Spring 2016

On the double chain ladder for reserve estimation with bootstrap applications

Larissa Schoepf

ON THE DOUBLE CHAIN LADDER FOR RESERVE ESTIMATION WITH BOOTSTRAP APPLICATIONS

by

LARISSA SCHOEPF

A THESIS

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN APPLIED MATHEMATICS
WITH STATISTICS EMPHASIS

2016

Approved by

Dr. Akim Adekpedjou, Advisor
Dr. Gregory Gelles
Dr. V.A. Samaranayake

ABSTRACT

To avoid insolvency, insurance companies must have enough reserves to fulfill their present and future commitment-refer to in this thesis as *outstanding claims* towards policyholders. This entails having an accurate and reliable estimate of funds necessary to cover those claims as they are presented. One of the major techniques used by practitioners and researchers is the *single chain ladder* method. However, though most popular and widely used, the method does not offer a good understanding of the distributional properties of the way claims evolve. In a series of recent papers, researchers have focused on two potential components of outstanding claims, namely: those that have incurred but not reported (IBNR), and those that are reported but not settled (RBNS). The deep analysis of those has led to improvements in the chain ladder technique leading to the so-called *double chain ladder* method in a reference to the two steps application of the single chain ladder. First to RBNS, and then to IBNR. Although this new technique of estimating outstanding claims is a significant improvement over the single chain ladder, there are still room for better. This thesis is based on the most up to date work in the area that is presented in a paper by Miranda, Nielsen, Verrall, and Wüthrich [13]. Using the machinery of stochastic processes, the authors outline how a possible inflation of the loss distribution over the years and distributional properties of future claims can be incorporated into the analysis leading to a better estimate of the reserves. We discuss in details those new breakthroughs, and, apply them to bootstrapped run-off triangle data. We assess the new methods with respect to the existing ones and provide a discussion and recommendation to practitioners.

ACKNOWLEDGMENTS

I would first like to thank my advisor Dr. Akim Adekpedjou. The door to his office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right direction whenever he thought I needed it.

I also would like to acknowledge my parents Ilona and Juergen Schoepf, my sister Alena Schoepf and my boyfriend Sebastian Muth for their support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Table | Page

1 INTRODUCTION

Insurance companies have to make sure that enough reserves are available to meet the demand of present and outstanding claims as a result of the occurrences of events as outlined in the contract between policyholders and insurers. Those events can be, but not limited to: properties losses, lump sum payment resulting from life insurance policies as a result of death, claims pertaining to health insurances policies, annuities or disabilities benefits. There are two major groups of insurance policies: property and casualty-so called non-life, and life insurance. This thesis focuses on the former. One buys a policy to cover for an unexpected or partial loss of a property due to accident, storm, damage, theft, vandalism etc... Because of the unpredictability of these occurences, insurers can suddenly face the possibility of paying claims as the policy dictates, and those can be very large in some cases. Therefore, a large amount of money needs to be reimbursed to policyholders under the terms of the contract. For instance, the hurricane Katrina back in 2006, or the storm in Joplin in 2011 have caused many people to lose their houses, businesses, cars etc... In both cases, we are talking about claims from policyholders in millions, if not billions. Though insurance companies can reinsure their insurance contracts to avoid situations like the previous two examples, however, most of the time, they need to have enough reserves to cover claims in order to ensure the financial stability of the company and its profits and losses accounts, since those depend on archives claims, but also on the forecasted claims yet to be settled.

Forecasting futures claims entails having better knowledge of history of claims in the company-that is past experiences are keys. Claims settled and claims filed but not yet settled are presented in the actuarial jargon in a triangle format, called *run-off triangle*. See Table 2.1 for an example. Any statistical method for estimating reserves is based on the triangle of data. For a detailed account and explanation on how the triangle evolve, we refer the reader to B. Ajne [1], A. Renshaw [17] and R. J. Verrall ([22], [23] and [24]).

There are various techniques in the literature for estimating outstanding claims, cf. again the previous references and other references therein. One of the methods extensively used by researchers and practitioners is the *single chain ladder* method, SCL from now on. This is the most celebrated and well known method of having a good understanding of outstanding liabilities in non-life insurance. It was originally a simple algorithm, appealing and one that gives reasonable estimates of outstanding claims. Later, it was connected to mathematical statistics by researchers who developed sound statistical methods for the SCL via maximum likelihood, regression models, Bayesian estimation etc., cf. Mack [8], Renshaw and Verrall [18]. See also Verrall ([25] and [26]), England and Verrall ([5] and [4]) for a detailed and comprehensive account of the various statistical methods for SCL. The models in those papers brought up many improvements on the SCL over the years and continue to make it the method mostly used for estimation of outstanding claims.

However, the SCL has some shortcomings. While it gives estimates of outstanding claims, better understanding of how claims evolve over the years is crucial. Observe that the SCL only forecasts outstanding claims, which include those that are reported but not settled-called RBNS, but does not include those that have occurred but are not yet reported-called IBNR. So the reserve should include both IBNR and RBNS, that is

$$\textbf{Reserve} = \textbf{IBNR} + \textbf{RBNS}.$$

A better understanding of these two parts will lead to a much better estimate of outstanding claims. In recognition of that, in a recent series of papers, Verrall et al. [28], Martínez-Miranda et al. ([14] and [16]) have deeply analyzed the claim generating process in order to have a better understanding of outstanding claims. Those improvements over the SCL are based on a two steps analysis of the data, each of which is a simple application of the former method because both sets of data can be represented in a run-off triangle format required for reserve estimation. The authors now call this *double chain ladder* (DCL) technique in reference to the two steps analysis. Those improvements were possible using the stochastic processes machinery that is important in the modeling and analysis of claims data.

Verall, Nielsen and Jessen [28] focused on the split between IBNR and RBNS delay. They used the run-off triangle of paid claims and the number of reported claims to propose a model that predicts IBNR and RBNS claims. The main focus lies on the two sources of delay and how to estimate the IBNR and RBNS claims seperately. Martínez-Miranda, Nielsen and Verall [12] focused more on the weak points of the SCL and the DCL method presented in Verall, Nielsen and Jessen [26]. They showed how alterations of the DCL method can produce a new method that is related to the Bornhuetter-Ferguson technique, cf. [19]. The Bornhuetter-Ferguson method is motivated by the lack of stability of the DCL method. Martínez-Miranda, Nielsen, Nielsen and Verall [14] showed how the DCL method is related to the SCL. It focuses on the estimation of the first moment parameters and the estimation of IBNR and RBNS claims. The paper on which this thesis is based shows that by making a particular choice about how the claims are estimated, the DCL method yields the same reserves than the classical SCL. This is mainly done by looking at the tail of claims estimated by the DCL method.

Those work brought significant improvements in run-off triangle analysis. However, some of the assumptions in the aforementioned papers related to DCL are not needed to estimate outstanding claims. This thesis is based, to the best of our knowledge, on the most up to date improvement of the DCL presented by Martínez-Miranda, Nielsen, Verrall, and Wüthrich [13]. The stochastic assumptions made in the manuscripts in the previous paragraph are important to understand the predictive distribution. But, Martínez-Miranda et al. [13] argues that having knowledge of the distribution of outstanding claims would translate into a better estimate. Specifically, they show that if prior knowledge is available about the future number of no-claims (zero-claims) and future loss distribution inflation rate, then those will affect the predicted distribution of outstanding claims. Therefore, if the issue is to qualify or improve best estimates, prior knowledge of zero-claims and development year severity inflation is not important. On the other hand, if the focus is the best estimate of outstanding claims, then one should (for example) consider underwriting year severity inflation. The two additional information can be easily incorporated into the well known DCL. This thesis is based on the same type of data as in Martínez-Miranda, Nielsen and Wüthrich [13], in the sense that it considers the two triangles used in DCL. Martínez-Miranda et al. [13] combines these two triangles with a third triangle on the number of payments. In their paper, they go through the full mathematical statistical modelling of the entire system behind the three triangles. It is essential to consider all available prior knowledge. Although those additions complicate matters, they add insight into the estimation procedures by properly taken into account the no-claims thereby having better understanding of the distributional properties of outstanding claims.

Following Martínez-Miranda et al. [13], we discuss the pro and cons of this new method, how to address properties of data surrounding the chain ladder prediction, and issues around the inclusion of claim severity inflation and distribution of claim settled without payment. Does this new statistical model estimate reserves in a better and more accurate way? Should insurance companies use the DCL method instead of the commonly used SCL to estimate reserves? This thesis introduces the SCL, explains in detail the DCL method and outlines the advantages and disadvantages of the two techniques.

The content of the thesis is as follow. In Section 2, we introduce the basic SCL and then show different approaches to estimate reserves using the SCL method as a basic foundation. We discuss the different statistical methods utilized and how the corresponding parameters are estimated. This section is just a brief summary of the technique. A more detailed account can be found in the master thesis of Netanya Martin [10] and other references presented there. Section 3 deals with the DCL method, beginning with the model itself and then proceed with how reserves are estimated via this technique. Moreover, we discuss in details how to incorporate

i) prior information alone

ii) future severity inflation alone and

iii) how to incorporate both i) and ii).

In Section 4, the bootstrap method is reviewed and how it can be used to bootstrap run-off triangles is presented. In Section 5, we compare the different techniques presented in this thesis. Discuss their advantages and disadvantages. Section 6 concludes the thesis work with a brief summary and an outline of potential dissertation research problems.

2 STATISTICAL METHODS FOR THE CHAIN LADDER TECHNIQUE

One of the most popular and mainly used method for estimating reserves is the single chain ladder method (SCL). In this section, we discuss different chain ladder techniques and other issues which can arise using these techniques. The outlined reasons will pave the way and motivate the use of the **Double Chain Ladder Method** which is an extension of the SCL method.

2.1 BASIC CHAIN LADDER

The unit is monitured for a certain type of event (accident, tornado, etc.) which occurs in year i. To describe the data, let i be the index for the accident year, with $i \in \{1, \ldots, m\}$. When the accident occurs, policyholders are expected to file a claim. In general, claims are not always reported right after events occur. Let j denote the claim filing year, also called delay year, with $j \in \{0, \ldots, m-1\}$. That is the j^{th} year that elapsed since the accident occured. In this section we assume that the data is in the form of a triangle. The maximal delay year is $m - 1$. We give in Table 2.1 an example of the type of data we are referring to. The data is taken from Taylor and Ashe [21]:

Table 2.1 The Taylor-Ashe Data (1983)

Z_{ij}	Delay yr 0	Delay yr 1	Delay yr 2	Delay yr 3	Delay yr 4	Delay yr 5	Delay yr 6	Delay yr 7	Delay yr 8	Delay yr 9
accident yr 1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
accident yr 2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	
accident yr 3	290507	1001799	926219	1016654	750816	146923	495992	280405		
accident yr 4	310608	1108250	776189	1562400	272482	352053	206286			
accident yr 5	443160	693190	991983	769488	504851	470639				
accident yr 6	396132	937085	847498	805037	705960					
accident yr 7	440832	847631	1131398	1063269						
accident yr 8	359480	1061648	1443370							
accident yr 9	376686	986608								
accident yr 10	344014									

The incremental claims are denoted by Z_{ij}, that is the amount of claims in dollar that occured in accident year i and filed after j years. Thus, the set of incremental claims is $\{Z_{ij} : i = 1, \ldots, m; j = 0, \ldots, m - i\}$. Table 2.2 is an example of the SCL triangle with m=5:

Table 2.2 Chain Ladder Triangle

Z_{10}	Z_{11}	Z_{12}	Z_{13}	Z_{14}
Z_{20}	Z_{21}	Z_{22}	Z_{23}	
Z_{30}	Z_{31}	Z_{32}		
Z_{40}	Z_{41}			
Z_{50}				

Z_{20} for instance, is the dollar value of a claim which occured in accident year 2 and reported in the same year as they occured. Z_{21}, on the other hand, describes the amount of claims occured in accident year 2 but filed one year later. That is, they occured in accident year 2 but were reported to the insurance company one year after they occured. The objective of the company is to forecast outstanding claims. Outstanding claims can be accidents which have not yet occured but need to be forecasted such that a company knows how much money they probably need going forward. In the table below, outstandig claims, calculated by the SCL method are colored in red:

Table 2.3 Chain Ladder Triangle with Outstanding Claims

Z_{10}	Z_{11}	Z_{12}	Z_{13}	Z_{14}
Z_{20}	Z_{21}	Z_{22}	Z_{23}	Z_{24}
Z_{30}	Z_{31}	Z_{32}	Z_{33}	Z_{34}
Z_{40}	Z_{41}	Z_{42}	Z_{43}	Z_{44}
Z_{50}	Z_{51}	Z_{52}	Z_{53}	Z_{54}

The chain ladder technique itself uses cumulative claims for estimation, given by

$$C_{ij} = \sum_{k=0}^{j} Z_{ik} \text{ for } i = 1, \ldots, m \; . \tag{2.1}$$

In the SCL method, development factors are used to forecast future claims. The development factors are denoted by λ_j and defined by

$$\lambda_j = \frac{\sum_{i=1}^{m-j} C_{ij}}{\sum_{i=1}^{m-j} C_{i,j-1}} \text{ for } j = 1, \ldots, m-1,$$

where C_{ij} is defined as in (2.1). So, given $C_{i0}, C_{i1}, \ldots, C_{i,j-1}$, the conditional expected cumulative claim is

$$\mathbb{E}\left[C_{ij} | C_{i0}, C_{i1}, \ldots, C_{i,j-1}\right] = \lambda_j \; C_{i,j-1} \text{ for } j = 1, \ldots, m-1.$$

Thus, the expected ultimate loss of all outstanding claims for accident year i is

$$\mathbb{E}\left[C_{i,m-1}\right] = \left(\prod_{j=m-i+1}^{m-1} \lambda_j \right) C_{i,m-i} \text{ for } i = 1, \ldots, m. \tag{2.2}$$

The estimation of claims using the SCL method produces estimates which have a column and a row effect. The parameter λ_j can be viewed as a column effect of claims when forecasting the ultimate loss $C_{i,m-1}$. The random variable $C_{i,m-i}$ on the other hand can be viewed as a row factor. The estimation of the ultimate claim $C_{i,m-1}$ is based on $C_{i,m-i}$ for every row i, and thus $C_{i,m-i}$ can be interpreted as a row effect in estimating the ultimate claim. The advantage of using development factors is that they are straightforward to calculate. Although the SCL method has a lot of advantages, it also has some disadvantages. For example:

i) The missing extension in calculating reserves beyond the latest delay year $m-1$.

ii) The SCL method is only an algorithm which produces estimates for outstanding claims. But there is no statistical model behind this estimation.

iii) There is no option in the SCL algorithm for including any tails, alterations or additional information.

If we estimate the outstanding claims in the lower right hand triangle, we are only estimating claims till the latest delay year $m-1$ (see Table 2.3). We don't look beyond the latest delay year since the estimation of the claims is only based on the cumulative claims $\{C_{ij} : i = 1, \ldots, m, j = 0, \ldots, m-i\}$ and thus, on the given incremental claims $\{Z_{ij} : i = 1, \ldots, m, j = 0, \ldots, m-i\}$. It would be in fact sometimes quite helpful to have such extension since claims can also be filed after $m-1$ years. In what follows we discuss different other techniques for reserve estimation.

2.2 LINEAR MODELS AND CHAIN LADDER

By estimating the development factors, we are taking into account column and row effects in the process of forecasting reserves. Since we want to have those effects included in our outstanding claims, we write our incremental claims in a multiplicative model, given by

$$\mathbb{E}\left[Z_{ij}\right] = U_i S_j, \tag{2.3}$$

where U_i is the parameter for row i which can be interpreted as the expected total claim for accident year i, and S_j for column j represents the expected proportion of the ultimate claim for each delay year j with the restriction $\sum_{j=0}^{m-1} S_j = 1$. It can be shown that S_j can be expressed by development factors since λ_j can also be interpreted as a column effect. U_i can be seen as the expected ultimate caim for row i with $U_i = \mathbb{E}\left[C_{i,m-1}\right]$. Kremer [7] showed, that one can write the expected proportion of the ultimate claim for delay year j by

$$\begin{cases} S_j = \frac{\lambda_j - 1}{\prod_{l=j}^{m-1} \lambda_l} \text{ for } j = 1, \ldots, m-1 \ , \\ S_0 = \frac{1}{\prod_{l=1}^{m-1} \lambda_l}. \end{cases}$$

To estimate U_i and S_j, we will use a log-linear model. To that end, we assume that the incremental claims Z_{ij} are lognormal distributed. Taking the logarithm on both sides of (2.3) yields

$$\mathbb{E}\left[Y_{ij}\right] = \mu + \alpha_i + \beta_j,$$

where Y_{ij}=log(Z_{ij}), μ is the overall mean, α_i is the row effect and β_j the column effect of the logged incremental claims. Thus,

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \tag{2.4}$$

where ϵ_{ij}, the error term, has mean 0 and variance σ^2. Since $\sum_{j=0}^{m-1} S_j = 1$, we obtain

$$\sum_{j=0}^{m-1} \mathbb{E}\left[Z_{ij}\right] = \sum_{j=0}^{m-1} U_i S_j = U_i \sum_{j=0}^{m-1} S_j = U_i.$$

So, Kremer [7] showed the relationship

$$U_i = e^{\alpha_i} e^{\mu} \sum_{j=0}^{m-1} e^{\beta_j}. \tag{2.5}$$

Equation (2.4) can be written in form of a log-linear model. To do this, some regularity conditions are needed. Let $\alpha_1 = \beta_0 = 0$ such that the following model has a non-singular design matrix X. Thus, we get the log-linear model

$$\underline{y} = X\underline{\beta} + \underline{\epsilon}, \tag{2.6}$$

where

$\underline{y}$	denotes the vector of the logged incremental claims $Y_{ij} = \log(Z_{ij})$ with $i = 1, \ldots, m; j = 0, \ldots, m - i$,
$\underline{X}$	is the design matrix, where each row contains the coefficients for μ, α_i and β_j to ensure that $\mathbb{E}[Y_{ij}] = \mu + \alpha_i + \beta_j$ holds,
$\underline{\beta}$	a parameter vector with $\underline{\beta} = [\mu; \alpha_2, \alpha_3, \ldots, \alpha_m; \beta_1, \beta_2, \ldots, \beta_{m-1}]$ and
$\underline{\epsilon}$	a vector of errors which are identically distributed with mean zero and variance σ^2.

We now give an example of (2.6):

Let $m = 3$, which means that the chain ladder triangle contains data for 3 years and hence, uses the incremental claims $Z_{10}, Z_{11}, \ldots, Z_{30}$. By taking the logarithm of those incremental claims we get $Y_{10}, Y_{11}, \ldots, Y_{30}$. Since $\alpha_1 = \beta_0 = 0$, the parameter $\underline{\beta}$ has the same form as in (2.6). With this information, we get the following equation:

$$\begin{bmatrix} y_{10} \\ y_{11} \\ y_{12} \\ y_{20} \\ y_{21} \\ y_{30} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{20} \\ \epsilon_{21} \\ \epsilon_{30} \end{bmatrix}.$$

By plugging in the given data from the run-off triangle, we can estimate the parameter $\underline{\beta}$. Hence, we get the estimates of the overall mean μ, the row effect α_i and the column effect β_j. With this estimates at hand are we able to forescast the incremental claims using (2.4). Kremer [7] also showed, that by hatting the parameters in equation (2.5), the estimated claims using U_i are very similar to the incremental claims estimated by the SCL method. One advantage of using the log-linear model instead of the SCL mehtod is that the log-linear model gives standard errors that can be used forecasting upper limits for our claims and facilitate inferences such as goodness of fit.

But we still can't extend our calculations to forecast claims further than delay year $m-1$ since the log-linear model is only based on the given incremental claims $\{Z_{ij}; i=1,\dots,m; j=0,\dots,m-i\}$ which means that this model can only estimate μ, α_i, β_j with $i=1,\dots,m$ and $j=0,\dots,m-1$ and thus, the incremental claims till accident year m and delay year $m-1$.

2.3 ESTIMATION OF RESERVES

The purpose of this subsection is to obtain estimates of our incremental claim Z_{ij}. To that end, we require some assumptions. First, let m be the number of accident years and n the number of observations, which in this case is $n{=}\frac{1}{2}m(m+1)$. We assume that the incremental claims Z_{ij} are lognormally distributed with $Z_{ij} \overset{iid}{\sim}$ lognormal with $\mathbb{E}\left[Z_{ij}\right] = \theta_{ij}\ \forall\ (i,j)$. Therefore,

$$Y_{ij} = \log\left(Z_{ij}\right) \sim \mathrm{N}\left(\mu, \sigma^2\right). \tag{2.7}$$

Using the lognormal distribution property, we have $\theta_{ij} = e^{\mu+\frac{1}{2}\sigma^2}$. From the log-linear model in (2.6) we can write $\underline{y} = X\underline{\beta} + \epsilon$ and therefore

$$\mathbb{E}\left[Y_{ij}\right] = X_{ij}\underline{\beta},$$
$$\mathrm{Var}\left(Y_{ij}\right) = \sigma^2,$$
$$\epsilon \overset{iid}{\sim} \mathrm{N}\left(0, \sigma^2\right),$$

with X_{ij} being the row of X regarding Y_{ij}. From the above sequence of equations we can estimate θ_{ij} and σ^2 by using $\theta_{ij} = e^{\;X_{ij}\underline{\beta} \,+\, \frac{1}{2}\sigma^2}$, which is obtained by replacing μ by $X_{ij}\underline{\beta}$. Observe that $\mu = \mathbb{E}\left[Y_{ij}\right]$ in (2.7). Thus we can derive the maximum likelihood estimator $\hat{\theta}_{ij}$ of θ_{ij} by

$$\hat{\theta}_{ij} = e^{\;X_{ij}\underline{\hat{\beta}} \,+\, \frac{1}{2}\hat{\sigma}^2}$$

with $\underline{\hat{\beta}} = \left(X'X\right)^{-1} X'\underline{y}$ the maximum likelihood estimator of $\underline{\beta}$ in the regression model and $\hat{\sigma}^2 = \frac{1}{n}\left(\underline{y} - X\hat{\beta}\right)'\left(\underline{y} - X\hat{\beta}\right)$ the maximum likelihood estimator of σ^2.

The expression of $\hat{\sigma}^2$ given above is biased. Since we need an unbiased estimator of θ_{ij}, we need to improve our estimator. Let $\widetilde{\theta}_{ij}$ be an unbiased estimate of θ_{ij}. Finney [6] introduced a function $g_l(t)$, which is used to obtain an unbiased estimate of θ_{ij}. Finney [6] showed that the unbiased estimate of $e^{\underline{Z}_i\underline{\beta} \,+\, a\sigma^2}$ is

$$e^{\underline{Z}_i\underline{\hat{\beta}}}\; g_l\left[\left(a - \frac{1}{2}\underline{Z}_i\left(X'X\right)^{-1}\underline{Z}_i'\right)s^2\right] \;,$$

where

$$\begin{aligned}
\underline{Z}_i &= \text{the } i^{th} \text{ row vector of the matrix } X \text{ of length } p, \text{ equal } X_{ij} \text{ in (2.6)},\\
g_l\left(t\right) &= \sum_{k=0}^{\infty} \frac{l^k\left(l+2k\right)}{l\left(l+2\right)\ldots\left(l+2k\right)} \cdot \frac{t^k}{k!}\;, \text{ a polynomial function,}\\
s^2 &= \frac{n}{n-p}\hat{\sigma}^2 \text{ an unbiased estimate of } \sigma^2,\\
l &= \left(n-p\right) \text{ the degrees of freedom associated with } s^2,\\
a &= \text{a constant.}
\end{aligned}$$

Thus, the unbiased estimate of θ_{ij} is

$$\widetilde{\theta}_{ij} = e^{\ X_{ij}\hat{\underline{\beta}}} g_l \left[\left(\frac{1}{2} - \frac{1}{2} X_{ij} \left(X'X\right)^{-1} X'_{ij} \right) s^2 \right] \ , \tag{2.8}$$

with l = n - (2m-1) and $s^2 = \frac{n}{n-(2m-1)}\hat{\sigma}^2$. With an unbiased estimate of θ_{ij} at hand, we can now calculate its unbiased variance. Denote the variance of the unbiased estimate of θ_{ij} by r^2_{ij}. Hence, we get

$$\begin{aligned} r^2_{ij} &= \text{Var}\left(\widetilde{\theta}_{ij}\right) \\ &= \mathbb{E}\left[\widetilde{\theta^2}_{ij}\right] - \left(\mathbb{E}\left[\widetilde{\theta}_{ij}\right]\right)^2 \\ &= \mathbb{E}\left[\widetilde{\theta^2}_{ij}\right] - \theta^2_{ij} \\ &= \mathbb{E}\left[\widetilde{\theta^2}_{ij}\right] - e^{2X_{ij}\underline{\beta}\ +\sigma^2}. \end{aligned} \tag{2.9}$$

Using (2.8), (2.9) and again the result in Finney [6], we obtain an unbiased estimate of the variance by

$$\begin{aligned} \widetilde{r}^2_{ij} &= \widetilde{\theta}^2_{ij} - e^{2X_{ij}\hat{\underline{\beta}}} g_l \left(\left(1 - 2X_{ij}\left(X'X\right)^{-1} X'_{ij}\right) s^2 \right) \\ &= e^{2X_{ij}\hat{\underline{\beta}}} \left[\left(g_l \left(\left(\frac{1}{2} - \frac{1}{2} X_{ij}\left(X'X\right)^{-1} X'_{ij} \right) s^2 \right) \right)^2 \right. \\ &\quad \left. - g_l \left(\left(1 - 2X_{ij}\left(X'X\right)^{-1} X'_{ij}\right) s^2 \right) \right], \end{aligned}$$

with $l = n-(2m-1)$ the degrees of freedom associated with s^2 and $s^2 = \frac{n}{n-(2m-1)}\ \hat{\sigma}^2$ an unbiased estimate of σ^2 since $p = (2m-1)$ is the number of parameters estimated. We only calculated the unbiased estimate of the individual incremental claims Z_{ij}, denoted by $\widetilde{\theta}_{ij}$. Our task is to estimate the total outstanding claims for each accident year i and each delay year j. By adding up the incremental claims over all j we will get the total outstanding claim for accident year i which will be denoted by

$$R_i = \sum_{j=m-i+1}^{m-1} Z_{ij}. \tag{2.10}$$

Since we know the unbiased estimate of Z_{ij}, we can write the unbiased estimate of R_i as $\widetilde{R}_i = \sum_{j=m-i+1}^{m-1} \widetilde{\theta}_{ij}$. Using (2.10), we have

$$\begin{aligned} \text{Var}\left(\widetilde{R}_i\right) &= \text{Var}\left[\sum_{j=m-i+1}^{m-1} \widetilde{\theta}_{ij}\right] \\ &= \sum_{j=m-i+1}^{m-1} \left[\text{Var}\left(\widetilde{\theta}_{ij}\right) + 2\sum_{k=j+1}^{m-1} \text{Cov}\left(\widetilde{\theta}_{ij}, \widetilde{\theta}_{ik}\right)\right]. \end{aligned}$$

To calculate the covariance, observe that

$$\begin{aligned} \text{Cov}\left(\widetilde{\theta}_{ij}, \widetilde{\theta}_{ik}\right) &= \mathbb{E}\left[\widetilde{\theta}_{ij}\widetilde{\theta}_{ik}\right] - \mathbb{E}\left[\widetilde{\theta}_{ij}\right]\mathbb{E}\left[\widetilde{\theta}_{ik}\right] \\ &= \mathbb{E}\left[\widetilde{\theta}_{ij}\widetilde{\theta}_{ik}\right] - \theta_{ij}\theta_{ik} \\ &= \mathbb{E}\left[\widetilde{\theta}_{ij}\widetilde{\theta}_{ik}\right] - e^{X_{ij}\underline{\beta}+\frac{1}{2}\sigma^2} e^{X_{ik}\underline{\beta}+\frac{1}{2}\sigma^2} \\ &= \mathbb{E}\left[\widetilde{\theta}_{ij}\widetilde{\theta}_{ik}\right] - e^{(X_{ij}+X_{ik})\underline{\beta}+\sigma^2}. \end{aligned}$$

Thus, the unbiased estimator of $\text{Cov}\left(\widetilde{\theta}_{ij}, \widetilde{\theta}_{ik}\right)$, still following Finney [6], denoted by $\widetilde{r}^2_{ijk}$, is

$$\begin{aligned} \widetilde{r}^2_{ijk} &= \widetilde{\theta}_{ij}\widetilde{\theta}_{ik} - e^{(X_{ij}+X_{ik})\underline{\beta}}\; g_l((1-\tfrac{1}{2}(X_{ij}+X_{ik})(X'X)^{-1}(X_{ij}+X_{ik})')s^2) \\ &= e^{(X_{ij}+X_{ik})\underline{\beta}}\left[g_l((\tfrac{1}{2}-\tfrac{1}{2}X_{ij}(X'X)^{-1}X'_{ij})s^2)\; g_l((\tfrac{1}{2}-\tfrac{1}{2}X_{ik}(X'X)^{-1}X'_{ik})s^2)\right. \\ &\quad \left. -g_l((1-\tfrac{1}{2}(X_{ij}+X_{ik})(X'X)^{-1}(X_{ij}+X_{ik})')s^2)\right]. \end{aligned}$$

Hence the unbiased estimate of $\text{Var}\left(\widetilde{R}_i\right)$ is

$$\sum_{j=m-i+1}^{m-1} \left[\widetilde{r}_{ij}^2 + 2 \sum_{j=j+1}^{m} \widetilde{r}_{ijk}^2 \right]. \tag{2.11}$$

The total outstanding claim for the whole triangle is denoted by

$$R = \sum_{i=2}^{m} R_i,$$

the sum over all outstanding claims for each accident year i except the first one since we know all claims for accident year 1. So, an unbiased estimate of the total outstanding claim is

$$\widetilde{R} = \sum_{i=2}^{m} \widetilde{R}_i = \sum_{i=2}^{m} \sum_{j=m-i+1}^{m-1} \widetilde{\theta}_{ij}. \tag{2.12}$$

The expression in (2.11) and (2.12) can be used to derive the confidence interval for estimated reserves.

2.4 PREDICTION OF CLAIMS INTERVALS

By obtaining an unbiased estimator for our total outstanding claims and its variance, it is often preferable to have a prediction interval for those claims. An insurance company wants to have an upper bound for outstanding claims such that they can safe enough money in case of a lot of events, like severe storms or accidents. In this case, they have to pay a lot of money to the policyholders. We will only look at the upper confidence bounds for outstanding claims. This upper bound represents an estimated value of outstanding claims which should never be exceeded by the true actual claims. Hence, we need an upper bound of estimated outstanding claims such that the actual outstanding claims do not exceed this upper bound or if it would exceed, it exceeds this bound with a very small probability. For calculating for instance a 95% confidence interval, we need to find a value k such that

$$\mathbb{P}\left(R \leq \widetilde{R} + k\right) = 0.95, \tag{2.13}$$

with R being the actual total outstanding claims for the triangle, $\widetilde{R}$ is the unbiased estimate as calculated above and k is a real number that need to be added to our unbiased estimate such that we can be 95% confident that our actual total outstanding claim will not exceed $\widetilde{R} + k$. The value of k can be viewed as an adjustment number. In this case we can write (2.13) as follows:

$$\mathbb{P}\left(R - \widetilde{R} \leq k\right) = 0.95\ .$$

To calculate k, we need to know the expectation and variance of $(R - \widetilde{R})$. Since $\widetilde{R}$ is an unbiased estimate of $\mathbb{E}\left[R\right]$, we can assume that $\widetilde{R}$ is independent of R and so we have

$$\mathbb{E}\left[R - \widetilde{R}\right] = \mathbb{E}\left[R\right] - \mathbb{E}\left[\widetilde{R}\right] = 0,$$

and

$$\begin{aligned} \mathrm{Var}(R - \widetilde{R}) &= \mathrm{Var}(R) + \mathrm{Var}(\widetilde{R}) - 2\ \mathrm{Cov}(R, \widetilde{R}) \\ &= \mathrm{Var}(R) + \mathrm{Var}(\widetilde{R}) \end{aligned}$$

by independence of $\widetilde{R}$ and R. R and $\widetilde{R}$ are assumed lognormally distributed and they will take large values since claims in an insurance company can be very high. In this case, it is okay to assume that $(R - \widetilde{R})$ is approximately normally distributed with mean 0 and variance $\mathrm{Var}(R) + \mathrm{Var}(\widetilde{R})$.

So, one can calculated k as follows:

$$
\begin{aligned}
& \mathbb{P}\left(R-\widetilde{R} \leq k\right)=0.95 \\
\Rightarrow \quad & \mathbb{P}\left(\frac{(R-\widetilde{R})-0}{\sqrt{\operatorname{Var}(R)+\operatorname{Var}(\widetilde{R})}} \leq \frac{k-0}{\sqrt{\operatorname{Var}(R)+\operatorname{Var}(\widetilde{R})}}\right)=0.95 \\
\Rightarrow \quad & \frac{k-0}{\sqrt{\operatorname{Var}(R)+\operatorname{Var}(\widetilde{R})}}=1.645 \\
\Rightarrow \quad & k=1.645 \sqrt{\operatorname{Var}(R)+\operatorname{Var}(\widetilde{R})}
\end{aligned}
$$

Thus, (2.13) becomes

$$
\mathbb{P}\left(R \leq \widetilde{R}+1.645 \sqrt{\operatorname{Var}(R)+\operatorname{Var}(\widetilde{R})}\right)=0.95,
$$

which means that an upper confidence bound for our total outstanding claims, such that only with a probability of 5% the actual total claims will exceed this upper bound, has been found. It only remains to find the value of $\operatorname{Var}(R)$ and $\operatorname{Var}(\widetilde{R})$. From the preceding calculations, we know what the unbiased estimator of $\operatorname{Var}(\widetilde{R})$ is $\widetilde{r}_{ij}^2$. However an unbiased estimator for the variance of R is unknown. To obtain that, we use the same approach as before together with Finney [6]. By independence, we obtain

$$
\begin{aligned}
\operatorname{Var}(R) &= \sum_{i=2}^{m} \sum_{j=m-i+1}^{m-1} \operatorname{Var}\left(Z_{ij}\right) \\
&= \sum_{i=2}^{m} \sum_{j=m-i+1}^{m-1} e^{2X_{ij}\underline{\beta}+\sigma^2}\left(e^{\sigma^2}-1\right) \\
&= \sum_{i=2}^{m} \sum_{j=m-i+1}^{m-1} e^{2X_{ij}\underline{\beta}+2\sigma^2}-e^{2X_{ij}\underline{\beta}+\sigma^2} .
\end{aligned}
$$

Using the results from above, we get that an unbiased estimate of $\text{Var}(Z_{ij})$ is

$$e^{2X_{ij}\underline{\hat{\beta}}}\left[g_l\left(\left(2-2X_{ij}\left(X'X\right)^{-1}X'_{ij}\right)s^2\right)-g_l\left(\left(1-2X_{ij}\left(X'X\right)^{-1}X'_{ij}\right)s^2\right)\right]$$

It is very useful to have unbiased estimates for our outstanding claims. We can construct prediction intervals and will have a good forecast of our actual claims. But there are still some disadvantages. Calculating unbiased estimates are often very complicated and tedious. Furthermore, no close expression or prior informations are available. To overcome this, we take a Bayesian approach. The use of Bayes estimates is also motivated by the fact that our model is based on the Bayesian theory.

Before proceeding using Bayesian estimation, we will look at another technqiue for theSCL model. If we are given the run-off triangle of different companies, it would be nice to assess how claims vary across those companies by looking how claims evolve. In some companies for example, claims are always filed directly after they occured and, other claims are reported 10-15 years after the event occured. With this information, one can compare companies by say, the types of insurance policies they offer or their database of policyholders. Moreover, the structure of the company can provide valuable information. So, we want to compare companies and look for a pattern in run-off triangles. It is much easier to use the maximum likelihood estimation instead of the SCL method or analysis of variance. In Equation (2.1) and (2.3), we used the development factor λ_j and also the parameter S_j to find the outstanding claims.

2.5 ESTIMATION OF DEVELOPMENT FACTORS

We want to compare different sets of data using several different triangles from different companies. The different values of the development factors λ_j can be used to compare how claims arise in a company. In the previous subsection, we showed that S_j can be written as a function of λ_j, that's

$$\begin{cases} S_j = \frac{\lambda_j - 1}{\prod_{l=j}^{m-1} \lambda_l} \text{ for } j = 1, \ldots, m-1 \text{ and} \\ S_0 = \frac{1}{\prod_{l=1}^{m-1} \lambda_l}. \end{cases} \tag{2.14}$$

Therefore, S_j can be written as a function of β_j, as well by

$$S_j = \frac{e^{\beta_j}}{\sum_{l=1}^{m-1} e^{\beta_l}} \text{ with } j = 0, \ldots, m-1 \text{ with } \beta_0 = 0. \tag{2.15}$$

Combining (2.14) and (2.15), we can write the development parameter λ_j as

$$\lambda_j = 1 \; + \; \frac{e^{\beta_j}}{\sum_{l=1}^{j-1} e^{\beta_l}} \text{ with } j = 1, \ldots, m-2. \tag{2.16}$$

Since we know the maximum likelihood estimator for $\underline{\beta}$ from the previous subsections, we can plug in $\hat{\underline{\beta}}$ into equation (2.15) and (2.16). That is because maximum likelihood estimates are invariant under parameter transformation. Thus, we get $\hat{\lambda}_j$ the maximum likelihood estimate for the parameter λ_j and $\hat{S}_j$ the maximum likelihood estimate for S_j with

$$\begin{aligned} \hat{S}_j &= \frac{e^{\hat{\beta}_j}}{\sum_{l=1}^{t-1} e^{\hat{\beta}_l}}, \\ \hat{\lambda}_j &= 1 \; + \; \frac{e^{\hat{\beta}_j}}{\sum_{l=1}^{j-1} e^{\hat{\beta}_l}}. \end{aligned}$$

In the previous subsections, the importance of having an estimate for the variance to check the goodness of fit was discussed. If we are given the variance-covariance matrix $\mathrm{V}(\underline{\beta})$, we can immediately estimate the variance-covariance matrix for the parameter

$\underline{\lambda} = (\lambda_1, \ldots, \lambda_m)$ and $\underline{S} = (S_1, \ldots, S_m)$ using the multivariate delta method

$$\begin{aligned} \mathrm{V}(\underline{\lambda}) &= \left(\frac{\delta\underline{\lambda}}{\delta\underline{\beta}}\right) \mathrm{V}(\underline{\beta}) \left(\frac{\delta\underline{\lambda}}{\delta\underline{\beta}}\right)', \\ \mathrm{V}(\underline{S}) &= \left(\frac{\delta\underline{S}}{\delta\underline{\beta}}\right) \mathrm{V}(\underline{\beta}) \left(\frac{\delta\underline{S}}{\delta\underline{\beta}}\right)'. \end{aligned}$$

The parameter λ_j describes how much claims are filed after j years and thus the number of outstanding claims depend on $\underline{\lambda}$ (2.2). The development factor $\underline{\lambda}$ for each company can be used to see if a company has more runoff in later years than other companies have. If λ_j for high j would be for example very large, then we can see that the runoff of a company's claim is very high which means that there are many claims filed in later years. The next subsection pertains to the estimate of the outstanding claims using the Bayesian technique.

2.6 BAYESIAN ESTIMATION OF CLAIMS

In this subsection we estimate total claims using the Bayesian estimation method. After obtaining Bayesian estimates for outstanding claims, we introduce the Bayesian estimation for our linear model from Subsection 2.2. For the Bayesian estimation for runoff triangles, we assume that

$$Z_{ij} \sim \text{ lognormal } (\theta, \sigma^2),$$

as before. Additionally, we assume that $\theta \in \Omega$ follows a certain prior distribution $\pi(\theta) = \mathbb{P}(\theta = k)$. In this case we have a normal distribution for our prior. A better way to have an idea about the behaviour of θ is to use available data. The posterior distribution of θ is proportional to a normal distribution since we have a conjugate prior. Following Bayesian estimation technique we can calculate the posterior distribution of θ given the data $D = \{x_1, \ldots, x_n\}$ by

$$\begin{aligned}\pi(\theta|D) &= \frac{f(D|\theta) \cdot \pi(\theta)}{f(D)} \\ &= \frac{L(x_1, \ldots, x_n|\theta) \cdot \pi(\theta)}{\int_\Omega L(x_1, \ldots, x_n|\theta) \cdot \pi(\theta)\ d\theta} \ \propto \ \mathrm{N}(m, r^2).\end{aligned}$$

Thus, we have

$$\begin{aligned}&\log Z_{ij} \mid \theta \sim N(\theta, \sigma^2) \text{ and} \\ &\theta \mid D \sim N(m, r^2).\end{aligned}$$

With this information at hand, we want to calculate the expectation and the variance of our claim Z_{ij}. Given σ^2 and r^2, we get

$$\begin{aligned}&\mathbb{E}\left[Z_{ij}|D\right] = e^{m+\frac{1}{2}\sigma^2+\frac{1}{2}r^2}, \\ &\mathrm{Var}\left(Z_{ij}|D\right) = e^{2m+\sigma^2+r^2}\left(e^{\sigma^2+r^2} - 1\right),\end{aligned}$$

using the hierarchical Bayes technique. Since we want to know what our ultimate total claim will be, the Bayes estimate of our outstanding claims is

$$\sum_{i=2}^{m} \sum_{j=m-i}^{m-1} \mathbb{E}\left[Z_{ij}|D\right]$$

and also the Bayes estimate of the variance is

$$R_i = \sum_{j>m-i}^{m} \left[\mathrm{Var}(Z_{ij}|D) + 2\sum_{k>j} \mathrm{Cov}(Z_{ij}, Z_{ik}|D)\right].$$

To apply the Bayes estimate to the linear model in (2.6), we use some facts from

Subsection 2.2. Recall that $Y_{ij} = X_{ij}\underline{\beta} + \epsilon$ and thus

$$\underline{y} \mid \underline{\beta} \;\sim\; \mathrm{N}(X\underline{\beta}, \Sigma),$$

with Σ being the variance-covariance matrix of $\underline{y}$. If we have additional information on $\underline{\beta}$, we could use a prior distribution, given by the practitioner, to calculate the posteriori distribution. Since the prior is normally distributed with parameters $X\underline{\beta}$ and Σ, we know that the posterior distribution is also normal. Thus,

$$\underline{\beta} \mid \underline{\theta} \;\sim\; \mathrm{N}(A\underline{\theta}, C),$$

with C being a diagonal matrix of variances and $\underline{\theta}$ a prior estimate. The Bayes estimate of $\underline{\beta}$ is denoted by $\underline{\widetilde{\beta}}$ and Verall [27] showed that it can be calculated using the equation

$$(\sigma^{-2}X'X + C^{-1})\underline{\widetilde{\beta}} = \sigma^{-2}X'X\underline{\hat{\beta}} + C^{-1}\underline{\theta}$$

and also the variance-covariance matrix of $\underline{\widetilde{\beta}}$ is obtained by

$$\mathrm{Var}(\underline{\widetilde{\beta}}) = \left[\sigma^{-2}X'X + C^{-1}\right]^{-1}$$

A nice solution of the Bayes estimate, in this case, is the credibility formula for the estimate $\underline{\widetilde{\beta}}$. We can write the estimate as

$$\underline{\widetilde{\beta}} = Z\underline{\hat{\beta}} + (I - Z)\underline{\theta}$$

with $Z = (\sigma^{-2}X'X + C^{-1})^{-1}\sigma^{-2}X'X$. Z being the credibility factor or the credibility matrix in our case and can be interpreted as the weight assigned to $\underline{\hat{\beta}}$ and $(1-Z)$ the weight assigned to the prior data $\underline{\theta}$. One can also estimate the parameter $\underline{\beta}$ by using empirical bayes estimates which will rely on a 3-stage model where we look at the row and the column parameter seperately. However we will not go into more detail here.

The Bayesian estimation is one of the best techniques for the SC Lmethod although all of the techniques outlined here can be used. The advantages in using the Bayesian estimates is the stability of the parameters. One reason why Bayesian estimate have a low standard error, is the amount of information used to obtain them. We include prior information and also use data from different rows and columns. Though they perform better than other estimates, we still don't look further than our m accident years. We have discussed many different techniques of the SCL method to forecast outstanding claims. We assumed lognormally distributed incremental claims and then applied methods like the linear model, unbiased and likelihood estimates and also Bayesian estimates. But there were some common disadvantages in all of this techniques such as

i) What happens if we would like to extend our calculation to more than our given m accident years? At this moment we are not able to do this without further information.

ii) There is a problem in the way the estimates are calculated. This techniques are not based on an underlying theory or look at the way how the claims arise and what the predictive distribution of our reserves are.

This is the reason why we will now look at another method for calculating the outstanding claims. This method is called the double chain ladder method and it looks at the way claims arise and how we can estimate reserves by looking at their distribution. The double chain ladder method also describes a way to extend our triangle, which was one of the things in the SCL method that was missing.

3 DOUBLE CHAIN LADDER

In Section 2, we have discussed how to forecast outstanding claims by using the chain ladder method. The disadvantages of using the chain ladder method are

i) The method does not estimate any reserves further than the maximal delay year of $m - 1$.

ii) There exists no theory behind the chain ladder method.

iii) There is no option to include additional information.

Estimating reserves using the SCL method does not take into account any underlying theory about the way claims arise. It is a straightforward algorithm that works fast and quite well, but it only projects one single triangle of aggregated data. The SCL method also cannot include additional information from the company properly, since it is only an algorithm that computes the estimated reserves. Because of its easy applicability, many people only use the SCL method. Since insurance companies also want to extend their forecasts for outstanding claims, tail factors can be used. Sometimes the SCL method only predicts reserves for 10 years, but some kinds of insurance can have claims filed after those 10 years, which has to be taken into account. Thus, if the development of the run-off triangle is not complete after the maximum delay year $m - 1$, a tail factor can be used to estimate the total outstanding claims including development years $j > m - 1$. A tail factor is a constant $k = 1 + \epsilon$, with $\epsilon \in [0.01, 0.06]$, which, if multiplied by the total outstanding claims, will result in a higher estimate for the total outstanding claims. In Mack [9], for instance, the tail factor based on some data was estimated to be $k = 1.05$, which means that in addition to the total outstanding claims, approximately 5% of those claims will be filed after $m - 1$ years. Using a tail factor is one option to avoid the problem of the SCL method not having any extension after delay year $m-1$ in its calculations. Although it is quite simple to use a tail factor, this estimation does not include the fact that claims after delay year $m-1$ can also have development inflation and that there is an evolution of claims after delay year $m-1$. It is definitely a problem that the SCL algorithm does not produce any tails and thus, estimates for high delay years are not exact. Another big disadvantage of the SCL is that the SCL is not a statistical model. It does not take into account any distributional properties about the claims or how adjustments or any additional information can be included in this framework properly. A statistical model that estimates outstanding claims by looking at their distribution, produces a tail, and includes additional information is the way to go.

In this section, we focus on a statistical model for estimating reserves, distributional properties of claims and prior information on claims that can be included in the DCL framework. The DCL method is closely related to the SCL method, though there are some differences:

- The DCL method includes the distribution of claims and looks at how claims evolve in order to estimate outstanding claims.
- The DCL method can be easily adapted to include other information as time goes by.
- The DCL method's use of tail facilitates reserve calculation for later years.

The goals of the DCL method are not to develop a new model to get different estimations for reserves. It is rather the goal to develop a model that procudes similar estimates like the SCL method but in addition to obtain the distribution of claims and estimate those claims.

3.1 DECOMPOSITION OF OUTSTANDING CLAIMS

If an accident occurs, claims are first filed and later paid. Sometimes claims will be filed years after they occur. Claims can also be paid years after they were filed. Thus, there exists a reporting delay and payment delay in estimating reserves. The following figure from Martínez et al. [15] shows that claims can occur after the reserves of a insurance company are set and sometimes claims occur before reserves are set but are paid afterwards.

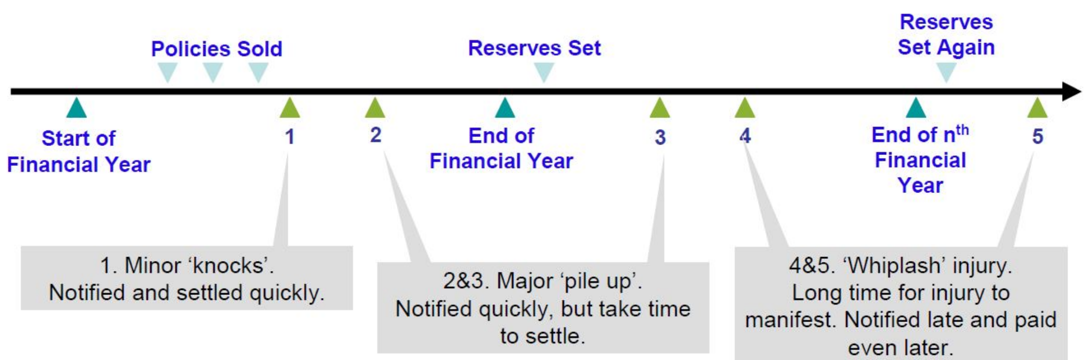

Figure 3.1 Stochastic Claims Reserving

The DCL method is based on two types of data in the form of run-off triangles. The first triangle contains data about the number of reported claims and the second triangle contains observations about the number of payments for each reported claim. The data is presented in a form of a triangle for $(i, j) \in I_m$, with $I_m = \{(i, j) : i = 1, \ldots, m; j = 0, \ldots, m-1; i + j \leq m\}$, where i denotes the accident year, j denotes the delay year, and m is the last observed accident year. For example, in Table 2.1 (The Taylor-Ashe Data), the last observed accident year was $m = 10$. We will not only consider predictions over the lower triangle, like in the SCL method, but we will also predict reserves over other sets of triangles. The following figure is taken from Martínez, Nielsen, and Verrall [11] and shows the possible index sets for predicting reserves: I is, in our case defined as I_m, the actual data which is available.

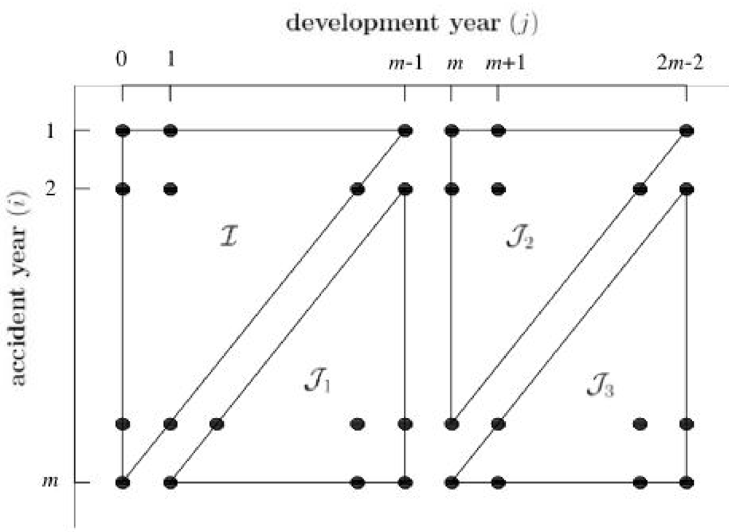

Figure 3.2 Index Sets For Aggregate Claims Data

The sets I_1, I_2, I_3 are defined as follows:

$$I_1 = \{i = 2, \ldots, m; j = 1, \ldots, m-1 \text{ so } i+j = m+1, \ldots, 2m-1\},$$

$$I_2 = \{i = 1, \ldots, m-1; j = m, \ldots, 2m-2 \text{ so } i+j = m+1, \ldots, 2m-1\},$$

$$I_3 = \{i = 2, \ldots, m; j = m, \ldots, 2m-2 \text{ so } i+j = 2m, \ldots, 3m-2\}.$$

When estimating reserves using the SCL method, we used the data for $(i, j) \in I_m$ to forecast reserves for only I_1, and we would need to use tail factors to estimate claims if we want to extend our estimation for I_2 and I_3. The DCL method automatically provides tail factors over $I_2 \cup I_3$, thus the DCL is consistent over all index sets in the process of estimating the reserves.

To proceed with the DCL technique, we introduce the following random variables:

- N_{ij}: the number of reported claims

  N_{ij} denotes the number of reported claims that occured in accident year i and filed after j years, where each claim can generate a number of payments.

- N_{ijl}^{paid}: the number of payments

  Those payments originated from the N_{ij} claims and were paid after l years, which means that with a payment delay of l years, where $l = 0, \ldots, m-1$. If we are only interested in the number of paid claims, we define $N_{ij}^{paid} = \sum_{l=0}^{j} N_{i,j-l,l}^{paid}$.

- $Y_{ijl}^{(k)}$: the individual severity claims

  The individual severity claims describe the individual settled payments, with $Y_{ijl}^{(1)}$ being the first payment and $Y_{ijl}^{(N_{ijl}^{paid})}$ being the last payment that originated from the N_{ij} claims and are paid after l years, where $k = 1, \ldots, N_{ijl}^{paid}$. We define $Y_{ij}^{(k)} = \sum_{l=0}^{j} Y_{i,j-l,l}^{(k)}$, where the individual payment of claims originated in accident year i and paid in year $i + j$.

- X_{ij}: the total payments:

  The amount of X_{ij} describes the total payment from claims that occured in year i and were paid in year $i + j$. Since the total payments are the sum of all individual payments, we can write X_{ij} as

$$
\begin{aligned}
X_{ij} &= \sum_{k=1}^{N_{ij}^{paid}} Y_{ij}^{(k)} \\
&= \sum_{l=0}^{j} \sum_{k=1}^{N_{i,j-l,l}^{paid}} Y_{i,j-l,l}^{(k)}.
\end{aligned}
$$

The reason why the DCL method is called *double chain ladder* is because the SCL method is performed twice for the following two run-off triangles:

$$(\chi_m, \Delta_m), \text{ where}$$

$$\chi_m = \{N_{ij} : (i,j) \in I_m\} \text{ is the number of reported claims and}$$

$$\Delta_m = \{X_{ij} : (i,j) \in I_m\} \text{ is the number of total payments.}$$

In Martínez, Nielsen, and Verrall [11] (Table 1 and Table 2) one can find an example of those run-off triangles:

Table 3.1 Aggregated Reported Claims

$i \setminus j$	0	1	2	3	4	5	6	7	8	9
1	6238	831	49	7	1	1	2	1	2	3
2	7773	1381	23	4	1	3	1	1	3	
3	10306	1093	17	5	2	0	2	2		
4	9639	995	17	6	1	5	4			
5	9511	1386	39	4	6	5				
6	10023	1342	31	16	9					
7	9834	1424	59	24						
8	10899	1503	84							
9	11954	1704								
10	10989									

Table 3.2 Aggregated Total Payments

$i \setminus j$	0	1	2	3	4	5	6	7	8	9
1	451288	339519	333371	144988	93243	45511	25217	20406	31482	1729
2	448627	512882	168467	130674	56044	33397	56071	26522	14346	
3	693574	497737	202272	120753	125046	37154	27608	17864		
4	652043	546406	244474	200896	106802	106753	63688			
5	566082	503970	217838	145181	165519	91313				
6	606606	562543	227374	153551	132743					
7	536976	472525	154205	150564						
8	554833	590880	300964							
9	537238	701111								
10	684944									

Table 3.1 shows the triangle χ_m that contains the data on the number of reported claims N_{ij}. Table 3.2 shows the triangle Δ_m with the data of total payments X_{ij}. It

is possible that sometimes claims are reported but not paid. Those will be denoted by RBNS claims (Reported But Not Settled), which can be estimated over $I_1 \cup I_2$ since we do know the number of claims in I_m and only have to forecast them using the SCL method. We know all of the claims that incurred and thus only have to estimate the delay payment. From the SCL method we know, that accidents sometimes occur in a given year but are not reported immediately. If accidents occur but claims are not reported, we denote those by IBNR claims (Incurred But Not Reported). Since we do not know how many claims occured and when they happened, the IBNR claims have to be estimated over $I_1 \cup I_2 \cup I_3$. Thus, we have to forescast how many accidents in the future will occur, and so we have not only to estimate the payment delay but we also have to estimate the reporting delay. The following figure from Martínez et al. [15] shows the difference between IBNR and RBNS claims.

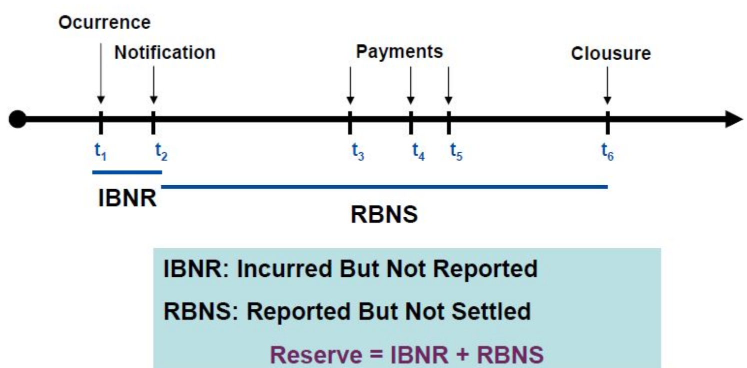

Figure 3.3 Decomposition of Outstanding Claims into IBNR and RBNS Claims

The IBNR and the RBNS claims sum to the total payments, so we can write the total payments X_{ij} as follows

$$X_{ij} = X_{ij}^{ibnr} + X_{ij}^{rbns}.$$

Since we are interested in estimating reserves and having better knowledge about the distributionl properties of claims, we need some further assumptions:

(D1) $N_{ij} \sim \text{Poi}(\alpha_i \beta_j)$ are independent random variables with the restriction that $\sum_{j=0}^{m-1} \beta_j = 1$. Since N_{ij} depends on accident year i and delay year j, the number of reported claims N_{ij} have a cross-classified mean $\mathbb{E}[N_{ij}] = \alpha_i \beta_j$, which means that the mean is evaluated for the accident year i and the delay year j at the same time.

(D2) $\left(N_{i,j,0}^{paid}, \ldots, N_{i,j,m-1}^{paid}\right) \sim \text{Multi}(N_{ij}; p_0, \ldots, p_{m-1})$ is a random vector, where $m-1$ is the maximum delay year and $p = (p_0, \ldots, p_{m-1})$ is a vector of delay probabilities such that $\sum_{l=0}^{m-1} p_l = 1$ and $0 \leq p_l \leq 1 \; \forall \; l = 0, \ldots, m-1$. The variable p_l is the probability that a claim will be paid with a delay of l years.

(D3) $Y_{i,j-l,l}^{(k)}$ are individual payments that are independent random variables with a mixed-type distribution. This means that the distribution of $Y_{i,j-l,l}^{(k)}$ has a discrete and a continuous component. The discrete part of the distribution is defined by

$$\mathbb{P}\left(Y_{i,j-l,l}^{(k)} = 0\right) = Q_i,$$

where Q_i is the probability of a zero-claim in accident year i. The continuous part of the distribution is defined by a conditional distribution, given Q_i, with $\mu_{ij} = \mu \cdot \gamma_i \cdot \delta_j$, the conditional mean with a common mean factor μ and two inflation factors γ_i and δ_j, depending on the accident year i and the delay year j. The conditional variance σ_{ij}^2 can be written as $\sigma_{ij}^2 = \sigma^2 \cdot \gamma_i^2 \cdot \delta_j^2$, where σ^2 is the common variance factor and γ_i^2 and δ_j^2 are the inflation factors.

(D4) The individual payments $Y_{ijl}^{(k)}$ are independent of the numbers of reported claims N_{ij}.

We know from the definition in (D1) that the expected value of the number of claims is defined by $\mathbb{E}\left[N_{ij}\right]=\alpha_i\beta_j$. With (D1)-(D4) in force, we can calculate the expected value of X_{ij}, thereby obtaining expected values of both run-off triangles χ_m and Δ_m. To calculate the expected value of the total payments, we first need to calculate the expected value of the individual payments $Y^{(k)}_{i,j-l,l}$:

$$\begin{aligned}
\mathbb{E}\left[Y^{(k)}_{i,j-l,l}\right] &= \mathbb{E}\left[Y^{(k)}_{i,j-l,l}|Y^{(k)}_{i,j-l,l}>0\right]\cdot\mathbb{P}\left(Y^{(k)}_{i,j-l,l}>0\right) \\
&\quad+\mathbb{E}\left[Y^{(k)}_{i,j-l,l}|Y^{(k)}_{i,j-l,l}=0\right]\cdot\mathbb{P}\left(Y^{(k)}_{i,j-l,l}=0\right) \\
&= \mu_{ij}\cdot(1-Q_i)+0\cdot Q_i \\
&= \mu_{ij}\cdot(1-Q_i) \\
&= \mu\cdot\gamma_i\cdot\delta_j\cdot(1-Q_i)
\end{aligned}$$

By the law of total variance, we can also define the variance of the individual payments by

$$\begin{aligned}
\operatorname{Var}\left(Y^{(k)}_{i,j-l,l}\right) &= \mathbb{E}\left[\operatorname{Var}\left(Y^{(k)}_{i,j-l,l}|Q_i\right)\right]+\operatorname{Var}\left(\mathbb{E}\left[Y^{(k)}_{i,j-l,l}|Q_i\right]\right) \\
&= \operatorname{Var}\left(Y^{(k)}_{i,j-l,l}|Y^{(k)}_{i,j-l,l}>0\right)\cdot\mathbb{P}\left(Y^{(k)}_{i,j-l,l}>0\right) \\
&\quad+\operatorname{Var}\left(Y^{(k)}_{i,j-l,l}|Y^{(k)}_{i,j-l,l}=0\right)\cdot\mathbb{P}\left(Y^{(k)}_{i,j-l,l}=0\right) \\
&\quad+\mathbb{E}\left[Y^{(k)}_{i,j-l,l}|Y^{(k)}_{i,j-l,l}>0\right]^2\cdot\mathbb{P}\left(Y^{(k)}_{i,j-l,l}>0\right)\cdot\left(1-\mathbb{P}\left(Y^{(k)}_{i,j-l,l}>0\right)\right) \\
&= \operatorname{Var}\left(Y^{(k)}_{i,j-l,l}|Y^{(k)}_{i,j-l,l}>0\right)\cdot\mathbb{P}\left(Y^{(k)}_{i,j-l,l}>0\right) \\
&\quad+\mathbb{E}\left[Y^{(k)}_{i,j-l,l}|Y^{(k)}_{i,j-l,l}>0\right]^2\cdot\mathbb{P}\left(Y^{(k)}_{i,j-l,l}>0\right)\cdot\left(1-\mathbb{P}\left(Y^{(k)}_{i,j-l,l}>0\right)\right) \\
&= \sigma^2_{ij}\cdot(1-Q_i)+\mu^2_{ij}\cdot Q_i\cdot(1-Q_i) \\
&= \sigma^2\cdot\gamma_i^2\cdot\delta_j^2\cdot(1-Q_i)+\mu^2\cdot\gamma_i^2\cdot\delta_j^2\cdot Q_i\cdot(1-Q_i) \\
&= (1-Q_i)\cdot\gamma_i^2\cdot\delta_j^2\cdot\left(\sigma^2+\mu^2\cdot Q_i\right).
\end{aligned}$$

To estimate the reserves, we need to know the conditional distibution of the outstanding claims. To that end, we need the first two moments of the total payments given the number of reported claims χ_m. The conditional expectation of the number of payments is

$$\begin{aligned}\mathbb{E}\left[N_{ij}^{paid}|\chi_m\right] &= \mathbb{E}\left[\sum_{l=0}^{j} N_{i,j-l,l}^{paid}|\chi_m\right] \\ &= \sum_{l=0}^{j} \mathbb{E}\left[N_{i,j-l,l}^{paid}|\chi_m\right] \\ &\overset{(D2)}{=} \sum_{l=0}^{j} N_{i,j-l} \cdot p_l.\end{aligned}$$

The conditional variance of the number of payments is

$$\begin{aligned}\mathrm{Var}\left(N_{ij}^{paid}|\chi_m\right) &= \mathrm{Var}\left(\sum_{l=0}^{j} N_{i,j-l,l}^{paid}|\chi_m\right) \\ &= \sum_{l=0}^{j} \mathrm{Var}\left(N_{i,j-l,l}^{paid}|\chi_m\right) \\ &\overset{(D2)}{=} \sum_{l=0}^{j} N_{i,j-l,l}^{paid} \cdot p_l \cdot (1-p_l).\end{aligned}$$

Assuming that the number of claims paid from various years are uncorrelated, define the conditional expectation of $Y_{ij}^{(k)}$ by $\mu_{ij} = \mu \cdot \gamma_i \cdot \delta_j \cdot (1 - Q_i)$ and its conditional variance by $\sigma_{ij}^2 = (1 - Q_i) \cdot \gamma_i^2 \cdot \delta_j^2 \cdot (\sigma^2 + \mu^2 \cdot Q_i)$, then obtain $\mathbb{E}\left[X_{ij}|\chi_m\right]$ using the iterated rule of expectation.

$$\begin{aligned}
\mathbb{E}\left[X_{ij}|\chi_m\right] &= \mathbb{E}\left[\mathbb{E}\left[X_{ij}|N_{ij}^{paid}\right]|\chi_m\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^{N_{ij}^{paid}} Y_{ij}^{(k)}|N_{ij}^{paid}\right]|\chi_m\right] \\
&\overset{(D3)}{=} \mathbb{E}\left[N_{ij}^{paid}\,\mathbb{E}\left[Y_{ij}^{(k)}\right]|\chi_m\right] \\
&\overset{(D4)}{=} \mathbb{E}\left[N_{ij}^{paid}|\chi_m\right]\mathbb{E}\left[Y_{ij}^{(k)}\right] \\
&= \sum_{l=0}^{j} N_{i,j-l}\cdot p_l\cdot\mu\cdot\gamma_i\cdot\delta_j\cdot(1-Q_i)\ .
\end{aligned}$$

Similarly, the variance of X_{ij} given χ_m can be approximated by

$$\begin{aligned}
\mathrm{Var}\left(X_{ij}|\chi_m\right) &= \mathbb{E}\left[\mathrm{Var}\left(X_{ij}|N_{ij}^{paid}\right)|\chi_m\right]+\mathrm{Var}\left(\mathbb{E}\left[X_{ij}|N_{ij}^{paid}\right]|\chi_m\right) \\
&= \mathbb{E}\left[\mathrm{Var}\left(\sum_{k=1}^{N_{ij}^{paid}} Y_{ij}^{(k)}|N_{ij}^{paid}\right)|\chi_m\right]+\mathrm{Var}\left(N_{ij}^{paid}\,\mathbb{E}\left[Y_{ij}^{(k)}\right]|\chi_m\right) \\
&= \mathbb{E}\left[N_{ij}^{paid}\,\mathrm{Var}\left(Y_{ij}^{(k)}\right)|\chi_m\right]+\mathrm{Var}\left(N_{ij}^{paid}\,\mathbb{E}\left[Y_{ij}^{(k)}\right]|\chi_m\right) \\
&\overset{(D4)}{=} \mathbb{E}\left[N_{ij}^{paid}|\chi_m\right]\mathrm{Var}\left(Y_{ij}^{(k)}\right)+\mathrm{Var}\left(N_{ij}^{paid}|\chi_m\right)\mathbb{E}\left[Y_{ij}^{(k)}\right]^2 \\
&= \mathbb{E}\left[N_{ij}^{paid}|\chi_m\right]\cdot\gamma_i^2\cdot\delta_j^2\cdot(1-Q_i)\cdot\left(\sigma^2+Q_i\mu^2\right) \\
&\quad +\mathrm{Var}\left(N_{ij}^{paid}|\chi_m\right)\cdot\mu^2\cdot\gamma_i^2\cdot\delta_j^2\cdot(1-Q_i)^2 \\
&= \sum_{l=0}^{j} N_{i,j-l}\cdot p_l\cdot\gamma_i^2\cdot\delta_j^2\cdot(1-Q_i)\cdot\left(\sigma^2+Q_i\mu^2\right) \\
&\quad +\sum_{l=0}^{j} N_{i,j-l}\cdot p_l\cdot(1-p_l)\cdot\mu^2\cdot\gamma_i^2\cdot\delta_j^2\cdot(1-Q_i)^2 \\
&\approx \gamma_i^2\cdot\delta_j^2\cdot(1-Q_i)\cdot(\sigma^2+\mu^2)\cdot\sum_{l=0}^{j} N_{i,j-l}\cdot p_l \\
&= \gamma_i\cdot\delta_j\cdot\frac{\sigma^2+\mu^2}{\mu}\,\mathbb{E}\left[X_{ij}|\chi_m\right] \\
&= \varphi_{ij}\,\mathbb{E}\left[X_{ij}|\chi_m\right],
\end{aligned}$$

where $\varphi_{ij} = \gamma_i \cdot \delta_j \cdot \varphi$ and $\varphi = \frac{\sigma^2+\mu^2}{\mu}$. Using the expected value of $N^{paid}_{i,j-l,l}$

$$\begin{aligned}
\mathbb{E}\left[N^{paid}_{i,j-l,l}\right] &\overset{(D2)}{=} \mathbb{E}\left[N_{i,j-l}\right] \cdot p_l \\
&= \alpha_i \cdot \beta_{j-l} \cdot p_l,
\end{aligned}$$

the expected value of the total payments is denoted by

$$\begin{aligned}
\mathbb{E}\left[X_{ij}\right] &= \mathbb{E}\left[\sum_{l=0}^{j} \sum_{k=1}^{N^{paid}_{i,j-l,l}} Y^{(k)}_{i,j-l,l}\right] \\
&\overset{(D3)}{=} \sum_{l=0}^{j} \mathbb{E}\left[\sum_{k=1}^{N^{paid}_{i,j-l,l}} Y^{(k)}_{i,j-l,l}\right] \\
&\overset{(D4)}{=} \sum_{l=0}^{j} \mathbb{E}\left[N^{paid}_{i,j-l,l}\right] \cdot \mathbb{E}\left[Y^{(k)}_{i,j-l,l}\right] \\
&= \sum_{l=0}^{j} \alpha_i \cdot \beta_{j-l} \cdot p_l \cdot \mu \cdot \gamma_i \cdot \delta_j \cdot (1 - Q_i) \\
&= (\alpha_i \cdot \mu \cdot \gamma_i \cdot (1 - Q_i)) \cdot \left(\delta_j \sum_{l=0}^{j} \beta_{j-l} \cdot p_l\right) \\
&= \widetilde{\alpha}_i \, \widetilde{\beta}_j, \qquad (3.1)
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\alpha}_i &= \alpha_i \cdot \mu \cdot \gamma_i \cdot (1 - Q_i) \text{ and} \\
\widetilde{\beta}_j &= \delta_j \sum_{l=0}^{j} \beta_{j-l} \cdot p_l.
\end{aligned}$$

3.2 THE DOUBLE CHAIN LADDER METHOD

In this subsection, we outline the DCL method. We want to estimate reserves and distibutional properties of claims. In this subsection we assume, that $\delta_j = 1$ and $Q_i = 0$. Estimating the outstanding claims under those assumptions is much easier, because there is no inflation effect for delay year j and there are no zero-claims in our data. At the end of this section, we will take a look at how one can incorporate the parameters δ_j and Q_i. But for now, we assume that $\delta_j = 1$ and $Q_i = 0$ and thus the expectation of $Y^{(k)}_{i,j-l,l}$ and X_{ij} becomes

$$\begin{aligned}
&\mathbb{E}\left[Y^{(k)}_{i,j-l,l}|Q_i\right] = \mu\cdot\gamma_i \text{ and}\\
&\mathbb{E}\left[X_{ij}\right] = \widetilde{\alpha}_i\widetilde{\beta}_j,
\end{aligned}$$

repectively, where $\widetilde{\alpha}_i = \alpha_i\mu\gamma_i$ and $\widetilde{\beta}_j = \sum_{l=0}^{j}\beta_{j-l}p_l$. The objective of the the DCL method is to estimate the parameters $\alpha_i, \beta_j, p_l, \gamma_i, \mu, \sigma^2$. The single chain ladder method can be applied twice on the triangle χ_m and Δ_m to that end. In Section 2, we have seen that the expectation of claims can be written as $\mathbb{E}\left[Z_{ij}\right] = U_iS_j$ in (2.3), with $S_j = \frac{\lambda_j - 1}{\prod_{l=j}^{m-1}\lambda_l}$ and $S_0 = \frac{1}{\prod_{l=1}^{m-1}\lambda_l}$ for $j = 1, \ldots, m-1$. Verall [27] showed that S_j is actually an estimate of the parameter $\hat{\beta}_j$. Thus,

$$\begin{aligned}
\hat{\beta}_j &= \frac{\hat{\lambda}_j - 1}{\prod_{l=j}^{m-1}\hat{\lambda}_l} \quad \forall\ j = 0, \ldots, m-1,\\
\hat{\beta}_0 &= \frac{1}{\prod_{l=1}^{m-1}\hat{\lambda}_l};\\
\hat{\alpha}_i &= \sum_{j=0}^{m-i} N_{ij} \prod_{j=m-i+1}^{m-1}\hat{\lambda}_j \quad \forall\ i = 1, \ldots, m,
\end{aligned} \tag{3.2}$$

with $\underline{\hat{\lambda}}$ being the the vector of estimated development factors, obtained by SCL method on the triangle χ_m. The same calculations can also be used for estimating the parameters $\hat{\widetilde{\alpha}}_i$ and $\hat{\widetilde{\beta}}_j$ by using the SCL method on the triangle of paid claims Δ_m. We will denote the estimated reserves using the SCL method by $\hat{X}^{SCL}_{ij}$. Thus, the estimated reserves can be written as $X^{SCL}_{ij} = \hat{\widetilde{\alpha}}_i\hat{\widetilde{\beta}}_j$.

There is another method for estimating the parameters (α_i, β_j). This method is called the method of moments. For estimating α_i and β_j using the method of moments, we can obtain

$$\sum_{k=0}^{m-i} \mathbb{E}\left[N_{ik}\right] = \alpha_i \cdot \sum_{k=0}^{m-i} \beta_k \ \forall\ i = 1, \ldots, m,$$
$$\sum_{k=1}^{m-j} \mathbb{E}\left[N_{kj}\right] = \beta_j \cdot \sum_{k=1}^{m-j} \alpha_k \ \forall\ j = 0, \ldots, m-1.$$

Every run-off triangle differs from one company to another or for each kind of insurance. To get the parameters α_i and β_j such that they reflect the triangle χ_m in the best way, it is helpful to use the given data in the upper triangle of χ_m instead of the expectation of the number of claims N_{ij}. Thus, by looking at the upper triangle, the expectation of the estimate $\hat{\alpha}_i$ and $\hat{\beta}_j$ are indeed the parameters α_i and β_j. In order to make sure that the parameters α_i and β_j are not biased, we can use the actual values of N_{ij} for $(i, j) \in I_m$ from the given triangle instead of using the expectation of N_{ij}. By adding up the rows and columns of our data given in the triangle χ_m, we obtain a system of linear equations. This system of linear equations can be solved for α_i and β_j to obtain the chain ladder estimates $\hat{\alpha}_i$ and $\hat{\beta}_j$ for the triangle χ_m. The same calculations applied to Δ_m yield

$$\sum_{k=0}^{m-i} \mathbb{E}\left[X_{ik}\right] = \widetilde{\alpha}_i \cdot \sum_{k=0}^{m-i} \widetilde{\beta}_k \ \forall\ i = 1, \ldots, m,$$
$$\sum_{k=1}^{m-j} \mathbb{E}\left[X_{kj}\right] = \widetilde{\beta}_j \cdot \sum_{k=1}^{m-j} \widetilde{\alpha}_k \ \forall\ j = 0, \ldots, m-1.$$

By solving the two equations above, we arrive at the method of moments estimate $\hat{\widetilde{\alpha}}_i$ and $\hat{\widetilde{\beta}}_j$. The next step in the DCL process is to estimate the parameters of the delay probability p_l using the following equation:

$$\widetilde{\beta}_j = \sum_{l=0}^{j} \beta_{j-l} \cdot p_l \;\forall\; j = 0, \ldots, m-1. \tag{3.3}$$

Estimates $\hat{\beta}_{j-l}$ and $\widetilde{\hat{\beta}}_j$ are plugged in (3.3) and then solved for p_l. Since the solution $\hat{p} = (\hat{p}_0, \ldots, \hat{p}_{m-1})$ of (3.3) does not satisfy the condition in (D2), we denote the solution of (3.3) by $\hat{\pi} = (\hat{\pi}_0, \ldots, \hat{\pi}_{m-1})$. The parameter $\hat{\pi}$ can be estimated solving the following system of linear equations:

$$\begin{bmatrix} \widetilde{\hat{\beta}}_0 \\ \vdots \\ \vdots \\ \widetilde{\hat{\beta}}_{m-1} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 & 0 & \cdots & 0 \\ \hat{\beta}_1 & \hat{\beta}_0 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \hat{\beta}_{m-1} & \cdots & \hat{\beta}_1 & \hat{\beta}_0 \end{bmatrix} \cdot \begin{bmatrix} \pi_0 \\ \vdots \\ \vdots \\ \pi_{m-1} \end{bmatrix}.$$

We need to adjust the solution $\hat{\pi}$ such that it satisfies the conditions $\sum_{l=0}^{m-1} p_l = 1$ and $0 \leq p_l \leq 1 \;\forall\; l = 0, \ldots, m-1$ without altering the delay property. Martínez-Miranda et al. [13] suggested a few procedures for adjustment and the following procedure is one of them:

1) Count the number of all delay probabilities $\hat{\pi}_l > 0$ such that $\sum_{l=0}^{d-1} \hat{\pi}_l < 1 \leq \sum_{l=0}^{d} \hat{\pi}_l$ is satisfied, where $d + 1 \leq m - 1$.
2) Set $\hat{p}_l$ equal to $\hat{\pi}_l$ for all $l \leq d - 1$ such that $\hat{p}_l = \hat{\pi}_l$, for $l = 0, \ldots, d - 1$.
3) Set $\hat{p}_d$ such that the conditions $\sum_{l=0}^{m-1} \hat{p}_l = 1$ and $0 \leq \hat{p}_l \leq 1 \;\forall\; l = 0, \ldots, m-1$ are satisfied. The result will be $\hat{p}_d = 1 - \sum_{l=0}^{d-1} \hat{p}_l$.
4) Set the rest of the probabilities equal to zero such that $\sum_{l=0}^{m-1} \hat{p}_l = 1$. Thus, $\hat{p}_{d+1} = \ldots = \hat{p}_{m-1} = 0$.

For adjusting those parameters, one always needs to look at the distribution and the properties of the delay function and use those properties to adjust the delay probabilities. There are no set techniques that will always work.

To estimate the parameters γ_i and μ, the equation $\widetilde{\alpha}_i = \alpha_i \cdot \mu \cdot \gamma_i$ can be solved for γ_i to get $\gamma_i = \frac{\widetilde{\alpha}_i}{\alpha_i \mu}$. Since this equation is over-parameterised, γ_i needs to be identifiable by setting γ_1 equal to one. Thus, $\mu = \frac{\widetilde{\alpha}_1}{\alpha_1}$. The estimate of γ_i is given by

$$\hat{\gamma}_i = \frac{\hat{\widetilde{\alpha}}_i}{\hat{\alpha}_i \hat{\mu}} \quad i = 1, \ldots, m. \tag{3.4}$$

By plugging $\hat{\mu} = \frac{\hat{\widetilde{\alpha}}_1}{\hat{\alpha}_1}$ into (3.4), the rest of the parameters γ_i can be estimated. Since the equation $\sum_{l=0}^{m-1} \beta_j = 1$ and the assumptions of the delay probabilities must be satisfied, the estimated mean needs to be corrected by dividing it by κ, where $\kappa = \sum_{j=0}^{m-1} \sum_{l=0}^{j} \hat{\beta}_{j-l} \hat{p}_l$. The corrected estimate of μ will be defined again as $\hat{\mu}$.

The last parameter left to be estimated in the DCL process is the variance σ^2. In Subsection 3.1, we showed that the conditional variance of the payments given χ_m is

$$\begin{aligned} \operatorname{Var}(X_{ij}|\chi_m) &\approx \gamma_i \frac{\sigma^2 + \mu^2}{\mu} \, \mathbb{E}[X_{ij}|\chi_m] \\ &= \varphi_i \, \mathbb{E}[X_{ij}|\chi_m], \end{aligned}$$

where $\varphi_i = \gamma_i \varphi$ and $\varphi = \frac{\sigma^2 + \mu^2}{\mu}$. The variance of the outstanding claims is proportional to its mean, which means that we can use the over-dispersed Poisson model for estimating the variance of the outstanding claims by solving the parameter φ for σ^2. Thus, the variance estimator is defined by $\hat{\sigma}^2 = \hat{\mu}\hat{\varphi} - \hat{\mu}^2$. The over-dispersion parameter φ can be estimated by

$$\hat{\varphi} = \frac{1}{n-m} \sum_{i,j \in I_m} \frac{\left(X_{ij} - \hat{X}_{ij}^{DCL}\right)^2}{\hat{X}_{ij}^{DCL} \hat{\gamma}_i},$$

with $n = m(m+1)/2$ and $\hat{X}_{ij}^{DCL} = \sum_{l=0}^{j} \hat{N}_{i,j-l} \hat{p}_l \hat{\mu} \hat{\gamma}_i$.

The final step of the DCL method is to estimate the outstanding claims. We will use the unconditional mean of the total payments in (3.1) by substituting in the unconditional mean of the reported claims:

$$\begin{aligned} \mathbb{E}[X_{ij}] &= \alpha_i \cdot \mu \cdot \gamma_i \cdot \sum_{l=0}^{j} \beta_{j-l} \cdot p_l \\ &= \mu \cdot \gamma_i \cdot \sum_{l=0}^{j} \alpha_i \cdot \beta_{j-l} \cdot p_l \\ &\overset{(D1)}{=} \mu \cdot \gamma_i \cdot \left(\sum_{l=0}^{j} \mathbb{E}[N_{i,j-l}] \cdot p_l \right) \\ &= \sum_{l=0}^{j} \mathbb{E}[N_{i,j-l}] \cdot p_l \cdot \mu \cdot \gamma_i. \end{aligned}$$

The estimated parameters $\hat{\theta} = (\hat{p}_l, \hat{\mu}, \hat{\gamma})$ can now be used to forecast the RBNS and IBNR claims. Thus, we get the estimate of the total outstanding claims for $i+j > m$:

$$\begin{aligned} \hat{X}_{ij}^{DCL} &= \hat{X}_{ij}^{rbns} + \hat{X}_{ij}^{ibnr} \\ &= \sum_{l=i-m+j}^{j} \hat{N}_{i,j-l} \hat{p}_l \hat{\mu} \hat{\gamma}_i + \sum_{l=\max\{0,j-m+1\}}^{i-m+j-1} \hat{N}_{i,j-l} \hat{p}_l \hat{\mu} \hat{\gamma}_i, \end{aligned} \tag{3.5}$$

where $\hat{N}_{ij} = \hat{\alpha}_i \hat{\beta}_j$. Since the RBNS claims have already been reported, there is another possibility for estimating the RBNS claims. It is possible to use either the fitted value $\hat{N}_{ij}$ like in (3.5) or the actual numbers of claims N_{ij}. The IBNR component always uses the fitted value $\hat{N}_{ij}$ since those claims have not been reported so we do not know the actual number of claims and thus have to use the estimate of the actual number of claims. By using the fitted value $\hat{N}_{ij}$ and the unadjusted delay probabilities for the IBNR and RBNS claims, we get the same estimate of outstanding claims as in the SCL method for $(i,j) \in I_1$:

$$\begin{aligned}
\hat{X}_{ij}^{rbns} + \hat{X}_{ij}^{ibnr} &= \sum_{l=i-m+j}^{j} \hat{N}_{i,j-l}\hat{\pi}_l\hat{\mu}\hat{\gamma}_i + \sum_{l=0}^{i-m+j-1} \hat{N}_{i,j-l}\hat{\pi}_l\hat{\mu}\hat{\gamma}_i \\
&= \sum_{l=0}^{j} \hat{N}_{i,j-l}\hat{\pi}_l\hat{\mu}\hat{\gamma}_i \\
&\overset{(D1)}{=} \sum_{l=0}^{j} \hat{\alpha}_i\hat{\beta}_{j-l}\hat{\pi}_l\hat{\mu}\hat{\gamma}_i \\
&= \hat{\alpha}_i\hat{\mu}\hat{\gamma}_i \sum_{l=0}^{j} \hat{\beta}_{j-l}\hat{\pi}_l \\
&= \hat{\tilde{\alpha}}_i \sum_{l=0}^{j} \hat{\beta}_{j-l}\hat{\pi}_l \\
&= \hat{\tilde{\alpha}}_i\hat{\tilde{\tilde{\beta}}}_j \\
&= \hat{X}_{ij}^{SCL} \; . \qquad (3.6)
\end{aligned}$$

In Tables 3.1 and 3.2, the two triangles Δ_m and χ_m are shown. By applying the DCL method, with $\hat{\mu}$ being the unadjusted estimate of μ and thus dividing $\hat{\mu}$ by $\kappa = 0.9994427$, we get the adjusted estimate $\hat{\mu} = 208.491$. The following table shows the estimated parameters:

Table 3.3 Estimated Parameters Using The DCL Method

$\widehat{\pi}_l$	$\widehat{p}_l$	$\widehat{\gamma}_i$
0.3649	0.3649	1
0.2924	0.2924	0.7562
0.1119	0.1119	0.7350
0.0839	0.0839	0.8908
0.0630	0.0630	0.7840
0.0332	0.0332	0.7790
0.0245	0.0245	0.6605
0.0121	0.0121	0.7370
0.0158	0.0141	0.6990
-0.0012		0.8198
$\widehat{\mu} = 208.3748$		
$\widehat{\sigma}^2 = 2010305.5$		

With those estimated parameters, we can compute the point forecast of the IBNR and RBNS claims. The point forecasts will be calculated using (3.5). The RBNS claims will be computed using the actual number of payments N_{ij} instead of the estimated number of payments $\hat{N}_{ij}$. We will illustrate the forecast of the cash flow by calendar year k with $k = 1, \ldots, 19$. The RBNS cash flow by calendar year is computed by summing up the point forecasts along the diagonals of $I_1 \cup I_2$ and the IBNR cash flow by summing up the point forecasts along the diagonals of $I_1 \cup I_2 \cup I_3$. The following table shows the point forecast of the cash flow of the RBNS and IBNR claims:

Table 3.4 Point Forecasts of Cash Flow by Calendar Year (Numbers In Thousands)

Future	RBNS	IBNR	Total
1	1261	97	1358
2	672	83	755
3	453	36	489
4	293	27	319
5	165	20	185
6	103	12	115
7	54	9	63
8	30	5	36
9	0	5	5
10		1	1
11		0.6	0.6
12		0.4	0.4
13		0.2	0.2
14		0.1	0.1
15		0.06	0.06
16		0.03	0.03
17		0.01	0.01
18		0.00	0.00
19		0.00	0.00
Total	3030	296	3326

3.3 INCORPORATING PRIOR KNOWLEDGE

In the previous derivations, we excluded the prior knowledge about zero-claims Q_i and about the claims development inflation δ_j. We now include the aforementioned prior information, in order to have better knowledge of reserves. We begin with the inclusion of

i) prior information about the claims development inflation,

ii) prior information about zero-claims and towards the end of the subsection

iii) prior information about both, the development inflation and the zero-claims.

3.3.1 Development Inflation. First, we want to include the prior knowledge of the claims development inflation factor δ_j. To do this, we restrict the probability of zero-claims Q_i to be zero for all $i = 1, \ldots, m$ and δ_j is unrestricted. Inclusion of inflation factors will be done by first dividing the data in the triangle Δ_m by δ_j.

Next, apply the DCL method to estimate the reserves and lastly, we multiply the inflation factor δ_j back to our reserve estimates. It is quite intuitive to first remove the prior inflation effect and after the DCL method multiply it back since we used the DCL method before, with the assumption that $\delta_j = 1$ which in this case will have the same results by using the DCL method before multiplying back the parameter δ_j .

To that end, let $\widetilde{X}_{ij} = \frac{X_{ij}}{\delta_j}$ be the total payments without the inflation effect, and let $\widetilde{\Delta}_m$ be our new triangle, with $\widetilde{\Delta}_m = \left\{\widetilde{X}_{ij}; (i,j) \in I_m\right\}$. The DCL method is applied to the triangles $(\chi_m, \widetilde{\Delta}_m)$. Despite the transformation, the assumptions (D1)-(D4) are still satisfied with Q_i being zero and δ_j equal to one. Those steps lead to the prediction of $\widetilde{X}_{ij}^{DCL}$. Since we want to have the predicted reserves with incorporating the development inflation, we multiply back the prior information δ_j to the predicted reserves $\widetilde{X}_{ij}^{DCL}$. Thus, the predicted reserve with prior information is denoted by

$$\widetilde{X}_{ij}^{DCLP} = \delta_j \cdot \widetilde{X}_{ij}^{DCL}, \text{ for } (i,j) \in I_1.$$

By including the development factor in our estimation, total reserves (sum of IBNR and RBNS claims) will not be altered very much from the estimated reserves without prior information. However, the differences can be seen in the values of IBNR and RBNS. Moreover, change can occur in claims further down (that's with high j) or at the beginning of the triangle construction. For instance, if δ_j is large for large j, that implies RBNS claims will increase and IBNR claims will decline as compared to the situation where prior knowledge is not included.

3.3.2 Zero-claims. This subsection pertains to the case where prior information on zero-claims is accounted for. We want to include the prior information about the number of zero-claims. In this case, it is not that simple to include prior information like for the development inflation but can still be done. We assume that

$\delta_j = 1$ and the probability of zero-claims is in $[0, 1]$, that's $0 \leq Q_i \leq 1$. Accounting for this kind of prior knowledge will not have any effect in the best estimation of the reserves. In Subsection 3.1, we derived an approximation of the conditional variance of the claims given the assumption that $Q_i = 0$. The conditional variance with $Q_i \neq 0$ is given by

$$\begin{aligned}
\text{Var } (X_{ij}|\chi_m) &= \gamma_i^2\delta_j^2(1-Q_i)\left(\sigma^2 + Q_i\mu^2\right)\sum_{l=0}^{j} N_{i,j-l,l}p_l \\
&\quad + \gamma_i^2\delta_j^2(1-Q_i)^2\mu^2\sum_{l=0}^{j} N_{i,j-l,l}p_l(1-p_l) \\
&\approx \gamma_i^2\delta_j^2(1-Q_i)(\sigma^2+\mu^2)\sum_{l=0}^{j} N_{i,j-l,l}p_l \\
&= \gamma_i\delta_j\frac{\sigma^2+\mu^2}{\mu}\mathbb{E}\left[X_{ij}|\chi_m\right] \\
&= \varphi_{ij}\mathbb{E}\left[X_{ij}|\chi_m\right], \qquad (3.7)
\end{aligned}$$

where $\varphi_{ij} = \gamma_i\delta_j\varphi$ and $\varphi = \frac{\sigma^2+\mu^2}{\mu}$. We assumed that $\delta_j = 1$ and thus, $\varphi_i = \gamma_i\varphi$ like before. Once again, one can use the over-dispersed Poisson model to approximate the parameters. The outstanding claims with the DCL method are estimated using a different inflation parameter γ_i given by

$$\gamma_i^{DCL}\frac{(1-Q_i)}{(1-Q_1)},$$

where γ_i^{DCL} is the inflation effect of $\widetilde{X}_{ij} = \frac{X_{ij}}{(1-Q_i)}$, the triangle of claims where the zero-claims effect is removed and $\frac{(1-Q_i)}{(1-Q_1)}$ the zero claim effect. Thus, the inflation parameter γ_i can be split into the inflation parameter without the zero-claims effect and the zero-claims effect itself. Since the estimation of the variance depends on the approximation in equation (3.7), we need to know how sensitive this approximation is with respect to the parameters Q_i and p_l. If the approximation is not always close to the true value, we could not rely on the estimate of the variance, which means that we could not rely on our estimated outstanding claims. To test the goodness of this approximation, we can test if the ratio of the true conditional variance and their approximation are nearly one. That's:

$$H_0:\ \psi = 1 \text{ vs. } H_1:\ \psi \neq 1,$$

where $\psi = \frac{\text{Var }(X_{ij}|\chi_m)}{\gamma_i^2\delta_j^2(1-Q_i)(\sigma^2+\mu^2)\sum_{l=0}^{j} N_{i,j-l,l}p_l}$ is the ratio of the true conditional variance and its approximation. Martínez-Miranda et al. [13] for example performed this test on some sample data and the resulting ratios ψ varied between 0.9960 and 0.9992 which is very close to one. Thus, Martínez-Miranda et al. [13] came to the conclusion that this approximation is good enough, since the ratios are close to one. The result of incorporating zero-claims does not have a big effect on the total reserves and it does not have an effect on the split of IBNR and RBNS claims.

3.3.3 Development Inflation and Zero-claims. After including the prior knowledge about the development inflation and the zero-claims seperately, we will include both together in our calculations. This can be done by combining the technique pertaining to each inclusion. In this case, let $Q_i \neq 0$ and $\delta_j \neq 1$. To estimate the reserves, the following steps apply:

1. Remove the development inflation effect and use $(\chi_m, \widetilde{\Delta}_m)$ for the DCL method, where $\widetilde{\Delta}_m = \left\{\widetilde{X}_{ij}; (i,j) \in I_m\right\}$ and $\chi_m = \{N_{ij}; (i,j) \in I_m\}$.

2. Estimate X_{ij}^{DCL} with incorporating Q_i like in Subsection 3.3.2.

3. Multiply back the development inflation effect. Thus, we get $X_{ij}^{DCLP} = \delta_j X_{ij}^{DCL}$.

The result of incorporating both prior information has a big effect in the split betweem IBNR and RBNS but not in the total reserves which is obviously caused by the prior information about the development inflation and not by the zero-claims probability.

Since one of the objective in this paper is to predict the distibution of claims and estimate a full cash flow, we can now approximate the aforementioned distribution by using parametric bootstrap technique.

4 THE BOOTSTRAP TECHNIQUE FOR THE DCL

B. Efron [2] first introduced the bootstrap method, which is a statistical technique using resampling to estimate the value of a parameter of sample data. After B. Efron introduced this method, it spread out within some decades and today it is a widely used method in statistics. An insurance company needs to know the number of outstanding claims but only has one data set of claims of their own company available, so it is not very accurate to calculate the reserves only based this sample data. The reason for this is that we don't know if this one sample is a good representation of all claims. Because of time and resource constaints, it is almost impossible to reach out to the entire population to estimate a parameter. The bootstrap method is a resampling technique that uses one sample data, resamples (with replacement) from the sample data and creates a large number of bootstrap samples. The parameter can be computed on each of those bootstrap samples and with this, we can get a good idea about the sampling distribution.

4.1 THE BOOTSTRAP METHOD

Statisticians primary task is to summarize a sample based on a study and generalize the findings to the parent population. This sample summary is called a statistic. Problems in statistics often involve estimating this unknown statistic. The main idea of the bootstrap method is to determine how accurate the estimated statistic is. This statistic can be the sample mean, median, standard deviation or quantiles. Estimating this statistic based on one sample data is not very accurate. The statistic will fluctuate from sample to sample, but statisticians want to know the statistic of the parent population in an overall sense. Simulating repeated samples of the same size from the population of interest a large number of times can be very expensive and time consuming. The bootstrap method is a technique based on resampling, where a big amount of computations are substituted in place of the sample data. With this replacement, we get a large number of samples. Thus, the bootstrap method uses the Monte Carlo approximation to get a predictive distribution of claims. Those samples are also called bootstrap samples. Thus, the statistic can then be estimated on each of those bootstrap samples and all possible values of the statistic can be expressed in form of a probability distribution, so called sample distribution or in this case, bootstrap distribution. It is not always easy to calculate for instance the standard error of an estimate θ. If for instance our parameter θ is the mean of the sample data, then computing the standard error of θ is very easy. In fact, the standard error is then denoted by

$$\hat{\sigma} = \sigma(X) = \left(\frac{\sum_{i=1}^{n} X_i - \overline{X}}{n} \right)^{\frac{1}{2}},$$

with $\overline{x}$ being the mean of the observation. But it is not always that easy. This is the reason, why the bootstrap method is a good technique to estimate the standard error or prediction error of the parameter θ. And as a side effect, the bootstrap method also gives us a predictive distribution of this parameter. There are various ways to bootstrap data. It can be parametric, nonparametric, semiparametric. We will only focus on the parametric bootstrap method.

Here is how the bootstrap works. Suppose we want to estimate a statistic θ (μ, σ^2, F, $\ldots$). One takes a random sample of size n from a population, say $X = (X_1, \ldots, X_n)$. The main steps of the parametric bootstrap method are shown below:

1. $X_1, \ldots, X_n \overset{iid}{\sim} F$, a random sample of size n, with F an unkown distribution from a parametric family.

2. We have a sample of observations $(x_1, \ldots, x_n) \in (X_1, \ldots, X_n)$.

3. Estimate the parameter $\hat{\theta}$ from the sample data $x = (x_1, \ldots, x_n)$.

4. Compute the empirical distribution $\hat{F}$ with probability mass $\frac{1}{n}$ on the observation $x = (x_1, \ldots, x_n)$, with

$$\hat{F}(x) = \frac{\sum_{i=1}^{n} \mathbb{1}(x_i \leq x)}{n}.$$

5. With fixed $\hat{F}$, simulate random samples of size n. Those simulated random samples will be denoted by $(x_1^*, \ldots, x_n^*) \in (X_1^*, \ldots, X_n^*)$, with $X_i^* \sim \hat{F}$ for $i = 1, \ldots, n$. Replace the original sample data by $x^* = (x_1^*, \ldots, x_n^*)$.

6. Create a large number (B) of bootstrap samples $(x^{*,1}, \ldots, x^{*,B})$ and estimate $\hat{\theta}^* = (\hat{\theta}^{*,1}, \ldots, \hat{\theta}^{*,B})$ based on the B bootstrap samples.

The bootstrapped parameter $\hat{\theta}^*$ can now be used for purposes like the approximation of the standard error, confidence intervals, the computation of the prediction error or computing the predictive distribution of the parameter θ. Thus, the advantages of the bootstrap method are:

i) It is a very simple and straighforward algorithm to estimate parameters using a small sample set of data.

ii) Since it generates a large number of bootstrap samples for the estimation, the results are very stable.

iii) The boostrap method is commonly used when the true distribution of the data is intractable or is of a complex form.

It turns out that insurance companies deal with a complex set of observation, which is very hard to deal with. So the bootstrap method will help in that context, that is obtaining distributional properties of claims. One of the objectives in this thesis is to find the distributional properties of claims. One way of achieving this goal is to use the bootstrap method to estimate a predictive distribution of reserves, with the inclusion of prior information. Using a large number of bootstrap samples (≥ 1000), the bootstrap method can estimate mean, prediction error and also some quantiles that will describe the distribution of reserves. In the following subsection, we will apply the bootstrap method on the DCL method and also use some data in form of a run-off triangle to estimate the predictive distribution of outstanding claims.

4.2 APPLICATION OF THE BOOTSTRAP METHOD

This subsection shows how the bootstrap method can be used to get the predictive distribution of reserves. Incorporating prior information, the main steps of estimating the distribution of the reserves are

1. Remove the development inflation effect and the knowledge on the number of zero-claims by dividing the claims X_{ij} by the inflation parameter δ_j and the number of zero claims $(1 - Q_i)$. Thus, the claims triangle excluding the prior information is denoted by $\widetilde{\Delta}_m = \left\{\widetilde{X}_{ij}; (i,j) \in I_m\right\}$, where $\widetilde{X}_{ij} = \frac{X_{ij}}{\delta_j(1-Q_i)}$.

2. Use the Bootstrap method on the triangles $(\chi_m, \widetilde{\Delta}_m)$ to simulate IBNR and RBNS claims including the information about the zero-claims. Thus, we get the bootstrapped IBNR predictions $\widetilde{X}_{ij}^{ibnr*}$ and the bootstrapped RBNS predictions $\widetilde{X}_{ij}^{rbns*}$.

3. Replace the development inflation effect by multiplying the inflation parameter δ_j back to the bootstrapped IBNR and RBNS predictions. So, the final IBNR and RBNS predictions are $X_{ij}^{ibnr*} = \delta_j \widetilde{X}_{ij}^{ibnr*}$ and $X_{ij}^{rbns*} = \delta_j \widetilde{X}_{ij}^{rbns*}$, repectively.

The IBNR and RBNS claims will be estimated seperately and added together towards the end to get the total bootstrapped reserves. To generate a large number of bootstrap samples, the Monte Carlo approximation is used. First, we use the DCL method to estimate θ using the triangles $(\chi_m, \widetilde{\Delta}_m)$, with $\widetilde{\Delta}_m = \left\{\widetilde{X}_{ij}; (i,j) \in I_m\right\}$. The estimated parameter $\hat{\theta}$ is then used by the bootstrap method to simulate the reserves. There are two different ways to simulate those reserves. The first one is by ignoring the uncertainty of the estimated parameters $\hat{\theta}$ and the second is by incorporating the uncertainty of $\hat{\theta}$. The following assumptions for the bootstrap method are needed:

- Assume (D1)-(D4) are still in force including the development inflations factor δ_j and the probability of zero-claims Q_i.
- δ_j and Q_i are known.
- $\theta = \left\{p_l, \mu_{ij} = \gamma_i \delta_j \mu, \sigma_{ij}^2 = \sigma_i^2 \gamma_i^2 \delta_j^2; i = 1, \ldots, m; l, j = 0, \ldots, m-1\right\}$.
- The maximal delay is $m - 1$ years.

Before we start simulating the claims, an estimate of the parameter θ is needed. Our estimation is based on the triangles $(\chi_m, \widetilde{\Delta}_m)$, with $\widetilde{\Delta}_m = \left\{\widetilde{X}_{ij} = \frac{X_{ij}}{\delta_j(1-Q_i)}; (i,j) \in I_m\right\}$. Thus, the estimate of θ is denoted by

$$\hat{\theta} = \left\{\hat{p}_l, \hat{\mu}_{ij} = \hat{\gamma}_i \delta_j \hat{\mu}, \hat{\sigma}_{ij}^2 = \hat{\sigma}_i^2 \hat{\gamma}_i^2 \delta_j^2, l = 0, \ldots, m-1, i = 1, \ldots, m\right\},$$

where $\hat{\sigma}_i^2 = (1 - Q_i)\hat{\sigma}^2 - Q_i \hat{\mu}$, the variance of the data including the probability of having a zero-claim and $\hat{\gamma}_i = \hat{\gamma}_i^{DCL}$.

Algorithm for RBNS claims:

The DCL method estimates the RBNS claims over the index sets $I_1 \cup I_2$ and thus, using the bootstrap method, the RBNS claims will also be bootstrapped over those sets. The following steps are done:

i) The DCL method will be used on our given run-off triangle $\widetilde{\Delta}_m$. Using equation (3.3) and (3.4), we get the estimated parameter $\hat{\theta}$.

ii) A new run-off triangle $\widetilde{\Delta}_m^*$ will be simulated using the assumptions in (D1)-(D4) and the estimated parameter $\hat{\theta}$.

iii) The parameter $\hat{\theta}$ will be bootstrapped using the DCL method on the bootstrapped triangle $\widetilde{\Delta}_m^*$. Estimating θ based on this triangle yields the bootstrapped parameter θ^*. This parameter will be used to calculate the RBNS predictions.

iv) The Monte Carlo approximation is used to repeat the steps i) - iii) B times to get the empirical bootstrap distribution of the RBNS predictions.

Figure 4.1 shows the main steps of the bootstrap method for the RBNS claims, including the uncertainty of the parameters but ignoring the development parameter δ_j and the probability of zero-claims Q_i. This figure shows that the original data (χ_m, Δ_m) is used to estimate the parameter $\hat{\theta}$ and with this parameter, the RBNS claims can be computed over $I_1 \cup I_2$. With the estimated prameter, the data can be bootstrapped and so (χ_m, Δ_m^*) can be used to calculate the bootstrapped parameter θ^* thereby obtaining the bootstrapped RBNS predictions over the index set $I_1 \cup I_2$. Repeating this procedure B times leads to the predictive bootstrap RBNS distribution. It is possible to estimate the IBNR and RBNS predictions with including the fact, that we don't know the true value of the parameter θ and thus, there is uncertainty in those calculations. To exclude the uncertainty of the parameter $\hat{\theta}$, the estimated parameter $\hat{\theta}$ will be directly used to calculate the RBNS claims without using the bootstrapped parameter θ^*.

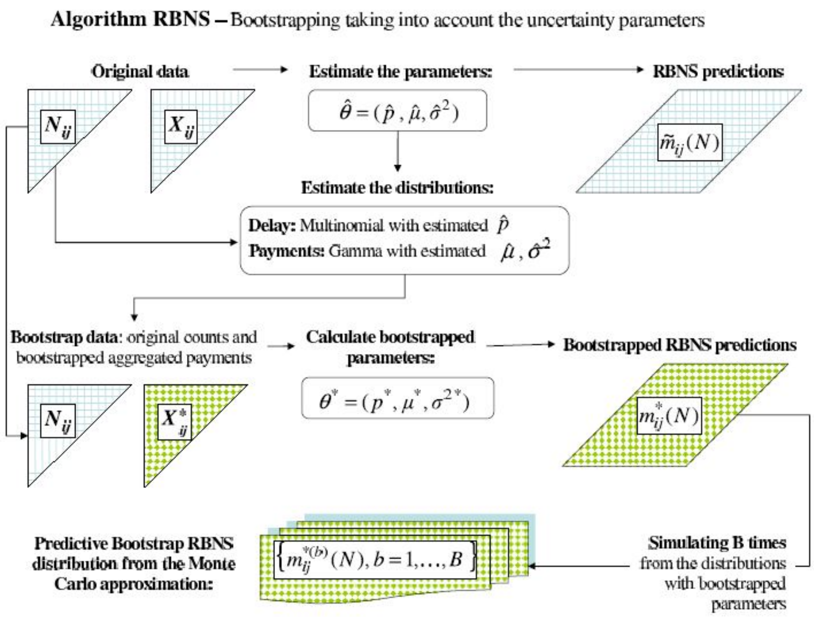

Figure 4.1 Bootrapping the RBNS Claims,
Source: Martínez-Miranda et al. [14]

The following algorithm shows the bootstrap method estimating the RBNS claims including the uncertainty of the parameter θ:

1. Estimation of the parameters and distribution:

   - Estimate $\hat{\theta}$ for the observed data $(\chi_m, \widetilde{\Delta}_m)$.
   - $\left(N^{paid}_{i,j,0}, \ldots, N^{paid}_{i,j,m-1}\right) \sim$ Mutli $(N_{ij}; \hat{p}_0, \ldots, \hat{p}_{m-1})\ \forall\ (i,j) \in I_m$.
   - $\left(Y^{(k)}_{i,j,l} > 0, l = 0, \ldots, m-1\right) \sim$ Gamma $\left(\hat{\lambda}_i, \hat{\kappa}_i\right)$, where the mean of the non-zero individual payments is $\mu_i = \hat{\gamma}_i\hat{\mu}$ and the variance is $\sigma_i^2 = \hat{\sigma}_i^2$. Thus, $\hat{\lambda}_i = \hat{\gamma}_i^2\hat{\mu}^2/\hat{\sigma}_i^2$ is the shape parameter and $\hat{\kappa}_i = \hat{\sigma}_i^2/\hat{\gamma}_i\hat{\mu}$ is the scale parameter.

2. Bootstrapping the data:

   Generate $\Delta_m^* = \left\{X_{ij}^*; (i,j) \in I_m\right\}$ given χ_m.

   - Simulate the payment delay $N_{i,j,l}^{paid*}$ by
     $\left(N_{i,j,0}^{paid*}, \ldots, N_{i,j,m-1}^{paid*}\right) \sim$ Mutli $(N_{ij}; \hat{p}_0, \ldots, \hat{p}_{m-1})$ $\forall$ $N_{ij}, (i,j) \in I_m$.
   - Simulate the number of non-zero payments N_{ij}^{paid*} by
     $N_{ij}^{paid*} \sim$ Bin $\left(\sum_{l=0}^{j} N_{i,j-l,l}^{paid*}, 1 - Q_i\right)$ $\forall$ $(i,j) \in I_m$.
   - Simulate X_{ij}^* $\forall$ $(i,j) \in I_m$ by
     $X_{ij}^* \sim$ Gamma $\left(N_{ij}^{paid*}\hat{\lambda}_i, \hat{\kappa}_i\right)$

3. Bootstrapping the parameters (include uncertainty of parameters):

   - $\widetilde{\Delta}_m^* = \left\{\widetilde{X}_{ij}^*; (i,j) \in I_m\right\}$, with $\widetilde{X}_{ij}^* = \frac{X_{ij}^*}{(1-Q_i)}$
   - $\chi_m = \{N_{ij}; (i,j) \in I_m\}$
   - Estimate θ based on $(\chi_m, \widetilde{\Delta}_m^*)$ and get a bootstrapped parameter θ^*.

4. Simulate the RBNS claims:

   - Simulate the payment delay $N_{i,j,l}^{rbns*}$ by
     $\left(N_{i,j,0}^{rbns*}, \ldots, N_{i,j,m-1}^{rbns*}\right) \sim$ Mutli $(N_{ij}; p_0^*, \ldots, p_{m-1}^*)$ $\forall$ $N_{ij}, (i,j) \in I_m$.
   - Simulate the number of non-zero payments N_{ij}^{rbns*} by
     $N_{ij}^{rbns*} \sim$ Bin $\left(\sum_{l=0}^{j} N_{i,j-l,l}^{rbns*}, 1 - Q_i\right)$ $\forall$ $(i,j) \in I_1 \cup I_2$.
   - Simulate X_{ij}^{rbns*} $\forall$ $(i,j) \in I_1 \cup I_2$ by
     $\widetilde{X}_{ij}^{rbns*} \sim$ Gamma $\left(N_{ij}^{rbns*}\lambda_i^*, \kappa_i^*\right)$, with $\lambda_i^* = \gamma_i^{*2}\mu^{*2}/\sigma_i^{*2}$ and $\kappa_i^* = \sigma_i^{*2}/\gamma_i^*\mu^*$.
     $X_{ij}^{rbns*} = \delta_j \widetilde{X}_{ij}^{rbns*}$.

5. Monte Carlo approximation:

   Repeat Step 2-4 B times and get the empirical distribution of RBNS claims $\left\{X_{ij}^{rbns*,b}; (i,j) \in I_1 \cup I_2, b = 1, \ldots, B\right\}$.

To ignore the uncertainty of the parameters, only step 1,4 and 5 are completed and thus, $\hat{\theta}$ from Step 1 is used in Step 4.

Algorithm for IBNR claims:

Bootstrapping the IBNR claims is different from bootstrapping RBNS claims. The first difference is that the DCL method estimates the IBNR claims over the index set $I_1 \cup I_2 \cup I_3$ and not only over $I_1 \cup I_2$ like the RBNS claims. And so the RBNS predictions using the bootstrap method will be over $I_1 \cup I_2 \cup I_3$. Another difference is that we don't know the number of claims in I_1 and thus also need to bootstrap the data in the triangle χ_m and not only the data in the triangle Δ_m like for the RBNS prediction. Figure 4.2 shows the main steps of the bootstrap method for the IBNR claims including the uncertainty of the parameters but again ignoring the additional information abour δ_j and Q_i.

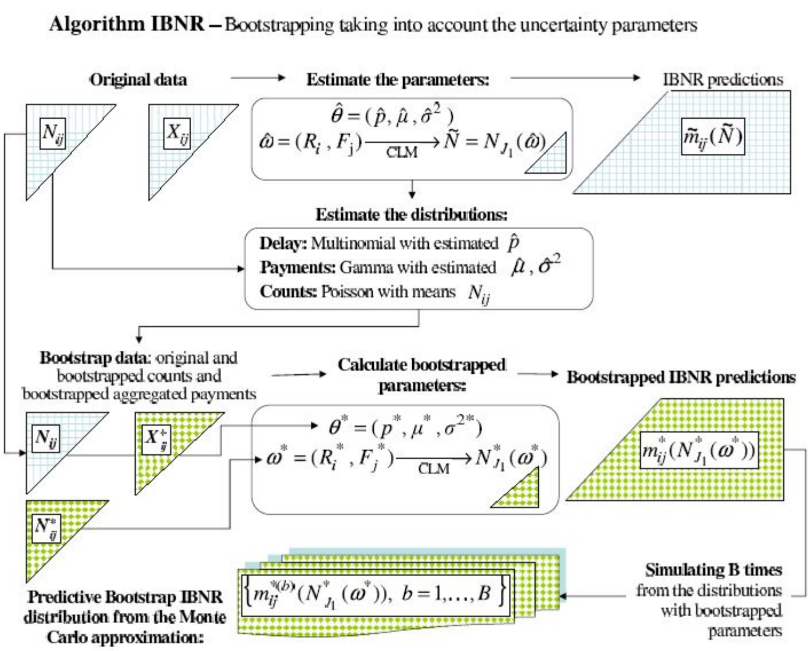

Figure 4.2 Bootrapping the IBNR Claims,
Source: Martínez-Miranda et al. [14]

Figure 4.2 shows that the first step in estimating the parameters is exactly the same as for RBNS claims. But with those parameters the IBNR claims are calculated over a bigger set, namely $I_1 \cup I_2 \cup I_3$. The original and the bootstrapped counts and the bootstrapped aggregated payments will be used in the bootstrap method for the IBNR prediction. The original counts and the bootstrapped payments are used to estimate the parameter θ^*, while the bootstrapped counts are used to predict the counts in I_1. The bootstrapped IBNR predictions are estimated over the index set $I_1 \cup I_2 \cup I_3$. Repeating the steps above B times, we get the bootstrapped IBNR distribution.

Thus, the following algorithm shows the bootstrap method estimating IBNR claims, including the uncertainty of parameters and also incorporating the information about δ_j and Q_i:

1. Estimation of parameters and distribution:

   - Estimate $\hat{\theta}$ for the observed data $\widetilde{\Delta}_m$.
   - Estimate $\hat{\alpha}$ and $\hat{\beta}$ for the observed data χ_m.
   - $\left(N^{paid*}_{i,j,0}, \ldots, N^{paid*}_{i,j,m-1}\right) \sim$ Mutli $(N_{ij}; \hat{p}_0, \ldots, \hat{p}_{m-1})\ \forall\ (i,j) \in I_m$.
   - $\left(Y^{(k)}_{i,j,l} > 0, l = 0, \ldots, m-1\right) \sim$ Gamma $\left(\hat{\lambda}_i, \hat{\kappa}_i\right)$.

2a. Bootstrapping the data $\Delta^*_m = \left\{X^*_{ij}; (i,j) \in I_m\right\}$:

   - Simulate the payment delay $N^{paid*}_{i,j,l}$ by
     $\left(N^{paid}_{i,j,0}, \ldots, N^{paid}_{i,j,m-1}\right) \sim$ Mutli $(N_{ij}; \hat{p}_0, \ldots, \hat{p}_{m-1})\ \forall\ N_{ij}, (i,j) \in I_m$.
   - Simulate the number of non-zero payments N^{paid*}_{ij} by
     $N^{paid*}_{ij} \sim$ Bin $\left(\sum_{l=0}^{j} N^{paid*}_{i,j-l,l}, 1 - Q_i\right)\ \forall\ (i,j) \in I_m$.
   - Simulate $X^*_{ij}\ \forall\ (i,j) \in I_m$ by
     $X^*_{ij} \sim$ Gamma $\left(N^{paid*}_{ij} \hat{\lambda}_i, \hat{\kappa}_i\right)$.

2b. Bootstrapping the data $\chi_m^* = \left\{N_{ij}^*; (i,j) \in I_m\right\}$:

- Simulate the number of claims N_{ij} by $N_{ij}^* \sim$ Poi $(\hat{\alpha}_i \hat{\beta}_j)$ $\forall$ $(i,j) \in I_1$.

3. Bootstrapping the parameters (include uncertainty of parameters):

- $\widetilde{\Delta}_m^* = \left\{\widetilde{X}_{ij}^*; (i,j) \in I_m\right\}$, with $\widetilde{X}_{ij}^* = \frac{X_{ij}^*}{(1-Q_i)}$
- $\chi_m = \{N_{ij}; (i,j) \in I_m\}$
- Estimate θ based on $(\chi_m, \widetilde{\Delta}_m^*)$ and get a bootstrapped parameter θ^*.

4. Simulating the IBNR claims:

- Simulate the payment delay $N_{i,j,l}^{ibnr*}$ by
  $\left(N_{i,j,0}^{ibnr*}, \ldots, N_{i,j,m-1}^{ibnr*}\right) \sim$ Mutli $(N_{ij}^*; p_0^*, \ldots, p_{m-1}^*)$ $\forall$ $N_{ij}^*, (i,j) \in I_1$.
- Simulate the number of non-zero payments N_{ij}^{ibnr*} by
  $N_{ij}^{ibnr*} \sim$ Bin $\left(\sum_{l=0}^{j} N_{i,j-l,l}^{ibnr*}, 1 - Q_i\right)$ $\forall$ $(i,j) \in I_1 \cup I_2 \cup I_3$.
- Simulate X_{ij}^{ibnr*} $\forall$ $(i,j) \in I_1 \cup I_2 \cup I_3$ by
  $\widetilde{X}_{ij}^{ibnr*} \sim$ Gamma $\left(N_{ij}^{ibnr*} \lambda_i^*, \kappa_i^*\right)$.
  $X_{ij}^{ibnr*} = \delta_j \widetilde{X}_{ij}^{ibnr*}$.

5. Monte Carlo approximation:

   Repeat Step 2-4 B times and get the empirical distribution of IBNR claims $\left\{X_{ij}^{rbns*,b}; (i,j) \in I_1 \cup I_2 \cup I_3, b = 1, \ldots, B\right\}$.

To ignore the uncertainty of the parameter θ, only steps 1, 4 and 5 are used and thus the boostrap method is based on $\hat{\theta}$ and not on θ^*. To apply the bootstrap method, we will use the triangles (χ_m, Δ_m) from Subsection 3.1. With the bootstrap method we are able to derive a predictive distribution of the RBNS and IBNR claims. In addition to that, we want to simulate triangles $(\chi_m^{'}, \Delta_m^{'})$ using the conditions (D1)-(D4) and then use the bootstrap method on those simulated triangles.

After performing the bootstrap method $B = 1000$ times, using the data of the triangles (χ_m, Δ_m) in Tables 3.1 and 3.2, including the uncertainty of the parameters, we obtain the following predictive distribution of the RBNS and IBNR claims:

Table 4.1 Distribution Forecast of RBNS and IBNR Claims on Original Data (Numbers in Thousands)

Distribution	RBNS	IBNR	Total
mean	3015	300	3315
pe	378	79	387
1%	2204	122	2495
5%	2446	181	2735
50%	2987	297	3290
95%	3660	437	3994
99%	4054	500	4339

The root mean square error, also known as the prediction error 'pe' is calculated using the following formula:

$$\begin{aligned} \text{RSME}\left(\hat{\theta}\right) &= \sqrt{\text{MSE}\left(\hat{\theta}\right)} \\ &= \sqrt{\mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right]}, \end{aligned}$$

where θ is the estimate for the reserves calculated from the original data and $\hat{\theta}$ is the bootstrapped estimate. Since we have no informtion about δ_j and Q_i, we ignore those parameters and set $\delta_j = 1$ and $Q_i = 0$. In Table 4.1, we see that the average claim is 3315000 and the prediction error 387000. The average claim is obtained by the following formula, where B is the number of repetition of the Monte

Carlo approximation and X_{ij}^{ibnr*} and X_{ij}^{rbns*} are the bootstrapped IBNR and RBNS predictions:

$$
\begin{aligned}
\mu(X_{ij}^{rbns*}) &= \frac{1}{B}\sum_{i=1}^{B} X_{i,j}^{rbns*,i} \ \forall (i,j) \in I_1 \cup I_2, \\
\mu(X_{ij}^{ibnr*}) &= \frac{1}{B}\sum_{i=1}^{B} X_{i,j}^{ibnr*,i} \ \forall (i,j) \in I_1 \cup I_2 \cup I_3.
\end{aligned}
$$

The average claim μ is denoted by 'mean' in Table 4.1. The probability that the total outstanding claims will be less than 4339000 is 99% and that the total outstanding claims are less than 2495000 is only 1%. With this predictive distribution, insurance companies can get a better idea of how the claims arise and how big their reserves should be to prevent insolvency. How much reserves a company in fact has, depends on the company itself. It may be helpful to look at the mean for the reserve and then with the predicive distribution decide how much money to put aside. An insurance company for instance can also use some risk measures like the Value at Risk (VaR) or the Tail Value at Risk (TVaR) to determine how much money they want to put aside. The VaR describes the amount of money required to ensure, with a very high probability p that a company does not get insolvent. The VaR is defined as follows:

$$
\mathbb{P}\left(X > \mathrm{VaR}_p(X)\right) = 1 - p,
$$

with $1-p$ the probability that a claim exceeds a certain amount and X the random variable for claims. Some companies want to be less risky and thus, take the $\mathrm{VaR}_{0.99}(X)$ of the bootstrapped data, where other companies may choose a lower probability, say $\mathrm{VaR}_{0.92}(X)$ or $\mathrm{VaR}_{0.95}(X)$. If a company chooses 92% instead of 99% for p, they will put less money aside but the probability that they have to pay more claims than they estimated is higher and thus it is more risky to get insolvent. The

Value at Risk for $p = 0.99$ and $p = 0.95$ in Table 4.1 is $\text{VaR}_{0.99}(X) = 4339000$ and $\text{VaR}_{0.95}(X) = 3994000$. But it is also possible to use the TVaR for estimating how much money a company should have as a reserve. The Tail Value at Risk measures the average claim for claims that exceed the Value at Risk and thus, is denoted by

$$\text{TVaR}_p(X) = \mathbb{E}\left[X | X > \text{VaR}_p(X)\right].$$

In this example, we get $\text{TVaR}_{0.99}(X) = 4532340$ and $\text{TVaR}_{0.95}(X) = 4173850$. One can see that the TVaR yields higher numbers than the VaR since it focuses on extreme and high claims instead on all claims, like the VaR for instance does.

It is not always good to only use the available data from the company and apply the bootstrap method on one single triangle. It is important to also use the bootstrap method on simulated data, which represents the general structure of claims. Using simulated data, companies with less data available can still get a good predictive distribution of claims. Using the assumptions in (D1)-(D4) one can simulate the triangles (χ_m, Δ_m) in the following way:

- $N_{ij} \sim \text{Poi}(\alpha_i \beta_j)$, with α_i the expected total ultimate claim amount for accident year i and β_j the expected proportion of the ultimate claim amount for delay year j. Those parameters can be estimated using the equation for $\hat{\alpha}_i$ and $\hat{\beta}_j$ in (3.2).

- $X_{ij} \sim \text{Gamma}\ (N_{ij}^{paid} \lambda_i, \kappa_i)$ like in the bootstrap method in Subsection 4.1.

To estimate the parameters $\alpha_i, \beta_j, N_{ij}^{paid}, \lambda_i$ and κ_i, we will use the triangles (χ_m, Δ_m) in Tables 3.1 and 3.2 as our underlying data. In (D1), we stated that the number of payments is Poisson distributed with mean $\alpha_i \beta_j$. Using this parameter on the triangles (χ_m, Δ_m), we get the following simulated run-off triangle χ_m:

Table 4.2 Simulated Triangle: Aggregated Incurred Counts

N_{ij}	0	1	2	3	4	5	6	7	8	9
1	6182	841	27	4	3	1	3	1	1	4
2	7867	1095	36	5	5	1	1	1	1	0
3	10007	1327	42	13	6	5	0	1	0	0
4	9361	1288	38	8	4	3	0	0	0	0
5	9761	1211	60	8	4	3	0	0	0	0
6	9953	1366	43	10	0	0	0	0	0	0
7	9773	1318	34	12	0	0	0	0	0	0
8	10730	1529	53	0	0	0	0	0	0	0
9	12052	1599	0	0	0	0	0	0	0	0
10	11045	0	0	0	0	0	0	0	0	0

Simulating the other triangle using a gamma distribution, leads to the following triangle:

Table 4.3 Simulated Triangle: Aggregated Payments

X_{ij}	0	1	2	3	4	5	6	7	8	9
1	451288	339519	333371	144988	93243	45511	25217	20406	31482	1729
2	448627	512882	168467	130674	56044	33397	56071	26522	14346	0
3	693574	497737	202272	120753	125046	37154	27608	17864	0	0
4	652043	546406	244474	200896	106802	106753	63688	0	0	0
5	566082	503970	217838	145181	165519	91313	0	0	0	0
6	606606	562543	227374	153551	132743	0	0	0	0	0
7	536976	472525	154205	150564	0	0	0	0	0	0
8	554833	590880	300964	0	0	0	0	0	0	0
9	537238	701111	0	0	0	0	0	0	0	0
10	684944	0	0	0	0	0	0	0	0	0

Applying the bootstrap method $B = 1000$ times on the simulated triangles in Tables 4.2 and 4.3, the predictive distribution of the RBNS and IBNR claims is shown in Table 4.4. This predictive distribution of the RBNS and IBNR claims using the simulated triangles is very similar to the distribution of the claims using the original data in Tables 3.1 and 3.2. The mean of the total claims, using the original data is 3315000 and the mean of the simulated triangles is 3322000. Thus the mean of both

Table 4.4 Distribution Forecast of RBNS and IBNR Claims on Simulated Data (Numbers in Thousands)

Distribution	RBNS	IBNR	TOTAL
mean	3027	295	3322
pe	377	77	384
1%	2219	134	2497
5%	2431	175	2713
50%	3016	290	3309
95%	3676	429	3980
99%	3973	496	4289

predictions are very similar. We can also see that both predictions have nearly the same prediction error. Using the original data, the prediction error is 378000 and using the simulated triangles, the prediction error is 384000.

The question that arises now is, which of the two approaches for estimating the total claims is better? This means, which of the two approaches gets a more accurate estimate for outstanding claims. Which one of those estimates has less error and also which of those approaches is easier to implement. In most cases the error is measured by the standard deviation, VaR or TVaR or root mean quare error (here prediction error). Thus, an insurance company wants to predict claims as exact as possible and so the prediction error of the bootstrapped parameter should as small as possible. Does the DCL method has a smaller error than the SCL since there is a underlying theory behind this method, or is the SCL outperforming since it is very easy to compute? Should an insurance company use the bootstrap method in addition to the DCL method to calculate the prediction error and get a predictive distribution? These issues will be discussed in the next section.

5 COMPARISON

In this section, the different techniques for estimating claims, discussed in this work, will be compared. This thesis gave an overview of the SCL technique and discussed in depth the DCL method. Additionally, the bootstrap method using the DCL method was presented. First, we compare the point forecast of the SCL method and that of the DCL method. Equation (3.5) was used to estimate the point forecast of claims and the output is shown in Table 5.1. In equation (3.5), N_{ij} instead of $\hat{N}_{ij}$

Table 5.1 Point Forecasts for DCL and SCL of Cash Flow by Calendar Year (Numbers in Thousands)

Future	RBNS	IBNR	TOTAL	SCL
1	1261	97	1358	1354
2	672	83	755	754
3	453	36	489	489
4	293	27	319	318
5	165	20	185	185
6	103	12	115	115
7	54	9	63	63
8	30	5	36	36
9	0	5	5	2
10		1	1	
11		0.6	0.6	
12		0.4	0.4	
13		0.2	0.2	
14		0.1	0.1	
15		0.06	0.06	
16		0.03	0.03	
17		0.01	0.01	
18		0.00	0.00	
19		0.00	0.00	
Total	3030	296	3326	3316

was used to estimate the RBNS claims. The point forecast of the total claim using the DCL method is similar to the point forecast using the SCL method. In (3.6), we showed that the DCL method estimates exactly the same outstanding claims for

$(i,j) \in I_1$ as the SCL does, if the estimated number of payments $\hat{N}_{ij}$ is used instead of $N_{i,j}$. Since the outstanding claims in the DCL in Table 5.1 are estimated with the actual number of payments N_{ij} instead of $\hat{N}_{ij}$, the point forecasts for the DCL is slightly different than the point forecast for the SCL. However, looking at future years $1,\dots,9$, the DCL and the SCL methods produce nearly the same point forecasts. The difference of the point forecast is for calendar years greater than 10. The DCL method produces a tail for calendar year $10,\dots,19$. That means that the SCL underestimates the claims filed after the maximal delay year $m-1=9$. Although there are not many claims estimated in later years, it is better not to underestimate the outstanding claims. Another difference in the point forecast is the decomposition of the claims into RBNS and IBNR claims produced by the DCL method. Table 5.1 shows that the DCL method produces the point forecasts by separating the reporting delay from the payment delay.

We also compare the difference between the SCL and DCL method using the bootstrap method. To that end, the bootstrap method was applied to the SCL method. The bootstrap method results using the SCL and DCL method, are given in the table below.

Table 5.2 Predictive Distribution of Claims using SCL and DCL (Numbers in Thousands)

Distribution	SCL	DCL
Mean	3539	3315
pe	297	387
1%	3233	2495
5%	3255	2735
50%	3536	3290
95%	3842	3994
99%	3872	4339

Table 5.2 shows that the mean of the outstanding claims for the SCL method is nearly the same as that for the DCL method. Namely, the mean for the SCL is 3539000 and that for the DCL is 3315000. We mentioned before that the SCL underestimates claims in later years. This can been seen clearly by looking at the 95% and the 99% quantile. Although the mean in both approaches is nearly the same, the DCL method yields higher estimates than the SCL methods. The 99% quantile for the SCL for instance is 3872000, whereas the 99% quantile for the DCL is 4339000. That is a difference of 467000. The prediction error for the SCL is 297000 and thus, it is much smaller than the prediction error for the DCL, which is 387000. This can be explained by the fact that the DCL estimates reserves till delay year 19, which yields in a higher probability for making prediction errors. Also by looking at Table 5.2, we can see that the 1% quantile for the SCL is 3233000 and the 99% quantile is 3872000 which has a difference of 639000. On the other hand, the difference of the 1% and the 99% quantile of the DCL is 1844000, which is much higher. The range of the estimated reserves using the DCL is much bigger than the range using the SCL.

We also compare the point forecasts using the DCL method with the predictive distribution using the bootstrap method on the DCL method. The DCL method only produces estimates for each calendar year but does not give any additional properties on the distribution of outstanding claims. Using the bootstrap method in addition to the DCL method, we can see that the mean in Table 4.1 is very similar to the total reserves of the point forecast. Thus, the mean produced by the bootstrap method is very similar to the reserves predicted by the DCL but in addition to that, the bootstrap method produces a predictive distribution. Especially having a prediction error can help to evaluate the mean and the quantiles and get a better understanding of how claims arise, as time goes by.

6 CONCLUSIONS

6.1 SUMMARY

As mentioned before, the SCL method is a very easy and staightforward method for estimating outstanding claims. Thus, it is okay for an insurance company to use this technique. But one disadvantage of the SCL method is that it does not differentiate between IBNR and RBNS claims. Moreover, it does not take into account the distribution of claims or how the claims arise. This technique also has no option for incorporating additional information or extensions beyond delay year $m-1$. The SCL method only takes the payment delay into account but not the reporting delay. This method assumes that no claims will occur after setting up the reserves. As discussed in Section 2, the SCL is not a reliable method for estimating reserves. If an insurance company wants to have a good estimation and also would like to have the chance to include any additional information like inflation, zero-claims or tails, then the DCL method should be recommended. One reason for using the DCL method is the tail produced using the DCL method. The SCL underestimates the reserves after $m-1$ years and thus, thereby increasing the likelihood of insolvency. In an insurance company, claims can sometimes be filed after 20, 25 years. If a insurance company only uses the SCL method to estimate their reserves and never estimate reserves after $m-1$ years, they may run out of money to pay due claims because the forecast does not include years $m, m+1, \ldots$ First, the DCL method uses two triangles and not only one like the SCL method. Additional to the claims, the DCL method also looks at the aggregate counts and thus, is able to predict reserves more accurately. Accounting for this additional triangle has another avantage. The aggregate counts can be used to make a split between RBNS and IBNR claims and thus, the source of the delay can be split into two different parts, the payment delay and the reporting delay.

So to conclude, the DCL method has many advantages that can help predict more accurately reserves and should be the one utilized by practitioners and researchers.

6.2 FURTHER RESEARCH

What happens if an insurance company knows that there will be zero-claims but they don't know the probability of those claims? Also, what happens if a company does not know the severity development inflation parameter but needs to include this parameter in its framework since there obvious is inflation. This paper did not cover how the inflation parameter or the zero-claim probability can be estimated if we don't know those parameters. Martinez Miranda et. al. [13] showed that it is possible with additional information to estimate those parameters.

There are also other things we did not look into in this thesis. For example many assumptions on page 386 in [13] can be weakened to make the DCL method a better model:

i) We can take N_{ij}, the number of reported claims, to be a poisson process with rate λ. It can happen, that the number of reported claims vary during the time of the year. In the winter, when there is snow on the streets or there is a blizzard, there will be probably more car accidents than at a time in the year where it is not that dangerous to drive. Also in the tornado season, more claims will be filed regarding damage on houses or cars. Thus, one can take a look at N_{ij} as a poisson process.

ii) One can also consider the case where the total payments X_{ij} follow a process that depend on t, for example a gamma process. The total payments can differ from time to time in a given year. Moreover, it can depend from region to region. For instance, there could be more payment in east coast and west coast in the

US as compared to the mid-west because both areas have more concentration of people and therefore more proportion insured hence leading to more claims.

iii) One can also take a look at the number of reported claims N_{ij} and the individual claims $Y_{ijl}^{(k)}$. In this thesis we assumed in (D4) that N_{ij} and $Y_{ijl}^{(k)}$ are independent of each other. But the number of reported claims and the individual claims can not always be assumed to be independent of each other. Is it possible to estimate reserves using dependence between N_{ij} and $Y_{ijl}^{(k)}$? That dependence can be modeled using frailty or other existing dependence models.

iv) Another potential research problem could be the research of the development factors itself. By looking at the development factors λ_j, we can see that those factors follow a certain pattern. A development factor λ_j is the ratio of the cumulative claims of development year j versus year $j-1$. The plot of the development factors in Tables 3.1 and 3.2 for $j \in \{0, \ldots, m-1\}$ is given in Figures 6.1 and 6.2. Clearly, as time goes by, the development factor λ_j stabilizes to 1. This means that in the first development years, many claims will be filed and settled but after a few years, nearly no more claims will be filed and settled. The factor λ_j being approximately equal to 1 means, that the cumulative claims in year j versus year $j-1$ are almost the same. This observation leads to the fact that reserves can be calculated till development year 6 and then use a tail factor, mentioned at the beginning of section 3, to make sure that an insurance company does not underestimate the reserves. Well, this requires future investigation.

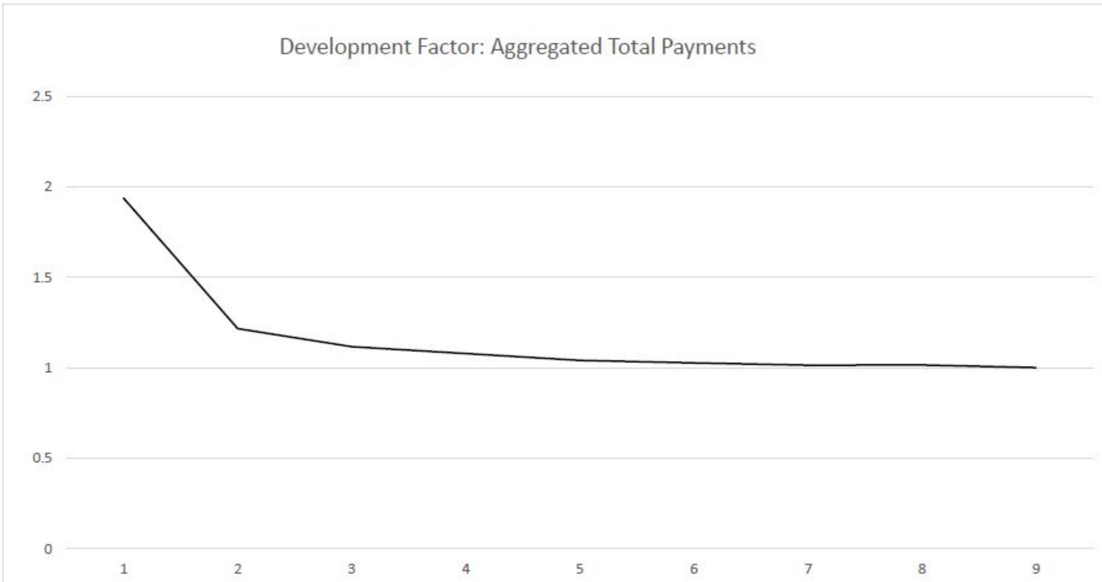

Figure 6.1 Development Factor of the Aggregated Total Payments

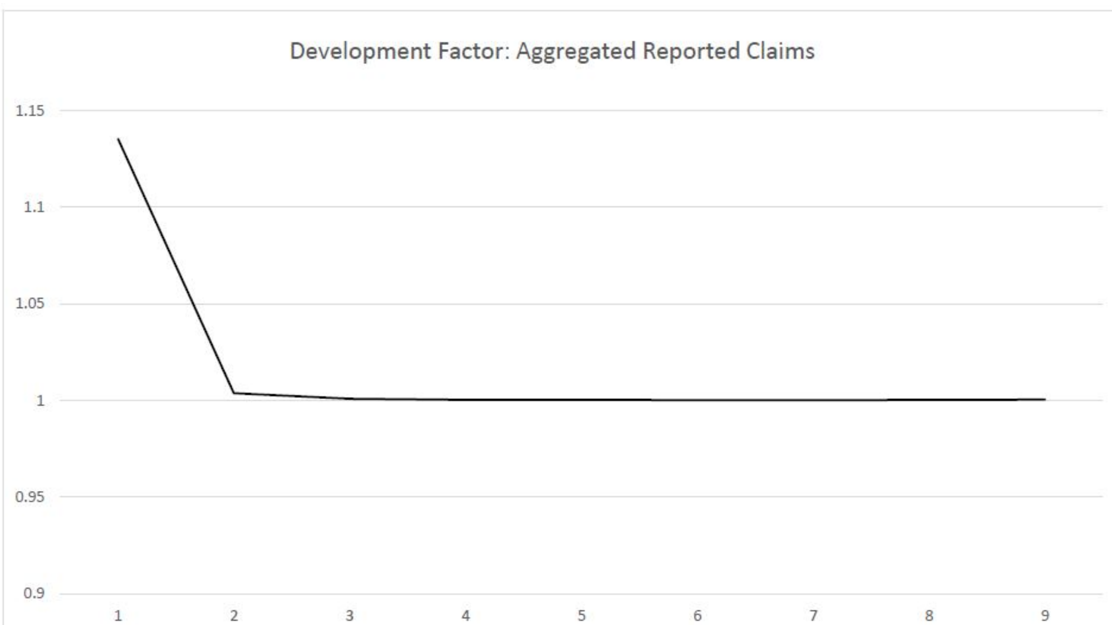

Figure 6.2 Development Factor of the Aggregated Reported Claims

BIBLIOGRAPHY

[1] B. Ajne. Exponential run-off. *Claims Reserving Manual*, 2, 1989.

[2] B. Efron. Bootstrap method: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

[3] B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, 1(1):54–77, 1986.

[4] P. England and R.J. Verall. Predicitve distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science*, 1(2):221–270, 2006.

[5] P.D. England and R.J. Verall. Stochastic claims reserving in general insurance (with discussion). *British Actuarial Journal*, 8:443–544, 2002.

[6] D.J. Finney. On the distribution of a variate whose logarithm is normally distributed. *JRSS*, 7:155–161, 1941.

[7] E. Kremer. IBNR-claims and the two-way model of ANOVA. *Scandinavian Actuarial Journal*, 1:47–55, 1982.

[8] T. Mack. A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *Astin Bulletin*, 39:35–60, 1991.

[9] T. Mack. The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor. *Astin Bulletin*, 29(2):361–366, 1999.

[10] Netanya Martin. Application of loglinear models to claims triangle runoff data, 2015. Master Thesis.

[11] María Dolores Martínez Miranda, Jens Perch Nielsen, and Richard Verrall. Double chain ladder. *Astin Bulletin*, 42(1):59–76, 2012.

[12] María Dolores Martínez-Miranda, Jens Perch Nielsen, and Richard Verrall. Double chain ladder and Bornhuetter-Ferguson. *North American Actuarial Journal*, 17(2):101–113, 2013.

[13] María Dolores Martínez Miranda, Jens Perch Nielsen, Richard Verrall, and Mario V. Wüthrich. Double chain ladder, claims development inflation and zero-claims. *Scandinavian Actuarial Journal*, 2015(5):383–405, 2015.

[14] M.D. Martínez-Miranda, B. Nielsen, J.P. Nielsen, and R. Verall. Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers. *Astin Bulletin*, 41(1):107–129, 2011.

[15] M.D. Martínez-Miranda, J.P. Nielsen, and R. Verall. Double chain ladder. University Lecture, 2011.

[16] M.D. Martínez-Miranda, J.P. Nielsen, and M.V. Wüthrich. Statistical modelling and forecasting in Non-life insurance. *Statistics and Operations Research Transactions*, 36(2):195–218, 2012.

[17] A. Renshaw. Chain ladder and interactive modelling. *J.I.A*, 116:559–587, 1989.

[18] A.E. Renshaw and R.J. Verall. A stochastic model underlying the chain-ladder technique. *B.A.J.*, 4(4):903–923, 1998.

[19] Klaus D. Schmidt. Bornhuetter-Ferguson as a general principle of loss reserving. Lehrstuhl für Versicherungsmathematik, Technische Universität Dresden.

[20] Kesar Singh and Minge Xie. Bootstrap: A statistical method. Rutgers University.

[21] G.C. Taylor and F.R. Ashe. Second moments of estimates of outstanding claims. *J. of Econometrics*, 23:37–61, 1993.

[22] R.J. Verall. A state space representation of the chain ladder linear model. *Journal of the Institute of Actuaries*, 116:589–610, 1989.

[23] R.J. Verall. Bayes and empirical bayes estimation for the chain ladder model. *Astin Bulletin*, 20(2):217–243, 1990.

[24] R.J. Verall. On the unbiased estimation of reserves from loglinear models. *Insurance: Methematics and Economics*, 10:75–80, 1991a.

[25] R.J. Verall. An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics*, 26(1):91–99, 2000.

[26] R.J. Verall, J.P. Nielsen, and A. Jessen. Including count data in claims reserving. *Astin Bulletin*, 40(2):871–887, 2010.

[27] Richard J Verrall. Statistical methods for the chain ladder technique. In *Casualty Actuarial Society Forum*, volume 1, pages 393–446, 1994.

[28] Richard J. Verrall, Jens Perch Nielsen, and Anders Hedegaard Jessen. Prediction of RBNS and IBNR claims using claim amounts and claim counts. *Astin Bulletin*, 40(2):871–887, 2013.

VITA

Larissa Schoepf was born in Heilbron, Germany, on August 21, 1991. In August 2014, she graduated from Ulm University with a B.S. in Mathematics and Economics. She interned at Allianz Life Insurance during the summer of 2014. She started her Master in Ulm University in 2014 and completed her first two semester there. In Fall 2015, she enrolled at Missouri S&T as a Master of Applied Mathematics student with a Statistics emphasis. She graduated from University Missouri of Science and Technology with a Masters in Applied Mathematics with an Emphasis in Statistics in May of 2016. After that, she expects to graduate from Ulm University in 2016 with a M.S. in Mathematics and Economics with an Emphasis in Actuarial Science.