Stochastic chain-ladder models in nonlife insurance

Ingrid Langfeldt Gould

Dissertation for the degree master of statistics

in insurance mathematics and finance

at the University of Bergen

November 2008

Acknowledgements

I would like to thank my supervisor Trygve Nilsen for finding an interesting and useful theme for my master's thesis. I would also like to thank Michael Rosendahl at TrygVesta for his help, and for supplying me with data for this thesis. Finally, I would like to thank my husband for always being the optimist he is, and for his constant encouragement to do my best.

Abstract

This thesis examines the stochastic models which reproduce chain-ladder estimates used in reserve estimation for nonlife insurance. The chain-ladder method provides no information regarding the variability of the outcome, thereby adding uncertainty to future claim estimations. Prediction errors can be found using a variety of stochastic chain-ladder models, but the different models are based on different assumptions. The relationship between some of these models was explored, and it was demonstrated how the models are defined for a run-off triangle of insurance claims. Two of these models, *Mack's model* and the *normal approximation to the negative binomial model*, were applied to a data set consisting of auto liability insurance claims. This was done in order to find the prediction error of their chain ladder estimates, as well as verify their ability to handle negative values. The two models used in the analysis were found to produce nearly identical prediction errors, and both were able to handle negative insurance claims, which were present in the data set. A number of similarities were found between the models, to the degree that the normal approximation to the negative binomial model should be considered as underlying Mack's model. However, since it is based on a generalized linear model, the normal approximation to the negative binomial model offers greater flexibility in applied calculations than Mack's model.

Keywords: Chain-ladder method, prediction error, run-off triangle, negative insurance claims

Table of contents

1. Introduction

1.1 Background

An insurance company has a portfolio of customers. Some of them will never make a claim, while others might make one or multiple claims. The insurer makes reserves to be able to cover these claims. In casualty insurance, the policy period is usually one year. After this year, the policy could either be renewed or terminated. If the policy is cancelled, this does not necessarily mean that the insurer's liability has ended, however. Since the insurance company has agreed to a defined policy period, all claims incurred within this period (and the policy conditions) are the insurer's responsibility. Among these are claims that have been reported but have not been settled (IBNS) and claims that have incurred but have not been reported (IBNR).

A claim adjuster at an insurance company should be able to determine approximately how much to set aside for IBNS-claims. However, IBNR-claims are far more difficult to assess. In some cases even the customer might not know that he or she has a claim to make. This could for example occur in cases of traumatic injuries such as whiplash, where the customer does not become aware of the severity of the injury until several weeks after the initial trauma. Another type of IBNR-claim could be water damage to a home, where the leak was not discovered before much later.

A common method used to estimate IBNR-claims is the *chain-ladder method*. This is based on an algorithm which makes a point estimate of future claims. The chain-ladder method is simple and logical, and is widely used in casualty insurance. Despite its popularity, there are weaknesses inherent to this method. Most importantly, it does not provide information regarding the variability of the outcome. With the processing power of today's computers, the simplicity of the method is no longer a valid argument. All the same, the chain-ladder method is frequently used by actuaries.

Improvements to the chain-ladder method have been made through the development of stochastic models which support the chain-ladder technique (England & Verral 2002;Hess

& Schmidt 2002;Mack 1994a;Mack 1994b;Neuhaus 2006;Renshaw 1998). Prediction errors can be obtained when a stochastic model is used, allowing greater knowledge of the reserve estimate.

1.2 Aims and outline

The main objective of this thesis is to demonstrate methods used to determine the variability of the outcome (prediction error) in a chain-ladder calculation. This will be achieved by describing the chain-ladder algorithm, reviewing the most important stochastic chain-ladder models, examining the connection between the stochastic models, fitting the models to run-off triangle of insurance claims, and applying two of these to a data set consisting of automobile insurance claims. The model assumptions in the two models will also be tested. The results from the analysis will be used to discuss the two stochastic models and the chain-ladder method.

1.3 Definitions, notation and limitations

A *stochastic chain-ladder model* is defined as a stochastic model that produces the same estimates of future claims as the chain-ladder method.

The chain-ladder method will be introduced using lower case letters. In this case, the chain-ladder method is considered a deterministic method where the variables are known. The stochastic chain-ladder models will generally use capital letters when the variables are to be considered as stochastic variables, and the known variables are written by using lower case letters. Estimators will generally be written with capital letters, and will be denoted with the hat operator.

There are numerous stochastic models that can be used to support the chain-ladder method. Only models that produce estimates equivalent to the chain-ladder method are included in this thesis. These are the multiplicative model, the Poisson model, the Negative

Binomial Model and Mack's model. Since the data set contained negative claims only two models will be used in the analysis. Only two of these models will be used in the analysis as a result of negative claims.

1.4 The chain-ladder algorithm

Incremental claims are defined by c_{ij} where *i* denotes the accident year and *j* the development year. Let d_{ij} denote the cumulative claims. The accident year is the year the accident occurs and the development year represents the reporting delay from when the claim occured. The cumulative claim d_{ij} is

$$d_{ij} = \sum_{k=1}^{j} c_{ik} \tag{1.1}$$

Observed claims can be illustrated as a run-off triangle, as illustrated in figure 1.

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & \\ c_{31} & c_{32} & & \\ c_{41} & & & \end{bmatrix} \qquad \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & \\ d_{31} & d_{32} & & \\ d_{41} & & & \end{bmatrix}$$

Figure 1: Two run-off triangles, where the left triangle displays the observed incremental claims, and the right triangle displays the observed cumulative claims. The rows display the accident year (i) and the columns display the development year (j), when n = 4. The claims in the north-western triangle are known values; the chain-ladder algorithm seeks to estimate future claims in the south-eastern (empty) triangle.

The individual development factor can be defined as

$$f_{ij} = \frac{d_{ij}}{d_{i,j-1}} \qquad for \quad 2 \le j \le n-i+1. \tag{1.2}$$

The observed values of f_{ij} can now be seen in figure 2 such as the ones shown for incremental and cumulative claims in figure 1. The unknown values for f_{ij} will leave empty spaces in the south-eastern triangle. Figure 1 has the dimensions 4x4, which will create a triangle of f_{ij} with the dimensions 3x3. It should be noted that because of the definition in (1.2) the first column in the run-off triangle of f_{ij} has column index 2.

$$\begin{bmatrix} f_{12} & f_{13} & f_{14} \\ f_{22} & f_{23} & \\ f_{32} & & \\ & & \end{bmatrix}$$

Figure 2: A run-off triangle of development factors f_{ij} which corresponds to a run-off triangle of claims with the dimensions 4x4.

The ultimate claim is for accident year $i \ge 2$ defined as

$$d_{in} = d_{i,n-i+1} \prod_{j=n-i+2}^{n} f_{ij} \, . \tag{1.3}$$

The individual development factors f_{ij} are not observable for $j \ge n-i+2$. They represent the south-eastern corner of figure 2. To be able to find the ultimate claim d_{in} the non-observable individual factors need to be estimated. An obvious approach would be to use the average of the observed development factors in development year j. This will produce identical individual development factors within development year j for the accident years in the south-eastern run-off triangle. However, the development factor used in the chain-ladder algorithm is not a simple average of the individual development factors. It is rather a weighted mean of the observed individual development factors $\hat{f}_j = \sum_{i=1}^{n-j+1} w_{ij} f_{ij}$, where w_{ij} denote the weights. Furthermore, this development factor is only a function of the development year j, and is therefore identical to development year j, for all accident years in the south-eastern run-off triangle. The hat operator is used since $\hat{f}_j$ is considered an estimator of the individual development factors. By choosing the appropriate weighting, it becomes clear that the development factor in the chain-ladder method is a weighted mean of the individual development factors. The *chain-ladder development factor* is

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}} = \sum_{i=1}^{n-j+1} \frac{d_{i,j-1}}{\sum_{h=1}^{n-j+1} d_{h,j-1}} \frac{d_{ij}}{d_{i,j-1}} = \sum_{i=1}^{n-j+1} \frac{d_{i,j-1}}{\sum_{h=1}^{n-j+1} d_{h,j-1}} f_{ij} = \sum_{i=1}^{n-j+1} w_{ij} f_{ij} \tag{1.4}$$

The individual development factor f_{ij} is weighted by the proportion of the claims in accident year i, in development year j-1. The grounds for using a weighted average will be discussed later. Since the chain-ladder development factor is central to the models described in this thesis, it is repeated:

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}}. \qquad \textit{for } j = 2,\ldots.n \tag{1.5}$$

The ultimate claim is the cumulative claim in the final development year. This is seen in the last column of the run-off triangle for cumulative claims (d_{ij}). The ultimate claims can now be calculated in the next simple step:

$$\hat{d}_{in} = d_{i,n-i+1} \prod_{j=n-i+2}^{n} \hat{f}_j \qquad \textit{for } i = 2,\ldots..,n \qquad (1.6)$$

Equations (1.5) and (1.6) form the basis of the chain-ladder technique. The last observed claim $d_{i,n-i+1}$ is used as a basis for all future estimations for accident year *i*. Implicitly, the previously observed claims that accident year are assumed to add no further information for the purpose of estimating future claims.

1.5 Use of stochastic models in the chain-ladder method

The primary weakness in the chain-ladder method is that it is a deterministic algorithm, which implies that nothing is known about the variability of the actual outcome. To amend this shortcoming, stochastic models have been developed which provide the same estimates as in the chain-ladder method. These models make it possible to find the variability of the estimate. A stochastic model can also be used to assess whether the chain-ladder method is suitable for a given data set. However, it is important to scrutinize the specific stochastic model chosen for the analysis, since each model is based on a number of assumptions (Verral 2000).

1.6 Formulating a stochastic model based on the chain-ladder method

Since the chain-ladder method is a deterministic method, a very simple stochastic model that is derived through the chain-ladder method is presented. Assume that claims D_{ij}, for $i = 1,\ldots,n$ and $j = 1,\ldots,n$ are stochastic variables, and are therefore written with the capital letter D_{ij}. The north-western triangle in figure 1 is a realization of the stochastic variables D_{ij}. f_j is considered as an unknown parameter. A linear relationship between the development years is assumed. For $2 \le j \le n$ the linear relationship is:

$$D_{ij} = D_{i,j-1} f_j \tag{1.7}$$

By calculating expectation on both sides of equation (1.7) the expression becomes

$$E(D_{ij}) = E(D_{i,j-1}) f_j \tag{1.8}$$

When predicting the ultimate claim (or just a claim several development years ahead) a formula corresponding to the chain-ladder method can be used:

$$E(D_{in}) = E(D_{i,n-i+1}) \prod_{j=n-i+2}^{n} f_j \tag{1.9}$$

In equation (1.9) the expectation of a previous claim $E\left(D_{i,n-i+1}\right)$ can be used to predict the future. The chain-ladder method, however, uses the last observed claim $d_{i,n-i+1}$ and not the expectation of it. The chain-ladder model assumes that the latest observation is more relevant than the expectation of it, and a stochastic model equivalent to the chain-ladder method can be derived by conditioning on the latest observed claim. Let $d_{i,j-1}$ be the last observable claim. If it is conditioned on $d_{i,j-1}$ in (1.8) the expression is:

$$E(D_{ij} \mid d_{i,j-1}) = d_{i,j-1} f_j \tag{1.10}$$

There have not yet been made any assumptions about the distribution of D_{ij} .The model presented in formula (1.10) is a simple stochastic model of the chain-ladder algorithm (Mack 1994b).

Introducing the stochastic variables D_{ij} some more notational points are now to be made. These will be used later when introducing the stochastic models. The run-off triangle for D_{ij} can be displayed with the stochastic variables D_{ij} for $1 \le i, j \le n$. It is not actually a triangle, since the empty places in the south-eastern triangle are also present.

$$\begin{bmatrix} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ D_{31} & D_{32} & D_{33} & D_{34} \\ D_{41} & D_{42} & D_{43} & D_{44} \end{bmatrix}$$

Figure 3: Run-off triangle of the cumulative claims as stochastic variables, when n = 4. The rows display the accident year (i) and the columns display the development year (j).

To make it easier to find the conditional expectations, the variables K_{ij} and K_j and *K* are introduced. Let *k* be the realization of the stochastic variable *K,* and $k = \{d_{ij}, i = 1,...,n, j = 1,...,n-i+1\}$. *K* is the information of the cumulative claims in the north-western corner of the run-off triangle. Let k_{ij} be the realization of the stochastic variable K_{ij} , and $k_{ij} = \{d_{i1},...,d_{ij}\}$ for accident year $i = 1,...,n$. Let k_j be the realization of the stochastic variable K_j and $k_j = \{d_{i1},...,d_{ij}, i = 1,...,n\}$.

1.7 Stochastic chain-ladder models

When finding a stochastic model that reproduces chain-ladder estimates, some assumptions must be made about the insurance claims. It is possible either to specify the distribution of the insurance claims, or merely state the two first moments (Verrall & England 2002).

The Poisson distribution may be appropriate when events are to be counted during an interval. During an insurance period accidents occur and claims are made. A number of authors propose a Poisson model in this situation (Hess & Schmidt 2002;Renshaw 1998;Verral 2000). Other distributions are closely linked to the Poisson distribution, and will therefore also be examined. These distributions are the negative binomial distribution, the multiplicative distribution and Mack's model (Verrall & England 2002). In contrast to the Poisson and negative binomial model, the multiplicative model and Mack's model only specify the first two moments.

2. Stochastic models

An underlying property to a claim is the amount of a claim. The number of claims is also relevant. An introduction to the *claim number and claim amount process* introduces this chapter. A few formulas relevant to these processes are presented and will also be used later in this thesis.

When presenting the stochastic models, the aim is to show that they indeed provide the same estimates as the chain-ladder method. There is also a close connection between the models, which will be demonstrated. The multiplicative model is presented first. Only the first moment, which has a multiplicative structure, is specified in the model. Some of the models to be presented later can be viewed as special cases of the multiplicative model. Also an alternative way of expressing the chain-ladder development factor arises from the multiplicative model and will be reviewed.

The Poisson model is a special case of the multiplicative model. It has the same multiplicative structure in the first moment. Using the maximum likelihood estimator creates the same development factor as the chain-ladder development factor and this will be proven. The relationship between the Poisson and the negative binomial model will be demonstrated using the notation for insurance claims.

Mack's model is the last model to be presented. Mack's assumptions state that the first moment is equivalent to the chain-ladder estimate, so the connection between the stochastic model and the chain-ladder method is trivial. In attempt to further understand Mack's model, the reasons behind the assumptions are explored.

2.1 Claim number and claim amount process

The incremental claim C_{ij} or the cumulative claim D_{ij} have not yet been specified any further. It may represent the number of claims an insurance company has received or can be the total amount tused to settle the insurance claims.

The total amount of claims is clearly also a function of the number of claims, which introduces the compound Poisson distribution. Let $N(t)$ be the number of claims which is a Poisson distributed variable, and it is a function of the continuous time t. $N(t)$ counts the number of claims in the interval $(0,t]$. $N(t)$ increases in steps, and is a non-decreasing function of time t. Let Y_k be the amount of claim number k. The total amount $X(t)$ of the $N(t)$ claims up to time t is

$$X(t) = \sum_{k=1}^{N(t)} Y_k \tag{2.1}$$

If Y_k is independent and identically distributed, then $X(t)$ follows a compound Poisson distribution. The expectation and variance can be found through calculations of *double expectation*:

$$E_X\left(X(t)\right) = E_N E_X\left(X(t)\middle| N(t)\right) = E_N\left(N(t)E_Y(Y_k)\right) = E_N\left(N(t)\right)E_Y(Y_k) \tag{2.2}$$

and *double variance*:

$$\begin{aligned} Var_X\left(X(t)\right) &= E_N\left(Var_X\left(X(t)\middle| N(t)\right)\right) + Var_N\left(E_X\left(X(t)\middle| N(t)\right)\right) \\ &= E_N\left(N(t)Var_Y(Y_k)\right) + Var_N\left(N(t)E_Y\left(Y_k\right)\right) \\ &= E_N\left(N(t)\right)Var_Y(Y_k) + \left(E_Y\left(Y_k\right)\right)^2 Var_N\left(N(t)\right) \end{aligned} \tag{2.3}$$

As $N(t)$ and $X(t)$ are function of the continuous time t, C_{ij} or D_{ij} are measures of either of these sizes at a specific time. Fixing the accident year, C_{ij} and D_{ij} only change between the development years. They can either be a measure of claim number or total claim amount. For later purposes, when C_{ij} or D_{ij} represents the total amount of claims, it will only be denoted the *amount of claims* and not the total amount of claims.

2.2 The multiplicative model and the chain-ladder method

The multiplicative model can be seen as underlying both Mack's model and the Poisson model. The multiplicative model is presented below, where the connection to the chain-ladder method is clarified. In this chapter the symbols x_i and y_j will be used. These are parameters in the multiplicative model (and not realizations of $X(t)$ and Y_k which were introduced in the previous chapter).

The multiplicative model is defined by the first moment, and for $1 \le i, j \le n$ it is

$$E(C_{ij}) = x_i y_j , \qquad (2.4)$$

where C_{ij} is a stochastic variable, x_i and y_j are unknown parameters, and $y_1 + y_2 + ... + y_n = 1$.

By the definition in (2.4) and the property that the sum of y_j equals one, gives that $x_i = E(D_{in})$. Expressed in words, (2.4) says that the expectation of the incremental claim can be written as a product of an accident year dependent parameter x_i and a development year dependent parameter y_j. Since x_i is the expected ultimate claim, it is logical that the sum of y_j is one. If C_{ij} represents the number of claims, y_j is the probability that a claim incurred in accident year i, is reported in development year j. This interpretation implicitly lays another restriction on y_j, $y_j \geq 0$ for *j= 1,...n.*

A very simple stochastic model of the chain-ladder method was derived in (1.8). Mack (1994) stated that this was equivalent to the multiplicative model. This can be proven by finding appropriate candidates for x_i and y_j.

By using (1.9) the expectation of the incremental claim can be written:

$$E(C_{ij}) = E(D_{ij}) - E(D_{i,j-1}) \tag{2.5}$$

$$= (f_{j+1}f_{j+2}\cdots f_n)^{-1}E(D_{in}) - (f_j f_{j+1}\cdots f_n)^{-1}E(D_{in})$$

$$= E(D_{in})\left((f_{j+1}f_{j+2}\cdots f_n)^{-1} - (f_j f_{j+1}\cdots f_n)^{-1}\right)$$

The next step is to recognize what the variables y_j must be so that (2.5) equals $x_i y_j$. The variable x_i has already been recognized, $x_i = E(D_{in})$, and clearly $y_j = (f_{j+1}f_{j+2}.....f_n)^{-1} - (f_j f_{j+1}....f_n)^{-1}$. For development year $2 \leq j < n$, the variables y_j are:

$$y_1 = (f_2 f_3....f_n)^{-1}$$

$$y_j = (f_{j+1}f_{j+2}.....f_n)^{-1} - (f_j f_{j+1}....f_n)^{-1} \tag{2.6}$$

$$y_n = 1 - (f_n)^{-1}$$

If the newly defined variables y_j meet the constraint $\sum_{j=1}^{n} y_j = 1$, they can be accepted. Summing up the terms in (2.6), a telescoping series is revealed, and using this property it is clear that $\sum_{j=1}^{n} y_j = 1$. Additionally $y_j \geq 0$ if $f_j \geq 1$ for *j = 1,...,n*. This definition of y_j seems to be a good choice. The cumulative claim in accident year i and development year j can be written as a sum of the incremental claims, and using the constraint laid upon y_j one can see that for accident year *i = 2,…,n*:

$$E(D_{in}) = x_i(y_1 + ... + y_n) \tag{2.7}$$

$$= x_i y_1 + x_i y_2 + ... + x_i y_n$$

$$= E(C_{i1}) + ... + E(C_{in})$$

By appropriately choosing x_i and y_j, it is clear that the simple stochastic model from chapter 1.4 is equivalent to the multiplicative model (Mack 1994b).

The development factor can be derived by rewriting expression (1.8) and using the identities from the multiplicative model. For $2 \leq j \leq n$ the expression is

$$f_j = \frac{E(D_{ij})}{E(D_{i,j-1})} \tag{2.8}$$

$$= \frac{x_i \left(y_1 + y_2 + ... + y_j \right)}{x_i \left(y_1 + y_2 .... + y_{j-1} \right)}$$

$$= \frac{y_1 + y_2 + ... + y_j}{y_1 + y_2 .... + y_{j-1}}$$

This development factor does not have the same appearance as the chain-ladder development factor, but it is the same. This can be proven by induction.

2.3 The Poisson model and the chain-ladder method

The Poisson model can be viewed as a special case of the multiplicative model. It has the same basic multiplicative structure of the first moment, but in addition a Poisson distribution of the incremental claims C_{ij} is assumed. Verral (2000) claimed that the Poisson model will produce exactly the same reserve estimates as the chain-ladder method. This is true when maximum likelihood estimators (MLE) are used , which will be proven.

C_{ij} are incremental claims, and let C_{ij} be independent Poisson distributed with $E\left(C_{ij}\right) = x_i y_j$, and $\sum_{j=1}^{n} y_j = 1$. From the multiplicative model the parameter x_i was determined; $x_i = E\left(D_{in}\right)$. x_i is the expected value of cumulative claims up to the latest development year observed so far.

The first moment can be parameterized as

$$E\left(C_{ij}\right)=x_i y_j = E\left(D_{in}\right) y_j = \frac{E\left(D_{i,n-i+1}\right) y_j}{\sum_{j=1}^{n-i+1} y_j} = \frac{z_i y_j}{s_{n-i+1}} \qquad (2.9)$$

where $z_i = E\left(D_{i,n-i+1}\right)$ and $s_k = \sum_{j=1}^{k} y_j$

Since y_j can be interpreted as the proportion of the ultimate claim in development year j, it is logical that $E\left(D_{i,n-i+1}\right)$ divided by the proportion of claims until $j = n-i+1$ equals $E\left(D_{in}\right)$.

Equation (2.9) can be written so that it is a formula for predicting the expectation of the ultimate claim $E\left(D_{in}\right)$. Approximating $E\left(D_{in}\right)$ with $\hat{D}_{in}$ the equation is:

$$\hat{D}_{in} = ED_{in} = x_i = \frac{z_i}{\sum_{k=1}^{n-i+1} y_k} = \frac{z_i}{1-\sum_{k=n-i+2}^{n} y_k}\,. \qquad (2.10)$$

Verral (2000) claims this is equivalent to the chain-ladder estimator:

$$\hat{D}_{n-j+1,n} = d_{n-j+1,j}\hat{f}_{j+1}\hat{f}_{j+2}.....\hat{f}_n \qquad \text{where } \hat{f}_j = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}}\,. \qquad (2.11)$$

To see that (2.10) and (2.11) are in fact equivalent, it is natural to look for estimators of the unknown parameters in (2.10). The maximum likelihood function will be used to find estimators. In this case the observations c_{ij} are considered known, and the parameters are considered as the variables. The maximum likelihood function can be written as:

$$L=\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{(z_i y_j / s_{n-i+1})^{c_{ij}} e^{-z_i y_j / s_{n-i+1}}}{c_{ij}!}\right) \tag{2.12}$$

It is the maximum likelihood function of a Poisson distributed variable with parameter $\frac{z_i y_j}{s_{n-i+1}}$. Further calculations show that this can be written as

$$L=\prod_{i=1}^{n}\left(\frac{z_i^{d_{i,n-i+1}} e^{-z_i}}{d_{i,n-i+1}!}\left(\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!}\prod_{j=1}^{n-i+1}\left(\frac{y_j}{s_{n-i+1}}\right)^{c_{ij}}\right)\right)=L_c L_d \tag{2.13}$$

where $L_c=\prod_{i=1}^{n}\left(\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!}\prod_{j=1}^{n-i+1}\left(\frac{y_j}{s_{n-i+1}}\right)^{c_{ij}}\right)$ and $L_d=\prod_{i=1}^{n}\left(\frac{z_i^{d_{i-n-i+1}} e^{-z_i}}{d_{i,n-i}!}\right)$. (2.14)

L_c is the conditional maximum likelihood function, where C_{ij} conditioned on $d_{i,n-i+1}$ is multinomially distributed with probabilities $\frac{y_j}{s_{n-i+1}}$ (see Appendix 2). The multinomial distribution is reasonable considering the possibility of a claim/or several claims being reported in increment *(i,j)*. The multinomial distribution represents the probability of

C_{ij} claims, which incurred in accident year i, will be reported in development year j. L_d is the maximum likelihood function where $D_{i,n-i+1}$ is Poisson distributed with mean z_i, and by this expression the maximum likelihood estimator (MLE) of z_i is found. The MLE of z_i is $d_{i,n-i+1}$, since $D_{i,n-i+1}$, is Poisson distributed.

Using the MLE of z_i the estimator of the ultimate claim becomes:

$$\hat{D}_{in} = \frac{d_{i,n-i+1}}{1 - \sum_{k=n-i+2}^{n} y_k} \tag{2.15}$$

For accident year *n-j+1* this expression is

$$\hat{D}_{n-j+1,n} = \frac{d_{n-j+1,j}}{1 - \sum_{k=j+1}^{n} y_k} \tag{2.16}$$

In expression (2.16) the only unknown parameter is y_k. This can be determined by finding the MLE by using L, but L_c may just as well be used. The logarithm of L_c is found, and the resulting expression is differentiated with respect to y_k, for $k = 1,\dots,n$. This needs to be done recursively, in a procedure described by Renshaw (1998). The parameter $\hat{y}_n$ is determined first, then $\hat{y}_{n-1}$ and so on. The calculations of finding $\hat{y}_n$ and the general formula for $\hat{y}_j$ are shown below:

$$\ln\left(L_c\right) = l_c \propto \sum_{i=1}^{n}\sum_{j=1}^{n-i+1} c_{ij}\log\left(\frac{y_j}{\sum_{k=1}^{n-i+1} y_k}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n-i+1} c_{ij}\left(\log y_j - \log\left(\sum_{k=1}^{n-i+1} y_k\right)\right)$$

$$\frac{\partial l_c}{\partial y_n} = 0 \Rightarrow \frac{c_{1n}}{\hat{y}_n} - \sum_{j=1}^{n}\frac{c_{1j}}{\sum_{k=1}^{n}\hat{y}_k} = \frac{c_{1n}}{\hat{y}_n} - \sum_{j=1}^{n}\frac{c_{1j}}{1} = 0$$

$$\Rightarrow \hat{y}_n = \frac{c_{1n}}{\sum_{j=1}^{n} c_{1j}} = \frac{c_{1n}}{d_{1n}} \tag{2.17}$$

$$\frac{\partial l_c}{\partial y_j} = 0 \Rightarrow \sum_{i=1}^{n-j+1}\left(\frac{c_{ij}}{\hat{y}_j} - \sum_{j=1}^{n-j+1}\frac{c_{ij}}{\sum_{k=1}^{n-i+1}\hat{y}_k}\right) = \frac{\sum_{k=1}^{n-i+1} c_{ij}}{\hat{y}_j} - \sum_{i=1}^{n-j+1}\left(\frac{d_{i,n-i+1}}{\sum_{k=1}^{n-i+1}\hat{y}_k}\right) = 0$$

$$\Rightarrow \hat{y}_j = \frac{\sum_{i=1}^{n-j+1} c_{ij}}{\sum_{i=1}^{n-j+1}\left(\frac{d_{i,n-i+1}}{\sum_{k=1}^{n-i+1}\hat{y}_k}\right)} = \frac{c_{1j} + \ldots + c_{n-j+1,j}}{d_{1n} + \frac{d_{2,n-1}}{1-\hat{y}_n} + \ldots + \frac{d_{n-j+1,j}}{1-\hat{y}_{j+1} - .. - \hat{y}_n}} \tag{2.18}$$

A maximum likelihood estimator of y_j, for $j = 1,\ldots,n$ is expressed in (2.18). The next step is to find an expression for the development factor $\hat{f}_j$, by using the MLE $\hat{y}_j$. By rearranging the chain-ladder equation in (2.11), it becomes an expression of the product of the development factors:

$$\hat{f}_{j+1}\hat{f}_{j+2}.....\hat{f}_n = \frac{\hat{D}_{n-j+1,n}}{d_{n-j+1,j}}$$

Inserting the expression for $\hat{D}_{n-j+1,n}$ from equation (2.16) and using the estimator $\hat{y}_j$ instead of y_j, the product of the development factors becomes:

$$\hat{f}_{j+1}\hat{f}_{j+2}....\hat{f}_n = \frac{1}{1-\hat{y}_{j+1}-\hat{y}_{j+2}-....-\hat{y}_n} \tag{2.19}$$

and

$$\hat{f}_j\hat{f}_{j+1}....\hat{f}_n = \frac{1}{1-\hat{y}_j-\hat{y}_{j+1}-....-\hat{y}_n} \tag{2.20}$$

By rearranging (2.19) an expression for $1-\hat{y}_{j+1}-\hat{y}_{j+2}-....-\hat{y}_n$ is derived, and this can be inserted in (2.20). Thus

$$\hat{f}_j\hat{f}_{j+1}....\hat{f}_n = \frac{1}{\frac{1}{\hat{f}_{j+1}\hat{f}_{j+2}....\hat{f}_n}-\hat{y}_j} \tag{2.21}$$

Finally an estimator of the development factor $\hat{f}_j$ is found

$$\hat{f}_j = \frac{1}{1-\hat{y}_j\hat{f}_{j+1}\hat{f}_{j+2}....\hat{f}_n} \tag{2.22}$$

Using the MLE of y_n from (2.17) the expression becomes

$$\hat{f}_n = \frac{1}{1-\hat{y}_n} = \frac{1}{1-\frac{c_{1n}}{d_{1n}}} = \frac{d_{1n}}{d_{1n}-c_{1n}} = \frac{d_{1n}}{d_{1,n-1}}. \tag{2.23}$$

The estimator obtained in (2.23) is the same as the chain-ladder estimator for $j = n$. To show that the rest of development factors in the Poisson model are the same as the chain-ladder development factors, induction can be used. Since it has been proven for $j = n$, the first part of the induction is completed. The next step is to find the general formula for $\hat{f}_j$. To do this the expression for $\hat{y}_j$ needs some simplification. Equation (2.18) gives an expression for $\hat{y}_j$ and the fractions in the denominator can be rewritten by using (2.19), (2.20) and equivalent. Thus:

$$\hat{y}_j = \frac{c_{1j} + c_{2j} + \ldots + c_{n-j+1,j}}{d_{1n} + d_{1,n-1}\hat{f}_n \ldots + d_{n-j+1,j}\hat{f}_{j+1}\hat{f}_{j+2}\ldots\hat{f}_n} \tag{2.24}$$

By examining the expression for $\hat{y}_j$ one can also see that it is the proportion of the ultimate claim. The numerator counts incremental claims over all observed accident years for development year j, and the denominator counts the estimated ultimate claims over same accident years. Equation (2.22) is a general expression for $\hat{f}_j$. The newly derived expression for $\hat{y}_j$ is inserted in (2.22). Thus

$$\hat{f}_j = \frac{1}{1-\frac{c_{1j} + c_{2j} + \ldots. + c_{n-j+1,j}}{d_{1n} + d_{2,n-1}\hat{f}_n + \ldots. + d_{n-j+1,j}\hat{f}_{j+1}\hat{f}_{j+2}\ldots\hat{f}_n}\hat{f}_{j+1}\hat{f}_{j+2}\ldots\hat{f}_n} \tag{2.25}$$

This is the general formula. It has already been proven that the estimator for $\hat{f}_n$ is the chain-ladder development factor.

As part of the induction it is assumed that for *k = j+1, …,n,* $\hat{f}_k$ equals the chain-ladder development factor. The last step is to prove that $\hat{f}_k$ equals the chain-ladder development factor for *k = j.*

The denominator in (2.25) needs to be simplified, which can be done by showing that

$$d_{1n} + d_{2,n-1}\hat{f}_n + ... + d_{n-j+1,j}\hat{f}_{j+1}\hat{f}_{j+2}...\hat{f}_n = \hat{f}_{j+1}\hat{f}_{j+2}...\hat{f}_n \sum_{i=1}^{n-j+1} d_{ij} \qquad (2.26)$$

This is true for *j = n-1*

$$d_{1n} + d_{2,n-1}\hat{f}_n = d_{1n} + d_{2,n-1}\frac{d_{1,n}}{d_{1,n-1}} = \frac{d_{1,n}}{d_{1,n-1}}\left(d_{1,n-1} + d_{2,n-1}\right) = \hat{f}_n\left(d_{1,n-1} + d_{2,n-1}\right)$$

Similarly for *j = n-2* the same relationship exist

$$d_{1n} + d_{2,n-1}\hat{f}_n + d_{3,n-2}\hat{f}_{n-1}\hat{f}_n = \hat{f}_n\left(d_{1,n-1} + d_{2,n-1} + d_{3,n-2}\hat{f}_{n-1}\right)$$

$$= \hat{f}_n\left(d_{1,n-1} + d_{2,n-1}\right)\frac{d_{1,n-1} + d_{2,n-1}}{d_{1,n-2} + d_{2,n-2}} + d_{3,n-2}\hat{f}_{n-1}$$

$$= \hat{f}_{n-1}\hat{f}_n\left(d_{1,n-2} + d_{2,n-2} + d_{3,n-2}\right)$$

By performing this *n-j* times (2.26) is proven, and the equation for $\hat{f}_j$ in (2.25) can be reduced to

$$\hat{f}_j = \frac{1}{1 - \dfrac{c_{1j} + c_{2j} + ... + c_{n-j+1,j}}{\hat{f}_{j+1}\hat{f}_{j+2}...\hat{f}_n \sum_{i=1}^{n-j+1} d_{ij}} \hat{f}_{j+1}\hat{f}_{j+2}...\hat{f}_n}$$

$$= \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{ij} - \sum_{i=1}^{n-j+1} c_{ij}} = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \tag{2.27}$$

The induction proof is fulfilled since $\hat{f}_j$ equals the chain-ladder development factor. It has been proved that using MLE in a Poisson model will produce exactly the same estimates as the chain-ladder method.

2.4 The Poisson model and its relation to the negative binomial model

The previous chapter started by considering C_{ij} as a Poisson random variable. This is also the case here, but in this case the intensity of the Poisson distribution will also be stochastic. Through the following definitions Verral (2000) made a recursive model that connected the Poisson model and the Negative Binomial model.

C_{ij} conditioned on $Z_{ij} = z_{ij}$ is Poisson distributed with mean $\dfrac{z_{ij} y_j}{s_j}$ where $z_{ij} = E\left(D_{ij}\right)$ and $s_j = \sum_{k=1}^{j} y_k$.

The variable Z_{ij} is denoted with the index j (in addition to index i) since this is a conditional model, where $z_{ij} = E\left(D_{ij}\right)$ changes with development year j. Before any assumptions are made about Z_{ij}, the relationship between Z_{ij} and $Z_{i,j-1}$ will be established:

$$Z_{ij} = E\left(D_{ij}\right) = E\left(D_{i,j-1}\right) + E\left(C_{ij}\right) = Z_{i,j-1} + \frac{Z_{ij} y_j}{s_j}$$

$$\Rightarrow Z_{ij} = \frac{Z_{i,j-1} s_j}{s_{j-1}} \tag{2.28}$$

Given this relationship the distribution of C_{ij} is:

C_{ij} conditioned on $z_{i,j-1}$ is Poisson distributed with mean $\dfrac{z_{i,j-1} y_j}{s_{j-1}}$

The parameter y_j can still be considered as the column parameter, and is the probability of a claim to be reported in development year j. The factor $\dfrac{z_{i,j-1}}{s_{j-1}}$ gives the expected ultimate claim.

The aim is to see that C_{ij} conditioned on the earlier observed claims $c_{i1},...,c_{i,j-1}$ is negative binomially distributed. In order to do this, it is necessary to make some assumptions about

$Z_{i,j-1}$. It is assumed that the distribution of $Z_{i,j-1}$ is known, so this model takes a Bayesian approach. In development year j, there are observations of claims up to development year *j-1*.

It is assumed that:

$Z_{i,j-1}$ conditioned on $c_{i1},...,c_{i,j-1}$ is gamma distributed with parameters α and β

By using standard Bayesian analysis one can find the distribution of $Z_{i,j-1}$ conditioned on $c_{i1},...,c_{i,j}$. In this case, the prior distribution $\pi_{Z_{i,j-1}|C_{i1},...,C_{i,j-1}}\left(z_{i,j-1}\left|c_{i1},...,c_{i,j-1}\right.\right)$ is the gamma distribution, the conditional distribution $f_{C_{ij}|Z_{i,j-1}}\left(c_{ij}\left|z_{i,j-1}\right.\right)$ is the Poisson distribution. The Bayesian formula is used to solve this problem is

$$\pi_{Z_{i,j-1}|C_{i1},...,C_{ij}}\left(z_{i,j-1}\left|c_{i1},...,c_{ij}\right.\right)=\frac{f_{C_{ij}|Z_{i,j-1}}\left(c_{ij}\left|z_{i,j-1}\right.\right)\pi_{Z_{i,j-1}|C_{i1},...,C_{i,j-1}}\left(z_{i,j-1}\left|c_{i1},...,c_{i,j-1}\right.\right)}{\int_0^\infty f_{C_{ij}|Z_{i,j-1}}\left(c_{ij}\left|z_{i,j-1}\right.\right)\pi_{Z_{i,j-1}|C_{i1},...,C_{i,j-1}}\left(z_{i,j-1}\left|c_{i1},...,c_{i,j-1}\right.\right)dz_{i,j-1}}$$

$$=\frac{\left(\left(z_{i,j-1}y_j/s_{j-1}\right)^{c_{ij}}/c_{ij}!\right)e^{z_{i,j-1}y_j/s_{j-1}}\frac{1}{\Gamma(\alpha)\beta^\alpha}\left(z_{i,j-1}\right)^{\alpha-1}\mathrm{e}^{z_{i,j-1}/\beta}}{\int_0^\infty\left(\left(z_{i,j-1}y_j/s_{j-1}\right)^{c_{ij}}/c_{ij}!\right)e^{z_{i,j-1}y_j/s_{j-1}}\frac{1}{\Gamma(\alpha)\beta^\alpha}\left(z_{i,j-1}\right)^{\alpha-1}\mathrm{e}^{z_{i,j-1}/\beta}dz_{i,j-1}}$$

$$=\frac{1}{\Gamma\left(\alpha+c_{ij}\right)}\left(\frac{y_j}{s_{j-1}}+\beta\right)^{\alpha+c_{ij}}\left(z_{i,j-1}\right)^{\alpha+c_{ij}-1}e^{-z_{i,j-1}\left(\frac{y_j}{s_{j-1}}+\beta\right)}$$

$$=\Gamma\left(\alpha+c_{ij},\frac{y_j}{s_{j-1}}+\beta\right) \qquad (2.29)$$

The distribution of $Z_{i,j-1}$ conditioned on $c_{i1},...,c_{ij}$ is found, and Verral (2000) proceeds by finding the distribution of $Z_{i,j}$ conditioned on $c_{i1},...,c_{ij}$. The relationship between $Z_{i,j-1}$ and Z_j is given by (2.28). By the simple transformation used below one can find

$$\pi_{Z_{ij}|C_{i1},...,C_{ij}}\left(z_{ij}\left|c_{i1},...,c_{ij}\right.\right)=\pi_{Z_{i,j-1}|C_{i1},...,C_{ij}}\left(z_{i,j-1}=\frac{s_{j-1}}{s_j}z_j\left|c_{i1},...,c_{ij}\right.\right)\frac{dz_{i,j-1}}{dz_{ij}}$$

$$=\frac{1}{\Gamma\left(\alpha+c_{ij}\right)}\left(\frac{s_{j-1}}{s_j}\left(\frac{y_j}{s_{j-1}}+\beta\right)\right)^{\alpha+c_{ij}}\left(z_{ij}\right)^{\alpha+c_{ij}-1}e^{-z_{ij}\left(\frac{y_j}{s_{j-1}}+\beta\right)}$$

$$=\Gamma\left(\alpha+c_{ij},\frac{s_{j-1}}{s_j}\left(\frac{y_j}{s_{j-1}}+\beta\right)\right) \tag{2.30}$$

The calculations above yield the distribution of Z_{ij} conditioned on $c_{i1},...,c_{ij}$ and next it is interesting to find the distribution for every j, where *j = 1,…,n*. It is natural to start by finding the distribution for *j =1*. To do this it is necessary to assume a prior distribution of Z_{i1}. Verral (2000) assumes that $\pi_{Z_{i1}}\left(z_{i1}\right)\propto\left(z_{i1}\right)^{-1}$. As in (2.29) the Bayesian formula can be used to find the distribution of z_{i1} conditioned on c_{i1}:

$$\pi_{Z_{i1}|C_{i1}}\left(z_{i1}\left|c_{i1}\right.\right)\propto\frac{\pi_{Z_{i1}}\left(z_{i1}\right)f_{C_{i1}|Z_{i1}}\left(c_{i1}\left|z_{i1}\right.\right)}{\int_0^\infty\pi_{Z_{i1}}\left(z_{i1}\right)f_{C_{i1}|Z_{i1}}\left(c_{i1}\left|z_{i1}\right.\right)dz_{i1}}=\frac{\frac{1}{z_{i1}}\frac{z_{i1}e^{-z_{i1}}}{c_{i1}!}}{\int_0^\infty\frac{1}{z_{i1}}\frac{z_{i1}e^{-z_{i1}}}{c_{i1}!}dz_{i1}}=\Gamma\left(c_{i1},1\right) \tag{2.31}$$

The general formula was found in (2.30), and the specific formula is now found for the first case, (2.31). This distribution is the case when j = 2, and $\alpha = c_{i1}$ and β =1 in (2.30). Since (2.30) will produce the distribution for j = 3, when α and β is known, it is only necessary to insert these values so one can see that

$$\pi_{Z_{i2}|C_{i1},C_{i2}}\left(z_{i2}\left|c_{i1},c_{i2}\right.\right) = \Gamma\left(c_{i1}+c_{i2},\left(1+\frac{y_j}{s_{j-1}}\right)\frac{s_{j-1}}{s_j}\right) = \Gamma\left(d_{i2},1\right) \tag{2.32}$$

To prove this for all j, induction can be used. The formula is assumed for *k = j-1*, that is $\pi_{Z_{i,j-1}|C_{i1},\dots,C_{i,j-1}}\left(z_{i,j-1}\left|c_{i1},\dots,c_{i,j-1}\right.\right) = \Gamma\left(d_{i,j-1},1\right)$. As done above, formula (2.30) can be used to prove it when *k = j*.

$$\pi_{Z_{ij}|C_{i1},\dots,C_{ij}}\left(z_{ij}\left|c_{i1},\dots,c_{ij}\right.\right) = \Gamma\left(d_{i,j-1}+c_{ij},\left(1+\frac{y_j}{s_{j-1}}\right)\frac{s_{j-1}}{s_j}\right) = \Gamma\left(d_{ij},1\right) \tag{2.33}$$

The run-off triangle only have known values in the north-western corner, and to predict the rest of the values of C_{ij}, it is desirable to find the distribution of C_{ij} conditioned on $c_{i1},\dots,c_{i,j-1}$. This can be found by this calculation:

$$f_{C_{ij}|C_{i1},\dots,C_{i,j-1}}\left(c_{ij}\left|c_{i1},\dots,c_{i,j-1}\right.\right) = \int f_{C_{ij}|Z_{i,j-1}}\left(c_{ij}\left|z_{i,j-1}\right.\right) f_{Z_{i,j-1}|C_{i1},\dots,C_{i,j-1}}\left(z_{i,j-1}\left|c_{i1},\dots,c_{i,j-1}\right.\right) dz_{i,j-1}$$

$$=\int_0^\infty \frac{\left(z_{i,j-1}y_j / s_{j-1}\right)^{c_{ij}} \mathrm{e}^{-\left(z_{i,j-1}y_j/s_{j-1}\right)}}{c_{ij}!}\frac{1}{\Gamma\left(d_{i,j-1}\right)}\left(z_{i,j-1}\right)^{d_{i,j-1}-1}\mathrm{e}^{-z_{i,j-1}}dz_{i,j-1}$$

$$=\frac{\left(y_j / s_{j-1}\right)^{c_{ij}}}{c_{ij}!\Gamma\left(d_{i,j-1}\right)}\frac{\Gamma\left(d_{i,j-1}+c_{ij}\right)}{\left(y_j / s_{j-1}+1\right)^{d_{i,j-1}+c_{ij}}}*$$

$$\int_0^\infty \frac{1}{\Gamma\left(d_{i,j-1}+c_{ij}\right)}\left(y_j / s_{j-1}+1\right)^{d_{i,j-1}+c_{ij}}\left(z_{i,j-1}\right)^{d_{i,j-1}+c_{ij}-1}\mathrm{e}^{-z_{i,j-1}\left(y_j/s_{j-1}+1\right)}dz_{i,j-1}$$

$$=\frac{\Gamma\left(d_{ij}\right)}{c_{ij}!\Gamma\left(d_{i,j-1}\right)}\frac{\left(y_j / s_{j-1}\right)^{c_{ij}}}{\left(s_j / s_{j-1}\right)^{d_{i,j-1}+c_{ij}}}$$

$$=\frac{\Gamma\left(d_{ij}\right)}{c_{ij}!\Gamma\left(d_{i,j-1}\right)}\left(\frac{s_{j-1}}{s_j}\right)^{d_{i,j-1}}\left(\frac{s_{j-1}}{s_j}\right)^{c_{ij}}\left(\frac{y_j}{s_{j-1}}\right)^{c_{ij}}$$

$$=\frac{\Gamma\left(d_{ij}\right)}{c_{ij}!\Gamma\left(d_{i,j-1}\right)}\left(\frac{s_{j-1}}{s_j}\right)^{d_{i,j-1}}\left(\frac{y_j}{s_j}\right)^{c_{ij}}$$

$$=\frac{\left(c_{i,j-1}+c_{ij}-1\right)!}{c_{ij}!\left(d_{i,j-1}-1\right)!}\left(\frac{s_{j-1}}{s_j}\right)^{d_{i,j-1}}\left(1-\frac{s_{j-1}}{s_j}\right)^{c_{ij}} \qquad (2.34)$$

Thus

C_{ij} conditioned on $c_{i1},...,c_{i,j-1}$ is negative binomial with mean $\frac{d_{i,j-1}y_j}{s_{j-1}}$ and variance $\frac{d_{i,j-1}s_j}{\left(s_{j-1}\right)^2}$

The chain-ladder development factors could also be expressed as a function of the column factors y_j, where

$$f_j = \frac{\sum_{k=1}^{j} y_k}{\sum_{k=1}^{j-1} y_k} = \frac{s_j}{s_{j-1}} \tag{2.35}$$

Accepting the definition in (2.35) reveals that the distribution of C_{ij} conditioned on $c_{i1},...,c_{i,j-1}$ can be written only as a function of observed cumulative claims, and the development factors. The distribution of C_{ij} conditioned on $c_{i1},...,c_{i,j-1}$ is

$$\frac{\left(d_{i,j-1}+c_{ij}-1\right)!}{c_{ij}!\left(d_{i,j-1}-1\right)!}\left(\frac{1}{f_j}\right)^{d_{i,j-1}}\left(1-\frac{1}{f_j}\right)^{c_{ij}} \tag{2.36}$$

where the mean and variance are $\left(f_j-1\right)d_{i,j-1}$ and $f_j\left(f_j-1\right)d_{i,j-1}$.

Since $D_{ij}=D_{i,j-1}+C_{ij}$, the distribution of D_{ij} conditioned on $c_{i1},...,c_{i,j-1}$ is also negative binomially distributed, and the distribution is:

$$\frac{\left(d_{i,j-1}+c_{ij}-1\right)!}{c_{ij}!\left(d_{i,j-1}-1\right)!}\left(\frac{1}{1+f_j}\right)^{d_{i,j-1}}\left(1-\frac{1}{1+f_j}\right)^{c_{ij}} \tag{2.37}$$

and can be written as

$$\frac{(d_{ij}-1)!}{(d_{ij}-d_{i,j-1})!(d_{i,j-1}-1)!}\left(\frac{1}{1+f_j}\right)^{d_{i,j-1}}\left(1-\frac{1}{1+f_j}\right)^{d_{ij}-d_{i,j-1}}$$

where the mean and variance is $f_j d_{i,j-1}$ and $f_j\left(f_j-1\right)d_{i,j-1}$

The formulas in (2.36) and (2.37) show that it is unnecessary to condition on all the earlier incremental claims $\left(c_{i1},...,c_{i,j-1}\right)$, the distribution of D_{ij} conditioned on $d_{i,j-1}$ is identical to (2.37).

2.5 Mack's model

Mack (1994b) proposed a distribution free stochastic model which produces equivalent results to the chain-ladder algorithm. As before, C_{ij} represents incremental change between development years j, and D_{ij} represents cumulative claims that occurred in accident year *i* and that are reported within development year *j*. The variable K_{ij} was defined in chapter 1.6, and the same definition is still valid. It is assumed that the first accident year is fully developed.

Mack made three assumptions to define this model. They are as follows:

1. There exist constants $f_2,...,f_n$ such that $E\left(D_{i,j}\middle|K_{i,j-1}=k_{i,j-1}\right)=f_j d_{i,j-1}$ for $j=2,..,n$

2. There exists constants $g_2,...,g_n$ such that $Var\left(D_{i,j}\middle|K_{i,j-1}=k_{i,j-1}\right)=g_j d_{i,j-1}$ for $j=2,..,n$

3. K_{in} and K_{kn} are stochastically independent for $i\neq k$.

(Mack 1994b)

The estimator $\hat{f}_j$ in Mack's model is the same as the chain-ladder development factor. The development factor $\hat{f}_j$ is the same for all accident years within development year *j*, and because of this an assumption of independence between the accident years is made.

The parameter g_j can be estimated by:

$$\hat{g}_j = \frac{1}{n-j} \sum_{i=1}^{n-j+1} d_{i,j-1} \left(\frac{d_{i,j}}{d_{i,j-1}} - \hat{f}_j \right)^2$$

Mack's model is defined only by the three assumptions above. By looking into the identities $\left(f_j, g_j\right)$ introduced in the assumptions it is possible to get a further understanding of the model. Mack presented this model in his paper (Mack 1994a), and the following results are from this article. The chain-ladder development factor will be examined first:

The development factors $\hat{f}_j$ are unbiased estimators of f_j. Using the rule of double expectation it is clear that

$$E\left(\hat{f}_j\right) = E\left(E\left(\hat{f}_j \middle| K_{j-1}\right)\right) = E\left(E \frac{\sum_{i=1}^{n-j+1} D_{i,j}}{\sum_{i=1}^{n-j+1} D_{i,j-1}} \middle| K_{j-1} \right) \qquad (2.38)$$

$$= E\left(\frac{1}{\sum_{i=1}^{n-j+1} D_{i,j-1}} \left(\sum_{i=1}^{n-j+1} E\left(D_{i,j} \middle| K_{j-1}\right) \right) \right)$$

$$= E\left(\frac{1}{\sum_{i=1}^{n-j+1} D_{i,j-1}} \left(\sum_{i=1}^{n-j+1} E\left(D_{i,j} \middle| K_{j-1}\right) \right) \right)$$

The last calculation used assumption 3. Since there is independence between K_{in} and K_{kn}, for $i \neq k$, it is only necessary to condition on the unknown values in the relevant accident year. Using assumption 1 it is easy to see that (2.38) equals

$$E\left(\frac{1}{\sum_{i=1}^{n-j+1} D_{i,j-1}} \left(\sum_{i=1}^{n-j+1} D_{i,j-1} f_j \right) \right) = f_j \tag{2.39}$$

This proves that $\hat{f}_j$ is unbiased (Mack 1994a).

The individual development factors are uncorrelated. This can be proved by showing that

$$E\left(\frac{D_{i,k+1}}{D_{ik}} \frac{D_{ik}}{D_{i,k-1}} \right) = E\left(\frac{D_{i,k+1}}{D_{ik}} \right) E\left(\frac{D_{ik}}{D_{i,k-1}} \right) \tag{2.40}$$

For $j \leq k$ it can be seen by using the rule of double expectation that

$$E\left(\frac{D_{i,k+1}}{D_{ij}}\right)=E\left(E\left(\frac{D_{i,k+1}}{D_{ij}}\middle|D_{i1},...,D_{ik}\right)\right)=E\left(\frac{1}{D_{ij}}E\left(D_{i,k+1}\middle|D_{i1},...,D_{ik}\right)\right) \qquad (2.41)$$

$$=E\left(\frac{1}{D_{ij}}D_{ik}f_{k+1}\right)=f_{k+1}E\left(\frac{D_{ik}}{D_{ij}}\right)$$

From the second to the third step in (2.41), the conditioning makes D_{ij} known for $j \le k$. The next step uses Mack's first assumption. When $j = k$, equation (2.41) is

$$E\left(\frac{D_{i,k+1}}{D_{ik}}\right)=f_{k+1}E\left(\frac{D_{ik}}{D_{ik}}\right)=f_{k+1}$$

(2.42)

When $j = k\text{-}1$ (2.41) is

$$E\left(\frac{D_{i,k+1}}{D_{i,k}}\frac{D_{i,k}}{D_{i,k-1}}\right)=f_{k+1}E\left(\frac{D_{ik}}{D_{i,k-1}}\right)=E\left(\frac{D_{i,k+1}}{D_{ik}}\right)E\left(\frac{D_{ik}}{D_{i,k-1}}\right) \qquad (2.43)$$

The first step in (2.43) used the identity found in (2.41), and the second step used the identity found in (2.42). This proves that the individual development factors are uncorrelated. This means that if it is natural to assume a small amount of claims after a development year with a large amount of claims, the chain-ladder development factor would not be suitable to predict future claims.

The chain-ladder development factor is a weighted mean of the individual development factors. It is unbiased, and a desirable quality of an unbiased estimator is small variance. The unweighted mean of the individual development factors is also unbiased, which implies that the reason for using the chain-ladder development factor is because of a smaller variance. Mack's second assumption determines the second moment. This is now explored:

When D_{ij} are considered to be stochastic variables for *i = 1,…,n* and *j = 1,…,n-i+1* also the development factor $\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}$ for *j = 2,…,n* are stochastic variables. The individual development factor is written with capital letter F_{ij} when it is considered a stochastic variable. The chain-ladder development factor is a weighted mean of the individual development factors, and in general this can be written like:

$$\hat{f}_j = \sum_{i=1}^{n-j+1} W_{ij} F_{ij} \qquad \text{where} \quad \sum_{i=1}^{n-j+1} W_{ij} = 1 \tag{2.44}$$

The individual development factors F_{ij} are assumed to be uncorrelated and unbiased for $1 \le i \le n$ .The variance of $\hat{f}_j$ conditioned on k_{j-1} is

$$Var\left(\hat{f}_j^w \middle| k_{j-1}\right) = Var\left(\sum_{i=1}^{n-j+1} W_{ij} F_{ij} \middle| k_{j-1}\right) = \sum_{i=1}^{n-j+1} w_{ij}^2 Var\left(F_{ij} \middle| k_{j-1}\right) \tag{2.45}$$

(2.45) is minimized with respect to W_{ij} where *j = 1,…,n* where the letter i denoting the accident year could have been removed since the variance is assumed to be equal for all accident years. The minimization must be done under the constraint on W_{ij} (see (2.44)), and the method of Lagrange multipliers can be used. The Lagrangian function is defined as $L(x,\lambda) = k(x) + \lambda g(x)$, where k is the function to be minimized with respect to x, and g is the constraint, and λ is the Lagrange multiplier. The minimum of (2.45) is:

$$\frac{\partial}{\partial w_i}\left(\sum_{i=1}^{n-j+1} w_{ij}^2 Var\left(F_{ij} \left| k_{j-1}\right.\right) + \lambda \left(1 - \sum_{i=1}^{n-j+1} w_{ij}\right)\right) = 0$$

This minimum of this function is found when the weights are inversely proportional to the variance of F_{ij}:

$$w_{ij} = \frac{\lambda}{2Var\left(F_{ij} \left| k_{j-1}\right.\right)} \tag{2.46}$$

The weight should be inversely proportional to the variance if minimum variance is a goal. In other words, the variance of the individual development factors should be inversely proportional to the weights. The weight of the chain-ladder development factor is $\frac{d_{i,j-1}}{\sum_{i=1}^{n-j+1} d_{i,j-1}}$.

Thus, the variance of the individual development factor is inversely proportional to $d_{i,j-1}$. The denominator in the fraction above can be replaced by a proportionality constant. Mack's third assumption can be rewritten so that it is clear that the chain-ladder factor actually is the estimator with minimal variance:

$$Var\left(\frac{D_{ij}}{D_{i,j-1}}\middle|k_{j-1}\right)=\frac{g_j}{d_{i,j-1}} \tag{2.47}$$

(2.47) is inversely proportional to the weight $d_{i,j-1}$, and is multiplied with a proportionality constant g_j.

The parameter g_j needs to be estimated. The proposed estimator is for *j = 2,…n*

$$\hat{g}_j=\frac{1}{n-j}\sum_{i=1}^{n-j+1}d_{i,j-1}\left(\frac{d_{ij}}{d_{i,j-1}}-\hat{f}_j\right)^2$$

This estimator is unbiased, and it will be proven that $E\left(\hat{g}_j\right)=g_j$. First the identity $(n-j)E\left(\hat{g}_j\middle|k_{j-1}\right)$ will be recovered, and this property can be used to see that $E\left(\hat{g}_j\right)=E\left(E\left(\hat{g}_j\middle|k_{j-1}\right)\right)=g_j$. The well known trick of adding and subtracting a constant will be used. In this case the constant f_j will be used.

$$\begin{aligned}(n-j)E\left(\hat{g}_j\left|k_{j-1}\right.\right) &= \sum_{i=1}^{n-j+1} d_{i,j-1}E\left(\left(\frac{D_{ij}}{D_{i,j-1}}-f_j-\hat{f}_j+f_j\right)^2\left|k_{j-1}\right.\right) \qquad (2.48)\\ &= \sum_{i=1}^{n-j+1} d_{i,j-1}E\left(\left(F_{ij}-f_j\right)^2\left|k_{j-1}\right.\right)\\ &-2\sum_{i=1}^{n-j+1} d_{i,j-1}E\left(\left(F_{ij}-f_j\right)\left(\hat{f}_j+f_j\right)\left|k_{j-1}\right.\right)\\ &+\sum_{i=1}^{n-j+1} d_{i,j-1}E\left(\left(\hat{f}_j+f_j\right)^2\left|k_{j-1}\right.\right)\\ &= \sum_{i=1}^{n-j+1} d_{i,j-1}Var\left(F_{ij}\left|k_{j-1}\right.\right)\\ &-2\sum_{i=1}^{n-j+1} d_{i,j-1}Cov\left(F_{ij},\hat{f}_j\left|k_{j-1}\right.\right)+\sum_{i=1}^{n-j+1} d_{i,j-1}Var\left(\hat{f}_j\left|k_{j-1}\right.\right)\end{aligned}$$

The chain-ladder factor is a weighted mean of the individual development factors. Using Mack's third assumption it is clear that $Cov\left(F_{kj},F_{lj}\right)=0$ for $k\neq l$, and because of this it can be seen that

$$Cov\left(F_{ij},\hat{f}_j\left|k_{j-1}\right.\right)=Cov\left(F_{ij},w_{ij}F_{ij}\left|k_{j-1}\right.\right)=Cov\left(F_{ij},\frac{D_{i,j-1}}{\sum_{l=1}^{n-j+1}D_{l,j-1}}F_{ij}\left|k_{j-1}\right.\right)=\frac{d_{i,j-1}}{\sum_{l=1}^{n-j+1}d_{l,j-1}}Var\left(F_{ij}\left|k_{j-1}\right.\right)$$

By using this property and Mack's first and second assumption equation (2.48) can be written as

$$(n-j)E\left(\hat{g}_j\left|k_{j-1}\right.\right)=\left(\sum_{i=1}^{n-j+1}g_j\right)-2g_j+g_j=(n-j)g_j$$

Using the rule of double expectations, it is clear that $\hat{g}_j$ is an unbiased estimator for g_j

$$E\left(\hat{g}_j\right)=E\left(E\left(\hat{g}_j\left|k_{j-1}\right.\right)\right)=E\left(g_j\right)=g_j$$

2.6 Mack's model and its connection to the compound Poisson distribution

Neuhaus (2006) states that if the incremental claims C_{ij}, conditioned on the development up to year $j-1$, is distributed as compound Poisson variables, this will imply the same model assumptions as Mack suggested. Let

C_{ij} conditioned on k_j be compound Poisson$\left(H_j\right)$ distributed (2.49)

and

$$f_j - 1 = \int_{-\infty}^{\infty} u dH_j(u) \text{ and } g_j = \int_{-\infty}^{\infty} u^2 dH_j(u). \tag{2.50}$$

U is the intensity of the claim, and U has the distribution H_j. It will now be proven that using (2.49) and (2.50) will lead to Mack's assumptions.

The definition of the cumulative and incremental claims gives

$$E\left(D_{ij} \middle| k_{i,j-1}\right) = E\left(D_{i,j-1} \middle| k_{i,j-1}\right) + E\left(C_{ij} \middle| k_{i,j-1}\right) = d_{i,j-1} + E\left(C_{ij} \middle| k_{i,j-1}\right). \tag{2.51}$$

By Mack's model we have that $E\left(D_{ij}\left|k_{i,j-1}\right.\right)=f_j d_{i,j-1}$ and this assumption combined with the formula above gives us

$$E(C_{ij}\left|k_{i,j-1}\right.)=(f_j-1)d_{i,j-1} \tag{2.52}$$

Since C_{ij} is a compound Poisson variable we have from (2.2) that $E\left(C_{ij}\right)=E\left(N\right)E\left(U\right)$, where N is the number of claims. We may condition on $k_{i,j-1}$, since the claim number process have independent increments and the claim number process is independent of the claim amount process. Since U has distribution H_j we find that

$$\begin{aligned} E\left(C_{ij}\left|k_{i,j-1}\right.\right) &= E\left(N\left|k_{i,j-1}\right.\right)E\left(U\left|k_{i,j-1}\right.\right) \\ &= E\left(N\left|k_{i,j-1}\right.\right)\int_{-\infty}^{\infty} UdH_j(U) \\ &= d_{i,j-1}(f_j-1) \end{aligned} \tag{2.53}$$

The last calculation is obtained since $\int_{-\infty}^{\infty} UdH_j(U)=f_j-1$ and by letting the Poisson parameter be proportional or equal to $d_{i,j-1}$. By adding $D_{i,j-1}$ on both sides of (2.53) we have confirmed assumption number 1.

Assumption nr 2 can be shown in a similar way. We have that

$$Var\left(D_{ij}\left|k_{i,j-1}\right.\right)=Var\left((D_{i,j-1}+C_{ij})\left|k_{i,j-1}\right.\right)=Var\left(C_{ij}\left|k_{i,j-1}\right.\right) \ . \tag{2.54}$$

The variance of a compound Poisson variable is stated in formula (2.3). If we let D_{ij} be the claim amount, λ is the parameter of the Poisson distributed variable N, and U is the size of a claim, we have that

$$Var\left(D_{ij}\right) = \lambda VarU + \lambda\left(EU\right)^2 = \lambda E(U^2) \tag{2.55}$$

In order to see the how the compound Poisson model and Mack's model are related, these two properties can be compared:

$$Var\left(D_{ij}\middle|k_{i,j-1}\right) = Var\left(C_{ij}\middle|k_{i,j-1}\right) = d_{i,j-1}g_j \tag{2.56}$$

$$\lambda E\left(U^2\right) = \lambda\int_{-\infty}^{\infty} U^2 dH_j(U) = \lambda g_j \tag{2.57}$$

The Poisson intensity is proportional or equal to $d_{i,j-1}$, and this shows that the compound Poisson model also satisfies Mack's second assumption.

2.7 Negative incremental claims

Negative incremental claims are a consequence of already reported claims which are being reduced or diminished. By this definition no cumulative claims can occur. Negative incremental claims can occur because of salvage, conservative case estimates or subrogation

(Kunkler 2006). Subrogation is a technique that insurance companies use when a claim has been covered, but there is a third party who can be held responsible for the claim. The insurance company makes a claim for compensation by the third party.

There have been proposed different solutions on how to handle negative incremental claims. One method involves adding a positive constant to all incremental claims. After the analysis is completed, the constant is subtracted. This method provides suitable results as long as there are not too many negative claims. On the other hand, this procedure makes the variability of the result depend on the constant added earlier, which cannot be considered reasonable (Kunkler 2006). If the negative claims are not manipulated as suggested above, the model to be used needs to handle negative claims. If the distribution is specified in the model, it needs to be defined for negative as well as for the positive numbers. A suitable candidate is the normal distribution, which is defined for both positive and negative numbers.

2.8 Predictions and prediction errors

The south-eastern corner of the run-off triangle is filled with point estimates, $\hat{D}_{ij}$. The last development year represents the ultimate claim $\hat{D}_{in}$ for $i = 2,...,n$. It is desirable to find a measure of the variability of this point estimate. The mean squared error (MSE) might be an appropriate measure. The formula for the MSE of $\hat{D}_{in}$ will be found below. Root mean squared error of prediction (RMSE) will be used as a measure of prediction error. MSE will also be referred to as the prediction variance.

There are already observed values in the north-western corner in the run-off triangle. The MSE should take these into account, and because of this the MSE is conditioned on k. To simplify notation, it will only be referred to as MSE of $\hat{D}_{in}$ (not the conditional MSE). The MSE of $\hat{D}_{in}$ is:

$$E\left(\left(D_{in}-\hat{D}_{in}\right)^2|k\right)=E\left(\left(D_{in}-E\left(D_{in}\right)+E\left(D_{in}\right)-\hat{D}_{in}\right)^2|k\right) \tag{2.58}$$

$$=Var\left(D_{in}|k\right)+E\left(E\left(D_{in}|k\right)-\hat{D}_{in}\right)^2 \tag{2.59}$$

$$=Var\left(D_{in}|k\right)+\left(E\left(D_{in}|k\right)-\hat{D}_{in}\right)^2 \tag{2.60}$$

Since $E\left(D_{in}|k\right)$ and $\hat{D}_{in}$ only are a scalars, the outer expectation is removed in the second term of (2.59), and it is only necessary to condition on k on the stochastic variable D_{in}. The first term in (2.60) is the variance around the true value D_{in}, and it will always be present. The second term in (2.60) is a measure on how much the predictor $\hat{D}_{in}$ misses its target $E\left(D_{in}|k\right)$, and is referred to as the estimation variance (Mack 1994a).

3. Analysis of data from auto liability insurance claims using stochastic chain ladder models

3.1 Sample

The data set used in the analysis is a set of claims data for auto liability insurance from TrygVesta. The insurance claims were organized by accident year and development year. It contained the number of reported claims, the amount of claims that had been paid, the number of RBNS- claims, the amount of reserves for RBNS-claims combined with paid claims and the number of settled claims.

The data set containing the number and amount of claims will be used when fitting the stochastic chain-ladder models. No analysis has been made on this set of data and no reserve has been added, and because of this, these observations seem to be the most appropriate for further analysis. By using these observations I will make predictions of total number and total amount of future claims. The data sets are presented in Appendix 1.

As mentioned in the introduction it is the IBNR-claims that are interesting to predict. The paid amount in an early development year is used to predict total future payments. The prediction of amount of claims contains both the RBNS-payments and the IBNR-payments. The insurance company might have been notified of a claim and have made a reserve estimate for this claim. Finding the IBNR-claims one simply has to withdraw the RBNS-claims. Since the insurance company has no information before the first notification of the claim, the number of claims from the set of data is equal to the number of IBNR-claims.

Given the run-off triangle as illustrated in figure 1, it is natural to calculate the rest of the triangle, but more importantly the ultimate claim $\hat{D}_{in}$. The insurance company needs a reserve, R_i, to cover future claims. We have that $\hat{R}_i = \hat{D}_{in} - D_{i,n-i+1}$. The estimates of R_i with corresponding prediction errors will be presented in tables.

3.2 Stochastic models in the analysis

Both sets of data contain negative incremental claims. This limits some of the stochastic models introduced earlier. Neither the Poisson model nor the negative binomial model can contain negative incremental claims if the regular maximum likelihood estimator of the parameters is used. As long as the sum of the incremental claims belonging to one development year is not negative, this problem can be solved by using a quasi log-likelihood (Renshaw 1998). In the data set to be used in this analysis, several negative claims occur, particularly in development year two. Even the sum of number of incremental claims turns out to be negative, which excludes the possibility of using the Poisson or the negative binomial model. The close connection between the Poisson and Negative Binomial model is described earlier, and a normal approximation to the Negative Binomial model can be used. This model can handle the negative numbers, and can also generate reserve estimates and prediction errors.

Mack's model is also used to find reserve estimates and prediction errors. Mack's model only makes assumption regarding the two first moments, and there seem to be no obvious reason why this would be a problem when negative claims occur as long as the cumulative claims are positive.

3.3 A critical view on the stochastic chain-ladder assumptions

The two models used in the analysis make similar assumptions regarding the two first moments. First, it is assumed that a linear relationship between an insurance claim in development year $j+1$ and a claim in development year j exists, and that the factor in the linear relationship is the chain-ladder factor. A corresponding linear relationship is assumed to exist for the second moment as well. Before applying the models to the data set, these assumptions will be explored.

3.3.1 The chain-ladder bias

The chain-ladder model assumes a linear relationship in claims between the development years where the chain-ladder development factor is used. The chain-ladder method contains no intercept, and in this situation it is interesting to see whether a linear model with intercept would predict future claims even better. Halliwell (2007) suggested that the bias of chain-ladder method could be tested by comparing the more general linear model where the intercept was not forced to pass through origin.

The linear relationship that is assumed in the two models is:

$$E\left(D_{ij} \middle| d_{i,j-1}\right) = f_j d_{i,j-1} \tag{3.1}$$

A more general linear model could be expressed like

$$D_{ij} = \beta_o + \beta_j d_{i,j-1} + \varepsilon \tag{3.2}$$

where D_{ij} is a stochastic variable, $d_{i,j-1}$ is considered known, β_0 and β_j are parameters which need to be estimated, ε is the error term and must follow the same distribution as D_{ij}. Three models will be fitted to the data. The difference between the three models is the change of estimators of β_0 and β_j. The estimators that will be used is the chain-ladder development factor $\hat{f}_j$ and the least square estimators $\hat{f}_0^{LS}$ and $\hat{f}_j^{LS}$. The three models are

- The chain-ladder model, $\hat{\beta}_o = 0$ and $\hat{\beta}_j = \hat{f}_j$
- The general model, $\hat{\beta}_o = \hat{f}_j^{LS}$ and $\hat{\beta}_j = \hat{f}_j^{LS}$

- The restricted model, $\hat{\beta}_o = 0$ and $\hat{\beta}_j = \hat{f}_j^{LS}$

The general model offers more flexibility because of the possibility of a second parameter, the intercept. It is logical to assume that this model can better be fitted to a data set than the two other models. This assumption is the basis for assessing the bias of the chain-ladder method.

The three models were fitted to the data set of the number and the amount of claims. The two data sets have 18 development years which can be compared with the previous development year. At least three observations are needed in each development year, so it is possible to make 16 plots, but the analysis with just a few observations are less trustworthy. The general linear model and the restricted linear model are made in R by using the following commands:

lm(developmentyear(*j+1*)~developmentyear(*j*)-1)

lm(developmentyear(*j+1*)~developmentyear(*j*))

The graphic results are presented below. Although the models were tested for all development years only a few plots are presented. The three different models ended up having almost identical estimates of the parameters in the plots of the following development years.

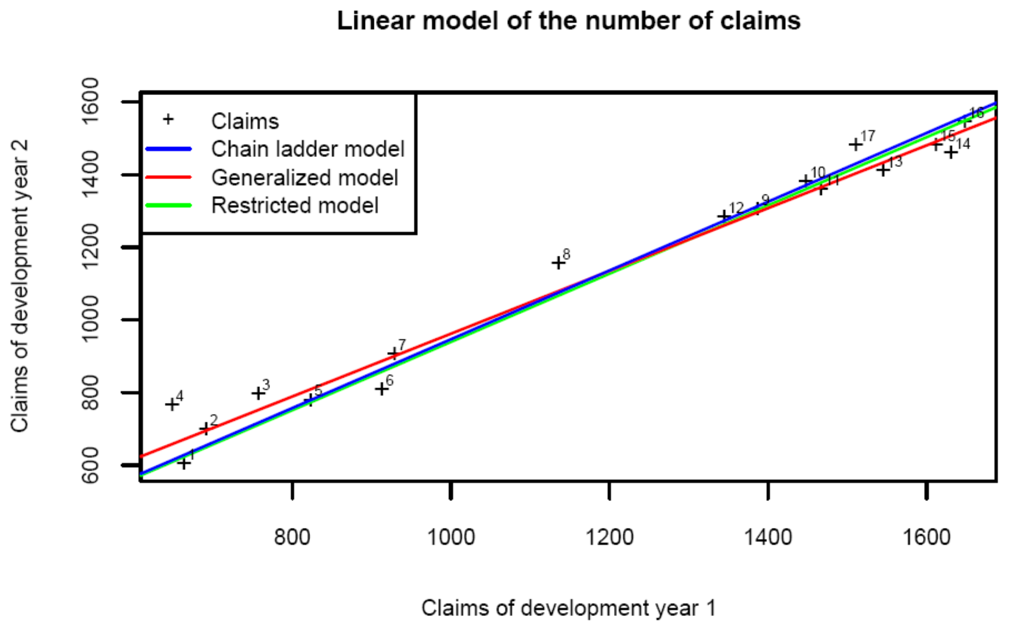
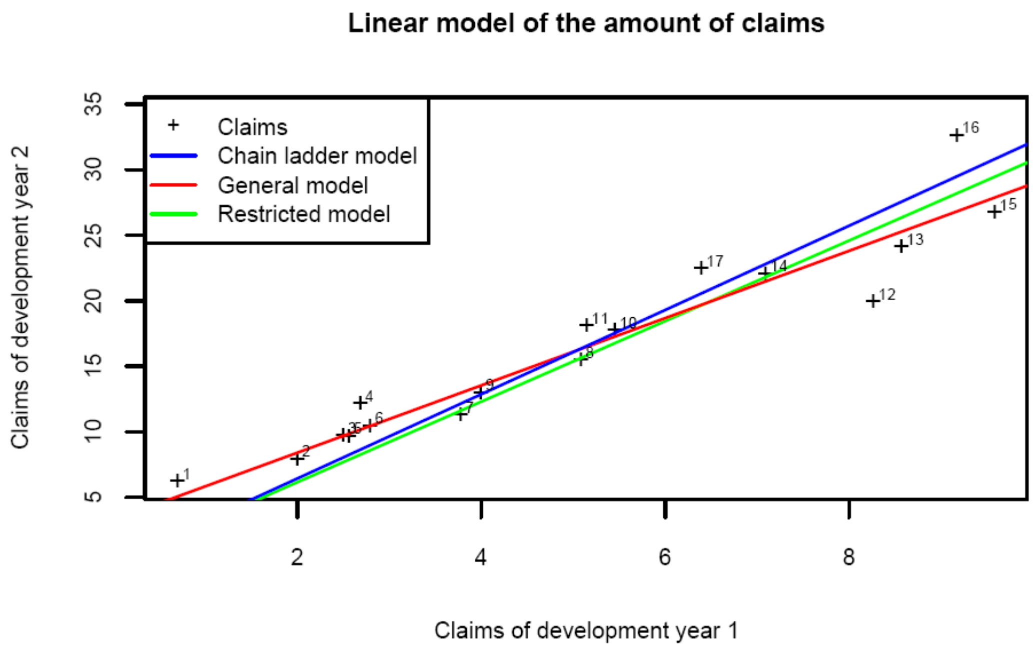
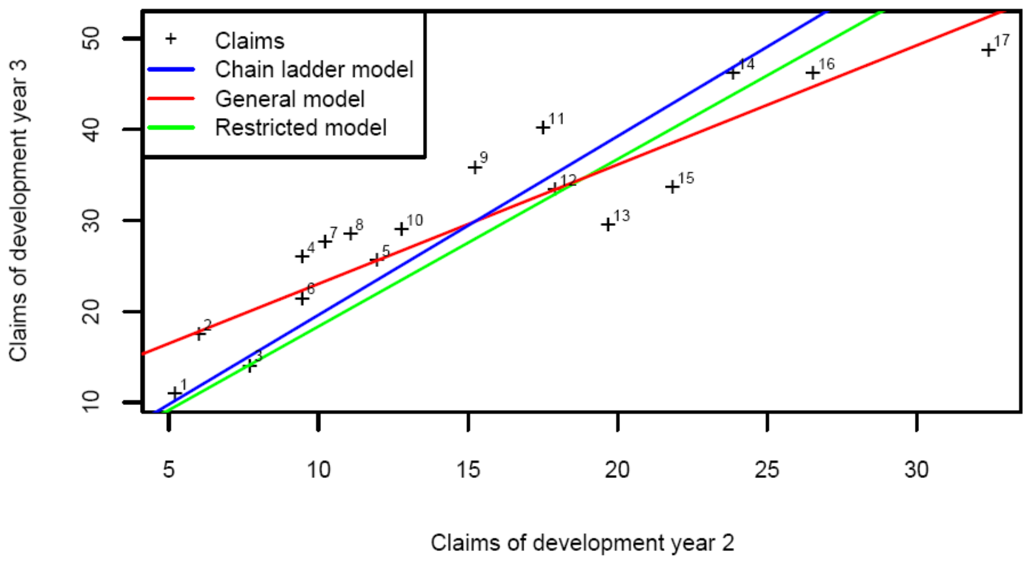


Figure 4. Three linear models fitted to the cumulative data of the number of claims. The data of development year 2 are plotted as a function of the date in development year 1.





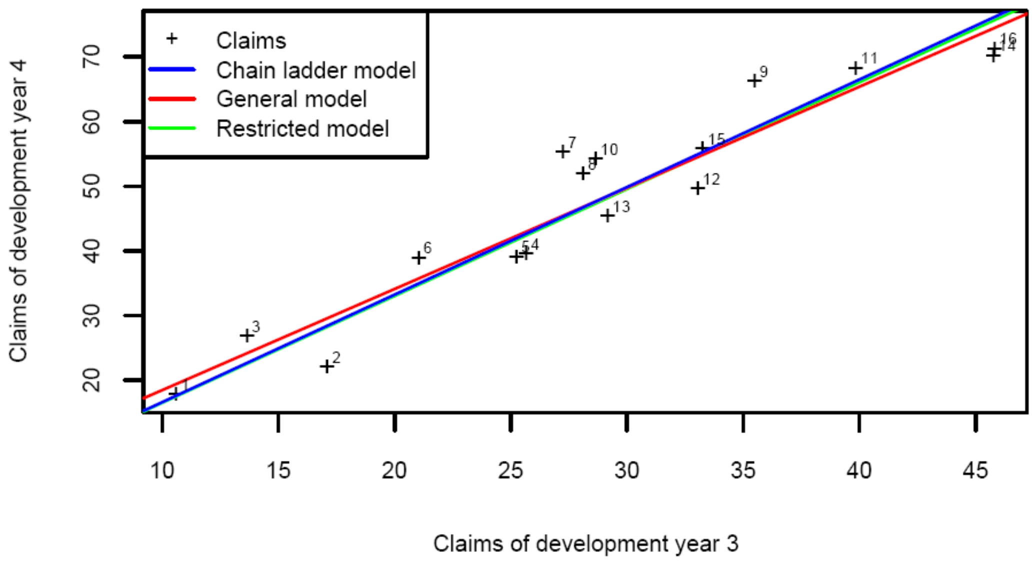


Figure 5. Three linear models are fitted to the cumulative data of the amount of claims. The data are plotted as a function of the previous development year. The three different plots are made for development year 2, 3 and 4.

The assumption regarding linearity of the first moment has not really been challenged. However, by examining the graphic results visually, it seems that a linear model of the form (3.2) fits the data sets well. If further investigation seemed necessary an analysis of variance could have been performed. The linear relationship in the first moment is accepted, and the focus of this analysis is the chain ladder bias.

The linear models of the number of claims are almost identical already in the first development year, but some differences can be mentioned. Both the chain-ladder and the restricted model are forced to pass through origin, and seem to result in a slight underestimation of the claims of the early accident years compared to the general model. It could also be an underestimation of the small claims, since claims in the early accident years seem to be smaller than claims in the late accident years. The derivative (gradient) of the chain-ladder model and the restricted model is greater to compensate for the positive intercept in the general model, and the two models might overestimate claims either in the late accident years or the greater claims than the general model.

To examine the chain-ladder bias, it is assumed that the general model is better than the two others. If β_0 is significantly different from zero, it would give reason to believe that the chain-ladder method is biased. To make inferences regarding the first parameter β_0 a distribution of the claims D_{ij} can be assumed. If D_{ij} is assumed to follow a normal distribution, the t-values of the parameter would a follow a students t-distribution with $n-2$ degrees of freedom. Using a level of significance of 0,1 development year 2 and 9 showed significant results for the first parameter β_0 (see Appendix 3). This implies that a linear model containing an intercept that is different from zero would fit the data even better for these development years.

The three plots for the amount of claims show that for a higher development year the models become more similar. After development year 3 they are almost identical. The linear models of the amount of claims show more diverging behaviour than the models for the number of claims. The same trend is apparent as for the number of claims. The general model has a positive intercept for all the three development years displayed above, which forces the chain-ladder model and the restricted model to compensate with a higher gradient. The gradient of the chain-ladder model even exceeds the gradient of the restricted model.

If the claims D_{ij} are assumed to follow a normal distribution, estimates of β_0 that are significantly different from zero were obtained for development year 2 and 3 (using a level of significance of 0,1). Only 5 of 16 parameter estimates were negative for the number of claims. Three negative estimates of the parameter β_0 were calculated for the data set of the amount of claims. This implies that the intercept should be positive. Whether it is underestimation of small claims and overestimation of larger claims, or underestimation in the early accident years and overestimation in the late accident years is difficult to determine from these analyses.

3.3.2 The variance of claims

The two models to be used in the analysis have the same formulation in the variance assumption:

$$Var\left(D_{ij}\left|k_{j-1}\right.\right)=c_j d_{i,j-1} \quad \text{where } c_j \text{ is a constant} \qquad (3.3)$$

The chain-ladder development factor is a weighted mean of the individual development factors, and it was proved in (2.46) that the variance of the chain-ladder development factor needed to be inversely proportional with the weights. The equivalent variance assumption of D_{ij} is given in (3.3). If another development factor is chosen, the variance assumption might need to be altered to attain minimum variance. Two other development factors are suggested as alternative development factors, and they are also a weighted mean of the individual development factors. These are

$$\hat{f}_j^{mean}=\frac{1}{n-j+1}\sum_{i=1}^{n-j+1} w_{ij} f_{ij}\,, \qquad \text{where } w_{ij}=1 \text{ for } i=1,\dots,n\text{-}j\text{+}1$$

$$\hat{f}_j^{ls}=\frac{\sum_{i=1}^{n-j+1} d_{i,j-1}d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}^2}=\sum_{i=1}^{n-j+1}\frac{d_{i,j-1}^2}{\sum_{i=1}^{n-j+1} d_{i,j-1}^2}\frac{d_{ij}}{d_{i,j-1}}\,,$$

$$= \sum_{i=1}^{n-j+1} \frac{d_{i,j-1}^2}{\sum_{i=1}^{n-j+1} d_{i,j-1}^2} f_{ij} \propto \sum_{i=1}^{n-j+1} w_{ij} f_{ij} \qquad \text{where } w_{ij} = d_{i,j-1}^2 \text{ for } i = 1,...,n\text{-}j+1$$

$\hat{f}_j^{mean}$ is the mean of the individual development factors and $\hat{f}_j^{ls}$ is the least square estimator of the individual development factors. A residual analysis can be performed using the different development factors and the belonging variance assumption that can be derived from (2.46).

The three residual plots become

$$r_{ij}^{mean} = \frac{d_{ij} - \hat{f}_j^{mean} d_{i,j-1}}{1}$$

$$r_{ij} = \frac{d_{ij} - \hat{f}_j d_{i,j-1}}{\sqrt{d_{i,j-1}}}$$

$$r_{ij}^{ls} = \frac{d_{ij} - \hat{f}_j^{ls} d_{i,j-1}}{d_{i,j-1}}$$

The three residual plots are produced for all j. r_{ij}^{mean} assumes D_{ij} has constant variance, r_{ij} assumes D_{ij} has variance proportional to $d_{i,j-1}$ and r_{ij}^{ls} assumes D_{ij} has variance proportional to $d_{i,j-1}^2$. If one of the residuals above seems to have a more random behaviour, the choice of development factor should be reconsidered.

These residual plots were created for the data set of number and amount of claims. Of the 19 development years in the data set, 14 have 6 or more observations and residual plots were created for these. It was searched for a possible development in the variance, a trend or difference between the development factors. The three residual plots for development year 2 and 3 are presented below for the number of claims. The residual plots are by visual examination identical for the three different development factors (only the scale of the y-axis deviates). Because of the likeness of the plots, only the results using r_{ij} is included in Appendix 4 for the rest of the development years for the number of claims and the amount of claims.

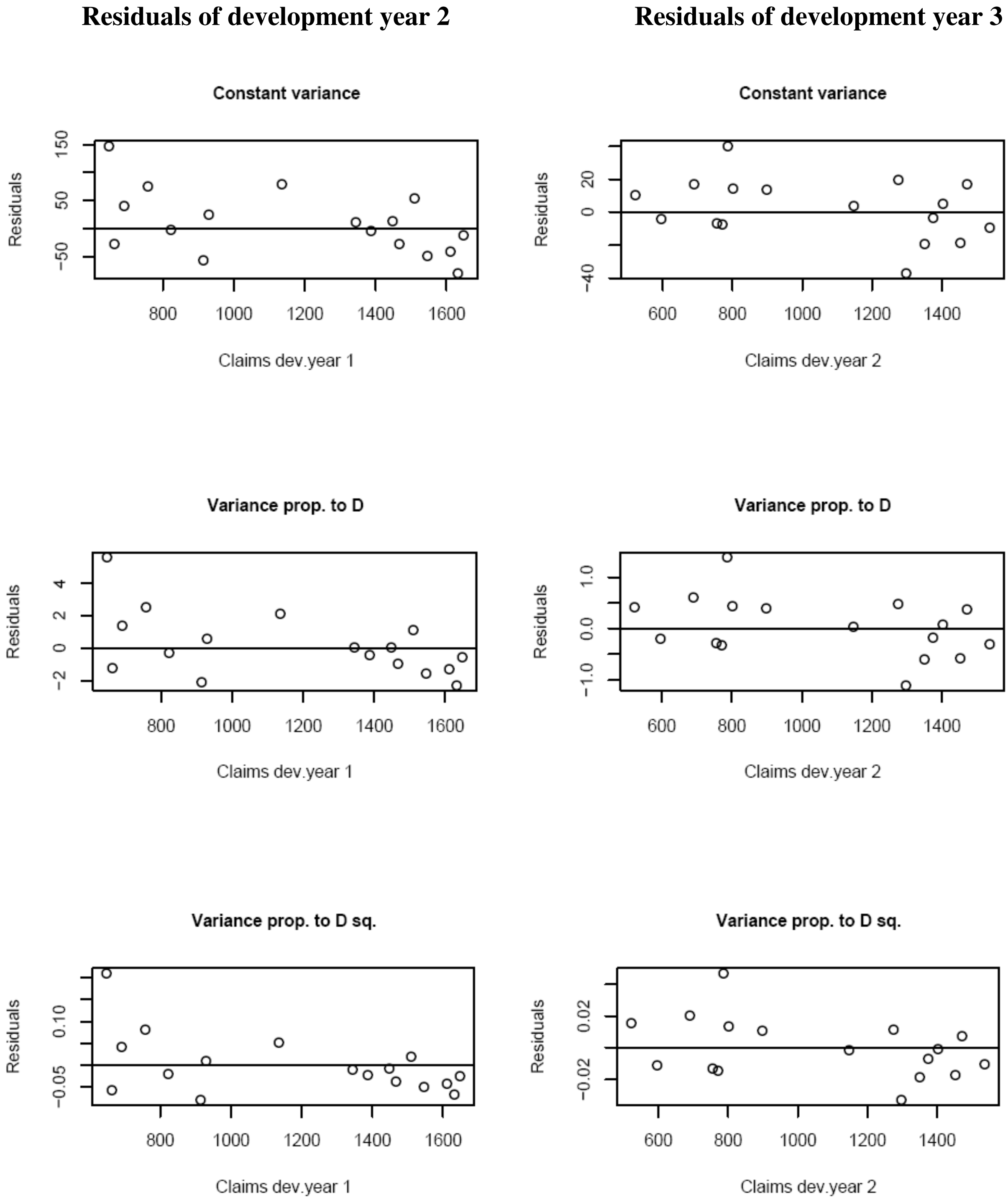


Figure 6. The three residual plots on the left side are r_{i2}^{mean}, r_{i2} and r_{i2}^{ls} plotted against claims of development year 1 for the number of claims. The three residual plots on the right side are r_{i3}^{mean}, r_{i3} and r_{i3}^{ls} plotted against claims of development year 2 for the number of claims. The variance assumption belonging to the residual is written above all the plots.

The residual plots for the number of claims are examined first. The three different development factors create strikingly similar residual plots. It seems there is no difference in the random behaviour of three development factors. There is an overrepresentation of positive residuals for small claims, and negative residuals for large claims. This suggests that there is a trend, which also was barely visible in the linear model created in the previous chapter. This trend is equally present for all development factors. After development year 9 the trend is no longer visible, but at that point the run-off triangle is almost fully developed. In development year 6, 7 and 8 the residuals are larger for small claims and the residuals are smaller for large claims, and this is equally present for the three development factors.

In the data set of the amount of claims the downward trend is visible for the first two development years and development year 6. Other than this the data set shows are more random behaviour in the residual plots. The exception is development year 7 which seems to show a non-random behaviour.

Neither of the development factors compute purely random residual plots. The trend could probably be removed by including a second parameter in the model, the intercept. Since there is almost no difference between the development factors $\hat{f}_j$, $\hat{f}_j^{mean}$, $\hat{f}_j^{ls}$, the usage of the chain-ladder development factor and its corresponding variance assumption will be continued.

3.4 Normal approximation to the negative binomial distribution

3.4.1 The model

A normal approximation to the negative binomial model was used to analyze that data set from TrygVesta. First the model is introduced generally. A linear and a generalized linear model will be fitted to the data in the process of predicting future claims. This will be presented in general formulas, but it will also be exemplified by using the dimensions of the data set from TrygVesta.

The negative binomial model was presented in (2.37), and the mean and variance of D_{ij} conditioned on $d_{i,j-1}$ was:

$$E\left(D_{ij}\left|d_{i,j-1}\right.\right)=f_j d_{i,j-1} \text{ and } Var\left(D_{ij}\left|d_{i,j-1}\right.\right)=f_j\left(f_j-1\right)d_{i,j-1} \qquad j=2,\ldots,n \tag{3.4}$$

Because of the negative incremental claims in the data set the development factor becomes smaller than one, and this produces negative variance. The model to be used needs to handle positive and negative values of incremental claims. The normal distribution is a possibility. Since the negative binomial model does not fit the data set, the conditional distribution of D_{ij} is instead assumed to approximately follow a normal distribution.

In this analysis the focus will be on the quantities f_{ij}, the individual development factors. In the model F_{ij} is considered a stochastic variable, and from the data set there are observed values of f_{ij} which are realizations of F_{ij}. f_{ij} is observed in the north-western corner of the run-off triangle. $\hat{f}_{ij}$ is the predicted value of f_{ij} which will replace the empty spots in the south-eastern corner of the run-off triangle. These are approximately independently and normally distributed within the development year j.

Let $F_{ij}=\dfrac{D_{ij}}{w_{ij}}$ where $w_{ij}=d_{i,j-1}$. The mean and variance are

$$E\left(\frac{D_{ij}}{w_{ij}}\left|d_{i,j-1}\right.\right)=f_j \qquad \text{and} \qquad Var\left(\frac{D_{ij}}{w_{ij}}\left|d_{i,j-1}\right.\right)=\frac{\phi_j}{w_{ij}} \tag{3.5}$$

The variable w_{ij} has been introduced because a weighted linear model will be used in the analysis of finding the unknown individual development factors. The weights in the analysis are $W_{ij} = \frac{w_{ij}}{\phi_j}$. The weights are inversely proportional with the variance, so that data with a greater variance is less weighted. The variance component depends on the development year, and will also need to be estimated.

The linear model is as follows

$$E\left(F_{ij}\right) = c + \alpha_{j-1} \qquad \text{for } j \geq 2, \qquad \text{and with a restriction } \alpha_1 = 0 \qquad (3.6)$$

It is assumed F_{ij} is independent, and it should be noted that the model does not condition on the latest observation in accident year i, $d_{i,n-i+1}$. Since F_{ij} is normally distributed, the link function is only the identity function. In order to find estimates of both f_{ij} and ϕ_j joint modelling can be used. This technique is described in (Renshaw 1994;Verrall & England 2002). The technique will be described here using the data set from TrygVesta.

Figure 1 displayed two run-off triangles for claims c_{ij} and d_{ij}, and the belonging run-off triangle for development factors, f_{ij}, had a smaller dimension by one. A run-off triangle of cumulative claims, d_{ij}, with dimension 19x19 has a corresponding run-off triangle of f_{ij} with dimension 18x18. Only the values in the north-western corner are known values, and these will be used as response variables as shown in the linear model in (3.6). There are 171 $\left(i + j \leq 19\right)$ known values of f_{ij}, and there are 153 $\left(j \geq 2, i + j < 18\right)$ values of $\hat{f}_{ij}$ to be predicted in the south-eastern corner of the run-off triangle. For development year j all predicted values, $\hat{f}_{ij}$, will be equal, and the subscript i could have been left out.

The two data sets from TrygVesta containing the number and the amount of claims both have an empty spot for accident year 1 and development year 1. This means that there are only 170 observations as opposed to 171 which one generally would have from a run-off triangle with dimension 19x19. This results in a missing observations in the vector **f** and one less row in the design matrix. When fitting the linear model in the statistical software programme R, f_{ij} is rearranged as a vector of dimension 170x1. The linear model in (3.6) can be written on vector form

$$E(\mathbf{F}) = \mathbf{X\theta} \tag{3.7}$$

where **F** is the response variable. **X** is the design matrix, and $\boldsymbol{\theta}$ is the parameter to be estimated. The vector **f** and the parameter, $\boldsymbol{\theta}$ will be

$$\mathbf{f}^{\mathbf{T}} = \begin{bmatrix} f_{2,2} & \cdots & f_{18,2} & f_{1,3} & \cdots & f_{17,3} & \cdots & f_{1,19} \end{bmatrix} \text{ and}$$
$$\boldsymbol{\theta}^{\mathbf{T}} = \begin{bmatrix} c & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \cdots & \cdots & \alpha_{18} \end{bmatrix} \tag{3.8}$$

The vector $\boldsymbol{\theta}$ has dimension 18x1. The design matrix can now be defined. In this format the linear model will have a design matrix of dimension 170x18. In the vector **f** there are 18 development years. The design matrix can be presented for each development year, and the corresponding dimension is written on the right hand side

Development year 1 $\quad [\mathbf{1} \;\; \mathbf{0} \;\; \mathbf{0} \;\; .. \;\; .. \;\; \mathbf{0}]_{(17\text{x}18)}$ (3.9)

Development year 2 $\quad [\mathbf{1} \;\; \mathbf{1} \;\; \mathbf{0} \;\; .. \;\; .. \;\; \mathbf{0}]_{(17\text{x}18)}$

Development year 3 $\quad [\mathbf{1} \;\; \mathbf{0} \;\; \mathbf{1} \;\; \mathbf{0} \;\; .. \;\; \mathbf{0}]_{(16\text{x}18)}$

….

Development year 18 $\quad [\mathbf{1} \;\; \mathbf{0} \;\; \mathbf{0} \;\; .. \;\; \mathbf{0} \;\; \mathbf{1}]_{(1\text{x}18)}$

The weights used in the linear model may be formulated as a vector. It will have a similar structure as **f**. Notice that in addition to the known values of w_{ij}, the variables ϕ_j are unknown, but they are only dependent on the development year. **W** can be written like this:

$$\mathbf{W}^{\mathbf{T}} = \begin{bmatrix} \frac{w_{1,2}}{\phi_2} & \cdots & \frac{w_{18,2}}{\phi_2} & \frac{w_{1,3}}{\phi_3} & \cdots & \frac{w_{17,3}}{\phi_3} & \cdots & \frac{w_{1,19}}{\phi_{19}} \end{bmatrix}$$

Arbitrary values for ϕ_j are chosen. To make it simple, the first set of ϕ_j are set equal to 1. The linear model in (3.7) can be solved in R by the command:

lm.wfit(**f, W, X**)

This command produces an estimate for the parameter $\boldsymbol{\theta}$, and by the linear combination an estimate for **f** can be obtained. Since the values of ϕ_j still are unknown, a second linear model needs to be fitted. The second model uses the residuals squared as the new response variables. Let r_{ij}^2 be the residuals squared, and they are defined as

$$r_{ij}^2 = w_{ij}(f_{ij} - \hat{f}_{ij})^2 \tag{3.11}$$

The generalized linear model to be fitted is

$$g\left(E\left(R_{ij}^2\right)\right) = c_2 + \gamma_{j-1} \quad \text{for } i = 1,...,n\text{-}j\text{+}1 \text{ and } j = 1,...,n \tag{3.12}$$

This is a "generalized" linear model since R_{ij}^2 can not be directly explained through a linear model. A link function (g), makes it possible to let R_{ij}^2 be explained through a linear model. This link function is closely related to the distribution of the response variable.

Since F_{ij} is normally distributed with mean f_j and variance $\sqrt{\frac{\phi_j}{w_{ij}}}$, $R_{ij} = \sqrt{w_{ij}}\left(f_{ij} - \hat{f}_{ij}\right)$ is normally distributed with $E\left(R_{ij}\right) = 0$ and $Var\left(R_{ij}\right) = \phi_j$. Thus $\frac{R_{ij}{}^2}{\phi_j} = w_{ij}\frac{\left(f_{ij} - \hat{f}_{ij}\right)^2}{\phi_j}$ is chi-squared distributed with $E\left(\frac{R_{ij}^2}{\phi_j}\right) = 1$ and $Var\left(\frac{R_{ij}^2}{\phi_j}\right) = 2$. The mean and variance of the response variable R_{ij}^2 is $E\left(R_{ij}^2\right) = \phi_j$ and $Var\left(R_{ij}^2\right) = 2\phi_j^2$. Let c be the chi-squared distribution of the variable $\frac{R_{ij}^2}{\phi_j}$, and let the new variable $Z = R_{ij}^2$ be distributed with function f. The distribution f can be found through a linear transformation of c

$$f_Z\left(z\right) = c_{R_{ij}^2/\phi_j}\left(\frac{z}{\phi_j}\right)\left|\frac{dr}{dz}\right| = \frac{1}{\Gamma\left(\frac{1}{2}\right)2^{1/2}}\left(\frac{z}{\phi_j}\right)^{1/2-1} e^{-\left(\frac{z}{2\phi_j}\right)}\frac{1}{\phi_j} \qquad (3.13)$$

$$= \frac{1}{\Gamma\left(\frac{1}{2}\right)\left(2\phi_j\right)^{1/2}} z^{-1/2} e^{-\left(\frac{z}{2\phi_j}\right)}$$

It is now clear that (3.13) is the gamma distribution with parameters $\frac{1}{2}$ and $2\phi_j$. The canonical link function of a gamma distribution is the inverse function (McCullagh & Nelder 1989). There are other possible link functions to the gamma distribution. This is the identity and the log function. Verral (2000) suggested using the log function. The different link functions were tested for the data sets from TrygVesta. Using the log link function indeed showed a linear relationship as opposed to the other link functions.

The generalized linear model written in vector form is

$$g\left[E(\mathbf{R}^2)\right] = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\theta}_1 \tag{3.14}$$

where g is the link function and $g(\boldsymbol{\mu}) = \log(\boldsymbol{\mu})$. The response variable $\mathbf{R}^2$ and the parameter vector $\boldsymbol{\theta}_1$, have the same format as the response variable $\mathbf{f}$ and the parameter vector $\boldsymbol{\theta}$ in (3.7). The identity matrix **X** is equivalent to the identity matrix in (3.7).

In R this can be done with the function:

glm.fit(**R,W,X,** Gamma(link = log)

New values for ϕ_j , *j = 2,...,18* can be obtained. ϕ_{19} can not be obtained since there is only one residual in the general linear model. The results that are presented later use two different options, $\phi_{19} = \phi_{18}$ and $\phi_{19} = \phi_{17}$.

The weight **W** is updated with new values of ϕ_j. Estimates of c and α_j *for j = 2,..18* are derived through the first linear model yet another time. The development factors can be calculated from these estimates. This is the joint modelling process.

The predicted values of f_{ij} can now be found through

$$\underset{(153\times1)}{\hat{\mathbf{f}}^{\mathbf{p}}} = \underset{(153\times18)}{\mathbf{X}^{\mathbf{p}}}\ \underset{(18\times1)}{\hat{\boldsymbol{\theta}}} \tag{3.15}$$

where $\mathbf{f}^{\mathbf{p}}$ is the vector of the predicted values of **f,** $\mathbf{X}^{\mathbf{p}}$ is the design matrix of the predicted development factors and $\hat{\boldsymbol{\theta}}$ is the vector of the parameter estimates. Like the observations of f_{ij}, the predicted values of f_{ij} will be lined up as a vector. The vector $\mathbf{f}^{\mathbf{p}}$ and $\hat{\boldsymbol{\theta}}$ can be written like

$$\left(\mathbf{f}^{\mathbf{p}}\right)^{\mathbf{T}} = \begin{bmatrix} f_{19,2} & f_{18,3} & f_{19,3} & f_{17,4} & f_{18,4} & f_{19,4} & \cdots & f_{19,19} \end{bmatrix} \text{ and}$$

$$\hat{\boldsymbol{\theta}}^{\mathbf{T}} = \begin{bmatrix} \hat{c} & \hat{\alpha}_2 & \hat{\alpha}_3 & \hat{\alpha}_4 & \hat{\alpha}_5 & \ldots & \ldots & \hat{\alpha}_{18} \end{bmatrix}$$

The design matrix for the different development years of the predicted values are presented below with the corresponding dimension written on the right hand side.

Development year 2 $\quad \begin{bmatrix}\mathbf{1} & \mathbf{1} & \mathbf{0} & .. & .. & \mathbf{0}\end{bmatrix}_{(1\times18)}$

Development year 3 $\quad \begin{bmatrix}\mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & .. & \mathbf{0}\end{bmatrix}_{(2\times18)}$

….

Development year 18 $\quad \begin{bmatrix}\mathbf{1} & \mathbf{0} & \mathbf{0} & .. & \mathbf{0} & \mathbf{1}\end{bmatrix}_{(17\times18)}$

It is desirable to find the standard errors of the parameters and of the development factors in the linear model. The theoretical calculation is shown below.

Let $\boldsymbol{\Sigma}^{-1}$ be a matrix with dimension 170x170. All the elements are zero except the diagonal which is the weight **W**. The variance of the parameter $\boldsymbol{\theta}$ can be found from the diagonal of the matrix $Var(\boldsymbol{\theta}) = \left(\mathbf{X}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}$. This is a matrix of dimension 18x18, and the square root of the diagonal produces the standard error.

It is also interesting to find the variance of **f**. The covariance matrix of **f**, with the corresponding dimensions written underneath the matrices, is

$$\underset{(153\times153)}{Cov\left(\hat{\mathbf{f}}\right)} = \underset{153\times18}{\mathbf{X}^{\mathbf{p}}} \underset{18\times18}{\left(\mathbf{X}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}} \underset{18\times153}{\left(\mathbf{X}^{\mathbf{p}}\right)^{\mathbf{T}}} \tag{3.16}$$

Joint modelling have produced estimates of the parameters in the first linear model, of the development factor f_j and of the variance component ϕ_j. It has been demonstrated how to find the standard errors of the parameters in the first linear model and the development factors. The results are presented below in tables 1-6.

Parameter	Estimate	Standard error
C	0,947	0,014
Alfa2	0,060	0,014
Alfa3	0,080	0,014
Alfa4	0,075	0,014
Alfa5	0,070	0,014
Alfa6	0,064	0,014
Alfa7	0,064	0,014
Alfa8	0,062	0,014
Alfa9	0,061	0,014
Alfa10	0,059	0,014
Alfa11	0,057	0,014
Alfa12	0,057	0,014
Alfa13	0,054	0,014
Alfa14	0,055	0,014
Alfa15	0,056	0,014
Alfa16	0,056	0,014
Alfa17	0,056	0,014
Alfa18	0,057	0,014

Table 1. Estimates of the parameters for the number of claims in the linear model.

Parameter	Estimate
Phi2	3,678
Phi3	0,320
Phi4	0,328
Phi5	0,125
Phi6	0,056
Phi7	0,012
Phi8	0,032
Phi9	0,017
Phi10	0,021
phi11	0,009
phi12	0,002
phi13	0,007
phi14	0,002
phi15	0,000
phi16	0,004
phi17	0,005
phi18	0,001
phi19	-

Table 2. Estimates of phi for the number of claims

Parameter	Estimate	Standard error
Dev. Factor 2	0,947	0,006
Dev. Factor 3	1,007	0,004
Dev. Factor 4	1,027	0,004
Dev. Factor 5	1,022	0,003
Dev. Factor 6	1,017	0,002
Dev. Factor 7	1,011	0,001
Dev. Factor 8	1,010	0,002
Dev. Factor 9	1,009	0,001
Dev. Factor 10	1,008	0,001
Dev. Factor 11	1,005	0,001
Dev. Factor 12	1,004	0,001
Dev. Factor 13	1,003	0,001
Dev. Factor 14	1,000	0,001
Dev. Factor 15	1,001	0,0003
Dev. Factor 16	1,002	0,001
Dev. Factor 17	1,003	0,002
Dev. Factor 18	1,002	0,001
Dev. Factor 19	1,003	0,001

Table 3. Estimates of the development factors and their standard errors for the number of claims

Parameter	Estimate	Standard error
C	3,215	0,163
Alfa2	-1,252	0,190
Alfa3	-1,553	0,169
Alfa4	-1,827	0,163
Alfa5	-1,976	0,163
Alfa6	-2,067	0,163
Alfa7	-2,133	0,163
Alfa8	-2,152	0,163
Alfa9	-2,183	0,163
Alfa10	-2,180	0,163
Alfa11	-2,193	0,163
Alfa12	-2,203	0,163
Alfa13	-2,193	0,163
Alfa14	-2,207	0,163
Alfa15	-2,211	0,163
Alfa16	-2,214	0,163
Alfa17	-2,207	0,163
Alfa18	-2,215	0,163

Table 4. Estimates of the parameters for the amount of claims in the linear model.

Parameter	Estimate
Phi2	2,266
Phi3	2,498
Phi4	0,934
Phi5	0,199
Phi6	0,208
Phi7	0,219
Phi8	0,099
Phi9	0,08
Phi10	0,021
phi11	0,029
phi12	0,021
phi13	0,006
phi14	0,001
phi15	0,003
phi16	0,010
phi17	0,0002
phi18	0,004
phi19	-

Table 5. Estimates of phi for the amount of claims

Parameter	Estimate	Standard error
Dev.factor 2	3,215	0,163
Dev.factor 3	1,963	0,098
Dev.factor 4	1,663	0,045
Dev.factor 5	1,388	0,017
Dev.factor 6	1,239	0,015
Dev.factor 7	1,148	0,015
Dev.factor 8	1,083	0,010
Dev.factor 9	1,063	0,009
Dev.factor 10	1,032	0,005
Dev.factor 11	1,036	0,006
Dev.factor 12	1,022	0,005
Dev.factor 13	1,013	0,003
Dev.factor 14	1,023	0,002
Dev.factor 15	1,008	0,003
Dev.factor 16	1,005	0,006
Dev.factor 17	1,002	0,001
Dev.factor 18	1,008	0,007
Dev.factor 19	1,000	0,011

Table 6.Estimates of the development factors and their standard errors for the amount of claims

3.4.2 Reserve predictions and prediction errors

When estimates of the development factors are found, claim estimates can be made. The empty spots in the run-off triangle can be estimated using the model in (3.4). The ultimate claim estimate $\hat{D}_{in}$ is calculated using the chain-ladder equation (1.6). The reserve estimate can be calculated since there is a simple connection between the reserve and the ultimate claim. The reserve is $R_i = D_{in} - d_{i,n-i+1}$, and equivalently the reserve estimate is $\hat{R}_i = \hat{D}_{in} - d_{i,n-i+1}$ for $i = 2,...,n$, where $\hat{D}_{in}$ has been calculated and $d_{i,n-i+1}$ is observed on the diagonal of the run-off triangle.

The variance factors (ϕ_j) have been found. The model in (3.4) determines the variance of the estimated claims $\hat{D}_{i,n-i+2}$ for $i = 2,...,n$. To find the prediction error and the reserve of the ultimate claim and some more calculations are needed. Equation (2.59) gave an expression for the MSE of $\hat{D}_{in}$:

$$MSE\left(\hat{D}_{in}\left|k\right.\right)=Var\left(D_{in}\left|k\right.\right)+E\left(E\left(D_{in}\left|k\right.\right)-\hat{D}_{in}\right)^2.$$

If independence between the accident years is assumed, it is not necessary to condition on all the observations. Furthermore, it is only the last observation in every accident year that is used in model (3.4). It suffices to condition on $d_{i,n-i+1}$ in this case. Approximating $E\left(\hat{D}_{in}\left|d_{i,n-i+1}\right.\right)$ with $E\left(D_{in}\left|d_{i,n-i+1}\right.\right)$ makes a new expression:

$$MSE\left(\hat{D}_{in}\left|d_{i,n-i+1}\right.\right)\approx Var\left(D_{in}\left|d_{i,n-i+1}\right.\right)+Var\left(\hat{D}_{in}\left|d_{i,n-i+1}\right.\right) \qquad (3.17)$$

The MSE of the reserve is the same as MSE of the ultimate claim. This is clear since

$Var\left(R_i\left|d_{i,n-i+1}\right.\right)=Var\left(D_{in}-D_{i,n-i+1}\left|d_{i,n-i+1}\right.\right)=Var\left(D_{in}\left|d_{i,n-i+1}\right.\right)$ and

$Var\left(\hat{R}_i\left|d_{i,n-i+1}\right.\right)=Var\left(\hat{D}_{in}-\hat{D}_{i,n-i+1}\left|d_{i,n-i+1}\right.\right)=Var\left(\hat{D}_{in}\left|d_{i,n-i+1}\right.\right)$

Thus $MSE\left(\hat{R}_i\left|d_{i,n-i+1}\right.\right)=MSE\left(\hat{D}_{in}\left|d_{i,n-i+1}\right.\right)\approx Var\left(D_{in}\left|d_{i,n-i+1}\right.\right)+Var\left(\hat{D}_{in}\left|d_{i,n-i+1}\right.\right)$ (3.18)

Verral (2000) denotes $Var\left(D_{in}\left|d_{i,n-i+1}\right.\right)$ as the process variance and $Var\left(\hat{D}_{in}\left|d_{i,n-i+1}\right.\right)$ as the estimation variance, and these terms will also be used here. To obtain the prediction error two recursive approaches will be used, which are presented in Verral (2000).

The estimation variance can be found for accident year *2,…,n*, and it is

$$Var\left(\hat{D}_{in}\left|d_{i,n-i+1}\right.\right)=Var\left(D_{i,n-i+1}\prod_{j=n-i+2}^{n}\hat{f}_j\left|d_{i,n-i+1}\right.\right) \tag{3.19}$$

$$= d_{i,n-i+1}^{2}Var\left(\prod_{j=n-i+2}^{n}\hat{f}_j\left|d_{i,n-i+1}\right.\right)$$

The second accident year can be found directly, since (3.19) only becomes $d_{2,n-1}^{2}Var\left(\hat{f}_n\left|d_{2,n-1}\right.\right)$. The estimation variance of the third accident year is more complicated since it is necessary to find the variance of a product of two development factors. The fourth accident year requires an estimate of the variance of the product of three development years, and so on. In order to find these variances independence or at least uncorrelated development factors must be assumed. Assuming independence or at least no correlation between the development factors the variance of the two last development factors is:

$$Var\left(\hat{f}_{n-1}\hat{f}_n\right)=\left(E\left[\hat{f}_{n-1}\right]\right)Var\left(\hat{f}_n\right)+\left(E\left[\hat{f}_n\right]\right)Var\left(\hat{f}_{n-1}\right)+Var\left(\hat{f}_{n-1}\right)Var\left(\hat{f}_n\right) \tag{3.20}$$

$$\approx\left(\hat{f}_{n-1}\right)^{2}Var\left(\hat{f}_n\right)+\left(\hat{f}_n\right)^{2}Var\left(\hat{f}_{n-1}\right)+Var\left(\hat{f}_{n-1}\right)Var\left(\hat{f}_n\right)$$

Appendix 5 proves the formula used in (3.20) . It is not the conditional variance $Var\left(\hat{f}_{n-1}\hat{f}_n \middle| d_{i,n-i+1}\right)$ that has been recovered but it is $Var\left(\hat{f}_{n-1}\hat{f}_n\right)$. In the actual calculation of the prediction error it is the numerical result of (3.16) that will be used, which is the unconditional variance.

When finding the variance of the product of the last three development factors, the previous result (variance of two development factors) will be used. Thus

$$Var\left(\hat{f}_{n-2}\left[\hat{f}_{n-1}\hat{f}_n\right]\right) \approx \left(\hat{f}_{n-2}\right)^2 Var\left(\hat{f}_{n-1}\hat{f}_n\right) + \left(\hat{f}_{n-1}\hat{f}_n\right)^2 Var\left(\hat{f}_{n-2}\right) + Var\left(\hat{f}_{n-2}\right) Var\left(\hat{f}_{n-1}\hat{f}_n\right)$$

The last step is found when $Var\left(\prod_{j=2}^{n}\hat{f}_j\right)$ is found.

To find the process variance, $Var\left(D_{in} \middle| d_{i,n-i+1}\right)$ a recursive procedure can be used. This procedure uses the rule of double expectation and double variance. The model gives that $Var\left(D_{ij} \middle| d_{i,j-1}\right) = \phi_j d_{i,j-1}$, so the process variance for the next development year is already defined. Leaving out the subscript i, the process variance two steps ahead is

$$Var\left(D_{j+1} \middle| d_{j-1}\right) = E\left(Var\left[D_{j+1} \middle| D_j\right] \middle| d_{j-1}\right) + Var\left(E\left[D_{j+1} \middle| D_j\right] \middle| d_{j-1}\right)$$

$$= E\left(\phi_{j+1} D_j \middle| d_{j-1}\right) + Var\left(f_{j+1} D_j \middle| d_{j-1}\right)$$

$$= \phi_{j+1} E\left(D_j \middle| d_{j-1}\right) + \left(f_{j+1}\right)^2 Var\left(D_j \middle| d_{j-1}\right)$$

$$= \phi_{j+1} f_j d_{j-1} + \left(f_{j+1}\right)^2 \phi_j d_{j-1}$$

The process variance three steps ahead is

$$Var\left(D_{j+2}\middle|d_{j-1}\right)=E\left(Var\left[D_{j+2}\middle|D_{j+1}\right]\middle|d_{j-1}\right)+Var\left(E\left[D_{j+2}\middle|D_{j+1}\right]\middle|d_{j-1}\right)$$

$$\begin{aligned}
&=E\left(\phi_{j+2}D_{j+1}\middle|d_{j-1}\right)+Var\left(f_{j+2}D_{j+1}\middle|d_{j-1}\right)\\
&=\phi_{j+2}E\left(D_{j+1}\middle|d_{j-1}\right)+\left(f_{j+2}\right)^2 Var\left(D_{j+1}\middle|d_{j-1}\right)\\
&=\phi_{j+2}E\left(E\left(D_{j+1}\middle|D_j\right)\middle|d_{j-1}\right)+\left(f_{j+2}\right)^2 Var\left(D_{j+1}\middle|d_{j-1}\right)\\
&=\phi_{j+2}E\left(f_{j+1}D_j\middle|d_{j-1}\right)+\left(f_{j+2}\right)^2 Var\left(D_{j+1}\middle|d_{j-1}\right)\\
&=\phi_{j+2}f_{j+1}f_jd_{j-1}+\phi_{j+1}\left(f_{j+2}\right)^2 f_jd_{j-1}+\phi_j\left(f_{j+2}\right)^2\left(f_{j+1}\right)^2 d_{j-1}
\end{aligned}$$

This procedure can be performed for three years a head and so on. The intention of using this recursive approach is to find the variance of the ultimate claim. The second accident year needs no more than one step, the third accident year needs two steps of the recursive approach and so on.

The overall MSE of the reserve is the sum of the estimation and process variance, but also a covariance element is added because of the covariance between the estimated values. The overall estimation and process variance is:

$$MSE\left(\hat{R}\middle|k\right)=\sum_{i=2}^{n}Var\left(D_{in}\middle|d_{i,n-i+1}\right)+\sum_{i=2}^{n}Var\left(\hat{D}_{in}\middle|d_{i,n-i+1}\right)+2\sum_{\substack{i=2\\l>i}}^{n}Cov\left(\hat{D}_{in},\hat{D}_{ln}\middle|d_{i,n-i+1}\right)\quad(3.21)$$

The estimation and process variance was calculated for the data set from TrygVesta in R. Since the calculation of the estimation variance required uncorrelated development factors the covariance matrix of $\hat{\mathbf{f}}$ was examined. $\hat{\mathbf{f}}$ is a vector of 153 elements. The 153 elements can be placed in the south east corner of the run-off triangle, and the elements situated in the same development year are equal. It was checked that only the covariance elements of the matrix were different from zero, and the rest were zero. This was the case for both the number and the amount of claims.

The estimation variance was found by making a loop for every accident year in R. The recursive procedure used the previously discovered variance of a product. The second accident year required only one calculation while accident year 19 required the same calculations done in a loop 18 times until the variance of the predicted ultimate claim could be obtained. To find the process variance a loop was also made for this calculation. Like the estimation variance, the loop ran a single time for accident year 2, and 18 times for accident year 19 to obtain the variance of the ultimate claim.

The total estimation and process variance was calculated by summing up the estimation and process variance. Great care was taken when finding the last term in (3.21).

3.4.3 Results

Reserve estimates with their respective prediction errors are presented in tables 7 and 8.

Accident year	**Reserve**	**Prediction error**	**Prediction error %**
2	2,2696	0,901	39,698
3	5,0462	1,440	28,536
4	8,5419	3,031	35,484
5	9,3872	3,554	37,860
6	10,4740	3,594	34,314
7	12,3098	4,263	34,631
8	15,9718	5,265	32,964
9	24,7007	6,426	26,015
10	33,7194	7,738	22,948
11	47,9351	10,177	21,231
12	59,6699	11,453	19,194
13	73,0450	13,474	18,446
14	93,8263	14,769	15,741
15	121,8697	18,137	14,882
16	162,3809	24,390	15,020
17	208,6802	35,895	17,201
18	211,5965	43,583	20,597
19	118,2655	93,339	78,923
Overall	**1 219,6896**	**78,571**	**6,442**

Table 7. Reserve and prediction error for the number of claims

Accident year	Reserve	Prediction error	Prediction error %
2	0,000	0,756	315539,2
3	0,596	1,222	205,0647
4	0,980	1,605	163,8432
5	1,725	2,304	133,5831
6	2,302	2,163	93,96275
7	6,109	2,7573	45,13254
8	8,155	3,058	37,4985
9	12,553	3,954	31,49889
10	14,976	4,137	27,62489
11	22,616	5,126	22,66539
12	26,523	5,918	22,31235
13	31,296	6,754	21,58113
14	64,976	11,289	17,37403
15	67,443	11,523	17,08554
16	115,679	15,368	13,28507
17	163,470	24,899	15,23153
18	168,715	41,325	24,49396
19	171,178	52,306	30,55655
Overall	**879,291**	**56,541**	**6,430291**

Table 8. Reserve and prediction error for the amount of claims

3.5 Mack's model

3.5.1 The model

Mack's model consisted of three assumptions. The two first assumptions concerned the two first moments, and they are for $j = 2,\dots,n$:

$$E\left(D_{ij}\middle|K_{i,j-1}=k_{i,j-1}\right)=f_j d_{i,j-1}$$

$$Var\left(D_{ij}\middle|K_{i,j-1}=k_{i,j-1}\right)=g_j d_{i,j-1}$$

The parameters have estimators:

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \qquad \text{and} \qquad \hat{g}_j = \frac{1}{n-j} \sum_{i=1}^{n-j+1} d_{i,j-1} \left(\frac{d_{ij}}{d_{i,j-1}} - \hat{f}_j \right)^2$$

These estimators were found for the data set from TrygVesta. These calculations were done in a spread sheet in excel, and the results are presented in table 9, where $\hat{f}_j$ is denoted as the development factor and $\hat{g}_j$ as the variance factor. There is not enough information to calculate $\hat{g}_{19}$, and for later purposes $\hat{g}_{19}$ will be set equal to either $\hat{g}_{18}$ or $\hat{g}_{17}$.

	Number of claims		**Amount of claims**	
Accident year	**Dev.factor**	**Variance factor**	**Dev. Factor**	**Variance factor**
1	-	-	-	-
2	0,947	3,678	3,215	2,266
3	1,007	0,340	1,963	2,654
4	1,027	0,328	1,663	0,996
5	1,022	0,119	1,388	0,179
6	1,017	0,059	1,239	0,221
7	1,011	0,012	1,148	0,232
8	1,010	0,032	1,083	0,099
9	1,009	0,019	1,063	0,080
10	1,008	0,023	1,032	0,023
11	1,005	0,008	1,036	0,033
12	1,004	0,003	1,022	0,024
13	1,003	0,006	1,013	0,007
14	1,000	0,003	1,023	0,001
15	1,001	0,000	1,008	0,004
16	1,002	0,002	1,005	0,002
17	1,003	0,007	1,002	0,0002
18	1,002	0,001	1,008	0,004
19	1,003	-	1,000	-

Table 9. Estimates of the development factor $\hat{f}_j$ and variance factor $\hat{g}_j$

for the number and the amount of claims.

3.5.2 Reserve predictions and prediction errors

The development factor is equivalent to the chain-ladder development factor. Future claim estimates can be made by using Mack's first assumption, and the empty spots in the run-off triangle can be filled with estimated values. Reserve estimates are found like they were when using the model of normal approximation, that is $\hat{R}_i = \hat{D}_{in} - d_{i,n-i+1}$. Since the development factor is equivalent to the previous model, the results are obviously identical.

Mack's second assumption determines the variance of the estimated claims. To find the prediction error of the ultimate claim more calculations are needed. It is clear from chapter 3.4.2 that the prediction error of the reserve is the same as the prediction error of the estimated ultimate claim. The MSE of $\hat{R}_i$ is

$$MSE\left(\hat{R}_i \middle| k\right) \approx \hat{D}_{in}^2 \sum_{j=n-i+2}^{n} \frac{\hat{g}_j}{\hat{f}_j^2} \left(\frac{1}{\hat{D}_{i,j-1}} + \frac{1}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \right) \qquad (3.22)$$

The overall prediction error of the reserve is

$$MSE\left(\hat{R} \middle| k\right) \approx \sum_{i=2}^{n} \left\{ MSE\left(\hat{R}_i \middle| k\right) + \hat{D}_{in} \left(\sum_{l=i+1}^{n} \hat{D}_{\ln} \right) \left(\sum_{k=n-i+2}^{n} \frac{2\hat{g}_k}{\hat{f}_j^2 \sum_{l=1}^{n-k+1} d_{lk}} \right) \right\} \qquad (3.23)$$

It is a quite an extensive task to find the estimators in (3.22) and (3.23), and it has recently been done in another master thesis (Gangsøy 2008). The calculations are because of this only included in Appendix 6.

The calculations of finding $MSE\left(\hat{R}_i | k\right)$ and $MSE\left(\hat{R} | k\right)$ for the data set from TrygVesta were done in an Excel spread sheet. The reserve estimates with the respective prediction errors are presented below in tables 10,11,12 and 13. The overall reserve, which simply is the sum of the reserves and the overall prediction error are also included.

For both the number of claims and the amount of claims, there are two tables. Since the variance factor $\hat{g}_{19}$ can not be calculated, it has been set equal to $\hat{g}_{18}$ and $\hat{g}_{17}$, and there is one table for each approximation.

3.5.3 Results

Reserve estimates with their respective prediction errors are presented in tables 10-13.

Accident year	**Reserve**	**Prediction error**	**Prediction error %**
1	0,000	-	-
2	2,270	0,870	38,333
3	5,046	1,390	27,545
4	8,542	3,440	40,272
5	9,387	3,680	39,202
6	10,474	3,690	35,230
7	12,310	4,400	35,744
8	15,972	5,250	32,870
9	24,701	6,430	26,032
10	33,719	7,660	22,717
11	47,935	10,320	21,529
12	59,670	11,700	19,608
13	73,045	13,720	18,783
14	93,826	14,980	15,966
15	121,870	18,470	15,156
16	162,381	24,440	15,051
17	208,680	35,900	17,203
18	211,596	44,060	20,823
19	118,265	95,550	80,793
Overall	**1219,690**	**139,140**	**11,408**

Table 10. Reserve estimates for the number of claims, and their prediction errors. For accident year 19 $g_{19} = g_{18}$.

Accident year	Reserve	Prediction error	Prediction error %
1	0,000	-	-
2	2,270	3,030	133,503
3	5,046	3,840	76,096
4	8,542	5,250	61,462
5	9,387	5,240	55,821
6	10,474	5,240	50,029
7	12,310	6,030	48,985
8	15,972	6,750	42,262
9	24,701	8,230	33,319
10	33,719	9,370	27,788
11	47,935	11,840	24,700
12	59,670	13,020	21,820
13	73,045	14,830	20,303
14	93,826	16,120	17,181
15	121,870	19,440	15,951
16	162,381	25,250	15,550
17	208,680	36,500	17,491
18	211,596	44,510	21,035
19	118,265	95,730	80,945
Overall	**1219,690**	**158,930**	**13,030**

Table 11. Reserve estimates for the number of claims, and their prediction errors. For accident year 19 $g_{19} = g_{17}$.

Accident year	Reserve	Prediction error	Prediction error %
1	0,000	-	-
2	0,000	0,680	283818,339
3	0,596	1,110	186,270
4	0,980	1,460	149,041
5	1,725	1,800	104,362
6	2,302	1,760	76,456
7	6,109	2,280	37,320
8	8,155	2,620	32,128
9	12,553	3,640	28,997
10	14,976	3,980	26,577
11	22,616	5,000	22,108
12	26,523	5,830	21,981
13	31,296	6,700	21,409
14	64,976	11,350	17,468
15	67,443	11,690	17,333
16	115,679	15,340	13,261
17	163,470	25,330	15,495
18	168,715	42,420	25,143
19	171,178	53,110	31,026
Overall	**879,291**	**89,860**	**10,220**

Table 12. Reserve estimates for the amount of claims, and their prediction errors. For accident year 19 $g_{19} = g_{18}$.

Accident year	**Reserve**	**Prediction error**	**Prediction error %**
1	0,000	-	-
2	0,000	0,160	66780,786
3	0,596	0,710	119,146
4	0,980	0,930	94,937
5	1,725	1,230	71,314
6	2,302	1,330	57,776
7	6,109	1,720	28,154
8	8,155	2,100	25,751
9	12,553	3,190	25,413
10	14,976	3,680	24,573
11	22,616	4,670	20,649
12	26,523	5,630	21,227
13	31,296	6,560	20,961
14	64,976	11,180	17,206
15	67,443	11,580	17,170
16	115,679	15,210	13,148
17	163,470	25,230	15,434
18	168,715	42,370	25,113
19	171,178	53,070	31,003
Overall	**879,291**	**86,340**	**9,819**

Table 13. Reserve estimates for the amount of claims and their prediction errors. For accident year 19 $g_{19} = g_{17}$.

4. Discussion

The Poisson model is a special case of the multiplicative model, and it was shown that using the maximum likelihood estimator in the Poisson model was equal to the chain-ladder method (Verral 2000). The negative binomial model can be derived from the Poisson model by letting the intensity be a stochastic variable as well as the claim (Verral 2000). When a normal approximation to the negative binomial model and Mack's model were fitted to the data, both models produced identical development factors. These were also both identical to the chain-ladder development factors.

Among the models that were introduced only one of them could handle negative incremental claims, Mack's model. As an alternative to the negative binomial model, a normal approximation was used since this would solve the problem with negative incremental claims. It is less attractive to use this approximation since more parameters need to be estimated (the variance factors). Mack's model only defines the two first moments, while the normal approximation to the negative binomial model also defines the individual development factors to be normally distributed. It is possible to create confidence intervals using the normal approximation, while Mack's model requires further assumptions to do this.

The two models used in the analysis both assume a symmetrical distribution around the mean. The two models were chosen because of their capability of handling negative incremental claims, and not because it is assumed that the claims indeed are symmetrically distribute around the mean. This has not been explored in this thesis.

The chain-ladder method is a linear model, and can be viewed as linear regression when the regression line is forced through origin. It seems that in the early development years a model also including an intercept different from zero would fit the model even better. The result of forcing the regression line through origin is that claims in the early development years are underestimated, and claims in the late development years are overestimated. This trend was also apparent when examining the residuals, and it was more dominant for the number of claims than for the amount of claims. The variance assumption in Mack's model can be viewed as a choice of an unbiased estimator carrying minimal variance (Mack 1994a). This assumption was tested by comparing it with other development factors which need other variance assumptions to attain unbiased estimators with minimal variance. Neither of the least square estimator or the simple average estimator proved any more random behaviour in the residual plots.

Estimates of the development factor, reserve estimates and prediction error were calculated using the normal approximation and Mack's model. The first moment of the two models was identical. The second moment had a different letter giving the variance component. The results show very similar estimates for the two variance components, but they are not identical. The variance component g_j in Mack's model was proved to be unbiased, and it can be viewed as an unbiased, weighted average of the residuals (Mack 1994a). The variance component ϕ_j in the normal approximation was found by fitting a generalized linear model to the squared residuals. It is simply an average of the residuals, but it is not unbiased. The variance factor g_j would have been biased as well if the factor $\frac{1}{n-j}$ had not subtracted the estimated parameters in the denominator.

The reserve and the prediction error in the two models were found algebraically. The reserve and prediction error estimates have also been obtained empirically using both models. The reserve estimates grow larger for higher accident years, since there are a growing number of undeveloped years. Naturally the prediction errors also grow larger for higher accident years. The empirical results are almost identical for the two models, and the difference can be assumed to be a cause of two different variance factors. To obtain exactly the same result, an unbiased version of ϕ_j must be used. Because of the similarities between the two models the normal approximation to the negative binomial model can be assumed to underlie Mack's model. The normal approximation to the negative binomial model uses a generalized linear model in the estimation, and this approach offers more flexibility in the analysis than Mack's model. A generalized linear model could have been fitted to the Poisson and the negative model if it had not been for the presence of negative incremental claims.

The variance factor of the last development year needed to be approximated. Mack's model was used twice using two different approximations. The difference was small, but this could have been tested further by using other approximations and a greater difference might have appeared.

There are some weaknesses in the chain-ladder method. The estimators of the last development factors are calculated using only a small number of observations. Furthermore, the last accident years require predictions of many development years ahead. This makes the ultimate claim prediction uncertain, and this is evident in the prediction error.

The data set from TrygVesta showed a large number of negative incremental claims in development year two. This could be a consequence of a large number of reported claims in the development year ahead, and would indicate that the individual development factors are correlated. Mack's model implies uncorrelated individual development factors, and the normal approximation assumes independent individual development factors. If this is not the case, the models are not appropriate for the data set.

It seems that the models have detected that negative incremental claims will occur after the first development year for accident year 19, since the reserve is smaller than for accident year 18. More empirical research should be done to reveal whether the individual development factors between development years truly are uncorrelated when there is a large frequency of negative incremental claims.

5. Conclusions

This thesis has showed theoretically that the multiplicative model, the Poisson model, the negative binomial model and Mack's model produce equivalent results to the chain-ladder method. A normal approximation to the negative model and Mack's model are two possible models when there are negative incremental claims in a data set. The two models create almost identical results, and the normal approximation can be seen as an underlying model of Mack's model.

6. References

England, P. D. & Verral, R. J. Stochastic Claims Reserving in General Insurance. Presented to the Institute of Actuaries 28.january 2002 .

Hess, K. Th. & Schmidt, K. D. 2002, "A comparison of models for the chain-ladder method", Insurance: Mathematics and Economics, vol. 31, pp. 351-364.

Mack, T. 1994a, *Measuring the variability of chain-ladder reserve estimates*, Spring Forum.

Mack, T. 1994b, "Which stochastic model is underlying the chain-ladder method?", Insurance: Mathematics and Economics, vol. 15, pp. 133-138.

Neuhaus, W. 2006, Estimating Outstanding Claims in General Insurance. (Unpublished Work)

Renshaw, A. E. & Verrall, R. J. 1998, "A stochastic model underlying the chain-ladder technique", British Actuarial Journal, vol. 4, no. 4, pp. 903-923.

Verral, R. J. An investigation into stochastic claims reserving models and the chain-ladder technique. Insurance: Mathematics and Economics 26, 91-99. 2000.

Verrall, R. J. & England, P. D. Stochastic Claims Reserving in General Insurance. Presented to the Institute of Actuaries. 2002.

Appendix I

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	-	522	535	536	538	544	547	549	554	558	564	567	566	567	567	570	571	573	575
2	663	596	595	592	607	610	618	625	633	634	641	643	646	647	648	649	649	650	
3	691	690	710	749	765	775	788	804	818	826	829	833	840	838	839	841	846		
4	757	787	831	855	880	899	908	929	939	945	953	958	959	960	962	961			
5	648	756	753	795	815	833	848	859	864	881	881	884	888	887	888				
6	823	771	767	803	825	852	855	859	871	871	874	878	881	883					
7	913	801	819	871	920	953	961	974	986	994	1001	1001	1002						
8	929	897	915	954	987	1000	1009	1012	1019	1027	1031	1034							
9	1136	1147	1156	1207	1231	1248	1264	1272	1282	1291	1297								
10	1387	1298	1267	1289	1301	1322	1338	1353	1359	1369									
11	1448	1374	1377	1403	1425	1447	1460	1473	1478										
12	1467	1351	1338	1352	1386	1409	1426	1438											
13	1345	1275	1301	1337	1360	1382	1396												
14	1545	1402	1414	1427	1457	1472													
15	1631	1452	1440	1466	1486														
16	1612	1473	1497	1532															
17	1648	1537	1535																
18	1511	1474																	
19	1431																		

Table 14. Run-off triangle of the number of claims from auto liability insurance from TrygVesta. The rows display the accident years (AY), and the columns display the development years (DY).

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	0,000	5,200	10,600	17,300	21,133	27,125	29,825	30,604	30,944	31,896	33,383	33,967	34,067	35,008	35,359	36,644	36,642	37,392	37,393
2	0,700	6,000	17,100	21,557	27,980	30,980	34,236	38,795	43,551	47,320	48,807	50,881	51,496	52,805	52,646	52,695	52,695	52,699	
3	2,000	7,700	13,654	26,334	38,734	48,492	52,685	57,677	61,664	62,646	64,204	66,754	68,493	69,570	70,418	70,217	70,462		
4	2,500	9,458	25,658	39,001	50,045	63,106	71,721	79,218	85,239	89,252	90,710	92,559	93,815	96,383	97,981	98,052			
5	2,688	11,958	25,241	38,562	53,836	70,607	83,326	92,246	100,784	104,161	107,791	111,542	114,025	116,686	117,068				
6	2,560	9,448	21,037	38,454	49,875	67,396	76,950	81,942	87,534	90,925	96,038	96,957	98,121	100,063					
7	2,795	10,220	27,252	54,768	81,070	100,423	115,359	118,617	121,727	126,110	132,068	131,688	131,801						
8	3,778	11,072	28,103	51,360	73,364	95,403	108,446	117,982	121,342	124,977	132,322	136,538							
9	5,083	15,237	35,488	65,717	92,923	106,712	116,284	132,801	146,637	149,505	151,035								
10	3,996	12,764	28,641	53,651	75,726	90,144	109,356	115,232	120,466	122,910									
11	5,451	17,508	39,835	67,750	89,010	112,132	123,015	134,509	142,955										
12	5,146	17,896	33,035	49,087	68,793	84,154	107,485	114,596											
13	8,264	19,681	29,160	44,865	63,715	80,831	93,745												
14	8,569	23,860	45,769	69,704	101,389	122,228													
15	7,088	21,831	33,254	55,304	75,132														
16	9,580	26,534	45,800	70,765															
17	9,172	32,395	48,358																
18	6,393	22,201																	
19	6,423																		

Table 15. Run-off triangle of the amount of claims from auto liability insurance from TrygVesta. The rows display the accident years (AY), and the columns display the development years (DY).

Appendix II

Show that L_C, the maximum likelihood function of C_{ij} conditioned upon $d_{i,n-i+1}$, is multinomial distributed with parameter $\frac{y_j}{s_{n-i+1}}$.

We had from formula (2.14) that

$$L_C = \prod_{i=1}^{n} \left[\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!} \prod_{j=1}^{n} \left(\frac{y_j}{s_{n-i+1}} \right)^{c_{ij}} \right] = \prod_{i=1}^{n} \left[\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!} \prod_{j=1}^{n} \left(y_{(i)j} \right)^{c_{ij}} \right]$$

where $y_{(i)j}$ is the probability for a claim that incurred in year i, will be reported in year j.

Let C_{ij}, for *i = 1,...,n* and *j = 1,....n-i+1* be independent Poisson random variables, with expectation $y_{i(j)} = \frac{y_j}{\sum_{k=1}^{n-i+1} y_k}$. Since this is a parameter, we give this term new letters, just to make it look more familiar $y_{i(j)} = p_{i(j)} = \frac{p_j}{\sum_{k=1}^{n-i+1} p_k}$.

We then have that D_{ij}, for *i =1,...n* and *j = 1,....,n-i+1* are independent Poisson random variables, with expectation $(p_{i(1)} + p_{i(2)} + ... + p_{i(n-i+1)})$. This is a result of D_{ij} being a sum of Poisson random variables.

The conditional distribution is as follows

$$f_{C_{ij}|D_{i,n-i+1}}\left(c_{ij}\middle|d_{i,n-i+1}\right) = \left(\prod_{j=1}^{n-i+1}\frac{p_j^{c_{ij}}e^{-p_j}}{c_{ij}\,!}\right) \Big/ \left(\frac{\left(p_1 + ... + p_{n-i+1}\right)^{d_{i,n-i+1}} e^{-(p_1+..+p_{n-i+1})}}{p_{i,n-i+1}\,!}\right)$$

$$= \frac{d_{i,n-i+1}\,!}{\prod_{j=1}^{n-i+1} c_{ij}\,!}\left(\frac{p_1}{\sum_{j=1}^{n-i+1} p_i}\right)^{c_{i1}} \cdots\cdots\cdots \left(\frac{p_{n-i+1}}{\sum_{j=1}^{n-i+1} p_i}\right)^{c_{i,n-i+1}}$$

$$= \frac{d_{i,n-i+1}\,!}{\prod_{j=1}^{n-i+1} c_{ij}\,!} p_{i(1)} p_{i(2)} \cdots p_{i(n-i+1)}$$

The last expression we now recognize as the multinomial distribution for C_{ij} conditioned on $d_{i,n-i+1}$.

Appendix III

Dev. Year	Estimate	St. Error	t-value
1	0,939	0,011	82,24
2	1,005	0,004	250,6
3	1,025	0,004	254,2
4	1,021	0,003	384,1
5	1,017	0,002	556,8
6	1,011	0,001	1154
7	1,010	0,002	655,6
8	1,008	0,001	773,7
9	1,008	0,001	700,4
10	1,005	0,001	978,7
11	1,003	0,001	1499
12	1,003	0,001	881,6
13	1,000	0,001	1306
14	1,001	0	3309
15	1,001	0,001	803,6
16	1,003	0,002	532,9

Table 16. Estimates, standard errors and t-values of the parameter for the restricted linear model for the total amount of claims.

Dev. Year	Parameter	Estimate	St. Error	t-value
1	Beta 0	97,727	42,424	2,304
1	Beta 1	0,864	0,034	25,293
2	Beta 0	22,412	14,179	1,581
2	Beta 1	0,986	0,013	77,776
3	Beta 0	23,791	14,077	1,69
3	Beta 1	1,004	0,013	78,02
4	Beta 0	13,346	23,791	1,365
4	Beta 1	1,009	0,013	112,328
5	Beta0	5,894	9,779	0,835
5	Beta 1	1,011	0,009	154,749
6	Beta0	-0,115	7,056	-0,033
6	Beta 1	1,011	0,007	308,659
7	Beta 0	4,167	5,868	0,71
7	Beta 1	1,006	0,005	178,95
8	Beta 0	10,442	3,702	2,821
8	Beta 1	0,998	0,004	272,847
9	Beta 0	-0,571	5,921	-0,096
9	Beta 1	1,008	0,006	163,894
10	Beta 0	4,888	4,112	1,189
10	Beta 1	1,000	0,005	222,491
11	Beta 0	3,154	3,193	0,988
11	Beta 1	1,000	0,004	169,611
12	Beta 0	-0,111	5,865	-0,019
12	Beta 1	1,003	0,007	143,126
13	Beta 0	1,866	3,856	0,484
13	Beta 1	0,998	0,005	209,414
14	Beta 0	-1,773	1,070	-1,657
14	Beta 1	1,003	0,001	744,479
15	Beta 0	6,736	3,330	20,023
15	Beta 1	0,993	0,004	229,461
16	Beta 0	-9,541	6,050	-1,577
16	Beta 1	1,007	0,009	116,967

Table 17. Estimates, standard errors and t-values of the parameter for the general model for the number of claims.

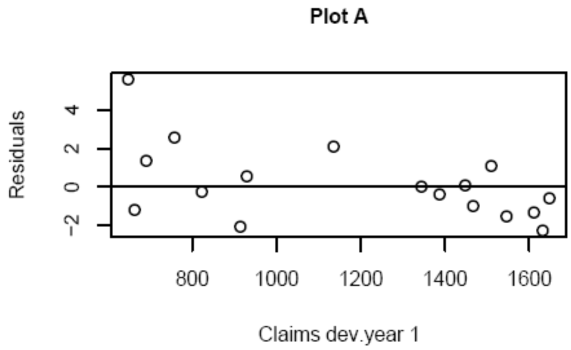
Dev. Year	Estimate	St. Error	t-value
1	3,074	0,115	26,828
2	1,837	0,090	20,322
3	1,652	0,043	38,511
4	1,395	0,016	86,281
5	1,233	0,015	82,725
6	1,148	0,016	73,027
7	1,083	0,010	104,990
8	1,062	0,009	116,790
9	1,030	0,004	258,571
10	1,035	0,007	153,678
11	1,020	0,006	170,537
12	1,012	0,004	277,756
13	1,023	0,002	613,585
14	1,008	0,003	302,115
15	1,002	0,005	186,429
16	1,002	0,001	810,478

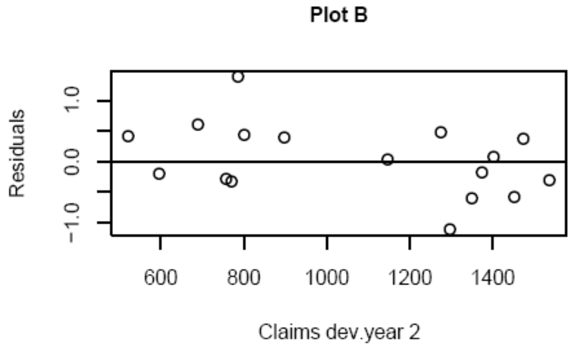
Table 18. Estimates, standard errors and t-values of the parameter for the restricted linear model for the total amount of claims.

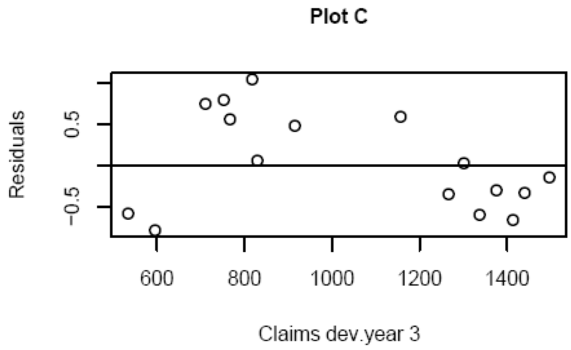
Dev. Year	Parameter	Estimate	St. Error	t-value
1	Beta 0	3,268	1,174	2,783
1	Beta 1	2,568	0,206	12,477
2	Beta 0	9,949	2,525	3,940
2	Beta 1	1,310	0,149	8,786
3	Beta 0	2,849	4,074	0,699
3	Beta 1	1,563	0,134	11,658
4	Beta 0	-2,822	2,436	-1,158
4	Beta 1	1,449	0,050	29,031
5	Beta0	3,189	2,884	1,106
5	Beta 1	1,189	0,043	27,850
6	Beta0	0,299	3,968	0,075
6	Beta 1	1,144	0,050	22,964
7	Beta 0	0,450	2,861	0,157
7	Beta 1	1,078	0,031	34,409
8	Beta 0	0,706	2,591	0,273
8	Beta 1	1,055	0,027	39,665
9	Beta 0	1,673	0,992	1,687
9	Beta 1	1,014	0,010	100,891
10	Beta 0	0,491	1,907	0,258
10	Beta 1	1,030	0,019	53,463
11	Beta 0	1,196	1,611	0,743
11	Beta 1	1,008	0,017	59,245
12	Beta 0	0,507	0,971	0,522
12	Beta 1	1,007	0,011	92,721
13	Beta 0	0,052	0,444	0,117
13	Beta 1	1,022	0,005	187,784
14	Beta 0	-0,045	0,833	-0,053
14	Beta 1	1,009	0,010	96,506
15	Beta 0	1,389	0,883	1,572
15	Beta 1	0,983	0,013	75,856
16	Beta 0	-0,316	0,219	-1,446
16	Beta 1	1,007	0,004	252,931

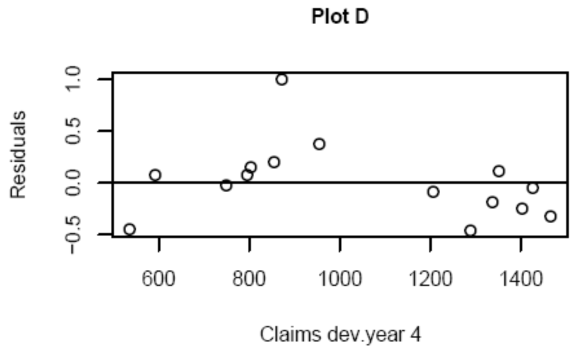
Table 19. Estimates, standard errors and t-values of the parameter for the restricted linear model for the number of claims.

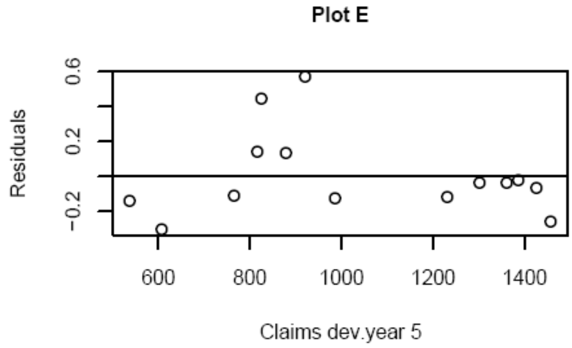
Appendix IV

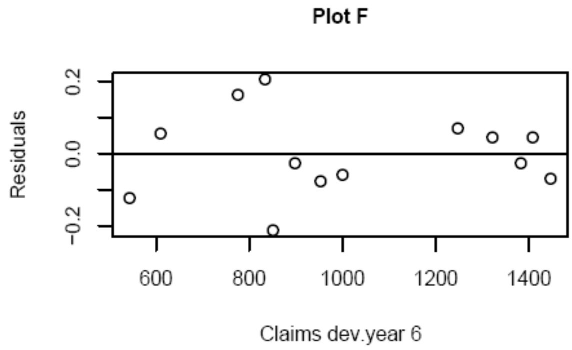

Plot A
Residuals
−2 0 2 4
800 1000 1200 1400 1600
Claims dev.year 1

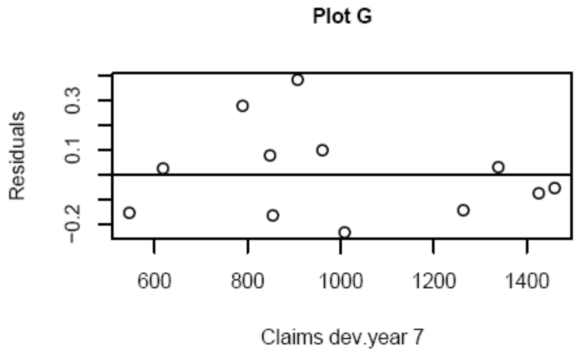

Plot B
Residuals
−1.0 0.0 1.0
600 800 1000 1200 1400
Claims dev.year 2

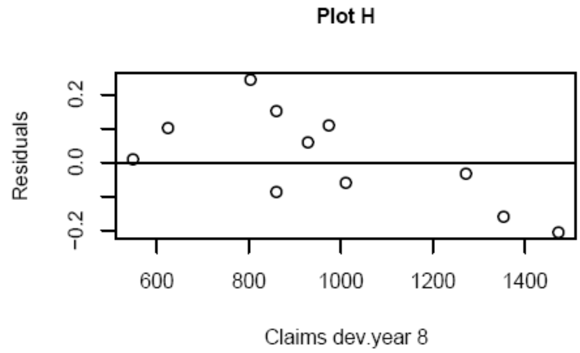

Plot C
Residuals
−0.5 0.5
600 800 1000 1200 1400
Claims dev.year 3

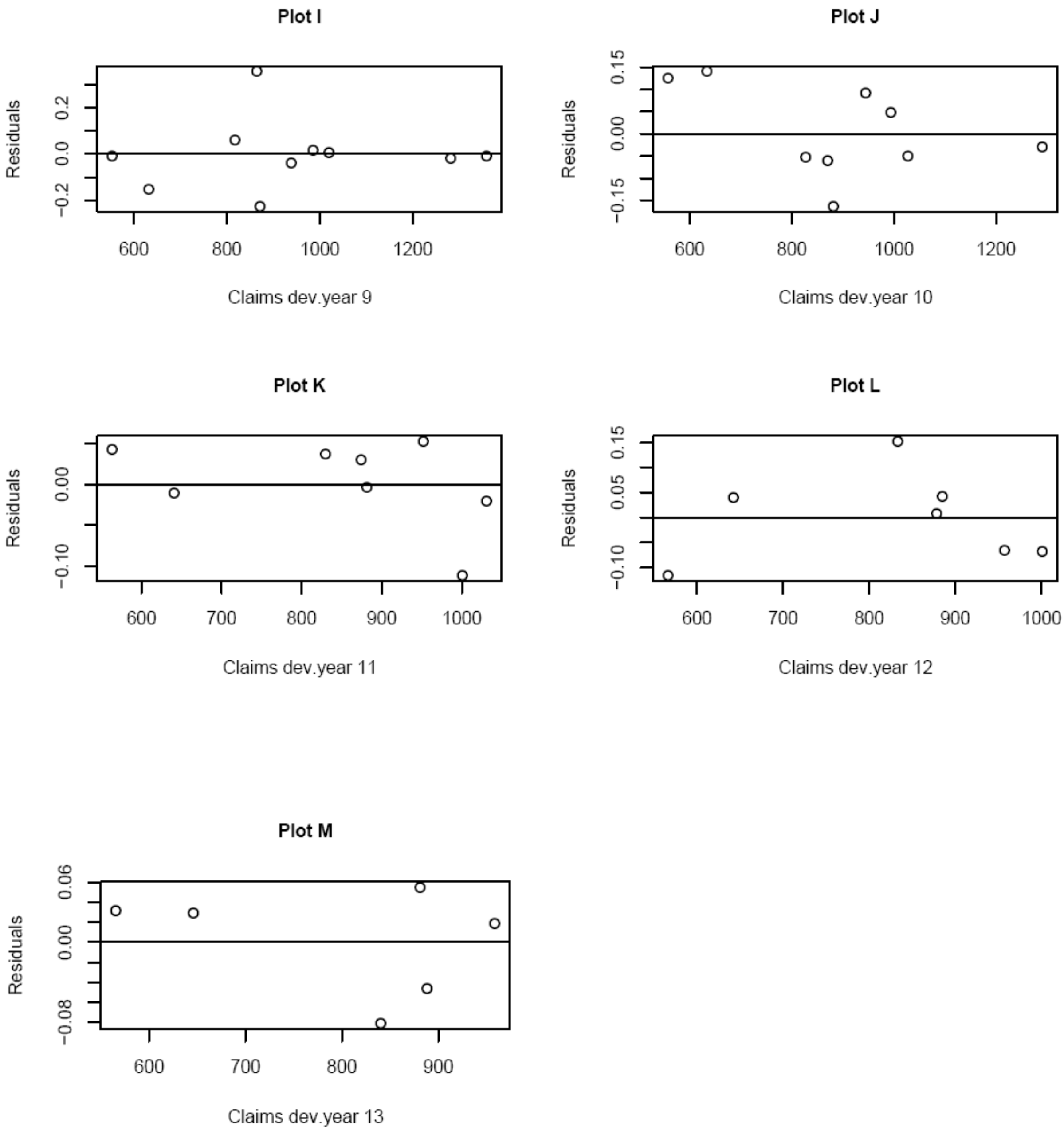

Plot D
Residuals
−0.5 0.0 0.5 1.0
600 800 1000 1200 1400
Claims dev.year 4


Plot E
Residuals
−0.2 0.2 0.6
600 800 1000 1200 1400
Claims dev.year 5

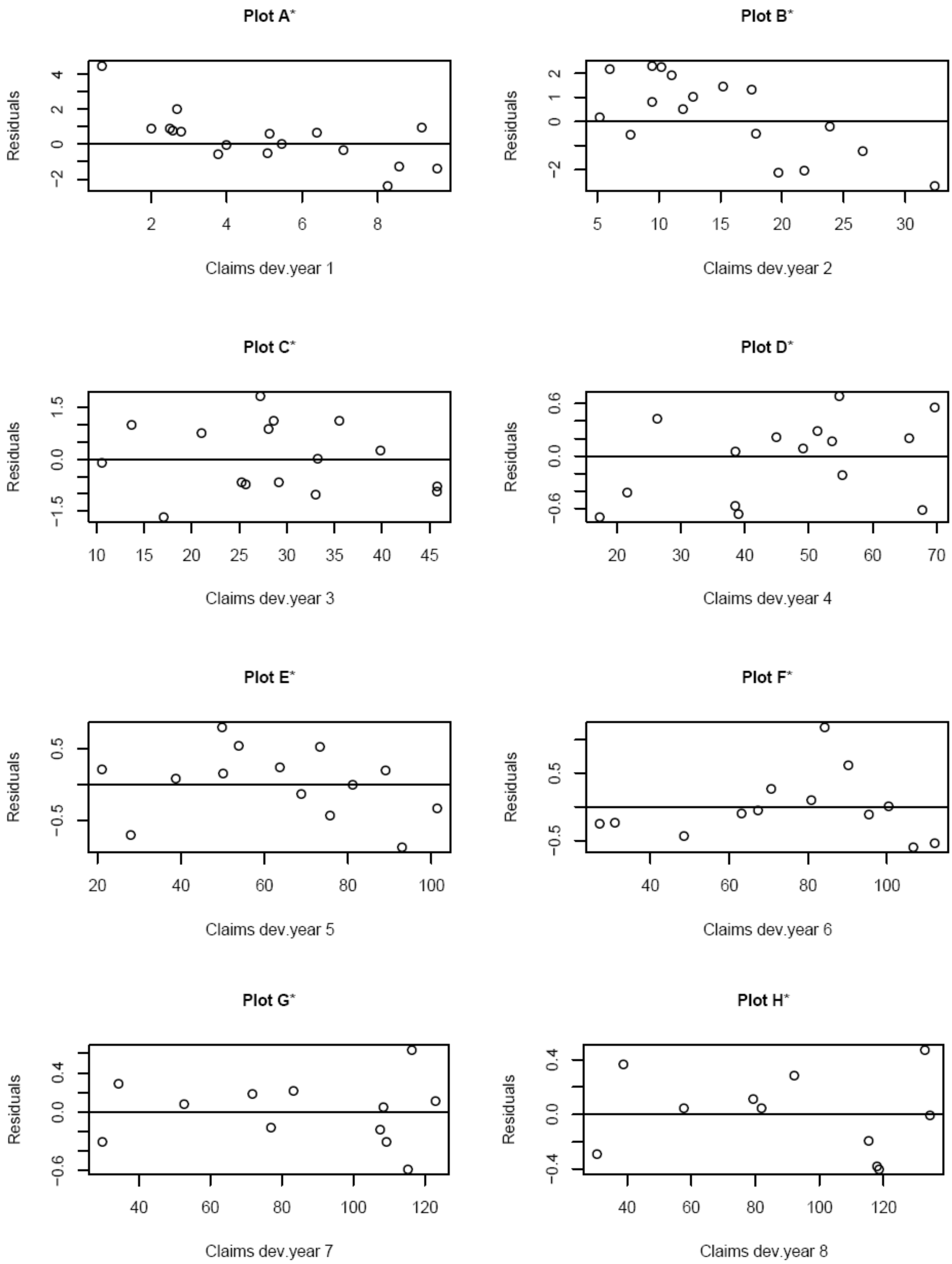

Plot F
Residuals
−0.2 0.0 0.2
600 800 1000 1200 1400
Claims dev.year 6

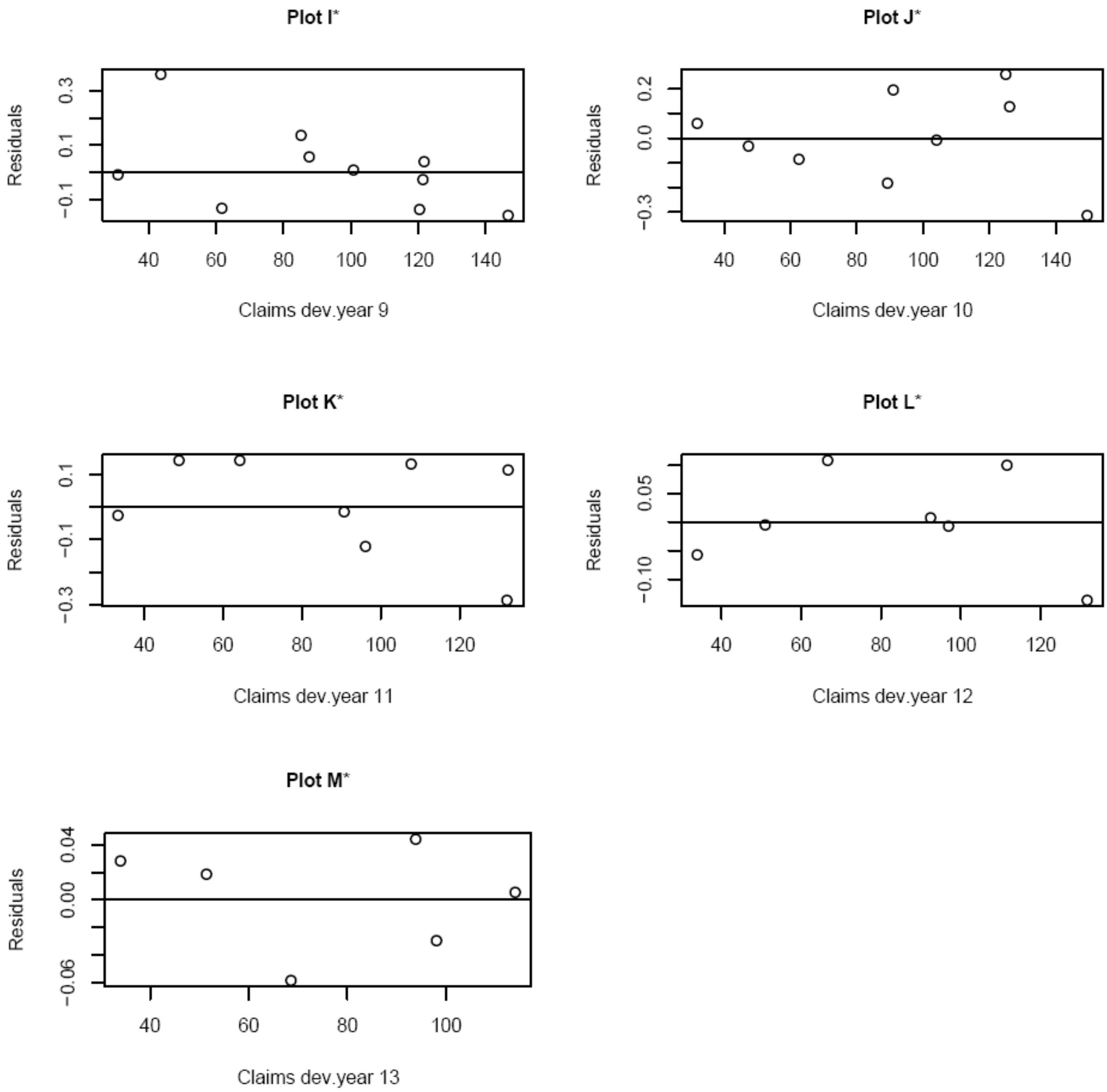

Plot G
Residuals
−0.2 0.1 0.3
600 800 1000 1200 1400
Claims dev.year 7


Plot H
Residuals
−0.2 0.0 0.2
600 800 1000 1200 1400
Claims dev.year 8



Plot A to M. Residuals, r_{ij} plotted against the the claims of development year j-1, $d_{i,j-1}$ for the number of claims. Plot A to M display the residual plots for development year 2 to development year 14, as a function of the claims in the previous development year.


Plot A*
Residuals
Claims dev.year 1
Plot B*
Residuals
Claims dev.year 2
Plot C*
Residuals
Claims dev.year 3
Plot D*
Residuals
Claims dev.year 4
Plot E*
Residuals
Claims dev.year 5
Plot F*
Residuals
Claims dev.year 6
Plot G*
Residuals
Claims dev.year 7
Plot H*
Residuals
Claims dev.year 8



Plot A* to M*. Residuals, r_{ij} plotted against the claims of development year j-1, $d_{i,j-1}$ for the amount of claims. Plot A* to M* display the residual plots for development year 2 to development year 14, as a function of the claims in the previous development year.

Appendix V

Let X and Y be independent random variables. The formula to be proven is:

$$Var(XY) = \left(E[X]^2\right)Var(Y) + E\left([Y]^2\right)Var(X) + Var(Y)Var(X)$$

Let $\delta X = \frac{X - EX}{EX}$ and $\delta Y = \frac{Y - EY}{EY}$

It can then be seen that:

$$XY - EXEY = EXEY\left([\delta X + 1][\delta Y + 1] - 1\right) = EXEY\left(\delta X + \delta Y + \delta X \, \delta Y\right)$$

The first relation is easily proved by using the definitions for δX and δY. To find the $Var(XY)$ it is possible to use the well known identity, which for the variable X is $Var(X) = E(X - EX)^2$. Thus

$$Var(XY) = E\left([XY - EXEY]^2\right) = E\left([EXEY]^2[\delta X + \delta Y + \delta X \, \delta Y]^2\right)$$

$$= [EXEY]^2 \, E\left[\frac{X - EX}{EX} + \frac{Y - EY}{EY} + \frac{(X - EX)(Y - EY)}{EXEY}\right]^2$$

$$= [EXEY]^2 \, E\left[\left(\frac{X - EX}{EX}\right)^2 + \left(\frac{Y - EY}{EY}\right)^2 + \left(\frac{(X - EX)(Y - EY)}{EXEY}\right)^2\right.$$

$$+2\left(\frac{X-EX}{EX}\right)\left(\frac{Y-EY}{EY}\right)+2\left(\frac{X-EX}{EX}\right)^2\left(\frac{Y-EY}{EY}\right)+2\left(\frac{Y-EY}{EY}\right)^2\left(\frac{X-EX}{EX}\right)\Bigg]$$

The three last elements disappear because of independence between X and Y, and the expression becomes:

$$Var(XY)=E\left([XY-EXEY]^2\right)=(EX)^2 Var(Y)+(EY)^2 Var(X)+Var(X)Var(Y)$$

(Goodman 1960)

Appendix VI

This proof is reproduced from Mack (1994a).

Given Mack's three assumptions

1. There exist constants $f_2, f_3, ..., f_n$ such that $E\left(D_{ij} \middle| k_{i,j-1}\right) = f_j d_{i,j-1}$ for $j = 2,...n$

2. There exists constants $g_2, g_3, ..., g_n$ such that $Var\left(D_{ij} \middle| k_{i,j-1}\right) = g_j d_{i,j-1}$ for $j = 2,....n$

3. K_{in} and K_{kn} are stochastically independent for $i \neq k$.

The MSE of the reserve is

$$MSE\left(\hat{R}_i \middle| k\right) = \ = \hat{D}_{in}^2 \sum_{j=n-i+2}^{n} \frac{\hat{g}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{D}_{i,j-1}} + \frac{1}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \right) \quad (A.1)$$

Proof:

The MSE of the reserve is the same as the MSE of the ultimate claim $\hat{D}_{in}$, this result is proved in (3.17) and (3.18) .

$$MSE\left(\hat{R}_i \middle| k\right) = MSE\left(\hat{D}_{in} \middle| k\right) \quad (A.2)$$

The aim is to find the MSE of the reserve, it is equivalent to find the MSE of the ultimate claim $\hat{D}_{in}$ which will be done here. The prediction variance of the ultimate claim was in (2.60) found to be

$$MSE\left(\hat{D}_{in}\left|k\right.\right)=Var\left(D_{in}\left|k\right.\right)+\left(E\left(D_{in}\left|k\right.\right)-\hat{D}_{in}\right)^{2} \qquad \text{(A.3)}$$

By Mack's third assumption it suffices to condition on the observations within the accident year i. It is only necessary to find $Var\left(D_{in}\left|k_{i,n-i+1}\right.\right)$ and $\left(E\left(D_{in}\left|k_{i,n-i+1}\right.\right)-\hat{D}_{in}\right)^{2}$.

The term $Var\left(D_{in}\left|k_{i,n-i+1}\right.\right)$ will be considered first. This can be written as:

$$Var\left(D_{in}\left|k_{i,n-i+1}\right.\right)=E\left(\left(D_{in}\right)^{2}\left|k_{i,n-i+1}\right.\right)-\left(E\left(D_{in}\left|k_{i,n-i+1}\right.\right)\right)^{2} \qquad \text{(A.4)}$$

The second term of (A.4) will be determined first. By using the rule of expectation the expected values of the claims in the south-eastern corner of the run-off triangle ($j\geq n-i+1$) are determined. The expected claims of the two last development years in accident year i are

$$E\left(D_{in}\left|k_{i,n-i+1}\right.\right)=E\left(E\left(D_{in}\left|D_{i,n-1}\right.\right)\left|k_{i,n-i+1}\right.\right)=f_{n}E\left(D_{i,n-1}\left|k_{i,n-i+1}\right.\right) \quad \text{and}$$

$$E\left(D_{i,n-1}\left|k_{i,n-i+1}\right.\right)=E\left(E\left(D_{i,n-1}\left|D_{i,n-2}\right.\right)\left|k_{i,n-i+1}\right.\right)=f_{n-1}E\left(D_{i,n-2}\left|k_{i,n-i+1}\right.\right) \qquad \text{A.5)}$$

Inserting the first expression derived in (A.5) in the next a new expression is obtained

$$E\left(D_{in}\left|k_{i,n-i+1}\right.\right)=f_{n-1}f_nE\left(D_{i,n-2}\left|k_{i,n-i+1}\right.\right)$$

Performing this step several times makes it possible to find $E\left(D_{in}\left|k_{i,n-i+1}\right.\right)$ or the expectation of any future claim in the south-eastern corner of the run-off triangle. Generally for accident year $i=2,...,n$ and development year $j\geq n-i+2$ the formula is:

$$E\left(D_{ij}\left|k_{i,n-i+1}\right.\right)=f_{n-i+2}...f_{j-1}f_jE\left(D_{i,n-i+1}\left|k_{i,n-i+1}\right.\right)=f_{n-i+2}...f_{j-1}f_jd_{i,n-i+1} \qquad \text{(A.6)}$$

The general formula for $E\left(D_{ij}\left|k_{i,n-i+1}\right.\right)$ is recovered, and it is trivial to determine $E\left(D_{in}\left|k_{i,n-i+1}\right.\right)$ and $\left(E\left(D_{in}\left|k_{i,n-i+1}\right.\right)\right)^2$. The second part of (A.4) is determined.

The next step is to find $E\left(\left(D_{in}\right)^2\left|k_{i,n-i+1}\right.\right)$. To recover $E\left(\left(D_{in}\right)^2\left|k_{i,n-i+1}\right.\right)$ the identity $E\left(\left(D_{ij}\right)^2\left|k_{i,n-i+1}\right.\right)$ will be calculated first by using the rule of double expectation and Mack's two first assumptions.

$$E\left(D_{ij}^2\left|k_{i,n-i+1}\right.\right)=E\left(E\left(D_{ij}^2\left|D_{i,j-1}\right.\right)\left|k_{i,n-i+1}\right.\right) \qquad \text{(A.7)}$$

$$= E\left\{\left(Var\left(D_{ij} \middle| D_{i,j-1}\right)\right) + \left(E\left(D_{ij} \middle| D_{i,j-1}\right)\right)^2 \middle| k_{i,n-i+1}\right\}$$

$$= E\left\{g_j D_{i,j-1} + \left(f_j D_{i,j-1}\right)^2 \middle| k_{i,n-i+1}\right\}$$

$$= g_j E\left(D_{i,j-1} \middle| k_{i,n-i+1}\right) + f_j^2 E\left(D_{i,j-1}^2 \middle| k_{i,n-i+1}\right)$$

$E\left(D_{ij}^2 \middle| k_{i,n-i+1}\right)$ can be determined by using (A.7) and (A.5). The calculations can be seen below, and the formulas used are written on the right hand side.

$$E\left(D_{in}^2 \middle| k_{i,n-i+1}\right) = g_n E\left(D_{i,n-1} \middle| k_{i,n-i+1}\right) + f_n^2 E\left(D_{i,n-1}^2 \middle| k_{i,n-i+1}\right) \quad \text{(A.7)}$$

$$= g_n \left(f_{n-i+2} f_{n-i+3} \cdots f_{n-1} d_{i,n-i+1}\right) + \quad \text{(A.5)}$$

$$= g_{n-1} E\left(D_{i,n-2} \middle| k_{i,n-i+1}\right) f_n^2 + E\left(D_{i,n-2}^2 \middle| k_{i,n-i+1}\right) f_{n-1}^2 f_n^2$$

(A.7)

$$= g_n \left(f_{n-i+2} f_{n-i+3} \cdots f_{n-1} d_{i,n-i+1}\right) + g_{n-1} d_{i,n-i+1} f_{n-i+2} f_{n-i+3} \cdots f_{n-2} f_n^2 \quad \text{(A.5)}$$

$$+ g_{n-2} E\left(D_{i,n-3} \middle| k_{i,n-i+1}\right) f_{n-1}^2 f_n^2 + E\left(D_{i,n-3}^2 \middle| k_{i,n-i+1}\right) f_{n-2}^2 f_{n-1}^2 f_n^2 \quad \text{(A.7)}$$

$$= g_n \left(f_{n-i+2} f_{n-i+3} \cdots f_{n-1} d_{i,n-i+1}\right) + g_{n-1} d_{i,n-i+1} f_{n-i+2} f_{n-i+3} \cdots f_{n-2} f_n^2$$

$$+ g_{n-2} d_{i,n-i+1} f_{n-i+2} f_{n-i+3} \cdots f_{n-3} f_{n-1}^2 f_n^2 \quad \text{(A.5)}$$

$$+ g_{n-3} E\left(D_{i,n-4} \middle| k_{i,n-i+1}\right) f_{n-2}^2 f_{n-1}^2 f_n^2 \quad \text{(A.7)}$$

$$+ E\left(D_{i,n-4}^2 \middle| k_{i,n-i+1}\right) f_{n-3}^2 f_{n-2}^2 f_{n-1}^2 f_n^2$$

etc.

The results from (A.5) and (A.7) are used until the last step when it is clear that $E\left(D_{i,n-i+1}^2 \middle| k_{i,n-i+1}\right) = d_{i,n-i+1}^2$ since $k_{i,n-i+1}$ is known. This can be written as:

$$E\left(D_{in}^2 \middle| k_{i,n-i+1}\right) = d_{i,n-i+1} \sum_{j=n-i+2}^{n} \left(f_{n-i+2} f_{n-i+3} \cdots f_{j-1} g_j f_{j+1}^2 \cdots f_n^2 \right) + d_{i,n-i+1}^2 f_{n-i+2}^2 \cdots f_n^2 \qquad \text{(A.8)}$$

We have established estimators for $E\left(D_{in}^2 \middle| k_{i,n-i+1}\right)$ and $E\left(D_{in} \middle| k_{i,n-i+1}\right)$, and (A.4) can be written like

$$Var\left(D_{in} \middle| k_{i,n-i+1}\right) = E\left(D_{in}^2 \middle| k_{i,n-i+1}\right) - \left(E\left(D_{in} \middle| k_{i,n-i+1}\right)\right)^2$$

$$= d_{i,n-i+1} \sum_{j=n-i+2}^{n} \left(f_{n-i+2} \cdots f_{j-1} g_j f_{j+1}^2 \cdots f_n^2 \right)$$
$$+ d_{i,n-i+1}^2 f_{n-i+2}^2 \cdots f_n^2 - \left(d_{i,n-i+1} f_{n-i+2} \cdots f_n \right)^2$$

$$= d_{i,n-i+1} \sum_{j=n-i+2}^{n} f_{n-i+2} \cdots f_{j-1} g_j f_{j+1}^2 \cdots f_n^2 \qquad \text{(A.9)}$$

By rewriting (A.9) first and then replacing the parameters g_j and f_j with $\hat{g}_j$ and $\hat{f}_j$ the process variance is:

$$Var\left(D_{in} \middle| k_{i,n-i+1}\right) = d_{i,n-i+1}^2 \hat{f}_{n-i+2}^2 \hat{f}_{n-i+3}^2 \cdots \hat{f}_n^2 \sum_{j=n-i+2}^{n} \frac{f_{n-i+2} \cdots f_{j-1} g_j f_{j+1}^2 \cdots f_n^2}{d_{i,n-i+1} \hat{f}_{n-i+2}^2 \hat{f}_{n-i+3}^2 \cdots \hat{f}_n^2}$$

$$= d_{i,n-i+1}^2 \hat{f}_{n-i+2}^2 \hat{f}_{n-i+3}^2 \cdots \hat{f}_n^2 \sum_{j=n-i+2}^{n} \frac{\hat{g}_j}{d_{i,n-i+1} \hat{f}_{n-i+2} \hat{f}_{n-i+3} \cdots \hat{f}_{j-1} \hat{f}_j^2}$$

$$= \hat{D}_{in}^2 \sum_{j=n-i+2}^{n} \frac{\left(\frac{\hat{g}_j}{\hat{f}_j^2}\right)}{\hat{D}_{i,j-1}} \qquad \text{(A.10)}$$

To reach the last expression, (A.10), we have used the fact that $\hat{D}_{in} = D_{i,j-1} \hat{f}_j \ldots \hat{f}_n$, when $D_{i,j-1}$ is estimated, and $\hat{D}_{in} = d_{i,n-i+1} \hat{f}_{n-i+2} \ldots \hat{f}_n$ where $d_{i,n-i+1}$ is considered known.

The first part of (A.2) has been found, and the next we are interested in finding $\left(E\left(D_{in} \middle| k_{i,n-i+1}\right) - \hat{D}_{in}\right)^2$. By using the formula found in (A.6) we have that

$$\left(E\left(D_{in} \middle| k_{i,n-i+1}\right) - \hat{D}_{in}\right)^2 = \left(\left(d_{i,n-i+1} f_{n-i+2} \ldots f_n\right) - \left(d_{i,n-i+1} \hat{f}_{n-i+2} \ldots \hat{f}_n\right)\right)^2 \qquad \text{(A.11)}$$

$$= d_{i,n-i+1}^2 \left(f_{n-i+2} \ldots f_n - \hat{f}_{n-i+2} \ldots \hat{f}_n\right)^2$$

Unlike what was done from (A.9) to (A.10), it is not a good idea to replace the parameter f_j with $\hat{f}_j$. If this had been done it is implicitly assumed that the estimator $\hat{f}_j$ actually is the same as the true value f_j, but it is more realistic that there is a difference between the estimator and the parameter. To solve this problem Mack (1994a) introduced a new identity, F. This F has nothing to do with the individual development factor. It is defined as

$$F = f_{n-i+2}...f_n - \hat{f}_{n-i+2}...\hat{f}_n = S_{n-i+2} + ... + S_n \tag{A.12}$$

where

$$S_j = \hat{f}_{n-i+2}...\hat{f}_{j-1} f_j f_{j+1}...f_n - \hat{f}_{n-i+2}...\hat{f}_{j-1}\hat{f}_j f_{j+1}...f_n \tag{A.13}$$

$$= \hat{f}_{n-i+2}...\hat{f}_{j-1}\left(f_j - \hat{f}_j\right) f_{j+1}...f_n$$

The new identity F squared can be written like:

$$F^2 = \left(S_{n-i+2} + ... + S_n\right)^2 = \sum_{j=n-i+2}^{n} S_j^2 + 2 \sum_{\substack{j,k=n-i+2 \\ k<j}}^{n} \sum_{k=n-i+2}^{n} S_k S_j \tag{A.14}$$

S_j^2 and $S_j S_k$ can be approximated with $E\left(S_j^2 \middle| k_{j-1}\right)$ and $E\left(S_j S_k \middle| k_{j-1}\right)$. By using this approximation the observations are taken into account, this would not be the case when approximating $E\left(\hat{f}_j\right)$ to f_j. Since $\hat{f}_j$ is an unbiased estimator (see chapter 2.6), $E\left(\left(f_j - \hat{f}_j\right)^2 \middle| k_{j-1}\right) = 0$ and also $E\left(S_j S_l \middle| k_{l-1}\right) = 0$. To see this clearly the calculations are done underneath, where still *j*<*l*:

$$E\left(S_j S_l \middle| k_{l-1}\right) = E\left(\left(\hat{f}_{n-i+2}...\hat{f}_{j-1}\left(f_j - \hat{f}_j\right) f_{j+1}...f_n\right)\left(\hat{f}_{n-i+2}...\hat{f}_{l-1}\left(f_l - \hat{f}_l\right) f_{l+1}...f_n\right) \middle| k_{l-1}\right)$$

$$= \left(\hat{f}_{n-i+2} \cdots \hat{f}_{j-1}\left(f_j - \hat{f}_j\right) f_{j+1} \cdots f_n\right) E\left(\left(\hat{f}_{n-i+2} \cdots \hat{f}_{l-1}\left(f_l - \hat{f}_l\right) f_{l+1} \cdots f_n\right) \middle| k_{l-1}\right).$$
$$= 0$$

The identity $E\left(S_j^2 \middle| k_{l-1}\right)$ can be found by first examining $E\left(\left(f_j - \hat{f}_j\right)^2 \middle| k_{j-1}\right)$:

$$E\left(\left(f_j - \hat{f}_j\right)^2 \middle| k_{j-1}\right) = Var\left(\hat{f}_j \middle| k_{j-1}\right) = Var\left(\frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}} \middle| k_{j-1}\right) \tag{A.15}$$

$$= \frac{1}{\left(\sum_{i=1}^{n-j+1} d_{i,j-1}\right)^2} Var\left(\sum_{i=1}^{n-j+1} D_{ij} \middle| k_{j-1}\right)$$

$$= \frac{1}{\left(\sum_{i=1}^{n-j+1} d_{i,j-1}\right)^2} \sum_{i=1}^{n-j+1} d_{i,j-1} g_j$$

$$= \frac{g_j}{\left(\sum_{i=1}^{n-j+1} d_{i,j-1}\right)} \tag{A.16}$$

The expression in (A.16) was only a part of what is needed to find $E\left(S_j^2 \middle| k_{l-1}\right)$, but by using (A.16) it is clear that $E\left(S_j^2 \middle| k_{l-1}\right)$ can be expressed as:

$$E\left(S_j^2\left|k_{j-1}\right.\right)=\hat{f}_{n-i+2}^2\cdots\hat{f}_{j-1}^2E\left(\left(f_j-\hat{f}_j\right)^2\left|k_{j-1}\right.\right)f_{j+1}^2\cdots f_n^2=\hat{f}_{n-i+2}^2\cdots\hat{f}_{j-1}^2\frac{g_j}{\sum_{i=1}^{n-j+1}d_{i,j-1}}f_{j+1}^2\cdots f_n^2$$

$\hat{f}_l$ for *l<j* us a scalar because it is conditioned on k_{j-1}.

By replacing the parameters f_j and g_j with $\hat{f}_j$ and $\hat{g}_j$ in the expression above an estimator for $\left(E\left(D_{in}\left|k_{i,n-i+1}\right.\right)-\hat{D}_{in}\right)^2$ is derived:

$$\left(E\left(D_{in}\left|k_{i,n-i+1}\right.\right)-\hat{D}_{in}\right)^2=d_{i,n-i+1}^2F^2=d_{i,n-i+1}^2\sum_{i=n-i+2}^{n}E\left(S_j^2\left|k_{j-1}\right.\right) \tag{A.17}$$

$$=d_{i,n-i+1}^2\hat{f}_{n-i+2}^2\cdots\hat{f}_n^2\sum_{j=n-i+2}^{n}\frac{\left(\dfrac{\hat{g}_j}{\hat{f}_j^2}\right)}{\sum_{i=1}^{n-j+1}d_{i,j-1}}$$

$$=\hat{D}_{in}^2\sum_{j=n-i+2}^{n}\frac{\left(\dfrac{\hat{g}_j}{\hat{f}_j^2}\right)}{\sum_{i=1}^{n-j+1}d_{i,j-1}}$$

Finally, the estimator of the prediction variance is available. By using the formulas from (A.2), (A.10) and (A.17) one can see that the prediction variance of $\hat{D}_{in}$ is:

$$MSE\left(\hat{R}_i\right)=Var\left(D_{in}\left|K\right.\right)+\left(E\left(D_{in}\left|K\right.\right)-\hat{D}_{in}\right)^2$$

$$= \hat{D}_{in}^2 \sum_{j=n-i+2}^{n} \frac{\left(\frac{\hat{g}_j}{\hat{f}_j^2}\right)}{\hat{D}_{i,j-1}} + \hat{D}_{in}^2 \sum_{j=n-i+2}^{n} \frac{\left(\frac{\hat{g}_j}{\hat{f}_j^2}\right)}{\sum_{i=1}^{n-j+1} d_{i,j-1}}$$

$$= \hat{D}_{in}^2 \sum_{j=n-i+2}^{n} \frac{\hat{g}_j}{\hat{f}_j^2} \left(\frac{1}{\hat{D}_{i,j-1}} + \frac{1}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \right) \tag{A.18}$$

In addition to the prediction variance of $\hat{D}_{in}$ it is essential to find the prediction variance of the total reserve estimate.

The explanations are shorter when proving this formula since the same calculations have been done when finding the prediction variance of every accident year. Instead, the already established identities of formulas that are being used will be written on the right hand side of the calculation. The identity to be proven is

$$MSE\left(\hat{R} \middle| k\right) = \sum_{i=1}^{n} \left\{ MSE\left(\hat{R}_i \middle| k\right) + \hat{D}_{in} \left(\sum_{l=i+1}^{n} \hat{D}_{\ln} \right) \left(\sum_{k=n-i+2}^{n} \frac{2\hat{g}_k}{\hat{f}_j^2 \sum_{l=1}^{n-k+1} d_{lk}} \right) \right\}$$

Proof:

$$MSE\left(\hat{R} \middle| k\right) = MSE\left(\sum_{i=2}^{n} \hat{R}_i \middle| k \right) \tag{A.19}$$

$$= E\left(\left(\sum_{i=2}^{n} \hat{R}_i - \sum_{i=2}^{n} R_i \right)^2 \middle| k \right)$$

$$= E\left(\left(\sum_{i=2}^{n}\hat{D}_{in} - \sum_{i=2}^{n} D_{in}\right)^2 \middle| k\right) \tag{3.18}$$

$$= Var\left(\sum_{i=2}^{n} D_{in} \middle| k\right) + \left(E\left(\sum_{i=2}^{n} D_{in} \middle| k\right) - \sum_{i=2}^{n}\hat{D}_{in}\right)^2 \tag{2.60}$$

$$= (1) + (2)$$

The two expressions (1) and (2) will be determined separately.

$$(1) = Var\left(\sum_{i=2}^{n} D_{in} \middle| k\right) \overset{independence}{=} \sum_{i=2}^{n} Var\left(D_{in} \middle| k\right) \tag{A.10}$$

$$(2) = \left(E\left(\sum_{i=2}^{n} D_{in} \middle| k\right) - \sum_{i=2}^{n}\hat{D}_{in}\right)^2 = \left(\sum_{i=1}^{n}\left(E\left(D_{in} \middle| k\right) - \hat{D}_{in}\right)\right)^2$$

$$= \left(\sum_{i=2}^{n}\left(d_{i,n-i+1} f_{n-i+2} \dots f_n - d_{i,n-i+1}\hat{f}_{n-i+2}\dots\hat{f}_n\right)\right)^2 \tag{A.11}$$

$$= \left(\sum_{i=2}^{n} d_{i,n-i+1}\left(f_{n-i+2}\dots f_n - \hat{f}_{n-i+2}\dots\hat{f}_n\right)\right)^2$$

$$= \left(\sum_{i=2}^{n} d_{i,n-i+1} F_i\right)^2$$

$$= \sum_{i=2}^{n} d_{i,n-i+1}^2 F_i^2 + 2\sum_{i=2}^{n}\sum_{\substack{l=2 \\ i<l}}^{n} d_{i,n-i+1} d_{l,n-l+1} F_i F_l$$

The variable F was introduced in (A.12) and it was defined as $F=\left(f_{n-i+2}\cdots f_n-\hat{f}_{n-i+2}\cdots\hat{f}_n\right)$.

In this case this variable is needed for $i = 2,\ldots,n$, so a subscript i is included, the variable becomes

$$F_i=\left(f_{n-i+2}\cdots f_n-\hat{f}_{n-i+2}\cdots\hat{f}_n\right).$$

To find a simpler expression for (2) an estimator for F_iF_l needs to be determined. This is done by using the same procedure as in (A.12-A.16). The estimator for F_iF_l is

$$\sum_{q=n-i+1}^{n}\hat{f}_{n-l+2}\cdots\hat{f}_{n-l+1}\hat{f}^2_{n-l+2}\cdots\hat{f}^2_{q-1}\frac{\hat{g}_q\hat{f}^2_{q+1}\cdots\hat{f}^2_n}{\sum_{i=1}^{n-q+1}d_{i,q-1}} \qquad \text{(A.12-A.16)}$$

The two identities (1) and (2) is added and the estimator above is used. Remembering the expression of $MSE\left(\hat{R}_i|k\right)$ it is clear that:

$$MSE\left(\hat{R}|k\right)=(1)+(2)=\sum_{i=2}^{n}Var\left(D_{in}|k\right)+\sum_{i=2}^{n}d^2_{i,n-i+1}F_i^2+2\sum_{i=2}^{n}\sum_{\substack{l=2\\ i<l}}^{n}d_{i,n-i+1}d_{l,n-l+1}F_iF_l$$

$$= \sum_{i=1}^{n} \left\{ MSE\left(\hat{R}_i \middle| k\right) + \hat{D}_{in} \left(\sum_{l=i+1}^{n} \hat{D}_{\ln} \right) \left(\sum_{k=n-i+2}^{n} \frac{2\hat{g}_k}{\hat{f}_j^{\,2} \sum_{l=1}^{n-k+1} d_{lk}} \right) \right\}$$