss distributions in the insurance industry. FFT works with discrete severity and based on the discrete Fourier transformation. It can be defined as follows.

If F:$[0, L] \longrightarrow C$ be a Riemann Integrable function with $f(0) = f(L)$. The k^{th} complex fourier- coefficient of f is defined as

$$
\begin{aligned}
f_k &= \frac{1}{L}\int\limits_{0}^{L} f(x)e^{-2\pi i\frac{k}{L}X}dx \\
f(x) &= \sum_{k=-\infty}^{\infty} \overline{f}(k)e^{2\pi i\frac{k}{L}X}
\end{aligned}
\tag{3.141}
$$

3.5.3.6 Discrete Fourier Transform

Let $x = (x_0, ..., x_{N-1})$ then the Discrete Fourier Transform of x is defined as

$$\overline{x}_k = \frac{1}{N}\sum_{j=0}^{\infty} X_K e^{2\pi i\frac{k}{L}j} \tag{3.142}$$

$x(k)$ is the DFT of the N-point sequence $x(n)$

$$
\begin{aligned}
x(k) &= \sum_{n=0}^{N-1} x(n)e^{2\pi i\frac{k}{L}n} \\
x(k) &= \sum_{n=0}^{N-1} x(n)W_N^{nk}
\end{aligned}
\tag{3.143}
$$

$$W_N = cos\left(\frac{2\pi}{N}\right) - isin\left(\frac{2\pi}{N}\right) \tag{3.144}$$

The characteristic function (CF) of any random variable X completely defines its probability distribution on a real line and it is given by the following formula:

$$f_X(u) = E\left[e^{iuX}\right] = \int\limits_{-\infty}^{\infty} e^{iuX} f_X(x)dx \tag{3.145}$$

Where u is a real number, i is the imaginary unit, and E denotes the expected value, $f_X(x)$ denotes the probability density function (PDF). The characteristic function is defined on the whole number line. There is a bijection between CDF and CFs in the sense that two distinct probability distributions can never share the same characteristic function.

Given a CF ϕ, it is possible to reconstruct the corresponding CDF.

$$f_X(y) - f_X(x) = lim_{\tau\to\infty} \frac{1}{2\pi} \int_{-\tau}^{+\tau} \frac{e^{-iux} - e^{-iuy}}{iu} \Phi_x(u) du \tag{3.146}$$

The inversion of the equation above is

$$f_X(x) = \frac{1}{2} + \frac{1}{2\pi} \int_{0}^{\infty} \frac{e^{iux}\Phi_x(u) - e^{-iux}\Phi_x(u)}{iu} du \tag{3.147}$$

where

$$\begin{aligned}
e^{iux} &= \cos ux + i \sin ux \\
I &= \int_{0}^{\infty} \frac{e^{iux}\Phi_x(-u) - e^{-iux}\Phi_x(u)}{iu} du \\
&= \int_{0}^{\infty} \int_{-\infty}^{\infty} \frac{e^{iux}\Phi_x(-u) - e^{-iux}\Phi_x(u)}{iu} du \\
&= \int_{0}^{\infty} \int_{-\infty}^{\infty} \frac{2 \sin u(x-z)}{u} dF(Z) du \\
&= \pi(2F(x) - 1) \\
&= \pi(2\left(\frac{1}{2} + \frac{1}{2\pi} I\right) - 1)
\end{aligned} \tag{3.148}$$

In our case the characteristic function of the severity density $f(x)$ is defined as

$$y = \int_{-\infty}^{\infty} f(x)e^{-itx}dx \tag{3.149}$$

The probability distribution of the frequency data set is discrete. The probability generating function.

$$Pr\left[N = k\right]$$
$$y = \sum_{k=0}^{\infty} S^k P_k$$

For a compound loss Z model, the characteristic function of a compound loss Z function denoted as $X(t)$ can be expressed through the probability generating function of the frequency distribution and characteristic function of the severity distribution.

$$X(t) = \sum_{k=0}^{\infty} (\varphi(t))^k P_k$$
$$y = \sum_{k=0}^{\infty} S^k P_k$$

If N is poisson distributed Poi(N).

$$\begin{aligned} X(t) &= \sum_{k=0}^{\infty} (\varphi(t))^k \frac{e^{-\lambda}\lambda^k}{k!} \\ &= e^{\lambda\varphi(t)-\lambda} \end{aligned} \tag{3.150}$$

N is from a negative binomial distribution

$$\begin{aligned} X(t) &= \sum_{k=0}^{\infty} (\varphi(t))^k \binom{k+m-1}{k} (1-p)^k p^m \\ &= \left(\frac{1}{1-(1-p)\varphi(t)} \right)^m \end{aligned} \tag{3.151}$$

Given the characteristic function, the density of aggregate loss Z can be calculated using the FFT.

$$h(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(t) e^{-itz} dt$$

For non-negative random variable Z

$$\begin{aligned} h(z) &= \frac{1}{2\pi} \int_{0}^{\infty} Re(x(t)) \cos(tz) dt \\ H(z) &= \frac{1}{2\pi} \int_{0}^{\infty} Re(x(t)) \frac{\sin(tz)}{t} dt \end{aligned} \tag{3.152}$$

Chapter 4

Data Analysis

The approach taken in the data analysis was first to structure data in a format it can be easily triangulated. The loss date or accident date, the notification dates and the valuation date or pay dates were important data fields in enabling computations of lag periods. The claims data is segmented into two data sets the paid and the incurred data set. The genesis of the analysis begins from the chain ladder deterministic model and build to statistical models and eventually to the stochastic methods. The assumption that the proportional development of claims from one development period to the next is the same for all origin years is not violated in all the models. To accomplish all the tasks the models are implemented in R a rich language for statistical modelling and data manipulations that allows fast prototyping and easily interfaces with other applications.

4.1 Chain Ladder

Historical claims data is often presented in form of a triangle structure, showing the development of claims over time for each exposure (origin) period. In this analysis we consider the origin year to be the loss or accident year, this can be defined in otherwise for example the year the policy was written or earned. The origins period can be yearly, quarterly or monthly, the choice of the origin period depends on outcome of an analysis of claims occurrence to partial and closure turn around times.

We will refer to the development period of an origin period is this paper as age or lag. Data on the diagonals represent payments in the same calendar period. The triangles emanate from data of individual policies that has been aggregated from homogeneous lines of business, division levels or perils. Below is a sample of the data ready for modelling in R.

TABLE 4.1: Structured Claims Data

w	d	premium	cpdloss	incloss
1	1	5812	952	1722
2	1	4908	849	1581
3	1	5454	983	1834
4	1	5165	1657	2305
5	1	5214	932	1832
6	1	5230	1162	2289
7	1	4992	1478	2881
8	1	5466	1240	2489
9	1	5226	1326	2541
10	1	4962	1413	2203

Where the first column represents the accident year, the second column represents the development periods, the third holds the earned premiums in the periods while the last two columns represent the paid and incurred losses respectively.

4.1.1 Incremental Triangle

The incremental triangle considered in this project was based on paid claims. England and Verrall(2000) point out that the using incremental losses requires one to use paid rather than incurred losses since the most severity distributions are defined on non negative losses. Incurred losses include estimates by claims adjusters that can be adjusted downward. Negative incremental paid losses occasionally occur because of salvage and subrogation, but a feature of the GLM and severity distribution parameter estimation procedure allows for negative incremental losses as long as all column sums of the loss triangle remain positive.

Table 4.2: Incremental Triangle

	d									
w	1	2	3	4	5	6	7	8	9	10
1	1,722	2,108	- 227	232	38	22	23	-	- 1	-
2	1,581	611	336	5	- 5	2	4	7	- 3	
3	1,834	1,175	479	512	105	- 18	25	58		
4	2,305	1,168	240	305	277	39	9			
5	1,832	793	461	407	28	42				
6	2,289	871	- 6	50	- 14					
7	2,881	1,373	587	335						
8	2,489	467	426							
9	2,541	766								
10	2,203									

4.1.2 Cumulative Triangle

From the rectangular data above that consists of 55 entries of incremental losses we structure it to a triangle and use it as our training data. The chain ladder approches require that the losses are considers as a cumulation.

As defined in chapter 3 w represents the accident years while d represents the development lags.

TABLE 4.3: Cumulative Triangle

	d									
w	1	2	3	4	5	6	7	8	9	10
1	1,722	3,830	3,603	3,835	3,873	3,895	3,918	3,918	3,917	3,917
2	1,581	2,192	2,528	2,533	2,528	2,530	2,534	2,541	2,538	NA
3	1,834	3,009	3,488	4,000	4,105	4,087	4,112	4,170	NA	NA
4	2,305	3,473	3,713	4,018	4,295	4,334	4,343	NA	NA	NA
5	1,832	2,625	3,086	3,493	3,521	3,563	NA	NA	NA	NA
6	2,289	3,160	3,154	3,204	3,190	NA	NA	NA	NA	NA
7	2,881	4,254	4,841	5,176	NA	NA	NA	NA	NA	NA
8	2,489	2,956	3,382	NA	NA	NA	NA	NA	NA	NA
9	2,541	3,307	NA	NA	NA	NA	NA	NA	NA	NA
10	2,203	NA	NA	NA	NA	NA	NA	NA	NA	NA

4.1.2.1 Classical Chain Ladder Analysis

This run off triangle shows the known values of loss from each origin year as of the end of the origin year and annual evaluations thereafter.For example, the known values of loss originating from the 8 exposure period are 2,489, 2,956, and 3,382 as of year ends 1, 2, and 3, respectively. The most recent diagonal, i.e., the vector 3,917, 2,538 . . . 2,263 from the upper right to the lower left, shows the most current evaluation available. The column headings show the years of the observations in the column relative to the beginning of the exposure period. The objective of a reserving exercise is to forecast the future claims development in the bottom right corner of the triangle and potential further developments beyond development year 10. Eventually all claims for a given origin period will be settled, but it is not always obvious to judge how many years or even decades it will take. We speak of long and short tail business contingent to the time it takes to pay all claims.

Table 4.4: Age To Age Factors

	1 - 2	2 - 3	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8	8 - 9	9 - 10
1	2.224	0.941	1.064	1.010	1.006	1.006	1.000	1.000	1.000
2	1.386	1.153	1.002	0.998	1.001	1.002	1.003	0.999	
3	1.641	1.159	1.147	1.026	0.996	1.006	1.014		
4	1.507	1.069	1.082	1.069	1.009	1.002			
5	1.433	1.176	1.132	1.008	1.012				
6	1.381	0.998	1.016	0.996					
7	1.477	1.138	1.069						
8	1.188	1.144							
9	1.301								
smpl	1.504	1.097	1.073	1.018	1.005	1.004	1.006	0.999	1.000
vwtd	1.479	1.09	1.076	1.02	1.005	1.004	1.006	0.999	1.000

4.1.2.2 Link Ratios Analysis

Most commonly as a first step, the age-to-age link ratios are calculated as the volume weighted average development ratios of a cumulative loss development triangle from one development period to the next. The bottom rows represent the simple average factors as well as the volume weighted factors.

In 1993 Thomas Mack introduced a method of estimating the standard errors of the Chain Ladder forecast without assuming a distribution under three conditions. Mack's(1992) chain ladder method calculates the standard error for the reserves estimates. The method works for a cumulative triangle if the following assumptions hold.

$$\begin{aligned} E\left[\frac{C_{w,d+1}}{C_{w,k}}|C_{w,1}, C_{w,2}, ..., C_{w,d}\right] &= \Lambda_d \\ Var\left[\frac{C_{w,d+1}}{C_{w,k}}|C_{w,1}, C_{w,2}, ..., C_{w,d}\right] &= \frac{\sigma_d^2}{m_{w,d}C_{w,d}} \end{aligned} \tag{4.1}$$

$C_{w,1}, ..., C_{w,n}, C_{d,1}, ..., C_{d,n}$ are independent for origin periods $w \neq d$. If the assumptions above hold, the Mack Chain Ladder model gives an unbiased estimator for IBNR (Incurred But Not Reported) claims. The Mack Chain Ladder model can therefore be regarded as a weighted linear regression through the origin for each lag period.

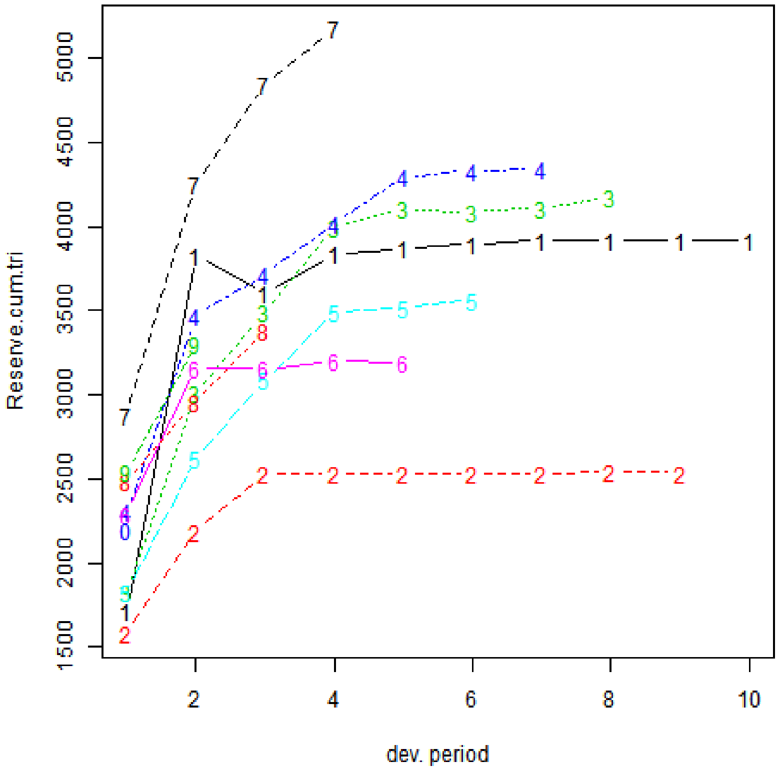

FIGURE 4.1: Claims development pattern (per origin period).

TABLE 4.5: Regression Statistics

Origin Years	Link Ratios	Residuals	R Squared
1	1.4792	12.9274	0.9696
2	1.0900	4.9917	0.9943
3	1.0756	2.9784	0.9981
4	1.0203	1.6610	0.9994
5	1.0047	0.4023	1.0000
6	1.0041	0.1450	1.0000
7	1.0062	0.4677	1.0000
8	0.9994	0.0363	1.0000
9	1.0000	NA	1.0000

To check that Mack's assumption are valid we review the residual plots, if there are no possible trends then the model is valid. For each origin year we can examine the

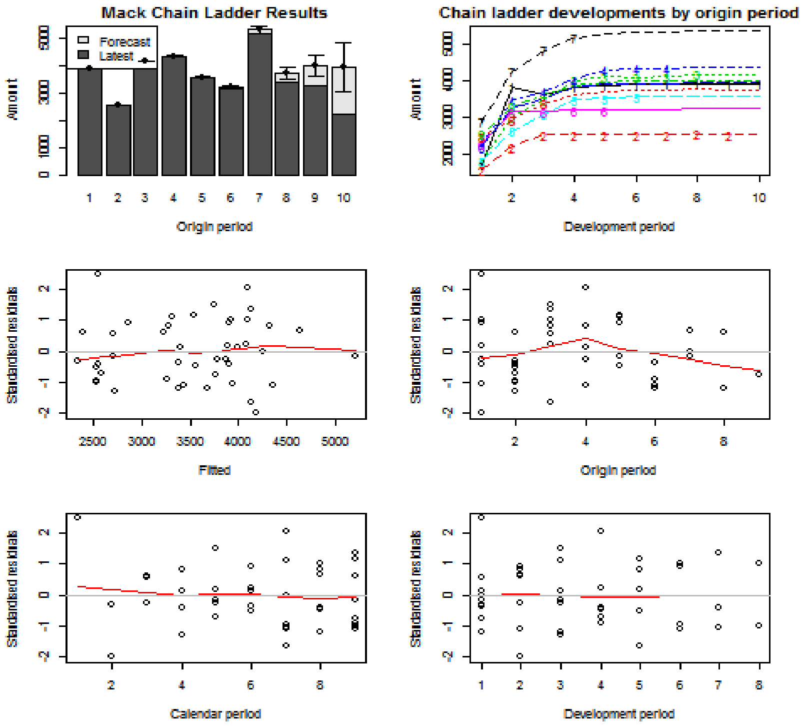

FIGURE 4.2: Residual Plot (Test validity of the model).

standard errors, the expectation is that for the less developed years the standard error is deviations are high compared to the fully run off years.

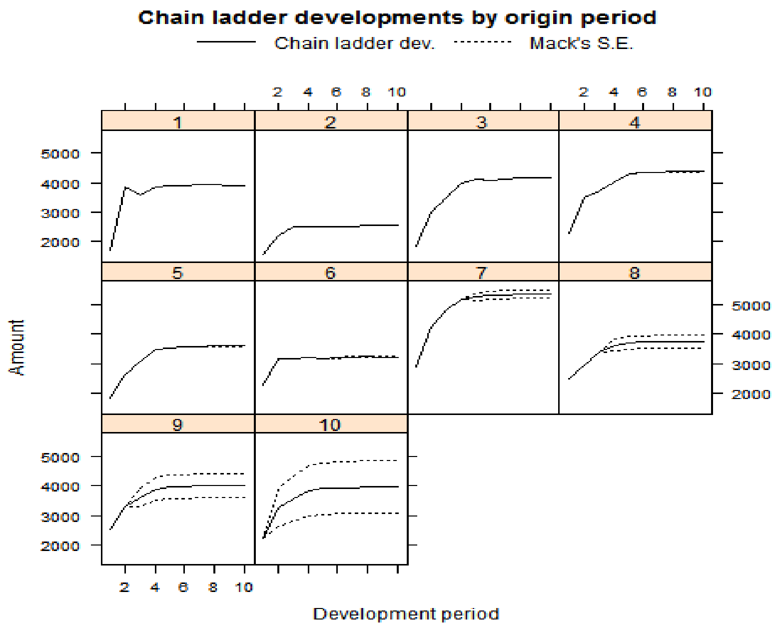

FIGURE 4.3: Standard Error Plot (Per Origin Period).

4.1.2.3 Bootstrap Chain-Ladder

One objective of this paper of to measure the process variation, following the paper by by England and Verrall(2001) we use the bootstrapping and simulation approach and address this in two stages. In the first stage we use the ordinary chain-ladder method and apply it to the cumulative claims triangle. From this we calculate the scaled Pearson residuals which we bootstrap R times to forecast future incremental claims payments via the standard chain-ladder method. In the second stage we simulate the process error with the bootstrap value as the mean and using the process distribution assumed. The set of reserves obtained in this way forms the predictive distribution, from which summary statistics such as mean, prediction error or quantiles can be derived. We assume that the losses follow a gamma distribution. The bootstrap procedure is performed by completing

Table 4.6: Bootstrap Chain Ladder

	Latest	Mean Ult	Mean IBNR	IBNR.S.E	IBNR 75%	IBNR 95%
1	3,917	3,917	0	0	0	0
2	2,538	2,538	0	0	0	0
3	4,170	4,166	-4	42	0	19
4	4,343	4,365	22	88	31	166
5	3,563	3,588	25	102	46	193
6	3,190	3,233	43	111	68	218
7	5,176	5,352	176	203	273	557
8	3,382	3,765	383	295	530	937
9	3,307	4,005	698	368	904	1,363
10	2,203	3,962	1,759	737	2,190	3,092
	Total		3,102		4,043	6,545

the following steps, which can be performed without difficulty in R:

- Obtain the standard Chain-Ladder development factors from cumulative data.
- Obtain cumulative fitted values for the past triangle by backwards recursion, starting with the observed cumulative paid to date in the latest diagonal.
- Obtain incremental fitted values for the past triangle by differencing.

- Calculate the unscaled Pearson residuals for the past triangle.
- Calculate the Pearson scale parameter.

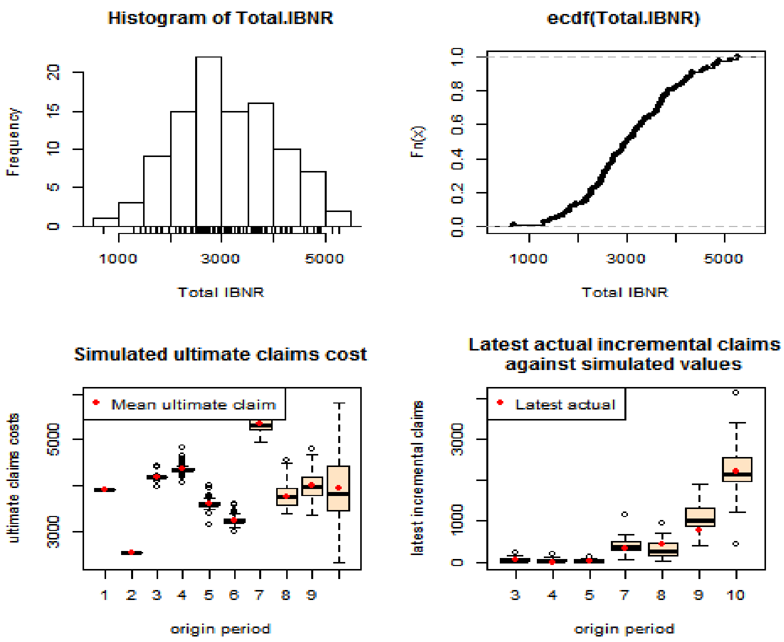

FIGURE 4.4: Bootstrap Plot (Simulation Analysis).

The distribution of the IBNR appears to follow a log-normal distribution.

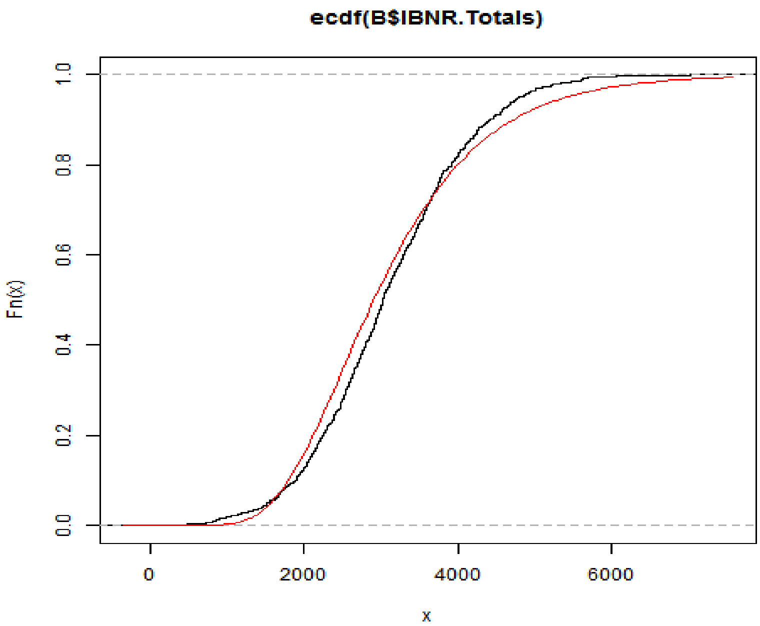

FIGURE 4.5: IBNR apperars to follow (log normal distribution).

4.1.2.4 LogNormal Model for Link Ratios

We assume that the age-to-age development factors are log normally distributed. The product of independent and log normal random variables is also log normal, which implies that age-to-ultimate loss development factors are log normal. Because the product of a constant and a log normal random variable is log normal, we have the cumulative paid loss ratio at all ages and the estimated parameters of the matching age-to-ultimate factor, we can therefore determine the parameter estimates of the ultimate loss ratio. Using these parameters we can estimate the IBNR and the expected loss ratio as well as confidence intervals around that estimate. The log normal parameters μ and σ of the age to age factors can be estimated by a variety of methods. In this analysis we apply the unbiased estimators. $\overline{y}$ is a maximum likelihood estimator.

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i)$$
$$\sigma^2 = \sum_{i=1}^{n} \frac{(y_i - \overline{y})^2}{n-1} \tag{4.2}$$

The table below shows the upper and lower bound for age to age factor. We set The

Table 4.7: Age-to-Age factors CI

Age-to-Age factors					
				2[2]*Log Normal 95% CI	
	Est mean	Est SD	Mean LDF	Lower Bound	Upper Bound
9–Ult*	0.000	0.001	1.000	0.999	1.001
LDF 8 - 9	-0.001	0.001	0.999	0.998	1.001
LDF 7 - 8	0.006	0.007	1.006	0.991	1.020
LDF 6 - 7	0.004	0.002	1.004	0.999	1.009
LDF 5 - 6	0.005	0.006	1.005	0.992	1.017
LDF 4 - 5	0.017	0.026	1.018	0.966	1.072
LDF 3 - 4	0.070	0.050	1.073	0.972	1.183
LDF 2 - 3	0.090	0.082	1.098	0.932	1.284
LDF 1 - 2	0.393	0.177	1.505	1.046	2.098

table below shows the 0.95 confidence interval for the age to ultimate factors assuming a log normal distribution.

TABLE 4.8: Age-to-Ultimate Factors Analysis

Age-to-Ultimate factors					
				2[2]*Log Normal 95% CI	
	Est mean	Est SD	Mean LDF	Lower Bound	Upper Bound
9–Ult*	0.000	0.001	1.000	0.999	1.001
8–Ult*	-0.001	0.001	0.999	0.997	1.001
7–Ult*	0.005	0.007	1.005	0.990	1.020
6–Ult*	0.009	0.008	1.009	0.993	1.024
5–Ult*	0.013	0.010	1.014	0.993	1.034
4–Ult*	0.031	0.028	1.032	0.975	1.090
3–Ult*	0.100	0.058	1.107	0.987	1.238
2–Ult*	0.190	0.100	1.216	0.994	1.471
1–Ult*	0.583	0.204	1.830	1.202	2.671

$$Dev_d = \beta(Lag/10|\mu, \sigma) - \beta(Lag - 1/10|\mu, \sigma) \quad (4.3)$$

4.1.2.5 Clark and Capecod Models

Where $\beta(Lag/10|\mu, \sigma)$ is the cumulative probability of a log normal distribution with unknown parameters μ and σ estimated as unbiased estimates. Other examples in this multitude include the models in Meyers(2007), who uses a model with constraints on the Dev_d parameters, and Clark(2003), who uses the Loglogistic and Weibull distributions to project Dev_d parameters into the future. Using a longitudinal analysis approach, Clark(2003) assumes that losses develop according to a theoretical growth curve. The LDF method is a special case of this approach where the growth curve can be considered to be either a step function or piecewise linear. Clark envisions a growth curve as measuring the percent of ultimate loss that can be expected to have emerged as of each age of an origin period. The LDF method assumes that the ultimate losses in each origin period are separate and unrelated. The goal of the method, therefore, is to estimate parameters for the ultimate losses and for the growth curve in order to maximize the likelihood of having observed the data in the triangle.

The CapeCod method assumes that the a prior expected ultimate losses in each origin year are the product of earned premium that year and a theoretical loss ratio. The CapeCod method, therefore, need estimate potentially far fewer parameters:for the growth function and for the theoretical loss ratio.

One of the side benefits of using maximum likelihood to estimate parameters is that its associated asymptotic theory provides uncertainty estimates for the parameters. Observing that the reserve estimates by origin year are functions of the estimated parameters, uncertainty estimates of these functional values are calculated according to the Delta method, which is essentially a linearisation of the problem based on a Taylor series expansion. The unknown parameters in this model are ELR_w (w=1,2,:::,10), representing the expected loss ratio for accident year AY,and Dev_d (d=1,2,:::,10), representing the incremental paid loss development factor. TheDev_d parameters are constrained to sum to one. The structure of the parameters is similar to the one above this stochastic model allows the ELR_w parameters to vary by accident year.

$$E[Loss_{w,d}] = Premiums_w * ELR_w * Dev_d \tag{4.4}$$

4.2 Stochastic Analysis

Let $X_{w,d}$ be a random variable for the loss in the training data (w,d). We describe the distribution of $X_{w,d}$ by the collective risk model, which can be described by the following simulation.

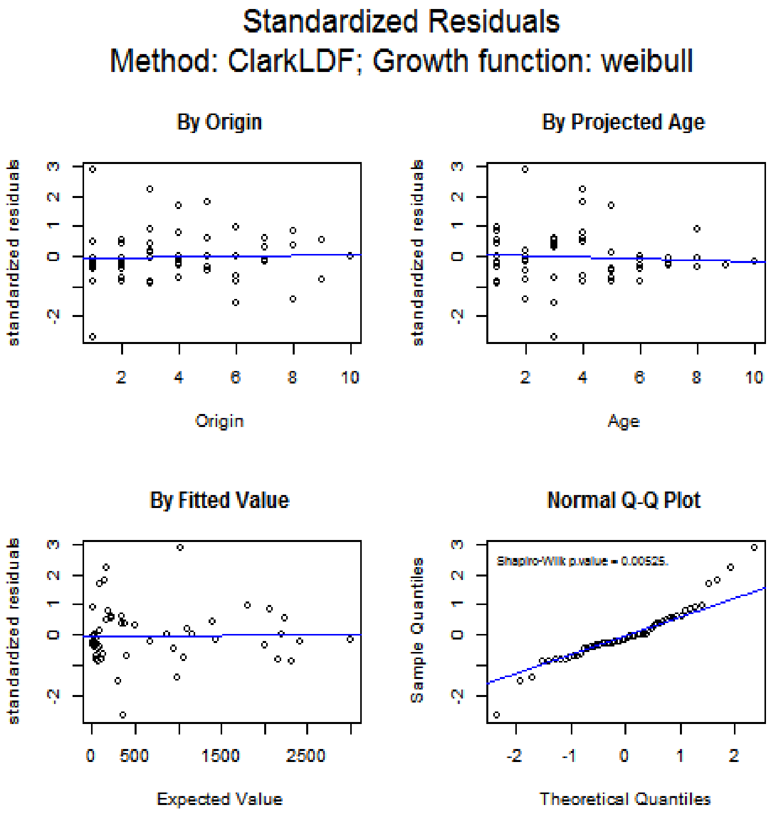

FIGURE 4.6: ClarkLDF Growth function: weibull(ClarkLDF Standardized Residuals Plots).

- Select a random claim count, $N_{w,d}$, d from a Poisson distribution with mean $\lambda_{w,d}$
- For i= 1, 2,..., $N_{w,d}$ select a random claim amount, $Z_{d,i}$
- Set $X_{w,d} = \sum_{i=1}^{N} Z_{d,i}$

The claim severity distribution of Z_d are given in this paper we the gamma distribution with the cumulative distribution function.

$$F(z) = \frac{\gamma(b, az)}{\Gamma(b)} \tag{4.5}$$

Set b = 2 for all settlement lags and a to vary with the settlement lags. The average severity increases with the settlement lag, which is consistent with the common observation that larger claims tend to take longer to settle. To summarize, we have two models (the Cape Cod and the log normal) that give $E[X_{w,d}]$ in terms of the unknown parameters $E[X_{w,d}]$ and Dev_d. We also assume that the claim severity distributions of Z_d are known. Then for any selected ELR_w and Dev_d, we can describe the distribution of $X_{w,d}$ by the following steps.

- Calculate $\lambda_{w,d} = \frac{E(X_{w,d})}{E(Z_d)} = \frac{Premiums_w * ELR_w * Dev_d}{E(Z_d)}$
- Generate the distribution of $X_{w,d}$ using the steps above.

4.2.0.6 The Posterior Distribution of Model Parameters

Let $L(X|ELR_w, Dev_d)$ be the likelihood of X given the parameters ELR_w, Dev_d. Using Bayes' Theorem, we can calculate the probability of the parameters ELR_w, Dev_d

$$Pr\left[(ELR_w, Dev_d)|X\right] \propto L(X|ELR_w, Dev_d) Pr\left[(ELR_w, Dev_d)\right] \tag{4.6}$$

Meyers(2007) calculates the likelihood of X by finely discretizing the claim severity distributions and using the Fast Fourier Transform (FFT) to calculate the entire aggregate loss distribution. The likelihood can be approximated using the Tweedie distribution in place of the collective risk model described in Simulation Algorithm. The Tweedie distribution can be viewed as a collective risk model with a Poisson claim count distribution and a gamma claim severity distribution.

We let λ be the mean of the Poisson distribution and θ_t and α_t be the scale and shape parameters of the gamma claim severity distribution.

$$f(x) = \Gamma(\alpha_t) \frac{e^{-\frac{x}{\theta_t}}}{x} \left(\frac{x}{\theta_t}\right)^{\alpha_t} \tag{4.7}$$

The expected claim severity is given by $\theta_t \alpha_t$ and the the claim severity variance is $\theta_t \alpha_t^2$.

4.2.0.7 Results

The results below are based on the basic chain ladder model. The two sets of results are based on the simple link ratios and the volume weighted link ratios. The chain-

TABLE 4.9: Simple and Volume weighted Results

	Diagonal	SimpleDF	VolumeDF	EstdUlt1	EstdUlt2	IBNR1	IBNR2
1	3,917	1.000	1.000	3,917	3,917	0	0
2	2,538	1.000	1.000	2,538	2,538	0	0
3	4,170	0.999	0.999	4,167	4,167	-3	-3
4	4,343	1.005	1.006	4,364	4,367	21	24
5	3,563	1.009	1.010	3,594	3,597	32	34
6	3,190	1.013	1.014	3,233	3,236	43	46
7	5,176	1.032	1.035	5,339	5,358	163	182
8	3,382	1.107	1.113	3,744	3,765	362	383
9	3,307	1.215	1.214	4,017	4,013	710	706
10	2,203	1.827	1.795	4,025	3,955	1,822	1,752
Total	35,789		NA	38,939	38,914	3,150	3,125

ladder model for volume weighted average link ratios is expressed as a weighted linear regression through the origin. Below is are the summary results of the regression model. The origin can be adjusted depending on the assumptions. Mack chain ladder approach

TABLE 4.10: Regression Statistics

Regression Statistics			
w	Link Ratios	Standard Error	R squared
1	1.4792	12.9274	0.9696
2	1.0900	4.9917	0.9943
3	1.0756	2.9784	0.9981
4	1.0203	1.6610	0.9994
5	1.0047	0.4023	1.0000
6	1.0041	0.1450	1.0000
7	1.0062	0.4677	1.0000
8	0.9994	0.0363	1.0000
9	1.0000		1.0000

forecasts future claims developments based on a historical cumulative claims development triangle and estimates the standard error around the estimates. If the classical chain ladder assumptions hold the present model can be regarded as a weighted linear regression through the origin for each development period. Under the bootstrap

Table 4.11: Mack Model Results

w	Latest	Dev.To.Date	Ultimate	IBNR	Mack.S.E	CV(IBNR)
1	3,917	1.0000	3,917	0	0	NA
2	2,538	1.0000	2,538	0	0	Inf
3	4,170	1.0006	4,167	-3	3	-1.1698
4	4,343	0.9945	4,367	24	37	1.5291
5	3,563	0.9904	3,597	34	34	0.9842
6	3,190	0.9858	3,236	46	40	0.8741
7	5,176	0.9661	5,358	182	146	0.8042
8	3,382	0.8982	3,765	383	225	0.5870
9	3,307	0.8240	4,013	706	412	0.5834
10	2,203	0.5570	3,955	1,752	878	0.5011
Totals	35,789	0.9197	38,914	3,125	1,057	0.3381

procedure it is possible to provide a predictive distribution of reserves or IBNRs for a cumulative claims development triangle. This approach uses a two-stage bootstrapping simulation approach. We apply the ordinary chain-ladder methods to the cumulative claims triangle. From this we calculate the scaled Pearson residuals which we bootstrap R times to forecast future incremental claims payments via the standard chain-ladder method. The process error can be simulated with the bootstrap value as the mean and using the process distribution assumed. The set of reserves obtained in this way forms the predictive distribution. For paid triangles, the distributions predicted by both the Mack and the bootstrap models tend to produce expected loss estimates that are too high. The explanation to this is:

- The claims handling department has experienced changes that are not observable at the current time. Improved efficiency in claims settlement and establishment of fraud risk frameworks that seem to be effective could be considered as possible changes.

- Model validation through qq plots.

TABLE 4.12: Bootstrap ODP and Gamma Models

Overdispersed Poisson distribution						
w	Latest	Mean Ultimate	Mean IBNR	SD IBNR	IBNR 75%	IBNR 95%
1	3,917	3,917	0	0	0	0
2	2,538	2,538	0	0	0	0
3	4,170	4,168	-2	41	0	17
4	4,343	4,362	19	89	34	159
5	3,563	3,590	27	103	45	215
6	3,190	3,232	42	111	64	254
7	5,176	5,354	178	210	276	580
8	3,382	3,774	392	283	534	895
9	3,307	4,042	735	407	950	1,520
10	2,203	3,988	1,785	721	2,198	3,077
Totals	35,789	38,965	3,176	1,020	3,796	4,927
Gamma Assumption						
w	Latest	Mean Ultimate	Mean IBNR	SD IBNR	IBNR 75%	IBNR 95%
1	3917	3,917	0	0	0	0
2	2538	2,538	0	0	0	0
3	4170	4,168	-2	42	0	22
4	4343	4,372	29	91	37	214
5	3563	3,593	30	101	44	204
6	3190	3,238	48	116	73	258
7	5176	5,356	180	204	290	555
8	3382	3,757	375	257	516	841
9	3307	4,015	708	392	922	1,450
10	2203	3,946	1,743	717	2,137	2,989
Totals	35789	38,900	3,111	1,014	3,671	4,957

The Cap Code approach assumes that the incremental losses across development periods in a loss triangle are independent. The other assumption is that the expected value of an incremental loss is equal to the theoretical expected loss ratio (ELR) times the on-level premium for the origin year times the change in the theoretical underlying growth function in our case we assume a weibull growth curve over the development period. To complete the model we wrap the expected values within an Overdispersed Poisson (ODP) process where the "scale factor" σ^2 is assumed to be a known constant for all development periods.

The parameters of Cape Cod method are therefore: ELR, and ω and θ. Finally, we use the maximum likelihood to parametrize the model that is uses the ODP process to estimate Process Risk, and uses the Cramer-Rao theorem and the delta method to estimate Parameter Risk.

Table 4.13: Cape Cod Results

w	Latest	Premium	ELR	GrowthFactor	Mean IBNR	UltValue	StdError	CV
1	3,917	5,812	74%	0.0023	10	3,927	38.317	3.798
2	2,538	4,908	74%	0.0036	13	2,551	44.026	3.319
3	4,170	5,454	74%	0.0058	23	4,193	59.299	2.541
4	4,343	5,165	74%	0.0093	36	4,379	73.930	2.071
5	3,563	5,214	74%	0.0154	60	3,623	96.669	1.621
6	3,190	5,230	74%	0.0263	102	3,292	127.642	1.251
7	5,176	4,992	74%	0.0467	173	5,349	166.206	0.962
8	3,382	5,466	74%	0.0875	355	3,737	238.900	0.674
9	3,307	5,226	74%	0.1792	694	4,001	327.675	0.472
10	2,203	4,962	74%	0.4437	1,633	3,836	494.365	0.303
Total	35,789	52,429			3,099	38,888	835.083	0.269

Collective Risk Model

The proposed model aims at testing the predictive power of the loss reserving models considered above. For the paid claims the methods considered above tend to understate the range of possible outcomes.

We use the Bayesian Markov Chain Monte Carlo model to improve the predictive power by recongnizing some elements implicit in the historical data.

From our Bayesian Framework:

$$Posterior = Prior \times Likelihood \tag{4.8}$$

We choose a prior and posterior distribution from the same family to obtain a conjugate distribution. Our prior is a poisson distribution while our likelihood is a gamma distribution.

4.2.0.8 Factor Model - Cape Code Model

The unknown parameters in this model are the expected loss ratios for the accident years and the development lags with parameters constrained to one.

$$E\left[Loss_{w,d}\right] = Premium_w \times ELR_w \times Dev_d \tag{4.9}$$

4.2.0.9 Log normal Model

In our earlier analysis we saw that the development factors parametrized on unbiased estimates of a log normal distribution yield similar results as those of the basic chain ladder,

$$Dev_{lag} = y\left(\frac{Lag}{10}|\mu,\sigma\right) - y\left(\frac{Lag-1}{10}|\mu,\sigma\right) \tag{4.10}$$

where $B(x|\mu,\sigma)$ is the cumulative probability of a log normal distribution with unknown parameters μ and σ. This model replaces the ten development lags parameters. We describe the distribution of $X_{w,d}$ by a collective risk model

4.2.0.10 Simulation

- Select a random claim count, $N_{w,d}$ from a poison distribution with mean $\lambda_{w,d}$.
- For $i = 1, 2, ..., N_{w,d}$ select a random claim $Z_{d,i}$
- Set $X_{w,d} = \sum_{i=1}^{N_{w,d}} Z_{d,i}$

From the model we assume that we have a claim severity distribution is known.

$$F_{\alpha,\lambda} = \int_0^{\Lambda} \frac{\lambda^{\alpha-1}e^{-\lambda}}{\Gamma(\alpha)} d\lambda$$

$$F_{\alpha,\lambda} = P(\Lambda \leq \lambda) = F\left(\alpha, \frac{\lambda}{\beta}\right) \tag{4.11}$$

Set $\alpha = 2$ and let β vary with settlement lags. Subject the claims to a limit of say one million.

With the Log Normal and Cape Code models we can obtain $E\left[X_{w,d}\right]$ in terms of unknown parameters namely the Expected Loss Ratio and the Development Lags. If the claims severity is known then for any selected ELR and Dev_{lag} we can describe the distribution of $X_{w,d}$ as

$$\lambda_{w,d} = \frac{E\left[X_{w,d}\right]}{E\left[Z_d\right]} \tag{4.12}$$

This is equivalent to

$$\lambda_{w,d} = \frac{Premiums_w \times ELR_w \times Dev_{lag}}{E\left[Z_d\right]} \tag{4.13}$$

We can calculate the likelihood of the losses by finely disretizing the claim severity distribution using FFT and calculate the entire aggregate distribution by multiplying the FFTs of the distribution.

Alternatively, we could use a tweedie distribution as a collective risk model with a poisson claim count and a gamma severity distribution. This takes us back to the Bayesian analysis discussed earlier. Given the gamma as the conditional distribution and the poisson as the prior distribution we can obtain the posterior distribution as

$$f(x) = \int_0^{\infty} f(x|\lambda) \times g(\lambda) d\lambda \tag{4.14}$$

The solutions to such equations may not have a closed form the metropolis algorithm comes in hardy to solve such equations especially when λ has many dimensions. Markov Chain Monte-Carlo method produces sample parameters that describe the posterior distribution. The Markov Chain will typically converge regardless of the starting point if it is ergodic. Monte-Carol is just a computer algorithm.

Table 4.14: Collective Risk Model Results

w	LCL.Est	LCL.S.E	LCL.CV	Mack.Est	Mack.S.E	Mack.CV	Actual
1	3,930	72	0.0183	3,917	0	0.0000	3,917
2	2,542	58	0.0228	2,538	0	0.0000	2,532
3	4,112	104	0.0253	4,167	3	0.0007	4,279
4	4,316	120	0.0278	4,367	37	0.0085	4,341
5	3,561	109	0.0306	3,597	34	0.0095	3,587
6	3,329	128	0.0384	3,236	40	0.0124	3,268
7	5,350	271	0.0507	5,358	146	0.0272	5,684
8	3,763	269	0.0715	3,765	225	0.0598	4,128
9	4,175	615	0.1473	4,013	412	0.1027	4,144
10	4,281	1,220	0.2850	3,955	878	0.2220	4,181
w=2:10	**35,429**	**1,570**	**0.0443**	**34,997**	**1,057**	**0.0302**	**36,144**

Chapter 5

Conclusions and Recommendations

5.1 Conclusions

The collective risk model provides a better framework for the range of reasonable estimates to quantify the uncertainty in estimation of a loss reserves. One source of this uncertainty comes from the statistical uncertainty of the parameter estimates, especially if one was to draw repeated random samples of data from the same underlying loss generating process, the collective risk model answers the question of how variable the estimates are from the mean loss reserve.

The problem of predicting the sum of the future outcomes is addressed by the Metropolis-Hastings algorithm having selected the Expected Loss Ratio and the Development Factors for the lags from iterations of a Metropolis-Hastings algorithm, the calculations of the posterior distribution of outcomes becomes conceptually easy by repeated use of the simulation algorithm.

The likelihood of the training data can calculated by finely discretizing the claim severity distributions and using the Fast Fourier Transform however this can be time consuming using a Tweedie distribution as a collective risk model with a Poisson claim count distribution and a gamma claim severity distribution leads to much faster running times for the algorithms and convergence of the process.

The collective risk model calculates the maximum likelihood estimates by searching over the space of ELR and the development factors subject to the constraint that the Dev_{lag} and up to 1.

The approach described in this project is a challenger to the practising actuaries and aims in striking a balance between importance of judgement in setting reserves and the experience gained by examining the reserves from different insurers. It incorporates prior knowledge through selection of an appropriate prior.

FFT has been used to calculate the predictive distribution. While it is very technical and hard to implement, it is faster and it produces more accurate results (relative to the model assumptions) making it worth for the industry to consider in the set of models to evaluate liabilities when the total outstanding claim size is thought to follow a Compound Poisson gamma distribution.

5.2 Recommendations

The stochastic claims reserving methods considered in this thesis predict the lower (unknown) triangle and assess the uncertainty of this prediction. For instance, Mack's uncertainty formula quantifies the total prediction uncertainty of the chain-ladder predictor over the entire run-off of the outstanding claims. Modern solvency considerations, such as Solvency II, require a second view of claims reserving uncertainty.

This second view is a short-term view because it requires assessments of the one year changes of the claims predictions when one updates the available information at the end of each accounting year. To achieve this one could force a constraint so that the last Dev_{Lag} parameters decrease by a constant factor, and incorporate possibility of the earlier origin period to continue developing.

If a claim severity distribution was not available, one could use the Tweedie distribution and treat the parameters as unknown. Exploring multivariate chain ladders allows multiple reserving triangles to be modelled and developed simultaneously. The advantage of the multivariate modelling is that correlations among different triangles can be modelled, this leads to more accurate uncertainty assessments.

The collective risk model doesn't incorporate all the information regarding the policy, in addition to the expected variation with operational time a seasonal effect could be inherent. Such complexity are better addressed using GLMs. Under the collective risk model it is difficult to extrapolate data points, GLM provides the framework within which such exploration can be carried out efficiently.

Finally incorporating the value at risk (VaR) and tail value at risk (TVaR) as risk measure at a confidence level, especially when modelling catastrophic events that can take a significant share of the risk capital of a company could play an important role in improving the technical liability estimates.

References

1. Alessandro Carrato, Markus Gesmann, Dan Murphy, Mario Wuthrich and Wayne. Zhang, Claims reserving with R, ChainLadder-0.2.0 Package Vignette , March 4, (2015)

2. Barnett, G., Zehnwirth, Best estimates for reserves, CAS Forum (2000)

3. Bornhuetter, R. L. and R. E. Ferguson, The actuary and IBNR, In Proceedings of the Casualty Actuarial Society, Volume 59, pp. 181-195, (1972).

4. Buhlmann, H., Schnieper, R. and E. Straub, Claims Reserves in Casualty Insurance based on a Probabilistic Model, (1980).

5. David R. Clark, LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach, Casualty Actuarial Society, (2003). CAS Fall Forum.

6. England, P. D. and R. J. Verrall, A Flexible Framework for Stochastic Claims Reserving, Proceedings of the Casualty Actuarial Society 88, (2001)

7. England, P. D. and R. J. Verrall, Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving, Insurance: Mathematics and Economics 25, (1999) pp. 281-293.

8. England, P. D. and R. J. Verrall, Stochastic Claims Reserving in General Insurance, British Actuarial Journal, (2002)

9. Glenn G. Meyers, Estimating Predictive Distributions for Loss Reserve Models, Variance Fall (2007), 1(2):248-272

10. Glenn G. Meyers, Stochastic Loss Reserving with the Collective Risk Model,Casualty Actuarial Society E-Forum, Fall (2008).

11. Glenn Meyers, The Leveled Chain Ladder Model for Stochastic Loss Reserving, Presented at the ASTIN Colloquim October 2, (2012).

12. Hayne, R. M, Measurement of Reserve Variability, Casualty Actuarial Society Forum 1, Fall 2003, pp. 141-172.

13. Hertig, J., A statistical approach to IBNR-Reserves in Marine Reinsurance, (1985)

14. Hurlimann, W., Credible loss ratio claims reserves: the Benktander, Neuhaus and Mack methods revisited, (2005)

15. Kremer, E., IBNR claims and the two way model of ANOVA, (1982)

16. Mack, Thomas., Measuring the Variability of the Chain Ladder Reserve Estimates, Casualty Forum (1994)

17. Mack.T, Distributionfree calculation of the standard error of chainladder reserve ASTIN Bull. 23, 213-225 (1993)

18. Mario V. Wuthrich and Michael Merz, Stochastic Claims Reserving Methods in Non-Life Insurance, (2006).

19. Mark R. Shapland, and Jessica (Weng Kah) Leong, Bootstrap Modeling: Beyond the Basics,Casualty Actuarial Society E-Forum (2010)

20. Martinez-Miranda, M.D., Nielsen, B., Nielsen, J.P. and Verrall, R, Cash ow simulation for a model of outstanding liabilities based on claim amounts and claim numbers, STIN Bulletin,41 (1), 107-129 (2011)

21. Meyers, Glenn G., and Peng Shi, The Retrospective Testing of Stochastic Loss Reserve Models, Casualty Actuarial Society E-Forum(Summer) 1-37, (2011)

22. Oran Brigham, The Fast Fourier Transform And Its Applications, Signal Processing Series, Prentice-Hall 81, (1988)

23. Pavel V. Shevchenko, Calculation of aggregate loss distributions, (2010).

24. Renshaw, A.E., Verrall, R.J, A stochastic model underlying the chain ladder technique.(1998)

25. Richard. J. Verrall, Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion, Casualty Actuarial Society 32, (2007), 443-458.

26. Schmidt, K. D, Methods and Models of Loss Reserving Based on Run-O Triangles, Casualty Actuarial SocietyFor um, Fall (2006).

27. Taylor, G. C, Claims Reserving In Non-Life Insurance, Claims Reserving In Non-Life Insurance, (1986).

28. Taylor, G. C. and G. McGuire, Loss reserving GLMs, (2000)

29. Thomas Mack, Distribution-free calculation of the standard error of chain ladder reserve estimates, Journal of Financial and Quantitative Analysis 22, Sept. (1987), 277-283

30. Thomas Mack, The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor, Astin Bulletin, Vol.29(2):261 - 266, (1999).

31. Verrall, Nielsen, J.P. and Jessen, A, Prediction of RBNS and IBNR claims using claim amounts and claim counts,ASTIN Bulletin, 40(2), 871 - 887 (2010)

32. Verrall, R. J. and England, P. D., An investigation into stochastic claims reserving models and the chain ladder technique., British Actuarial Journal,(2000)

33. Verrall, R.J, An investigation into stochastic claims reserving models and the chainladder technique, (2000).

34. Wacek, Michael G, The Path of the Ultimate Loss Ratio Estimate, Variance,Fall 2007.

Appendix A

Appendix: Process, Parameter and Model Risk

Stochastic modelling aims at addressing the process risk, parameter risk and model risk. One way to describe process and parameter risk is to consider the relationship for a random variable X conditioned on a parameter say θ.

$$Var\,(X) = E_{\theta}\,[[Var(X|\theta)]] + Var_{\theta}\,[E(X|\theta)] \tag{A.1}$$

The left hand side of the equation represents the total risk.

The first term represents the average variance of the outcomes from the expected result.

The second term represents the parameter risk the variance due to the many possible parameters in the posterior distribution. This is also referred to as the range of reasonable estimates.

The MCMC model sample simulates N number of θ parameters in the CCL model.

Hence:

$$TotalRisk = Var\left[\sum_{w=1}^{10} C_{w,10}\right] \tag{A.2}$$

Using the formula for the mean of a lognormal distribution we can calculate.

$$ParameterRisk = Var_{\theta}\left[E\left[\sum_{w=1}^{10} C_{w,10}|\theta\right]\right] = Var\left[\sum_{w=1}^{10} e^{\mu_{w,10}+\frac{\sigma_{10}^{2}}{2}}\right] \tag{A.3}$$

Model risk is the risk that one did not select the right model. We can formulate a model as a weighted average of the candidate models, with the weights as parameters. If the posterior distribution of the weights assigned to each model has significant variability, this is an indication of model risk. Viewed in this light, model risk is a special case of parameter risk. If we run the model over many simulations the parameter risk will shrink towards zero and any remaining risk, such as model risk, will be interpreted as process risk. The total risk is more preferred as this is the only risk we can test by looking at the actual outcomes.