Title of paper: A Dynamic Claims Reserving Model

Author: Christian Roholte Larsen, Larsen&Partners Ltd.
Email: Christian.Roholte@Larsen.dk

Abstract:

Traditional Chain Ladder models are based on a few cells in an upper triangle and often give inaccurate projections of the reserve. Traditional stochastic models are based on the same few summaries and in addition are based on the often unrealistic assumption of independence between the aggregate incremental values. In this paper a set of stochastic models with weaker assumptions based on the individual claims development are described. These models can include information about settlement and can handle seasonal effects, changes in mix of business and claim types as well as changes in mix of claim size. It is demonstrated how the distribution of the process can be specified and especially how the distribution of the reserve Gross and Net of Reinsurance can be determined. The method is illustrated with an example.

Keywords:

Claims reserving, Chain Ladder, Reinsurance, Marked Poisson Process, Decomposition, Markov Chain, Logistic Regression, Generalised Linear Models, Generalised Pareto Distribution, Business Mix, Bootstrap.

1. Introduction

In this paper a particular approach to claims reserving is taken where relatively complete information of individual claims is used.

In the traditional claims reserving methods the information available, often concerning thousands of claims, is reduced to very few numbers in the upper triangle. Since the traditional models assume that the past development pattern will continue in the future it is not surprising that the models are inaccurate when there are changes of business mix in the portfolio.

Traditional stochastic models provide methods for estimating the variability of the projection i.e. the process variation. However, the projection of the uncertainty is often based on the assumption that the incremental amounts in the 'upper triangle' are stochastically independent which is normally not fulfilled. Even if the assumption is fulfilled the models are often inaccurate and therefore also the estimate of the future randomness.

Projections of Reserve Net of Reinsurance become straightforward in the model presented since the model provides distribution projections per claim, including IBNR claims. Furthermore, the projections from the different layers add up to the total projection. This is not always the case when traditional 'netting down' is carried out.

The presented model projects the effect of changes in portfolio size, changes of mix of business, changes of mix of claim types, seasonal effects and changes of empirical claims distribution. The pure period inflation can be estimated and isolated from inflation caused by changes in portfolio mix.

The model deals specifically with the fact that the development of large claims often is very different from the development of other claims. In traditional reserving models based on aggregated data a few large claims have a disproportionate impact on the development factors and projections. For example if we see relatively many large claims in the most recent policy year then the projection for this year will be too high if the large claims are developed more quickly than other claims.

The model addresses the common situation where the incremental amounts concerning a specific claim are not stochastically independent and is dynamic in the sense that the distributions of future incremental amounts are stochastically dependent on the past incremental amounts.

While the model presented in this paper is based on Incurred Amounts it could equally be based on Payments.

The model can take the claims settlement into consideration and this will be discussed briefly.

Although the model is complex compared to traditional reserving models, each model component is manageable and the parameters can be estimated using traditional smoothing methods such as Generalised Pareto Models or Generalised Linear Models. These procedures are available in a number of software packages, including Larsen&Partners' solution for claims reserving.

The fundamental problem with traditional reserving models, including stochastic models, is that only the average development factors are modelled, not the distribution of the *individual* development factors. There is, therefore, a need for a reserving model providing more accurate and consistent projections for each XL layer which takes changes in business mix and claims distribution into consideration.

The method suggested includes an estimation of the multi dimensional distribution of the future incurred amounts per development period given the current incurred amount and other information concerning the claim, such as claim type and policy information. Based on this distribution the mean of the future changes, i.e. the required reserve in excess of the individual reserve, and even the distribution of this or any function of it or of the total projection can be calculated. Where a theoretical calculation in best case would be extremely complicated a projection via simulation can be obtained.

The method will provide measures of the variability of outstanding claims, in addition to simply calculating a "best estimate". This is particularly useful since the British Institute and Faculty of Actuaries define a best estimate as `the expected value of the distribution of possible outcomes of the unpaid liabilities', without providing guidance as to how such a distribution should be obtained.

The paper is structured as follows: In section 2 the basic model is formulated as a Marked Poisson process and the stochastic reserve is defined. In section 3 a dynamic application is made by assuming that the conditional distribution of the yearly incremental amount given the accumulated amounts in previous periods is only dependent on a function of these amounts. It is shown that the conditional distribution of the reserve, given the current stage of the process, can be regarded as a sum of stochastically independent variables. In section 4 an example is given of how the distributions can be specified using Generalised Linear Models and other traditional smoothing methods and in section 5 examples are outlined based on policy and claims data from a Marine portfolio. In section 6 and 7 the method of estimating the distribution of the reserve via simulation and quantification of the estimation uncertainty by Bootstrapping is discussed.

2. The basic model

2.1. Notation and basic model assumptions

Norberg (1999) defines a Marked Poisson process as follows:

A *claim* is a pair $C=(T,Z)$, where T is the time of occurrence of the claim and Z is the so-called mark describing its development from the time of occurrence until the time of final settlement.

The *claims process* is a random collection of claims $\{(T_l,Z_l)\}_{l=1,\dots,N}$, the index l indicating chronological order so that $0 < T_1 < T_2 < \dots$

It is assumed that the times are generated by an inhomogeneous Poisson process with intensity $w(t)$ at time $t > 0$.

It is assumed that the distribution of the mark is of the form $Z_l = Z_{Tl}$, where $\{Z_t\}_{t>0}$ is a family of random elements that are mutually independent and also independent of the Poisson process, and $Z_t \sim P_{Z|t}$.

The claim process is then called a *Marked Poisson process with intensity* $w(t)$ *and position-dependent marking* $P_{Z|t}$ and we write

$$\{(T_l,Z_l)\}_{l=1,\dots,N} \sim Po(w(t),P_{Z|t}\,;\, t>0).$$

We shall exclusively consider marks of the form

$$Z_l = (J_l,Y_{J,l},Y_{J+1,l}, \dots,Y_{D,l},G_l)$$

where l is the claim identification index, J_l is the reporting delay in years i.e. $J_l = 1$ if the claim l is reported within the calendar year of occurrence, $J_l = 2$ if reported the year after etc. $Y_{k,l}$ is the incurred amount in the development period k=J,..,D (we implicitly assume that the claims are settled after D development years) and G_l is a stochastic characteristic of the claim, for example claim-type and information from the policy the claim is covered under.

The claim identification index l will frequently be omitted.

We will also consider a partitioning of year 1 into q intervals $[s_m\ s_{m+1}[$, m=1,..,q where $0=s_1<..<s_{q+1}=1$. The partitioning could for example be quarters or even days. We will denote these periods by the seasonal periods. The interval $[i-1+s_m,i-1+s_{m+1}[$ will be denoted i_m or (i,m).

We assume that the intensity $w(t)$ is constant in year i, except for the same seasonal variation within the year, i.e. we assume that there is a function σ, σ(0)=1 on the interval [0,1[such that

$$w(t)= w_i\, \sigma(t\text{-}i), \text{ where } w_i=w(i) \text{ and } i=[t].$$

We further assume that $\sigma(t\text{-}i)=\sigma_m$ is constant for $t \in [s_m\ s_{m+1}[$, m=1,..,q, and have

$$w(t)= w_i\, \sigma_m, \text{ for } t \in i_m.$$

Claims settlement:

We will discuss the situation where the settlement of the claim is included in the mark by defining the indicator variables CC_k=I(claim is closed by the end of the development period k) and consider marks of the form

$$Z=(J,Y_J,Y_{J+1}, \dots,Y_D,CC_J,CC_{J+1}, \dots,CC_D,G),$$

2.2. The stochastic reserve

The stochastic outstanding claims reserve R_D (in excess of the individual case reserve) at the end of year D, is defined as the sum of all incremental amounts after time D concerning claims that have occurred by the end of year D i.e. as

$$R_D = \sum_{\substack{k>D-i+1 \\ l:Tl<D}} Y_{k,l}$$

and the stochastic IBNR reserve $IBNR_D$ at time D is defined as the sum of all incremental amounts in the future concerning claims that have occurred at time D but reported after time D i.e.

$$IBNR_D = \sum_{\substack{k\geq j>D-i+1 \\ l:Tl<D}} Y_{k,l}$$

We are interested in the conditional distribution of R_D and of $IBNR_D$ given the information at time t=D or more generally in the conditional distribution of the Marked Poisson process $Po(w(t),P_{Z|t}\,;\ D\geq t>0)$ given the process' value at the end of period D, i.e. at time t=D.

2.3. Decomposing the process

We now decompose the process by the values of (I,M,J,G) where (I,M) is the seasonal period of occurrence, J the development delay and G the characteristic. In other words, for each combination of i,m,j and g we consider the process where $T\in i_m$ and $Z=(j,Y_j,Y_{j+1},\ldots Y_D,g)$. The process is here called the (i,m,j,g)-component process and the claims are called the (i,m,j,g)-claims. The number of (i,m,j,g)-claims and the incremental amounts Y_k from the (i,m,j,g)-component processes will also be denoted $N_{i,m,j,g}$ (or N(i,m,j,g)) and $Y_{i,m,j,k,g,l}$ respectively. We then have the following expressions for the reserves:

$$R_D = \sum_{\substack{l=1,\ldots,N(i,m,j,g) \\ i=1,..,D,\,j=1,..,D \\ m=1,..,q,\,g\in G \\ k>D-i+1}} Y_{i,m,j,k,g,l} \quad \text{and} \quad IBNR_D = \sum_{\substack{l=1,..,N(i,m,j,g) \\ i=1,..,D, \\ m=1,..,q,\,g\in G \\ k\geq j>D-i+1}} Y_{i,m,j,k,g,l}$$

Below we will show that under certain assumptions the conditional distributions of the sums given the process' value at time D can be regarded as sums of stochastically independent variables.

It follows (Norberg (1999)) that the (i,m,j,g)-component processes are Marked Poisson processes, that they are independent and that the (i,m,j,g)-claims occur with an intensity which is the claim intensity multiplied by the probability that the claim is a (j,g)-claim i.e.

$$w_{i,m,j,g}(t) = 0,\ t \notin i_m \text{ and}$$

$$w_{i,m,j,g}(t) = w_i \sigma_m P_{Z|t}\{J=j, G=g\}$$

$$= w_i \sigma_m P_{Z|t}\{J=j|G=g\} P_{Z|t}\{G=g\},\ t \in i_m.$$

It also follows that the development of the (i,m,j,g)-claims follows the conditional distribution of the mark given it is a (i,m,j,g)-claim and that $N_{i,m,j,g}$ are independent and Poisson distributed and independent of the marks.

2.4. Business Mix assumption

Let $e_{i,m}$ be the exposure in the interval i_m and $e_{i,m,g}$ the exposure concerning the (i,m,j,g)-claims. While allowing for changes in business mix over the period [0,D] we assume that $P_{Z|t}\{G=g\}$, $t \in i_m$ is constant i.e. that the change in business mix through these shorter periods are negligible. We will then use the parameterisation

$$P_{Z|t}\{G=g\} = c\,(e_{i,m,g}/e_{i,m}) f_{I,M,G}(i,m,g),\ t \in i_m$$

where $c>0$, $f_{I,M,G}(i,m,g)>0$ and $f_{I,M,G}(0,0,g_0)=1$ for a *reference level* g_0 of G.

2.5. The 'Chain-Ladder' assumption

We now assume that $P_{Z|t}\{J=j|G=g\}$, $t \in i_m$, only depends on t through t-[t] i.e. that there exists a function $f'_{J,G}$ for which

$$P_{Z|t}\{J=j|G=g\} = f'_{J,G}(j,g,t-[t]),\ t \in i_m$$

This implies that the conditional distribution of the development delay J given G is independent of the year of occurrence i. This corresponds to the Chain-Ladder assumption. As a consequence the number, $N_{i,m,j,g}$, of (i,m,j,g)-claims is Poisson distributed with mean

$$E(N_{i,m,j,g}) = \int_{t\in(i,m)} e_{i,m} w_{i,m,j,g}(t)\,dt$$

$$= e_{i,m,g} c w_i \sigma_m f_{I,M,G}(i,m,g) \int_{t\in(i,m)} f'_{J,G}(j,g,t-[t])\,dt$$

$$= e_{i,m,g} c w_i \sigma_m f_{I,M,G}(i,m,g) f_{M,J,G}(j,m,g)$$

where $f_{M,J,G}(m,j,g) = \int_{t\in(i,m)} f'_{J,G}(j,g,t-[t])\,dt$.

It should be emphasised that the development pattern $f_{M,J,G}(m,j,g)$ concerning g-claims can be dependent on g and m but not on i.

3. A dynamic application

3.1 Assumptions concerning the mark

We now make the following two assumptions concerning the time dependent mark in the (i,m,j,g)-components processes:

1) The conditional distribution of $(Y_k|(Y_{k-1},\ldots,Y_j,j,g))$, k=j,..,D, only depends on $(Y_{k-1},\ldots,Y_j)$ through a function h of $(Y_{k-1},\ldots,Y_j)$ i.e. functions h exist so that

$$P_t(Y_k|Y_{k-1},\ldots,Y_j,j,g) \sim P_t(Y_k|\, h(Y_{k-1},\ldots,Y_j),j,g).$$

2) $P_t(Y_k|\, h(Y_{k-1},\ldots,Y_j),j,g)$ only depends on t through (i,m) i.e.

$$P_t(Y_k|\, h(Y_{k-1},\ldots,Y_j)\, j,g) = P(Y_k|\, h(Y_{k-1},\ldots,Y_j),i,m,j,g)$$

where P_t is the joint probability distribution of the mark Z_t. We have omitted the index (i,m,j,g) from the 'Y's and the function h.

It is seen that the assumptions are fulfilled if, for example, all (i,m)-claims occur at the beginning of the period i_m and the process (S_k), k=j,…,D defined by $S_k = Y_k+\ldots+Y_j$, is a Markov Chain.

Under the assumptions 1) and 2) we have for (i,m,j,g)-claims

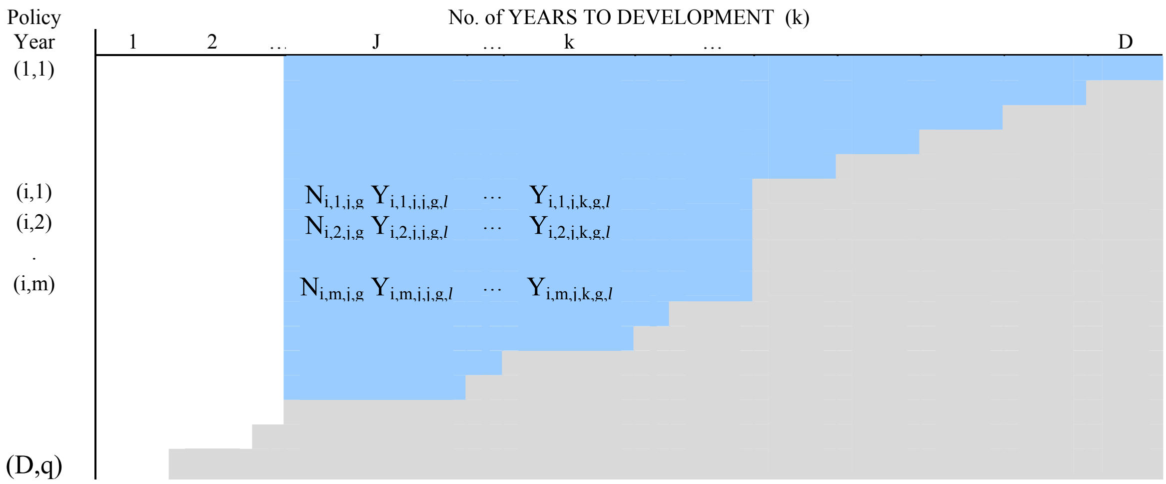
$$P_t(Y_D,Y_{D-1},\ldots,Y_j,j,g\,)= P_t(Y_D|Y_{D-1},\ldots,Y_j,j,g)\; P_t(Y_{D-1},\ldots,Y_j\,,j,g\,)=\ldots..$$

$$= P(Y_D|\, h(Y_{D-1},\ldots,Y_j),i,m,j,g)\; P(Y_{D-1}|\, h(Y_{D-2},\ldots,Y_j),i,m,j,g)\ldots\; P(Y_j,i,m,j,g).$$

The main advantage is that the conditional distributions are independent while still maintaining a possibility that the incremental amounts are dependent on the past developments.

Observation plan:

For i=1,..,D, m=1,..,q, j=1,...,D-i+1 and $g\in G$ we observe the number $N_{i,m,j,g}$ of claims reported and the incremental amounts $Y_{i,m,j,k,g,l}$,l=1,.., $N_{i,m,j,g}$ in the columns k=j,…,D-i+1, i.e. figures in the upper triangle.



The index (i,m,j,g,l) from $Y_{i,m,j,k,g,l}$ will frequently be excluded below.

It now follows that the likelihood, $L_{i,m,j,g}$,for the event (N,Y) for the (i,m,j,g)-component is

$$L_{i,m,j,g} = Po(N_{i,m,j,g}\,,\, e_{i,m,g} c w_i \sigma_m f_{I,M,G}(i,m,g) f_{M,J,G}(m,j,g)) \prod_{D-i+1>k\geq j,\, l=1,..,N(i,m,j,g)} P(Y_{i,m,j,k,g,l} \mid h(Y_{i,m,j,k-1,g,l},\ldots,Y_{i,m,j,j,g,l}),i,m,j,g)$$

Since the components are independent the total likelihood is the product of all the component likelihoods.

<u>Claims settlement:</u>

If the indicators CC_k for closed claim are included in the mark we could for example consider functions of the form $h(Y_{k-1}\ldots,Y_j,CC_{k-1},\ldots,CC_j)= (h_1(Y_{k-1},\ldots,Y_j,), h_2(CC_{k-1},\ldots,CC_j,))$.

3.2. The Distribution of the Reserve

It is seen that the joint distribution can be specified by for each component (i,m,j,g) specifying the independent Poisson distributions of $N_{i,m,j,g}$, and by specifying the distribution of Y_j and of Y_k given $h(Y_{k-1,\ldots,}\, Y_j)$, k=j+1,..,D which are all independent and also independent of the $N_{i,m,j,g}$.

The conditional distribution of the reserve:

Since the N's and Y's are independent so are the conditional distributions of the N's and Y's and the conditional distribution of the ($N_{i,m,j,g}$), j>D-i+1 is the original distribution since the N's are independent. It also follows (by successive conditioning) that the conditional distribution of the Y's are determined by the conditional distributions $P(Y_k \mid h(Y_{i,m,j,k-1,g},\ldots,Y_{i,m,j,j,g}),i,m,j,g)$, k>D-i+1, which are independent and also independent of the $N_{i,m,,j,g}$.

4. An example

Let $S_{k-1} = Y_{k-1}+\ldots+Y_j$ and $h(Y_{k-1,\ldots,}Y_j)=SG(S_{k-1})=SG_k$ where SG is a grouping of S_{k-1}, i.e. h is the dicretisized accumulated incurred amount at the beginning of period k.

We will for simplicity exclude the seasonal effect i.e. q=1 and omit the index m.

We now specify the distribution below assuming that $g=(g_1,g_{2,\ldots..},g_n)$ corresponding to n covariates.

4.1. The distribution of $N_{i,j,g}$

It is further assumed that there are no interactions i.e. that the mean has the form

$$E(N_{i,j,g1,g2,\ldots gn}) = e_{i,j,g1,g2,\ldots gn}\, cf_I(i)f_J(j)f_1(g_1)\ldots f_n(g_n)$$

where $f_I,f_J,f_{1,\ldots,}f_n$ are positive functions of the covariate levels, c>0 and where $e_{i,j,g1,g2,\ldots gn}$ is the exposure in number of insurance years. Please note that the exposure is independent of the reporting delay j and of any covariate from the Mark which does not originate from the policy covering the claim such as claim type.

For each covariate there is a *reference-level* for which the factor is 1. Therefore the mean concerning the reference cell (i.e. the cell consisting of the combination of all the reference-levels) is proportional to the exposure i.e. E(N) = ec for the reference cell.

Please note that in the situation where there are no covariates g the model gives the same estimates as the Chain Ladder model based on volume weighted averages. This follows from the fact that the sums $N'_{i.}$ and $N'_{.j}$ of the estimated values $N'_{i,j}$ in the upper triangle in both models are equal to the observed sums $N_{i.}$ and $N_{.j}$.

4.2. The distribution of $Y_{i,j,j,g}$

The probability for the event $\{Y_{i,j,j,g} = 0\}$ is positive and the distribution of $Y_{i,j,j,g}$ can be specified by the probability $P(Y_{i,j,j,g} = 0)$ and by the conditional distribution $P(Y_{i,j,j,g} \mid Y_{i,j,j,g} > 0)$. The conditional distribution could be assumed to be a Gamma distribution with a multiplicative mean structure and a variance that is proportional to the mean. However, this assumption is not suitable for large incremental incurred amounts and these are therefore modelled in isolation:

First we specify the probabilities for the three disjoint events $\{Y_{i,j,j,g} = 0\}$, $\{0 < Y_{i,j,j,g} < L\}$ and $\{L \leq Y_{i,j,j,g}$). They are uniquely determined by the conditional probabilities $p_{>0} = P(Y_{i,j,j,g} > 0)$ and $p_{>L} = P(Y_{i,j,j,g} > L \mid Y_{i,j,j,g} > 0)$ via the expressions

$$P(Y_{i,j,j,g} = 0) = 1-p_{>0},$$
$$P(0 < Y_{i,j,j,g} < L) = (1- p_{>L})p_{>0} \text{ and}$$
$$P(L \leq Y_{i,j,j,g}) = p_{>L}p_{>0}.$$

The probabilities $p_{>0}$ and $p_{>L}$ are both assumed to be of the form $1/(1+p)$ where

$$p = p_{i,j,j,g1,g2,\dots gn} = cf_I(i)f_J(j)f_1(g_1)\dots f_n(g_n).$$

This is a Logistic Regression Model with the logit function as link function.

Secondly we define the conditional distributions of $Y_{i,j,j,g}$ given the above events:

<u>$\{0 < Y_{i,j,j,g} < L\}$:</u>

The conditional distribution of $Y_{i,j,j,g}$ given $\{0 < Y_{i,j,j,g} < L\}$ is assumed to be Gamma distributed with mean and variance of the form

$$E(Y_{i,j,j,g1,g2,\dots gn}) = cf_I(i)f_J(j)f_1(g_1)\dots f_n(g_n) \text{ and}$$

$$V(Y_{i,j,j,g1,g2,\dots gn}) = E(Y_{i,j,j,g1,g2,\dots gn})^2\varphi$$

The support for the Gamma distribution is $\{0<y\}$ and therefore the choice of distribution is not entirely consistent. However, if L is large this is not necessarily a significant problem in practise.

<u>$\{L \leq Y_{i,j,j,g} \}$:</u>

The conditional distribution of $Y_{i,j,j,g}$ given $Y_{i,j,j,g} \geq L$ is assumed to be a Generalised Pareto Distribution i.e. the distribution function is of the form

$$F(y) = 1-[1+(y-L)/(\alpha\beta)]^{-\alpha} , \alpha>0, \beta>0.$$

4.3. Distributions of $Y_{i,j,k,g}$ given $S_{i,j,k-1,g}=0$, k=j+1,..D

These distributions are defined similarly to the distributions of $Y_{i,j,j,g}$ above, however a covariate concerning the development delay k is also incorporated, for example it is assumed that the conditional distribution of $Y_{i,j,k,g}$ given $\{0 < Y_{i,j,k,g} < L\}$ is Gamma with mean and variance of the form

$$E(Y_{i,j,k,g1,g2,\dots gn}) = cf_I(i)f_J(j)f_K(k)f_1(g_1)\dots f_n(g_n) \text{ and}$$

$$V(Y_{i,j,k,g1,g2,\dots gn}) = E(Y_{i,j,k,g1,g2,\dots gn})^2\varphi$$

In the example in section 5 below it is further assumed that the factors concerning I,J and the covariates G are the same for the distributions of $Y_{i,j,j,g}$ and for the conditional distribution of $Y_{i,j,k,g}$ given $S_{k-1}=0$ and that these distributions only differ via the factors concerning the development K. However, an extra covariate indicating whether or not the claim has been positive in the past is included.

4.4. Distribution of $Y_{i,j,k,g}$ given $S_{i,j,k-1,g} >0$ and SG_k, k=j+1,..D

This situation is different from the above since the incremental amounts can be negative. However, the total incurred amount is assumed to be non-negative i.e. $S_k=Y_j+...+Y_k\geq 0$ for $k\geq j$. The variable SG_k will be treated as a part of the covariates g.

First we specify the probability for the following five disjoint sets with joint probability 1: $\{Y_{i,j,k,g} = 0\}$, $\{0 < Y_{i,j,k,g} < L\}$, $\{L \leq Y_{i,j,k,g}\}$, $\{0> Y_{i,j,k,g} > -S_{i,j,k-1,g}\}$ and $\{Y_{i,j,k,g} = -S_{i,j,k-1,g}\}$.

It is seen that the probabilities are uniquely determined by the conditional probabilities $p_{>0} = P(Y_{i,,j,k,g} > 0 \mid S_{i,j,k-1,g}>0)$, $p_{>L} = P(Y_{i,j,k,g} > L \mid Y_{i,j,k,g} > 0, S_{i,j,k-1,g} >0)$, $p_{=0} = P(Y_{i,j,k,g} = 0 \mid Y_{i,j,k,g} \leq 0, S_{i,j,k-1,g} >0)$ and $p_{S>0} = P(Y_{i,j,k,g} > - S_{i,j,k-1,g} \mid Y_{i,j,k,g} < 0, S_{i,j,k-1,g} >0)$ since we have the expressions:

$$P(Y_{i,j,k,g} = 0 \mid S_{i,j,k-1,g} >0) = p_{=0}(1-p_{>0}),$$
$$P(0 < Y_{i,j,k,g} < L\mid S_{i,j,k-1,g} >0) = (1- p_{>L})p_{>0},$$
$$P(L \leq Y_{i,j,k,g} \mid S_{i,j,k-1,g} >0) = p_{>L}p_{>0},$$
$$P(0> Y_{i,j,k,g} > - S_{i,j,k-1,g} \mid S_{i,j,k-1,g} >0) = p_{S>0}(1-p_{>0})(1- p_{=0}) \text{ and}$$
$$P(Y_{i,j,k,g} = - S_{i,j,k-1,g} \mid S_{i,j,k-1,g} >0) = (1-p_{S>0})(1-p_{>0})(1- p_{=0})$$

We assume that $p_{=0}$,$p_{>0}$,$p_{>L}$and $p_{S>0}$ have the form 1/(1+p) where

$$p = p_{i,j,k,sg,g1,g2,\dots gn} = cf_I(i)f_J(j)f_K(k)f_{SG}(sg_k)f_1(g_1)\dots f_n(g_n)$$

Secondly we define the conditional distributions of $Y_{i,j,k,g}$ given the above events and SG_k:

$\{0< Y_{i,j,k,g} <L,S_{i,j,k-1,g} >0\}$

The conditional distribution of $Y_{i,j,k,g}$ given $(0< Y_{i,j,k,g} <L, S_{i,j,k-1,g} >0)$ is assumed to be Gamma distributed with mean and variance of the form

$$E(Y_{i,j,k,sg,g1,g2,\dots gn}) = cf_I(i)f_J(j)f_K(k)f_{SG}(sg_k)f_1(g_1))\dots f_n(g_n) \text{ and}$$

$$V(Y_{i,j,k,g1,g2,\dots gn}) = E(Y_{i,j,k,sg,g1,g2,\dots gn})^2\varphi,$$

where the covariate SG is incorporated.

<u>$\{Y_{i,j,k,g} \geq L \ , S_{i,j,k-1,g} > 0\}$</u>

The conditional distribution of $Y_{i,j,k,g}$ given ($Y_{i,j,k,g} \geq L$, $S_{i,j,k-1,g} > 0$, SG_k) is assumed to be a Generalised Pareto Distribution i.e. the density function is of the form

$$F(y) = 1-[1+(y-L)/(\alpha\beta)]^{-\alpha} , \alpha>0, \beta>0.$$

<u>$\{0 > Y_{i,j,k,g} > - S_{i,j,k-1,g} \ , S_{i,j,k-1,g} > 0\}$</u>

The range for $Y_{i,j,k,g}$ given ($0 > Y_{i,j,k,g} > - S_{i,j,k-1,g}$, $S_{i,j,k-1,g} > 0$) is obviously $]- S_{i,j,k-1,g} ,0[$ and therefore the range for $-\log((Y_{i,j,k,g} + S_{i,j,k-1,g})/S_{i,j,k-1,g})$ is $]0,\infty[$.

The conditional distribution of $Y_{i,j,k,g}$ given ($0 > Y_{i,j,k,g} > - S_{i,j,k-1,g}$, $S_{i,j,k-1,g} > 0$, SG_k) is specified by assuming that the distribution of $-\log((Y_{i,j,k,g} + S_{i,j,k-1,g})/S_{i,j,k-1,g})$ is Gamma distributed with mean and variance of the form as above.

<u>Remarks:</u>

Since we are only observing the (i,j,g)-component processes where $i+j \leq D+1$ interactions between I and J, I and K and J and K should not be considered since extrapolation to ‘future’ cells where $D < i+k$ would not be possible.

‘Inflation’ in the reporting intensity (or in the incurred amount) can be incorporated by substituting the I-factor f_I with a factor f_{I+J} (resp. f_{I+K}). In this way the pure period inflation can be quantified as well as the effect implied by changes in claims mix.

<u>Claims settlement:</u>

Let us briefly consider the situation where the indicators for closed claim CC_k are included in the mark and consider the functions

$$h(Y_{k-1}\dots,Y_j, CC_{k-1},\dots,CC_j) = ((Y_{k-1}+\dots+Y_j), CC_{k-1}).$$

The distributions can be specified by for example assuming that Y_k and CC_k are condition independent given $((Y_{k-1}+\dots+Y_j), CC_{k-1})$ and then specify the marginal distributions. The marginal distribution concerning Y_k and the ‘event’-probabilities can be specified in the same way as above where an extra covariate f_{CC} concerning CC is included in the Generalised Linear Models. The marginal distributions of CC_k given $((Y_{k-1}+\dots+Y_j), CC_{k-1}=1)$ and of CC_k given $((Y_{k-1}+\dots+Y_j), CC_{k-1}=0)$ can be modelled using Logistic Regression.

Distribution of the reserve:

It follows from section 3.2 that the conditional distribution of the reserve R_D, given the information at the time D, is specified via the distributions above.

5. Estimating the parameters

We will illustrate the model based on a Marine portfolio with policy and claims information available from the period 1992-2004.

5.1. Data

Data consists of the following:

Policy records	Claims records
Policy Id	Policy Id
Start date	Claim Id
End date	Claim date
	Reporting date
Vessel type	Claim type
Vessel Tonnage	Incurred amount
Class of Business	Transaction date

Based on this data the sufficient statistics are created:

N	Y
	l: claim id
i: year of occurrence	i: year of occurrence
j: reporting period	j: reporting period
e: exposure in years	k: development period
g_1: grouped claim type	g_1: grouped claim type
g_2: grouped vessel type	g_2: grouped vessel type
g_3: grouped vessel tonnage	g_3: grouped vessel tonnage
g_4: class of business	g_4: class of business
N: Number of claims	Y_k: incurred amount in the development period
	S_{k-1}: accumulated incurred amount at the beginning of the development period
	SG_k: the discretisized value of S_{k-1}.

The N-data has been created as follows: The exposure is first summarised by all combinations of (i,g_2,g_3,g_4). Then the exposure for the combination of (i,j,g_1,g_2,g_3,g_4) is defined as the exposure for the projection (i,g_2,g_3,g_4) i.e. the exposure is independent of reporting delay j and claim-type g_1.

The Y-data has been created as follows: For all observable combinations of i and k (i.e. where $i+k \leq D+1$) where there are no records of incurred amount a record is generated with $Y_k=0$ and thereafter the $S_{k-1}$and SG_k values are calculated.

5.2. Estimation method

The estimated parameters concerning the Poisson, Gamma and Logit models are the maximum likelihood estimates. The parameters concerning the Generalised Pareto distributions are fitted using non linear regression analysis where the 'distance between the empirical d.f and model d.f.' is minimised. The process of fitting the parameters in the model-components will be illustrated below by a few examples.

5.3. Example 1

In order to specify a reasonable model thorough empirical analyses are required. As a first example we will illustrate the impact of the SG-criteria on the likelihood that the incremental amount is greater than L=$500000 given that it is greater than 0 i.e

$$p_{>L} = P(Y_{i,j,k,g} > L \mid Y_{i,j,k,g} > 0,\ S_{i,j,k-1,g} > 0).$$

The probability $p_{>L}$ is of the form $1/(1+p)$ where

$$p = p_{i,j,k,sg,g1,g2,\ldots gn} = cf_I(i)f_J(j)f_K(k)f_{SG}(sg_k)f_1(g_1)\,)\ldots f_4(g_4).$$

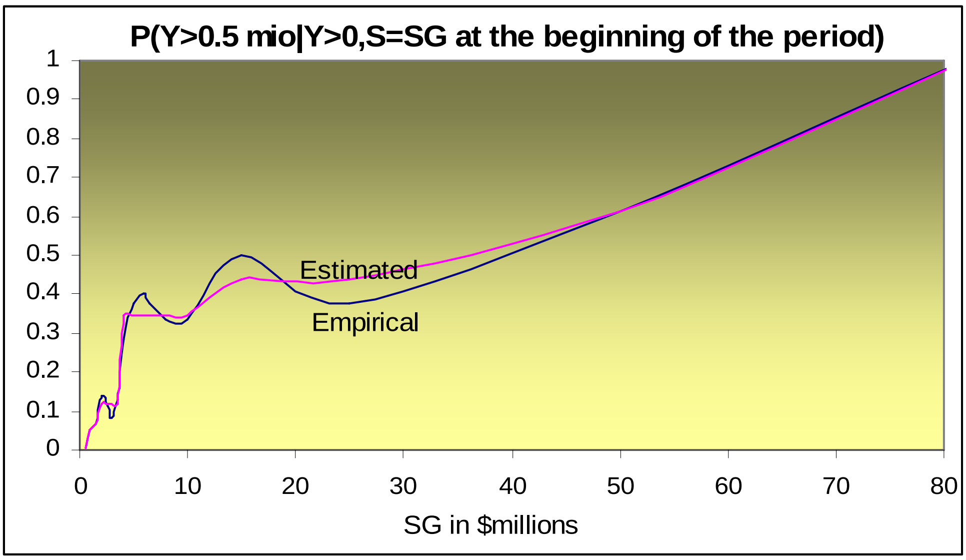
The sufficient statistic is the number of 'trials' T and number of 'hits' H:

$$T_{i,j,k,sg,g1,g2,g3,g4} = \sum I(Y_{i,j,k,sg,g1,g2,g3,g4} > 0,\ S_{i,j,k-1,g1,g2,g3,g4} > 0)$$

$$H_{i,j,k,sg,g1,g2,g3,g4} = \sum I(Y_{i,j,k,sg,g1,g2,g3,g4} > L,\ S_{i,j,k-1,g1,g2,g3,g4} > 0)$$

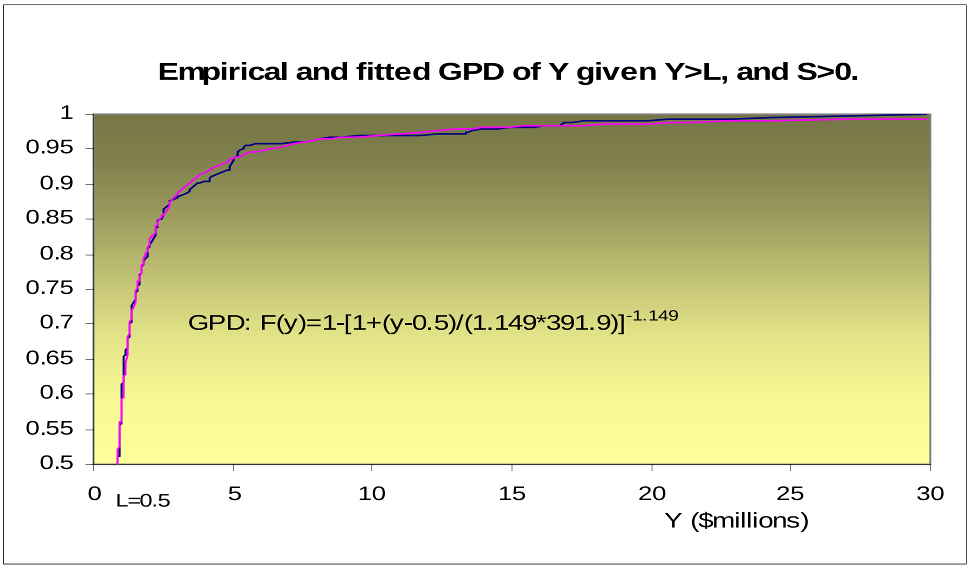
where the summary is over the claims identification *l*.

The observed hit-rate, H/T, and the estimated value for each level of the SG-criteria are outlined below. Apart from random fluctuations the hit-rate is increasing dramatically by SG-value i.e. for claims where the incurred amount at the beginning of the period is large there is a much higher likelihood that the incremental value is greater than $500.000 given that the incremental value is positive.



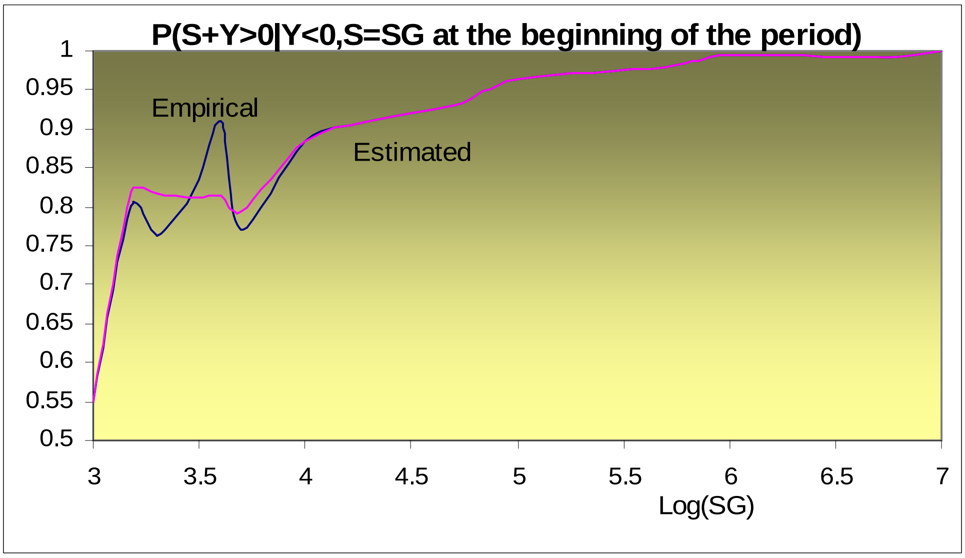
5.4. *Example 2*

We will now focus on the conditional distribution of $Y_{i,j,k,g}$ given that it is greater than L=\$500000 and given that $S_{i,j,k-1,g} > 0$. 242 yearly incremental amounts of this kind have been observed. A section of the empirical d.f. and the fitted Generalised Pareto d.f. are outlined below.



5.5. *Example 3*

As another example we will outline the empirical probability $p_{S>0} = P(Y_{i,j,k,g} > - S_{i,j,k-1,g} \mid Y_{i,j,k,g} < 0, S_{i,j,k-1,g} > 0)$ i.e. the empirical likelihood that a positive claim is not becoming a zero-claim given that the incremental amount is negative.



It is noted that this likelihood is close to 1 for the very large claims and therefore the likelihood that a positive claim becomes a zero-claim is close to 0 for the very large claims. The estimation method of $p_{S>0}$ is similar to the method concerning $p_{>L}$ as described above.

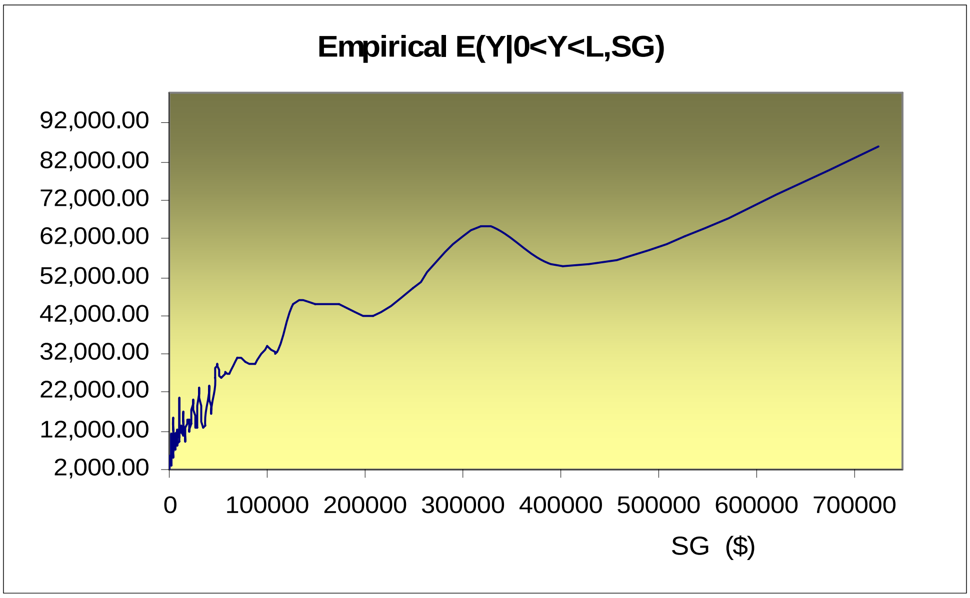
5.6. *Example 4*

In this last example we will look at the conditional distributions of $Y_{i,j,k,g}$ given the event $\{0< Y_{i,j,k,g} <L , S_{i,j,k-1,g} >0\}$ which is assumed to be Gamma with mean and variance of the form

$$E(Y_{i,j,k,g1,g2,\ldots g4}) = cf_I(i)f_J(j)f_K(k)f_{SG}(sg)f_1(g_1))\ldots f_4(g_4) \text{ and}$$

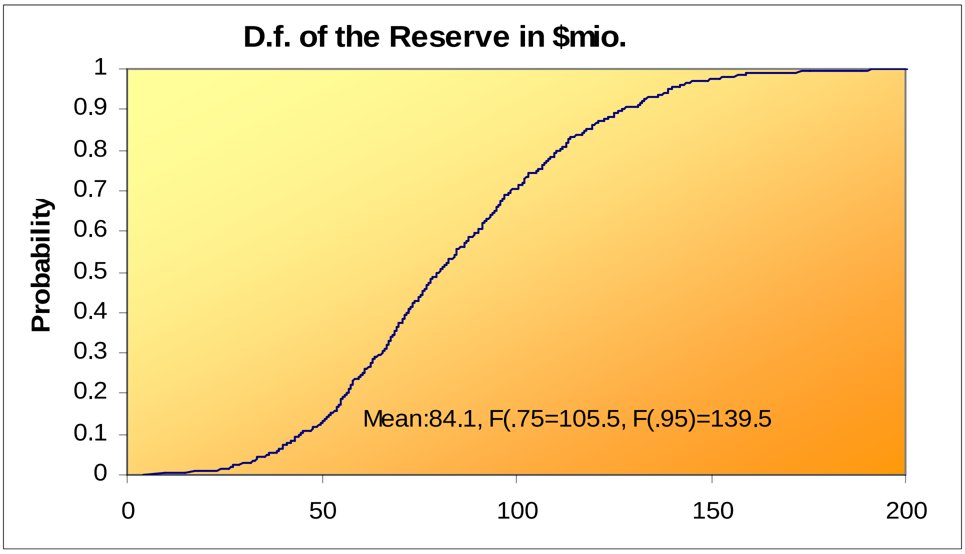
$$V(Y_{i,j,k,g1,g2,\ldots g4}) = E(Y_{i,j,k,g1,g2,\ldots g4})^2\varphi,$$

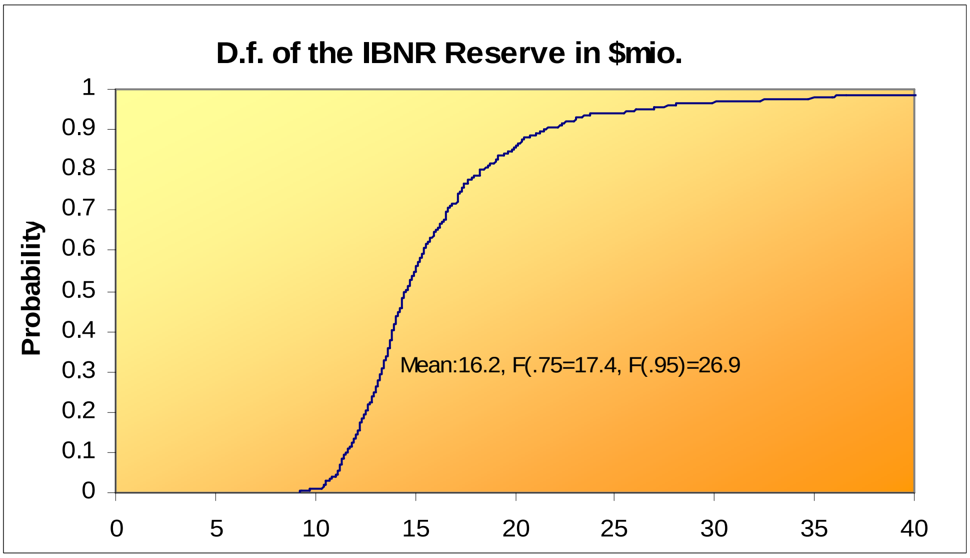
The empirical mean as a function of SG is outlined below. It is not surprising that the SG criteria is statistically significant. The greater the claim the greater the future average increase given that it is positive and less than \$500.000. However, this trend does not continue for SG>700000 where the mean is approximately \$85000 independent of SG.



6. Estimating the reserve distribution

Despite the fact that the joint distribution is fully specified by the one-dimensional conditional distributions, an exact calculation of the conditional distribution of the reserve given all information in the past is not simple. However, simulation of the process is straightforward, and therefore also the calculation of the distribution of the reserve and of all imaginable 'Net of Reinsurance' reserves and statistics of this. The distribution of the total reserve and the total IBNR reserve, based on 500 simulated 'ultimate' projections, are outlined below:





7. Estimation uncertainty

Since the observation is random, the estimated parameters, as functions of the observation, are also random. The implication is that the projected distribution of the reserve as described above is uncertain. To quantify this uncertainty the Bootstrap method could be applied since the claims are assumed to be stochastically independent.

The steps are as follows:

1) From a Poisson distribution with mean equal to the total number N of observed claims a number M is sampled.
2) M claims $\{(T_m,Z_m)\}_{m=1,\dots,M}$ }from the set of observed claims $\{(T_l,Z_l)\}_{l=1,\dots,N}$ }are sampled with replacement.
3) The model parameters {p} are estimated based on the sampled claims $\{(T_m,Z_m)\}_{m=1,\dots,M}$ }.
4) One reserve outcome is simulated according to the model and fitted parameters {p} (as described above) given the original observation.
5) Step 1-4 are repeated e.g. 500 times.

The resulting distribution would be a reasonable approximation to the total uncertainty i.e. the process variation as well as the estimation uncertainty if the number of repetitions in step 5) is sufficiently large and if the empirical distribution is 'close' to the distribution of the underlying process, i.e. if there are 'many' claims.

This solution would be practical if all the programs corresponding to the model components described in section 4 can be run automatically in a batch.

8. Conclusion

It is concluded that the mean of the outstanding claims liabilities, when detailed individual claims information is available, can be assessed more accurately by modelling the claims process as a Marked Poisson process than by using the traditional Chain Ladder models. Also the distribution of the outstanding claims liabilities and the combined process variation and estimation uncertainty can be assessed by a combination of Bootstrapping and simulation.

Acknowlegement

The author would like to thank Professor Ragnar Norberg for an interesting discussion about Marked Poisson processes.

9. References

Claims Reserving Manual, Volume I and II, *The (British) Faculty and Institute of Actuaries*
Dobson, A (1990), An introduction To Generalized Linear Models, London: *Chapman and Hall*
Efron, B., Tibshirani R.J., An introduction to the Bootstrap: *Chapman and Hall*
England, P.D., Verrall R.J., (2002), Stochastic claims reserving in general insurance, *British Actuarial Journal, 8 (3), 443-518.*
Hesselager, O. (1995). Modelling of discretized claim numbers in loss reserving, *ASTIN Bull. 25, No.2,119-135.*
Hoedemakers, T.,Beirlant, J., Goovaerts, M.J. and Dhaene, J., On the distribution of discounted loss reserves using generalized linear models, *Scandinavian Actuarial Journal No.1/2005*
Mack, T. (1993), Distribution-free calculation of the standard error of chain ladder reserve estimates, *ASTIN Bull. 23, No. 2, 213-225*
Norberg, R. (1986), A Contribution to Modelling of IBNR Claims, *Scan. Actuarial J. 1986: 155-203*
Norberg, R. (1993), Prediction of outstanding liabilities in non-life insurance, *ASTIN Bull. 23, No. 1, 95-115*
Norberg, R. (1999), Prediction of outstanding claims II: Model variations and extensions, *ASTIN Bull. 29, No. 1, 5-25*