A Nonlinear Regression Model of Incurred But Not Reported Losses

Scott Stelljes, ACAS, MAAA

Abstract

The process of loss development has been studied by casualty actuaries for many years. When an accident period is closed, the ultimate claim liabilities are unknown because many of the claims are still unreported and some that are reported remain unsettled. The difference between ultimate losses and reported losses is known as "Incurred But Not Reported" loss or IBNR. The reserve for IBNR losses is the largest liability on an insurer's balance sheet. Quantifying the uncertainty in estimates of IBNR is of great importance to the financial health of casualty insurance companies.

Most of the current methods for estimating ultimate losses focus on estimation of loss development factors which relate the emergence of losses to the amount of losses already reported. This paper presents a model for predicting incremental losses as a function of exposures, calendar period and development age.

A nonlinear regression model is used for estimating the 95% confidence interval of IBNR for an accident period. The model predicts the incremental pure premium for a development interval as a function of development age, calendar quarter and exposure. The estimated IBNR is the sum of forecasted incremental pure premiums. The regression model produces confidence interval estimates for the model parameters and for IBNR.

The regression model is applied to trended losses. We assume that the trend has been estimated by some reasonable time series method that produces confidence interval estimates of trend factors. Many good methods are available. We use the confidence interval estimate of the trend factors to adjust the IBNR estimates for uncertainty in loss trend.

The model presented here assumes normally distributed residuals. Although the underlying loss severities are probably not normal, the central limit theorem implies that this assumption would be appropriate if the number of claims is large. Thus, the model will most likely work well for high frequency lines of business such as personal auto.

We will present methods for estimating parameters, confidence intervals for the parameters, and the distribution of IBNR. These methods will be illustrated using simulated automobile bodily injury liability data. Model predictions will be compared to actual emerged losses.

Based on a comparison of predicted IBNR to the "actual" IBNR from the simulated data, the model appears to produce unbiased predictions and reasonable confidence interval estimates of IBNR. We conclude that the distribution of incremental pure premiums is close to normal and there is not a significant correlation between development age intervals. Thus, traditional regression methods can be used to estimate the distribution of forecasted incremental pure premiums and consequently, IBNR.

Keywords: Non-linear regression, IBNR, reserving.

1. INTRODUCTION

Many actuaries and their clients are unsatisfied with point estimates of IBNR reserves. Better decisions can be made if one has a range of possible outcomes and associated probabilities. Confidence interval estimates would satisfy this need. We will introduce a nonlinear regression model that produces confidence interval estimates of IBNR. The models are fitted to incremental pure premiums - the incremental change in case incurred (or paid) losses for an accident period during a development interval divided by the corresponding calendar period earned exposures. This approach was inspired by Buhlman's complementary loss ratio method as presented by Stanard [3].

1.1 Research Context

The context of this paper is reserving methods and reserving uncertainty and ranges.

1.2 Objective

The objective of this research is to produce a model of loss development that models losses as a function of exposures, can be applied to either paid or incurred losses, and produces a confidence interval estimate of IBNR.

The current literature includes some papers, e.g., Murphy [1] that present regression models to predict age-to-age loss development factors and measure the uncertainty in the predicted factors but there are very few that present models of loss dollars. Barnett and Zehnwirth [2] is an excellent example of a dollar based model, but it is applied to the logarithms of incremental losses and this becomes a problem when there is negative loss development. Recoveries lead to negative paid development and case reserve estimation errors can result in negative case development. In order to use a log link, it is necessary to discard information. Less information is discarded if the analysis is performed on paid losses but much of the data in the tail of a case incurred development triangle is negative. Many reserving actuaries believe that there is useful information in case incurred losses and they often compare estimates derived from paid and incurred data.

Furthermore, Narayan [4] remarks that dollar based regression models do not take into account changing levels of exposure. This is a serious flaw because the amount of loss in an accident period is highly correlated to the number of earned exposures.

Thus, there is a need for a dollar based regression model that can be applied without using a log link and that makes appropriate adjustments for changing levels of exposure.

In this paper, we present a nonlinear regression model that predicts incremental pure premiums as a function of development age. The model is applied to losses that have been adjusted for loss trend using a separate trend model. The trend model can be any time series model that produces confidence intervals for future trend factors. In the examples, we assume that future trend is represented by a geometric Brownian motion process but this is not necessarily the only model for future loss trend. Adjusting losses for trend is not necessary in a link ratio method because future development is predicted as a function of case or paid losses. The link ratios are multiplied by losses which are already stated at the appropriate cost level. The factors produced by our model are applied to exposures so it is necessary to adjust losses for trend.

The model presented in this paper does not require the use of any link function, so it can be applied either to paid or case incurred loss data. Furthermore, since we use pure premiums with exposure weights, the model relates losses to exposures.

1.3 Outline

The remainder of the paper proceeds as follows.

Section 2: Presentation of data. A simulated data set including a loss triangle and earned exposures is presented along with some observations. The nonlinear model is presented and the estimation of parameters is explained.

Section 3: The model is fitted to the simulated data and used to produce confidence interval estimates of ultimate incurred losses for each accident quarter. An analysis of residuals is presented.

Section 4: Conclusions.

Section 5: References.

2. BACKGROUND AND METHODS

A nonlinear regression model will be presented and used to analyze simulated loss development data. The model will be fitted to incremental pure premiums. The incremental pure premium for an accident quarter/development quarter is defined as the change in case incurred loss during the development quarter divided by the calendar quarter earned exposures.

In section 2.1, we will present the simulated loss development data. The data was simulated based on method 4 in Narayan [4] with some modifications. See Appendix B for a description of the method used to simulate the data. Narayan and other authors simulated thousands of sets of data for the purpose of comparing methods. We simulated a single triangle for the purpose of showing sample calculations. The simulation is not intended to validate the model. The simulated data is intended to resemble personal auto bodily injury data in accident quarter/development quarter format.

Section 2.2 is a presentation of the nonlinear regression model.

In section 2.3, we present the mathematics of estimating confidence intervals for the model parameters and IBNR.

2.1 Loss Development Data

Exhibit 2.1.1 shows a small portion of the simulated loss data in the traditional triangular array. The losses shown in Table 1 are cumulative case incurred losses. I.e., the amount shown for each development quarter is the sum of all paid losses from the beginning of the accident quarter through the end of the development quarter and the outstanding case reserves as of the end of the development quarter. The column to the left of the losses shows earned exposures. The second table shows the incremental pure premiums. These are the incremental losses divided by earned exposures. For example, the entry for accident quarter 1, development interval 1-2 is (1,713,179-1,244,722)/50,333.

EXHIBIT 2.1.1

Tabl 1. Cumulative Losses by Accident Quarter and Development Age

Accident	Earned	Development Age				
Quarter	Exposures	1	2	3	4	5
1	50,333	1,244,722	1,713,179	1,996,372	2,065,006	2,166,446
2	50,801	1,417,101	2,004,222	2,341,886	2,437,727	
3	51,187	1,143,473	1,646,289	2,130,201		
4	51,146	1,055,290	2,268,788			
5	51,527	1,508,450				

Table 2. Incremental Pure Premiums

Accident	Earned	Development Interval				
Quarter	Exposures	0-1	1-2	2-3	3-4	4-5
1	50,333	24.73	9.31	5.63	1.36	2.02
2	50,801	27.90	11.56	6.65	1.89	
3	51,187	22.34	9.82	9.45		
4	51,146	20.63	23.73			
5	51,527	29.27				

Exhibit 2.1.2 shows the averages and variances and Pearson correlations of incremental pure premiums by development age for some simulated data. The data exhibits a typical loss development pattern. We see that the average incremental pure premiums start high and decrease rapidly as the development age increases, converging to zero. There are some negative incremental losses resulting from recoveries, settling of claims for less than the case reserve, and reductions to case reserves. The table also shows that the variance decreases as development age increases. Thus, most of the uncertainty in loss development is in the early stages. The correlation matrix shows that the correlation of incremental pure premiums between different ages is usually insignificant.

EXHIBIT 2.1.2

Sample of Simulated Incremental Pure Premiums - Ages 1-7

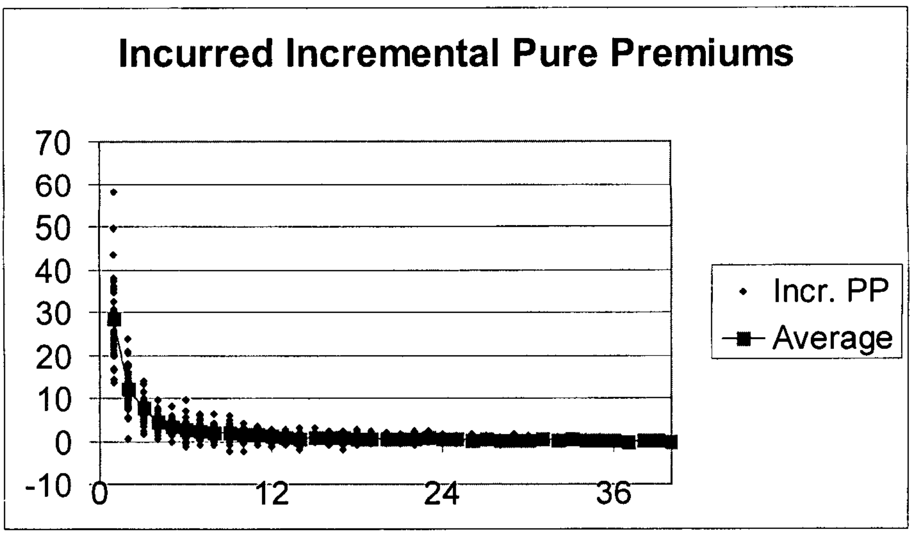
Age	0-1	1-2	2-3	3-4	4-5	5-6	6-7
Average	28.70	12.13	7.67	4.67	3.49	2.84	2.48
Variance	87.26	23.19	8.40	3.99	2.81	4.33	3.24

Pearson Correlations

	0-1	1-2	2-3	3-4	4-5	5-6	6-7
0-1	1.00	0.38	0.13	0.45	0.14	0.63	0.21
1-2		1.00	0.31	0.44	-0.01	0.29	0.25
2-3			1.00	0.15	0.11	-0.01	0.15
3-4				1.00	0.47	0.45	0.20
4-5					1.00	0.16	0.00
5-6						1.00	0.11
6-7							1.00

Exhibit 2.1.3 shows a scatter plot of the incremental pure premiums and the average incremental pure premiums by age.

EXHIBIT 2.1.3



In the scatter plot, the incremental pure premiums appear to be distributed around the average symmetrically. This and the fact that the correlations are not significant imply that the data fits the assumptions of regression models as stated in [5] reasonably well. The nonconstancy of the variances is a violation of the assumptions underlying ordinary regression but that problem can be solved by using a weighted regression model.

A weighted regression model is one in which a weight is assigned to each observation in the data. The more weight given to an observation, the more influence it has on the parameter estimates. We need to use a weight function that is inversely proportional to the variance of the data. It would also be advantageous to obtain exposure weighted parameter estimates. So, we will use weights that are a function of development age and exposures.

We will now define some of the variables that will be used in the analysis. First, the accident quarter will be represented by t which will take values of 1, 2, ..., 40. The development quarters will be represented by x which will be assigned the value of the development age (in quarters) at the end of the interval. For example, $x = 1$ will correspond to the 0-3 months development interval. Calendar quarters will be represented by u and will be calculated as $u = t + x - 1$. The incremental losses for accident quarter t and development interval x will be represented by $L_{t,x}$. Car months will be represented by c_t.

Appendix A shows the full set of simulated loss development data.

2.2 The Model

Our model of incremental pure premiums is a nonlinear regression model. Nonlinear regression models are statistical models of the form:

$$y = f\left(\vec{x}, \vec{\theta}\right) + \varepsilon \qquad (2.2.1)$$

In (2.2.1), $\vec{x}$ is a vector of predictor variables, $\vec{\theta}$ is a vector of parameters, f is a nonlinear function, and ε is a normal random variable with mean 0. Usually, ε is assumed to have a constant variance σ^2. If the variance of the error term is not constant, a weight function that is inversely proportional to the variance may be specified.

The parameters of a nonlinear regression model are estimated by solving the normal equations. This usually requires using a numerical method such as the Gauss-Newton algorithm.

There are many commercial statistical software packages available that will perform the calculations and also provide approximate confidence intervals for the parameters and for predicted observations. The SAS system was used to perform the calculations to estimate confidence intervals for the model parameters and predicted IBNR.

We fit the following model to the incremental incurred pure premium data:

$$y = \left[\alpha \exp(\beta x) + \gamma \exp(\delta x)\right] + \frac{1}{w}\varepsilon \qquad (2.2.2)$$

where $y = \frac{L_{t,x}}{c_t} \cdot \exp(ru)$ is the incremental pure premium for accident quarter t in development age interval x adjusted for loss trend. u represents the calendar quarter. r is the loss trend. α , β , γ , and δ are the model parameters. $w = x^{1.5} \cdot c_t$ is the weight function. This weight function was selected based on an analysis of the residuals from an unweighted regression model.

We assume here that the trend r has been estimated by some reasonable method and that we have confidence interval estimates for the trend factors that we will apply to the IBNR estimates. The confidence intervals for the IBNR estimates will be adjusted to reflect the uncertainty in the trend factors.

It is tempting to include loss trend as a fifth parameter in the model in order to obtain prediction intervals for trend-adjusted IBNR directly. The resulting model equation would be

$$y = \left[\alpha \exp(\beta x) + \gamma \exp(\delta x)\right] \cdot \exp(ru) + \frac{1}{w}\varepsilon$$

Unfortunately, there are two problems with this model. One is that the model sometimes produces unrealistic estimates of trend due to a lack of credibility. The other problem is that we would be extrapolating the model instead of interpolating it. Extrapolation can be misleading even in the case of linear models and it is strongly discouraged in the case of nonlinear models. Of course, we need to extrapolate the trend factors but there are mathematically sound time series models available for this purpose.

2.3 Estimation of Parameters

The SAS system used the Gauss-Newton method to estimate the least squares estimates of the model parameters. The following presentation of the mathematics of the Gauss-Newton method is based on Seber and Wild [5].

To estimate the least squares parameters, we need to minimize the sum of squared errors of the n observations:

$$S(\vec{\theta}) = \sum_{i=1}^{n} \left[y_i - f(x_i; \vec{\theta})\right]^2 \tag{2.3.1}$$

In the case of our model, $\vec{\theta} = (\alpha, \beta, \gamma, \delta)$ and $f(x; \vec{\theta}) = \alpha \exp(\beta x) + \gamma \exp(\delta x)$. We find the minimum of $S(\vec{\theta})$ by setting all of its partial derivatives to 0.

Minimizing the sum of squared errors is a straightforward procedure for linear models but when f is nonlinear we must use numerical methods to estimate the parameters. One commonly used method is the Gauss-Newton method which works well in the case of normally distributed residuals.

We define the following matrices: $F(\bar{\theta}) = \left[\left(\dfrac{\partial f(x_i;\bar{\theta})}{\partial \theta_j}\right)\right]$ and

$\mathbf{f}(\bar{\theta}) = \left(f(x_1;\bar{\theta}), f(x_2;\bar{\theta}), \ldots, f(x_n;\bar{\theta})\right)'$.

F is an $n \times p$ matrix where n is the number of observations and p is the number of parameters. $\mathbf{f}(\bar{\theta})$ has dimension $n \times 1$.

Suppose $\theta^{(a)}$ is an approximation to $\bar{\theta}$. We approximate $\mathbf{f}(\bar{\theta})$ by the first order terms of its Taylor series in a small neighborhood near $\theta^{(a)}$:

$$\mathrm{f}(\bar{\theta}) \approx \mathrm{f}(\theta^{(a)}) + F(\bar{\theta} - \theta^{(a)}) \tag{2.3.2}$$

The residual vector is $r(\bar{\theta}) = y - \mathrm{f}(\theta^{(a)}) \approx r(\theta^{(a)}) - F(\bar{\theta} - \theta^{(a)})$. Substituting $S(\bar{\theta}) = r'(\bar{\theta})r(\bar{\theta})$ leads to

$$S(\bar{\theta}) \approx r'(\theta^{(a)})r(\theta^{(a)}) - 2r'(\theta^{(a)})F(\bar{\theta} - \theta^{(a)}) + (\bar{\theta} - \theta^{(a)})' F'(\bar{\theta} - \theta^{(a)})F(\bar{\theta} - \theta^{(a)}) \tag{2.3.3}$$

The right hand side of (2.3.3) is minimized with respect to $\bar{\theta}$ when

$$\bar{\theta} - \theta^{(a)} = F'(\bar{\theta} - \theta^{(a)})F(\bar{\theta} - \theta^{(a)})r(\theta^{(a)}) = \delta^{(a)}$$

This produces iterative approximations of $\theta^{(a)}$:

$$\theta^{(a+1)} = \theta^{(a)} + \delta^{(a)} \tag{2.3.4}$$

To use the Gauss-Newton method, one must provide $\theta^{(0)}$, the initial approximation to $\bar{\theta}$. The algorithm will converge provided the first approximation is sufficiently close to the fitted value, $\hat{\theta}$.

After fitting data to the model presented in Section 2.2, we estimated confidence intervals for the parameters and for the predicted observations. Seber and Wild [5] present formulas for approximate confidence intervals for the model parameters and for a predicted observation.

The 95% confidence interval for parameter θ_i is given by

$$\hat{\theta}_i \pm (s \cdot c_{ii})^{1/2} \cdot t(N-P,.025) \qquad (2.3.5)$$

where s^2 is the mean square error and c_{ii} is the i^{th} diagonal element of $(F'WF)^{-1}$.

The 95% confidence interval for a predicted observation corresponding to age x_i is given by

$$\hat{y}_i \pm s \cdot \left(\frac{1}{w_i} + f_i'(F'WF)^{-1} f_i \right)^{1/2} \cdot t(N-P,.025) \qquad (2.3.6)$$

where f_i is the i^{th} row of F, i.e. the vector of estimated first derivatives evaluated at x_i and W is a $N \times N$ matrix with w_i as the i^{th} diagonal entry and all other entries equal to 0. $t(N-P,.025)$ is the value of Student's t distribution for $N-P$ degrees of freedom and probability .025.

The confidence intervals for predicted observations can be used to produce a confidence interval for IBNR. Based on the assumption that the incremental pure premiums for different development intervals are independent, the variance of IBNR pure premium is the sum of the variances of the incremental pure premiums for the remaining development intervals. From equation (2.3.6) we see that the variance of the incremental pure premium for one development interval is $s^2 \cdot \left(\frac{1}{w_i} + f_i'(F'WF)^{-1} f_i \right)$. The expected value of IBNR pure premium is the sum of the expected incremental pure premiums.

3. RESULTS AND DISCUSSION

The model presented in section 2 was fitted to the data presented in section 1. Only data for the latest 20 calendar quarters was used to estimate parameters. This is consistent with common actuarial practice of using recent calendar quarters rather than all of the available data so that predictions are responsive to recent changes in development patterns. We also used only data for age>1 since we do not need to estimate IBNR for that age interval. Thus, 590 observations were used to fit the model.

We used the estimated parameters to produce confidence interval estimates of IBNR for each accident quarter.

In section 3.1 we will show confidence intervals for the estimated parameters. The confidence intervals for predicted IBNR will be presented in section 3.2. In section 3.3 we present an analysis of the residuals.

3.1 Confidence Interval Estimates of Parameters

The estimated parameters and standard errors for our simulated data are:

$$\begin{aligned} \hat{\alpha} &= 3.1994, \ s(\hat{\alpha}) = 0.5807 \\ \hat{\beta} &= -0.0754, \ s(\hat{\beta}) = 0.0096 \\ \hat{\gamma} &= 29.4446, \ s(\hat{\gamma}) = 5.5549 \\ \hat{\delta} &= -0.5480, \ s(\hat{\delta}) = 0.0767 \end{aligned} \quad (3.1.1)$$

A 95% confidence interval for each parameter is of the form $\left(\hat{\theta} - s(\hat{\theta})t(.025, n-p), \hat{\theta} + s(\hat{\theta})t(.025, n-p)\right)$. There were 590 observations and we estimated 4 parameters. The resulting confidence intervals are:

$$\begin{aligned} \hat{\alpha}&: \ (2.0596, 4.3392) \\ \hat{\beta}&: \ (-0.0942, -0.0566) \\ \hat{\gamma}&: \ (18.5334, 40.3557) \\ \hat{\delta}&: \ (-0.6986, -0.3974) \end{aligned} \quad (3.1.2)$$

The Mean Square Error from the estimation is 2,987,236.

An advantage of having confidence interval estimates of the parameters is that when more data becomes available, we can test whether the current parameters should be rejected. We would reject the current estimates only if the new estimates lie outside the intervals in (3.1.2). This procedure will lead to more stable estimates of ultimate losses and IBNR.

3.2 Confidence Interval Estimates of IBNR

The estimation of IBNR was performed in two steps. First, we use equation (2.3.6) to calculate an expected value and standard error for the incremental pure premium for each development quarter until age 40 (for simplification, we assume that this is ultimate). This results in deflated IBNR estimates. The second step is to find a confidence interval for the inflation adjusted IBNR. This was done using a simulation.

<u>Step 1: Predicted incremental pure premiums</u>

The expected value of each predicted incremental pure premium is calculated by substituting the estimated parameters from (3.1.1) into the model equation, $\hat{y} = \hat{\alpha} \cdot \exp(\hat{\beta} x) + \hat{\gamma} \cdot \exp(\hat{\delta} x)$ where x is the age of the development quarter.

In order to estimate the standard errors, we need the matrix defined in section 2.3:

$$(F'WF)^{-1} = \begin{bmatrix} 0.000000112875 & -0.000000001771 & 0.000000486556 & -0.000000010876 \\ -0.000000001771 & 0.000000000031 & -0.000000006728 & 0.000000000158 \\ 0.000000486556 & -0.000000006728 & 0.000010329411 & -0.000000127258 \\ -0.000000010876 & 0.000000000158 & -0.000000127258 & 0.000000001968 \end{bmatrix}$$

As an example, we will calculate the IBNR prediction interval for accident quarter 2. $x = 40$ for the remaining development quarter. The expected IBNR pure premium is

$$3.1994 \times \exp(-0.0754 \times 40) + 29.446 \times \exp(-0.5480 \times 40)$$
$$= .15676.$$

We will need the above matrix and the derivatives of the model function evaluated at $x = 40$ to calculate the standard error of the predicted IBNR. The derivatives are:

$$\frac{\partial f}{\partial \alpha} = \exp(\beta x)$$
$$\frac{\partial f}{\partial \beta} = \alpha x \cdot \exp(\beta x)$$
$$\frac{\partial f}{\partial \gamma} = \exp(\delta x)$$
$$\frac{\partial f}{\partial \delta} = \gamma x \cdot \exp(\delta x)$$

Evaluating the derivatives at age 40 and the estimated parameters, we obtain:

$$\frac{\partial f}{\partial \alpha}(40) = 0.0490$$
$$\frac{\partial f}{\partial \beta}(40) = 6.2704$$
$$\frac{\partial f}{\partial \gamma}(40) = 3.02 \times 10^{-10}$$
$$\frac{\partial f}{\partial \delta}(40) = 3.56 \times 10^{-7}$$

Let the element in the jth row and kth column of $(F'WF)^{-1}$ be denoted m_{jk}. We calculate $f_i'(F'WF)^{-1} f_i$ from (2.3.6) as:

$$\begin{aligned}
&f_i'\left(F'WF\right)^{-1}f_i\\
&=m_{11}\frac{\partial f}{\partial \alpha}\cdot\frac{\partial f}{\partial \alpha}+m_{12}\frac{\partial f}{\partial \alpha}\cdot\frac{\partial f}{\partial \beta}+m_{13}\frac{\partial f}{\partial \alpha}\cdot\frac{\partial f}{\partial \gamma}+m_{14}\frac{\partial f}{\partial \alpha}\cdot\frac{\partial f}{\partial \delta}\\
&+m_{21}\frac{\partial f}{\partial \beta}\cdot\frac{\partial f}{\partial \alpha}+m_{22}\frac{\partial f}{\partial \beta}\cdot\frac{\partial f}{\partial \beta}+m_{23}\frac{\partial f}{\partial \beta}\cdot\frac{\partial f}{\partial \gamma}+m_{24}\frac{\partial f}{\partial \beta}\cdot\frac{\partial f}{\partial \delta}\\
&+m_{31}\frac{\partial f}{\partial \gamma}\cdot\frac{\partial f}{\partial \alpha}+m_{32}\frac{\partial f}{\partial \gamma}\cdot\frac{\partial f}{\partial \beta}+m_{33}\frac{\partial f}{\partial \gamma}\cdot\frac{\partial f}{\partial \gamma}+m_{34}\frac{\partial f}{\partial \gamma}\cdot\frac{\partial f}{\partial \delta}\\
&+m_{41}\frac{\partial f}{\partial \delta}\cdot\frac{\partial f}{\partial \alpha}+m_{42}\frac{\partial f}{\partial \delta}\cdot\frac{\partial f}{\partial \beta}+m_{43}\frac{\partial f}{\partial \delta}\cdot\frac{\partial f}{\partial \gamma}+m_{44}\frac{\partial f}{\partial \delta}\cdot\frac{\partial f}{\partial \delta}\\
&=3.88134\times 10^{-10}
\end{aligned}$$

The weight is $w_i = c_i x_i^{1.5} = 50,801\cdot 40^{1.5} = 12,851,749$. The mean square error is 2,987,236. $t\left(586,.05/2\right)=1.96402$. Substituting this information into equation (2.3.6) we obtain 0.94925 as the width of the 95% confidence interval for the IBNR for accident quarter 2. Thus, the confidence interval for the IBNR pure premium is $\left(-0.79249,1.10601\right)$. The confidence interval for the dollars of IBNR is $\left(-40259,56186\right)$. For an accident quarter with more than one development quarter remaining, we would need to repeat these calculations for each remaining development quarter and sum the estimated expected values. Next, the estimated IBNR will be adjusted for loss trend.

Step 2: Including trend

Because we fitted the model to losses trended to the current calendar quarter, the dollars need to be adjusted to future cost levels. We also need to adjust the width of the confidence intervals for the uncertainty in the trend.

The trend was estimated from a time series method. The estimated trend had a mean of .005 per calendar quarter with a standard deviation of $.004\sqrt{t}$ where t is the number of quarters projected. We assume that the trend process is a Geometric Brownian Motion.

There are a number of ways to find the simultaneous confidence interval for loss development and trend. For example, we could use a Bonferronni confidence interval but this would result in an excessively wide confidence interval. Instead, we performed a simulation to estimate the variance of inflation adjusted IBNR.

We simulated incremental pure premiums before adjusting for inflation from a normal distribution with mean $\hat{\alpha}\cdot\exp\left(\hat{\beta}x\right)+\hat{\gamma}\cdot\exp\left(\hat{\delta}x\right)$ and variance given by equation (2.3.6). We simulated trend factors for each calendar quarter as a Geometric Brownian Motion with drift .005 and volatility .004. The inflation adjusted incremental pure premiums were calculated as the product of the simulated unadjusted pure premiums and the simulated trend factors. Next, the incremental pure premiums were summed over all remaining development quarters to obtain IBNR pure premium. The simulation was repeated 10000 times and the mean and standard deviation of the IBNR was calculated for each accident quarter. IBNR pure premium multiplied by exposures produces IBNR dollars.

Table 3.2.1 shows the results of the simulation. The actual IBNR is the difference between the age 40 evaluation (which we treat as ultimate here) and the evaluation at the end of the 40th calendar quarter from the simulated loss development data. The expected total IBNR is 30105084. The standard deviation of the total IBNR is 1350093. The 95% confidence interval for total IBNR is (27458951 , 32751218). The actual total IBNR is 30120821.

Table 3.2.1

Accident Quarter	Exposures	Expected Value	Standard Deviation	95% Confidence Interval Lower	Upper	Actual IBNR
2	50,801	8,190	24,518	-39,864	56,244	-3,686
3	51,187	16,643	35,835	-53,593	86,879	20,450
4	51,146	26,310	44,192	-60,304	112,925	11,254
5	51,527	36,541	51,941	-65,262	138,344	73,738
6	52,348	49,099	58,839	-66,225	164,422	98,397
7	52,480	61,528	65,232	-66,325	189,381	37,099
8	53,148	75,340	71,800	-65,385	216,065	156,305
9	53,924	91,671	78,552	-62,287	245,629	237,876
10	54,403	109,065	85,433	-58,380	276,511	-95,408
11	54,557	124,874	91,436	-54,338	304,086	384,465
12	55,083	144,622	96,258	-44,040	333,284	260,118
13	55,292	168,450	103,341	-34,095	370,995	299,600
14	55,899	192,189	108,233	-19,944	404,322	175,632
15	56,067	215,948	115,108	-9,659	441,555	3,570
16	57,025	247,643	123,187	6,201	489,086	237,988
17	57,071	279,736	129,481	25,957	533,515	224,736
18	57,317	311,248	134,933	46,784	575,712	268,971
19	57,907	346,819	143,714	65,144	628,493	712,233
20	58,285	388,878	149,405	96,050	681,706	428,225
21	59,096	433,974	157,772	124,746	743,202	819,832
22	59,193	479,592	165,473	155,270	803,915	930,364
23	59,524	530,342	173,337	190,607	870,076	564,488
24	59,745	583,879	177,894	235,213	932,546	412,411
25	60,427	645,944	188,083	277,309	1,014,580	421,418
26	60,155	705,701	195,557	322,416	1,088,985	699,647
27	60,568	776,239	207,953	368,659	1,183,819	794,518
28	60,708	852,632	215,059	431,123	1,274,140	995,212
29	60,262	925,896	222,578	489,652	1,362,140	944,400
30	60,606	1,012,197	233,755	554,046	1,470,349	945,867
31	60,580	1,109,304	251,368	616,632	1,601,976	1,084,176
32	60,648	1,213,637	258,802	706,395	1,720,879	1,703,397
33	61,159	1,344,114	277,079	801,049	1,887,178	1,107,447
34	61,462	1,492,000	292,032	919,627	2,064,372	1,133,824
35	61,934	1,660,873	312,021	1,049,324	2,272,423	1,882,576
36	61,716	1,858,275	333,112	1,205,388	2,511,161	1,567,491
37	61,837	2,123,409	361,113	1,415,642	2,831,177	1,962,887
38	62,285	2,514,004	394,000	1,741,778	3,286,231	1,938,616
39	62,728	3,055,695	450,062	2,173,589	3,937,801	2,836,989
40	63,180	3,892,584	522,958	2,867,605	4,917,563	3,843,696

3.3 Analysis of Residuals

It is important to examine the residuals from a regression model to check the consistency of the data with the assumptions of the model. In this section we will look at plots of the residuals to look for patterns. We will also see the results of a Shapiro-Wilk test of normality, a histogram and a probability plot.

These tests are shown for demonstration purposes only. The data used to demonstrate the methodology in this paper is simulated and will pass the normality test. Real data might not pass tests of normality but if the deviation from normality is not too extreme, then the estimated confidence intervals are still reasonable.

The unmodified residuals, $r_i = y_i - \hat{y}_i$, do not have constant variance because the data do not have constant variance. The tests will be performed on studentized residuals, defined as $r_i / std(r_i)$. Seber and Wild [5] show the following formula for the standard errors of the residuals.

$$std(r_i) = s \cdot \left(\frac{1}{w_i} - f_i' \left(F'WF \right)^{-1} f_i \right) \qquad (3.3.1)$$

Exhibits 3.3.1 through 3.3.3 show the scatter plots of the studentized residuals against predicted value, development age, and calendar quarter. The plots do not show any obvious patterns and the studentized residuals seem to have constant variance. Thus, the weight function appears to be appropriate and there does not appear to be any reason to modify the model.

Exhibit 3.3.4 is a histogram of the studentized residuals. Exhibit 3.3.5 is a normal probability plot (calculated using methodology from [6]). The shape of the histogram appears to be consistent with a normal distribution. The probability plot is nearly linear which supports the assumption that the residuals have a normal distribution. A Shapiro-Wilk test was performed on the residuals and produced a statistic of 0.9983 with a p-value of 0.8445. Thus, we cannot reject the hypothesis that the residuals have a normal distribution.

Exhibit 3.3.1

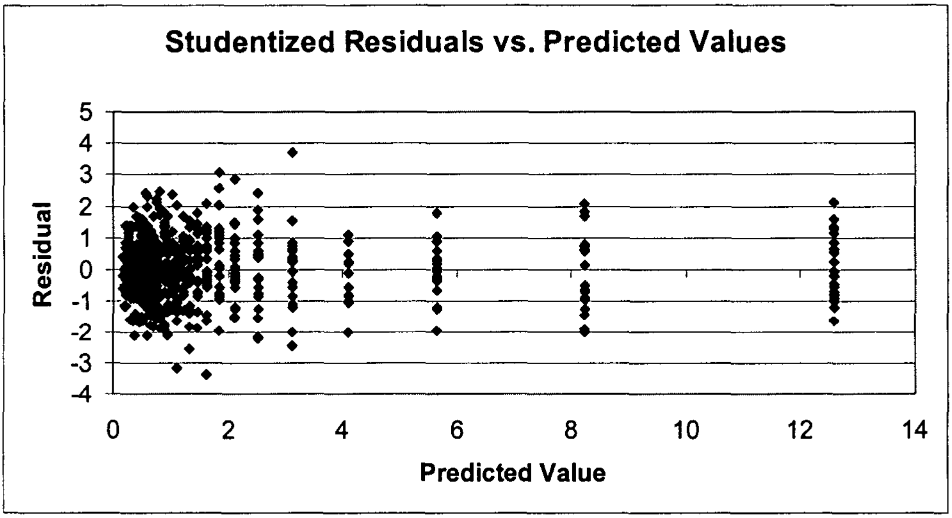


Exhibit 3.3.2

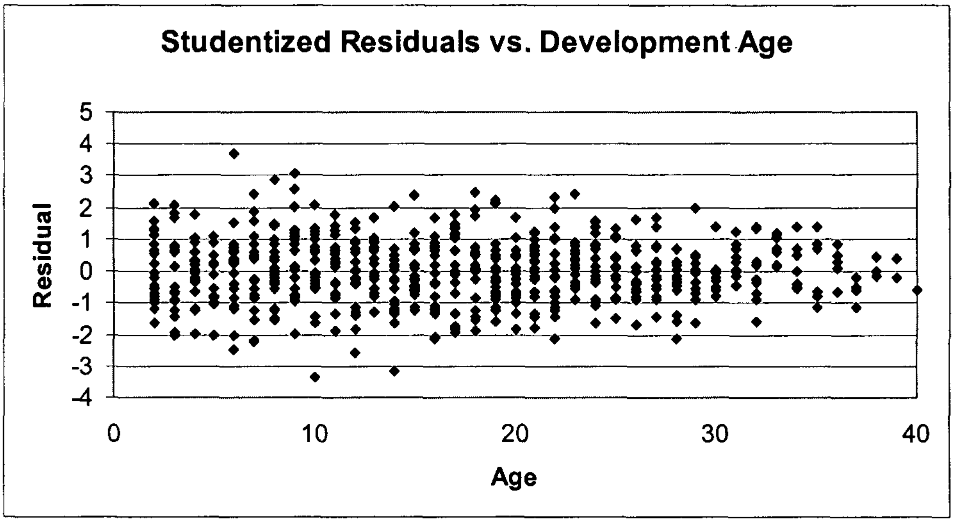


Exhibit 3.3.3

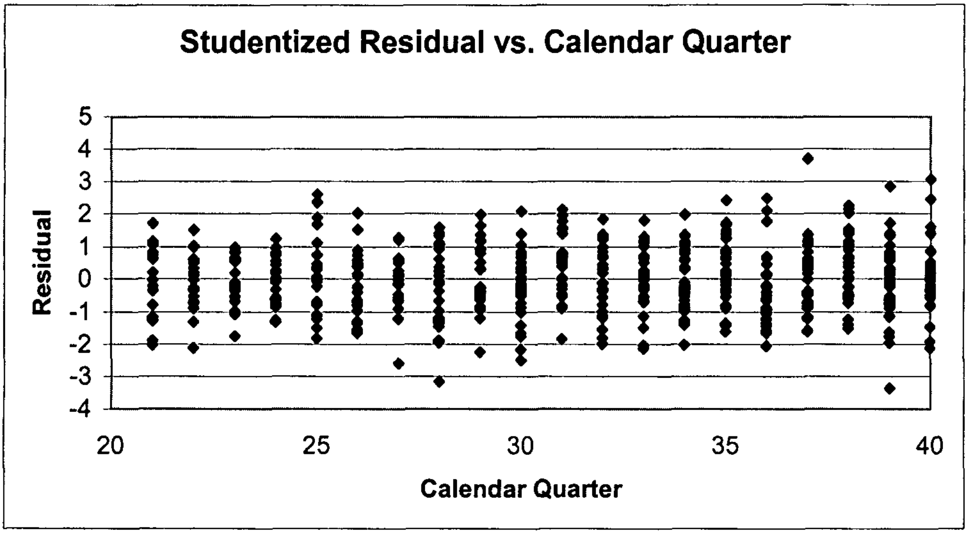


Exhibit 3.3.4

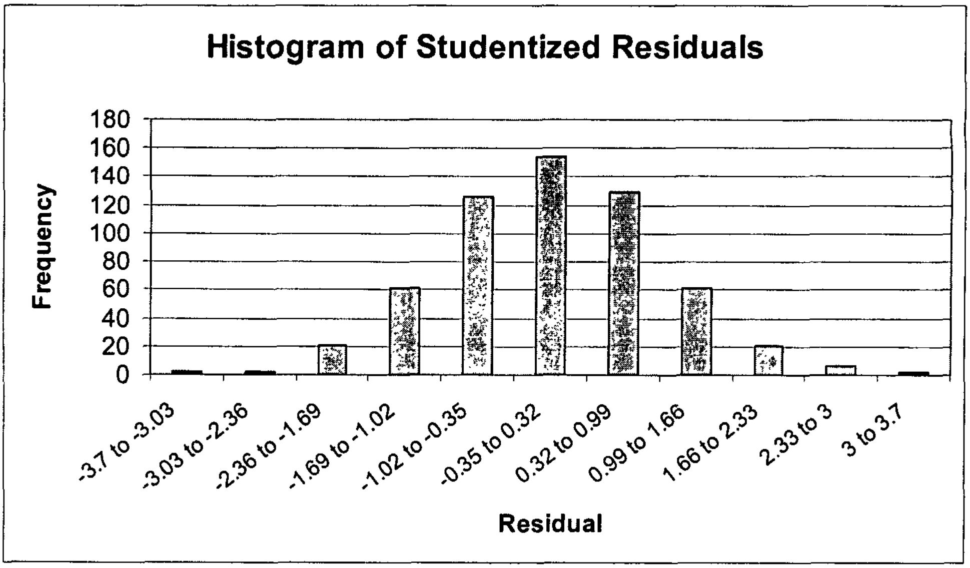
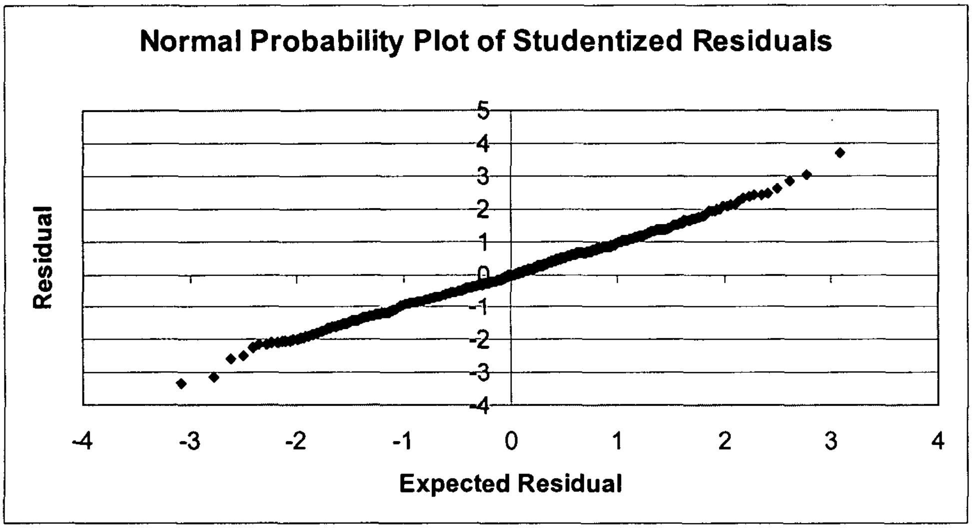


Exhibit 3.3.5



4. CONCLUSIONS

Our model has satisfied the objective stated in section 1.2. The model could be fitted either to paid or case incurred losses. Since the observations are incremental pure premiums and the weights are a function of exposures, the model makes appropriate adjustments for changing levels of exposures. By using nonlinear regression, we have avoided the need for a log link and we have been able to keep negative observations in the data. The model appears to produce unbiased estimates of IBNR and reasonable 95% confidence intervals.

The plots displayed in section 3.3 indicate that incremental pure premiums have an approximately normal distribution.

The assumptions we made work well with auto bodily injury data. We have assumed that the data satisfy the usual assumptions of nonlinear regression models including independent normal errors. We have also used a functional form that fits our data well but might not fit other lines. We would like to close with a few suggestions for fitting models to other lines.

The assumption of normal errors should be reasonable for high frequency lines of business. The assumption that the errors are uncorrelated should also be reasonable most of the time. If these assumptions are rejected, there are nonlinear models that may be used. Seber and Wild [5] discuss models with non-normal and autocorrelated errors.

Seber and Wild [5] has a chapter on growth models which lists many functional forms other than the form presented in this paper. Some of these models might fit the pure premiums of other lines of business. Some of the models could be applied to cumulative instead of incremental data.

Another class of models that will fit pure premium development is generalized linear models. In this type of model, the development age interval could be represented as a categorical variable. These models would allow the analyst to consider a great variety of error distributions and error correlation structures. One drawback to this approach is that there are more parameters to estimate which means that the confidence interval for IBNR will be wider. Dobson [7] is an excellent reference on generalized linear models.

Acknowledgment

The author acknowledges John Grogan for suggesting the function used to fit the data and Dale Porfilio for support and encouragement to submit this paper for publication.

5. REFERENCES

[1] D. Murphy, Unbiased Loss Development Factors, Casualty Actuarial Society Forum, 1994.
[2] G. Barnett and B. Zehnwirth, Best Estimates for Reserves, Casualty Actuarial Society Forum, 1998.
[3] J. Stanard, A Simulation Test of Prediction Errors of Loss Reserve Techniques, Proceedings of the Casualty Actuarial Society, 1985.
[4] P. Narayan, T. Warthen, A Comparative Study of the Performance of Loss Reserving Methods Through Simulation, Casualty Actuarial Society Forum, 1997.
[5] G. Seber, C. Wild, Nonlinear Regression, 1989 Wiley and Sons
[6] J. Neter, M. Kutner, C. Nachtsheim, W. Wasserman, Applied Linear Statistical Models, 1996, McGraw-Hill.
[7] A. Dobson, An Introduction to Generalized Linear Models Second Edition, 2001, Chapman and Hall/CRC

Appendix A – Simulated Loss Development Data – Earned Exposures and Incremental Case Incurred Losses

Accident		Development Quarter									
Quarter	Exposures	1	2	3	4	5	6	7	8	9	10
1	50,333	1,244,722	468,457	283,193	68,634	101,440	125,515	159,525	68,010	64,419	40,378
2	50,801	1,417,101	587,121	337,664	95,841	243,190	65,263	94,623	47,772	188,347	31,533
3	51,187	1,143,473	502,816	483,911	174,352	70,477	140,000	86,071	93,975	58,041	-13,132
4	51,146	1,055,290	1,213,499	496,421	393,358	173,547	116,010	101,531	115,414	31,102	64,744
5	51,527	1,508,450	834,730	302,691	409,622	423,103	143,476	67,506	146,727	146,017	38,057
6	52,348	1,192,515	1,074,133	478,010	172,336	69,719	-9,197	206,481	221,841	-2,220	170,525
7	52,480	1,067,318	438,952	368,962	214,718	274,665	86,113	39,208	121,683	62,838	8,398
8	53,148	758,275	513,455	255,088	159,199	60,798	102,129	97,269	107,077	134,067	86,292
9	53,924	1,664,156	31,358	491,390	196,506	202,088	27,387	48,404	-41,016	77,453	57,713
10	54,403	1,537,825	422,697	301,334	210,821	254,271	149,386	225,652	104,121	81,082	76,196
11	54,557	2,026,667	738,848	252,783	251,618	193,254	96,790	170,517	106,026	-121,070	152,682
12	55,083	1,296,855	495,927	439,098	125,722	96,000	111,872	213,356	188,106	123,309	67,820
13	55,292	1,995,401	821,508	552,503	226,522	145,695	389,146	333,936	118,922	136,678	104,544
14	55,899	2,078,843	786,272	529,533	329,974	179,953	135,260	180,141	106,218	165,608	63,277
15	56,067	1,952,667	859,868	632,198	264,522	231,370	216,887	22,205	117,571	81,594	149,627
16	57,025	1,258,033	650,068	387,309	183,986	152,797	244,063	68,088	103,095	56,465	80,728
17	57,071	1,627,621	320,911	303,800	327,057	236,332	161,152	205,081	147,898	288,069	127,213
18	57,317	1,681,507	446,643	359,407	248,809	270,162	229,530	58,483	-7,112	246,925	85,944
19	57,907	2,508,300	1,018,661	99,969	436,712	156,983	241,768	303,837	-9,729	194,850	181,037
20	58,285	1,238,641	812,792	542,250	329,100	246,551	61,085	173,928	17,813	183,213	64,235
21	59,096	1,793,043	482,793	546,164	313,044	353,857	327,614	90,275	235,255	32,150	-8,168
22	59,193	1,433,225	532,545	589,099	306,945	330,835	50,915	285,934	84,085	48,543	144,367
23	59,524	1,516,012	753,758	581,957	365,421	217,070	239,708	-63,906	191,485	107,079	181,666
24	59,745	1,803,164	519,924	295,180	225,283	222,089	122,650	-57,787	170,330	46,008	56,351
25	60,427	1,347,360	328,992	357,630	157,580	135,900	-80,274	117,487	208,666	121,095	179,925
26	60,155	810,643	604,364	214,555	155,748	114,459	93,877	397	95,471	40,225	82,597
27	60,568	1,850,892	980,308	613,338	369,051	298,488	272,379	196,454	107,684	175,612	251,126
28	60,708	3,006,298	1,044,056	843,024	581,194	261,303	209,512	255,504	255,482	95,327	-22,389
29	60,262	986,119	672,498	590,640	43,019	-8,877	-32,562	119,151	17,117	205,731	144,380
30	60,606	2,630,383	1,101,593	805,938	238,565	228,041	253,614	194,571	157,225	190,266	-153,391
31	60,580	1,515,313	601,512	511,685	304,718	142,590	143,733	206,446	79,617	-47,845	62,197
32	60,648	3,517,516	1,048,147	260,427	466,379	114,732	589,058	125,985	381,048	361,346	9,249
33	61,159	1,673,500	513,290	333,639	305,302	308,506	232,796	81,397	104,474	119,655	-44,314
34	61,462	1,207,813	739,162	524,302	392,092	363,797	230,703	398,177	149,413	19,749	166,569
35	61,934	2,202,629	528,671	378,846	162,033	150,420	225,454	174,285	201,873	193,199	7,944
36	61,716	1,051,422	470,986	415,687	433,785	152,892	263,161	151,337	-23,299	23,005	183,878
37	61,837	2,355,630	1,302,388	824,821	307,806	79,979	117,519	208,110	162,036	207,896	141,836
38	62,285	2,016,667	990,682	153,863	290,379	-10,658	120,430	7,588	208,320	71,027	123,875
39	62,728	1,468,675	925,175	157,502	266,297	252,103	380,364	226,500	76,920	43,621	26,222
40	63,180	1,952,713	712,475	446,253	551,239	361,511	276,575	355,898	-8,263	66,140	96,505

Accident		Development Quarter									
Quarter	Exposures	11	12	13	14	15	16	17	18	19	20
1	50,333	55,132	56,286	39,158	54,447	32,780	64,401	47,793	-30,285	60,984	-1,923
2	50,801	97,899	120,746	66,487	16,116	50,122	42,810	11,512	77,062	7,181	72,151
3	51,187	91,825	66,621	-46,239	10,535	-6,023	58,325	4,497	19,837	5,313	7,916
4	51,146	166,882	134,068	17,951	101,782	64,696	64,331	-98,606	-37,817	60,282	-3,207
5	51,527	119,089	-11,801	42,615	-73,816	72,867	-9,629	91,001	47,562	22,359	-16,392
6	52,348	16,662	112,356	-26,328	108,397	152,916	95,642	107,603	30,439	10,858	99,144
7	52,480	95,445	53,315	65,328	34,802	-4,403	-49,160	-32,400	40,083	-11,567	51,725
8	53,148	166,630	-14,316	19,746	-8,727	25,872	-1,430	53,704	73,004	-26,364	61,378
9	53,924	39,927	78,793	150,847	42,016	83,246	57,028	13,857	62,675	59,835	39,645
10	54,403	52,923	759	-8,435	45,026	37,059	126,682	-15,224	47,275	54,888	10,951
11	54,557	118,521	48,258	97,730	-6,982	168,831	30,198	100,201	-11,399	27,865	57,843
12	55,083	52,455	117,838	19,476	61,249	42,336	-6,884	-42,245	5,514	40,494	-37,779
13	55,292	108,005	101,991	50,016	-10,580	23,714	-14,118	101,221	66,648	131,158	33,186
14	55,899	31,057	-37,318	104,948	67,958	-7,386	95,217	-34,104	130,890	-6,796	28,246
15	56,067	94,894	-5,199	55,050	-107,620	33,005	35,708	113,029	-23,751	33,324	82,253
16	57,025	34,439	-82,833	-7,708	78,608	49,459	91,763	-36,547	48,994	3,417	39,090
17	57,071	119,713	127,702	120,055	98,655	33,349	36,053	79,890	72,189	80,971	2,954
18	57,317	-42,024	90,633	88,686	89,706	102,187	89,757	114,280	125,545	21,101	58,920
19	57,907	26,467	57,494	37,776	-1,643	120,996	-11,362	45,765	162,032	-7,833	9,218
20	58,285	54,460	168,172	60,942	33,469	43,582	95,786	136,815	7,129	146,101	8,598
21	59,096	108,307	118,119	130,671	12,719	66,407	-49,728	103,805	-23,377	13,446	24,913
22	59,193	170,747	121,252	122,821	-25,894	96,750	89,657	52,945	49,776	50,822	120,953
23	59,524	205,422	-10,765	87,080	7,915	20,942	82,590	86,042	-20,650	86,539	3,828
24	59,745	57,883	-9,148	57,563	76,990	72,755	53,851	37,035	75,261	-8,698	-11,311
25	60,427	98,081	67,455	30,353	184,721	26,902	16,334	62,868	80,394	2,462	19,806
26	60,155	-5,192	53,749	114,555	37,095	35,334	7,140	62,927	39,025	48,072	-549
27	60,568	166,693	82,674	70,339	70,978	79,356	-64,843	22,823	76,726	79,929	35,373
28	60,708	188,567	168,677	129,179	68,308	47,794	96,191	128,506	49,566	-28,134	58,887
29	60,262	104,592	80,640	95,315	34,245	48,974	81,604	39,399	32,106	55,537	-46,734
30	60,606	99,495	58,402	48,408	56,793	-4,451	19,091	-5,279	19,722	53,159	39,365
31	60,580	139,467	180,552	45,404	72,414	3,441	38,563	126,913	50,865	37,834	56,248
32	60,648	79,096	136,515	178,105	91,579	20,394	100,918	56,855	43,922	-7,463	34,194
33	61,159	54,143	-81,040	20,949	1,608	60,381	111,910	13,739	102,704	27,132	104,321
34	61,462	97,706	108,206	14,850	59,003	54,189	69,831	65,128	23,821	43,958	-11,047
35	61,934	59,384	69,087	148,809	136,211	71,394	4,055	126,075	52,993	84,082	56,630
36	61,716	139,013	56,147	92,609	125,058	7,067	90,151	101,031	27,566	-17,295	58,929
37	61,837	4,048	253,659	157,930	58,979	100,435	19,044	-20,740	51,891	112,978	-55,242
38	62,285	145,307	110,498	149,791	87,189	164,906	27,941	11,832	73,887	77,094	14,150
39	62,728	133,332	164,332	-10,897	106,455	136,006	141,784	83,994	79,801	71,479	-24,821
40	63,180	120,876	1,869	149,325	-46,560	52,798	85,751	68,371	100,236	-48,006	133,049

Accident Quarter	Exposures	Development Quarter 21	22	23	24	25	26	27	28	29	30
1	50,333	1,844	47,936	8,281	51,891	37,771	-4,045	27,774	14,035	871	56,156
2	50,801	21,390	63,117	23,327	-12,069	14,680	41,491	15,434	11,107	19,559	10,661
3	51,187	54,081	7,552	40,110	30,494	-5,797	730	29,749	25,467	30,419	7,913
4	51,146	76,821	112,076	30,277	20,160	57,926	74,676	-23,786	20,074	6,919	18,023
5	51,527	43,306	960	47,240	-7,478	-5,993	-31,465	45,344	28,740	26,218	10,956
6	52,348	8,403	24,816	14,994	66,326	7,418	22,099	3,600	-46,942	78,616	16,230
7	52,480	35,360	38,631	16,046	53,286	18,835	12,820	23,495	5,427	33,480	2,394
8	53,148	27,648	103,471	-2,524	-3,970	71,300	28,587	-3,460	9,452	-8,909	-6,737
9	53,924	54,178	71,192	59,018	52,434	-25,919	50,456	76,803	43,181	-2,099	17,733
10	54,403	33,351	24,839	42,521	26,870	17,470	10,409	-7,892	-29,828	2,882	200
11	54,557	65,973	-5,380	53,969	15,744	-3,427	4,913	8,390	-24,473	-32,538	62,557
12	55,083	63,685	-10,583	64,637	78,643	30,741	11,856	15,134	1,767	18,412	31,248
13	55,292	82,088	7,445	121,478	-32,097	41,168	47,156	49,125	9,276	44,273	23,737
14	55,899	50,701	-22,139	55,822	44,064	65,745	-5,697	71,653	59,301	-11,011	8,634
15	56,067	-18,731	-14,131	44,114	86,453	31,838	25,910	-15,473	17,799	-2,589	34,604
16	57,025	-22,239	89,127	13,948	20,393	-2,351	8,374	31,029	39,339	-8,451	1,222
17	57,071	57,279	36,946	39,534	90,362	14,387	-29,765	30,222	16,053	17,682	35,591
18	57,317	79,551	77,137	47,330	29,128	-40,416	46,964	9,795	23,656	43,627	-433
19	57,907	-36,342	-51,199	-629	38,859	-20,756	54,574	72,098	36,775	39,504	31,052
20	58,285	43,739	-8,134	54,269	25,913	-16,757	8,755	7,972	43,674	-3,448	64,314
21	59,096	75,294	-34,942	88,190	124,206	62,976	77,091	39,748	40,729	43,609	89,136
22	59,193	11,060	76,017	61,132	105,644	56,274	15,014	-2,897	80,213	53,917	118,331
23	59,524	29,866	45,800	38,868	68,925	7,687	-61,021	30,638	39,572	45,399	-11,739
24	59,745	46,041	33,062	16,682	40,849	-18,453	7,049	58,613	48,743	-17,040	8,158
25	60,427	25,489	-35,072	29,365	1,481	46,825	-43	39,986	80,497	51,650	-27,268
26	60,155	136,480	50,523	73,985	-15,999	21,991	43,033	32,821	8,902	29,994	41,090
27	60,568	-2,199	34,675	135,124	6,514	15,272	62,756	66,009	-10,230	-37,723	3,901
28	60,708	115,933	100,646	55,828	25,764	-3,515	9,366	-23,401	89,137	46,630	75,698
29	60,262	70,050	48,884	59,346	53,211	-3,141	6,048	29,235	13,746	38,350	43,614
30	60,606	100,823	-80,196	-23,695	19,793	20,686	-29,950	-5,204	99,580	36,328	56,872
31	60,580	42,546	19,448	19,949	-29,940	17,116	55,736	756	21,693	8,254	48,025
32	60,648	74,650	86,062	71,446	138,206	-8,941	75,564	27,495	84,913	-26,461	74,757
33	61,159	16,045	110,447	129,009	-45,715	68,666	7,394	20,046	33,159	7,386	18,884
34	61,462	4,309	-26,370	107,835	127,369	15,493	-50,769	-7,521	-25,623	-1,506	18,283
35	61,934	83,466	73,782	56,185	-32,328	-38,556	27,399	-11,618	54,166	26,555	-750
36	61,716	-27,140	93,574	66,551	13,086	30,072	-12,666	-11,496	-7,722	13,375	17,919
37	61,837	30,283	14,515	-30,671	-60,204	31,067	15,254	78,382	95,606	7,715	9,987
38	62,285	95,225	114,060	54,619	-67,884	7,563	-31,075	-36,590	9,379	78,245	14,113
39	62,728	4,900	-81	43,622	78,577	92,489	28,945	-16,724	67,108	17,473	-6,230
40	63,180	-21,730	80,710	55,218	15,476	39,584	3,858	18,112	22,462	13,209	33,635

Accident Quarter	Exposures	Development Quarter 31	32	33	34	35	36	37	38	39	40
1	50,333	34,276	3,835	43,711	-1,640	29,076	13,319	-2,630	8,678	3,188	-5,821
2	50,801	27,835	-5,203	29,996	-2,056	-5,783	18,323	-5,364	5,188	19,082	-3,686
3	51,187	17,345	51,332	18,327	1,471	34,161	24,299	-18,628	20,407	5,774	14,676
4	51,146	23,792	20,703	18,035	26,068	48,501	-5,982	5,236	30,317	-20,445	1,381
5	51,527	52,198	-30,522	43,073	32,051	-18,894	33,613	54,075	34,899	24,632	-39,868
6	52,348	35,800	-10,345	46,979	51,773	-9,409	-22,677	20,537	25,957	12,417	62,164
7	52,480	15,761	7,390	33,962	13,579	24,407	19,471	-3,988	-10,108	9,185	-1,869
8	53,148	3,334	24,196	20,303	24,861	40,319	12,113	-13,820	42,130	35,297	15,405
9	53,924	12,418	56,958	52,833	26,726	32,933	5,232	29,185	37,027	45,244	8,695
10	54,403	43,335	-31,285	-54,667	-16,646	15,567	-15,554	-9,313	-13,437	29,191	736
11	54,557	72,266	95,758	6,159	7,910	61,208	28,223	28,531	49,899	16,681	17,830
12	55,083	39,707	25,525	25,741	13,760	17,106	49,171	1,081	16,032	24,045	16,702
13	55,292	20,470	-17,379	26,954	5,100	45,000	38,726	-19,480	78,077	44,212	9,911
14	55,899	22,377	1,924	-9,766	47,354	18,223	454	6,891	20,907	2,586	7,758
15	56,067	-35,330	-18,262	-26,548	38,756	13,302	-9,275	-3,005	47,269	-34,312	-3,364
16	57,025	12,632	12,743	38,517	4,169	12,911	-1,785	-20,735	78,244	30,557	-779
17	57,071	38,585	45,591	11,315	-26,333	15,402	24,202	-25,715	11,080	9,528	36,913
18	57,317	20,486	28,330	4,385	-14,482	35,357	22,167	-4,089	2,423	1,184	60,888
19	57,907	68,328	82,874	121,115	-32,249	19,414	73,240	23,233	39,217	78,051	-12,467
20	58,285	42,726	28,751	58,837	7,617	7,172	72,846	1,544	-5,147	29,124	8,197
21	59,096	33,319	44,155	-39,904	29,104	50,092	10,920	13,243	23,352	35,124	14,389
22	59,193	-24,229	54,258	48,576	51,990	44,191	48,506	57,887	-6,941	-18,202	-21,329
23	59,524	48,463	18,459	7,095	13,631	6,314	16,901	46,450	-16,939	43,202	56,548
24	59,745	35,148	16,789	-7,315	-9,671	21,791	14,107	28,696	9,512	8,829	15,567
25	60,427	-16,598	26,696	11,564	14,065	-29,491	2,041	18,738	47,090	-8,041	-23,085
26	60,155	-8,915	52,310	915	813	56,718	-15,282	-26,165	20,384	20,458	18,977
27	60,568	22,553	42,332	9,009	-7,442	2,140	93,063	88,561	46,159	-5,060	-262
28	60,708	-14,237	65,453	25,751	12,368	-49,710	41,335	-49,919	-30,620	69,756	11,831
29	60,262	49,917	32,539	-6,986	48,452	15,625	28,630	15,743	23,349	-5,595	42,937
30	60,606	60,504	65,845	93,343	-27,623	-3,656	51,672	33,114	50,926	101,765	39,729
31	60,580	-47,046	46,028	24,302	56,096	-8,692	27,322	28,081	6,079	-23,284	20,008
32	60,648	63,634	122,152	-1,646	-37,185	-19,352	96,570	10,367	35,026	41,909	50,868
33	61,159	30,616	8,972	11,306	39,325	10,365	32,535	50,209	7,522	34,812	25,278
34	61,462	19,812	3,141	33,675	12,108	-21,363	18,639	44,897	46,331	-32,125	-14,164
35	61,934	-30,866	32,254	88,375	36,930	62,025	72,476	54,286	-50,512	12	-6,728
36	61,716	-35,620	-2,115	31,594	37,150	-2,481	26,166	14,732	19,708	19,340	5,106
37	61,837	35,752	60,881	-13,187	-21,121	39,280	-1,210	-23,822	11,761	42,508	39,751
38	62,285	-10,410	21,570	35,964	-13,033	-26,726	-20,093	-11,908	65,799	-11,499	-2,260
39	62,728	70,218	-25,147	46,379	5,606	39,349	-13,438	70,889	53,260	-18,834	-12,364
40	63,180	65,404	13,053	28,027	-40,448	-2,637	3,059	-7,238	41,295	-7,643	14,249

Appendix B – Simulation Model Used to Generate the Data

The simulation model used to generate the loss and exposure data is based on method 3 in Narayan [4] with some modifications. In this appendix, we will present an outline of the model and the SAS code used to produce the data. Note that the SAS program will not produce the same data every time it is run because the random number seeds were randomized.

Outline of the simulation methodology:

1. Initialize the values for exposures at 50,000 per quarter and the inflation index at unity.
2. For each of the 40 accident quarters:
   a. Generate a random number of exposures from a Brownian motion process.
   b. Generate a random frequency from a Normal distribution.
   c. Generate a random number of claims from a Poisson distribution with a parameter equal to the product of the exposures and the frequency.
   d. Generate an inflation index from a geometric Brownian motion.
   e. Initialize ultimate loss to zero. Then, for each claim
      i. Generate a random loss severity from a Lognormal distribution, multiply it by the inflation index and add it to the ultimate loss.
3. For each accident quarter,
   a. calculate 40 random increment factors from the formula: $incr = .33 \cdot age^{-1.25} + \left(.07 \cdot age^{-.7}\right) \cdot Normal\left(0,1\right)$. This is not guaranteed to add up to unity but the simulated values add up very close to unity. This procedure is similar to step (i) in Narayan's method 4 except that we are using a random decay pattern instead of a constant pattern.
   b. Multiply the ultimate loss by the increment factors to produce random incremental losses for 40 development quarters.

SAS code:

```
*random number seed;
%let seed=0;

*exposure parameters (Geometric Brownian Motion);
%let expostart = 50000;
%let grthmean = 0.005;
%let grthstdv = .005;

*frequency parameters (Normal);
%let frqmean = .01;
%let frqstdev = .001;

*untrended severity parameters (LogNormal);
%let mu = 8;
%let s = 1.4;

*inflation parameters (Geometric Brownian Motion);
%let cpi0 = 100;
%let cpimu = .006;
```

```
%let cpisig = .0035;

/* First data step - generate exposures and ultimate losses for 40
accident quarters.*/

data tri1;
      *initialize exposures and cpi;
      expos = &expostart;
      cpi = &cpi0;
      do aqtr=1 to 40;
            *generate exposures by brownian motion;
            expos = round(expos * (1 + &grthmean +
            &grthstdv*rannor(&seed)));

            *generate a normally distributed claim frequency;
            freq = &frqmean + &frqstdev*rannor(&seed);

            *generate a Poisson number of claims;
            clms = ranpoi(&seed,freq*expos);

            *generate an inflation index by geometric brownian motion;
            cpi = cpi*exp(&cpimu + &cpisig*rannor(&seed));

            *calculate aggregate loss (ultloss);
            ultloss = 0;
            do clmnum = 1 to clms;
            *calculate loss severity and add it to ultloss;
                        ultloss = ultloss +
            round(exp(&mu+&s*rannor(&seed))*cpi/&cpi0);
            end;
      output;
      end;


proc sort data=tri1; by aqtr;

/* Second data step - calculate incremental incurred losses for 40
development quarters for each accident quarter to produce a decumulated
loss development data set. */

data decumtri;
      set tri1;
      do age=1 to 40;
            decay = .33*age**(-1.25) + (.07*age**(-.7))*rannor(&seed);
            incr_inc = ultloss*decay;
            time = aqtr + age - 1;
            output;
      end;

run;
```

Multilevel Non-Linear Random Effects Claims Reserving Models And Data Variability Structures

Graciela Vera

Abstract

Characteristic of many reserving methods designed to analyse claims data aggregated by contract or sets of contracts, is the assumption that features typifying historical data are representative of the underwritten risk and of future losses likely to affect the contracts. Kremer (1982), Bornheutter and Ferguson (1972), de Alba (2002), and many others, consider models with development patterns common to all underwriting years and known mean-variance relationships. Data amenable to such assumptions are indeed rare. More usual are large variations in settlement speeds, exposure and claim volumes. Also typifying many published models are Incurred But Not Reported (*IBNR*) predictions limited to periods with known claims, frequently adjusted with "tail factors" generated from market statistics. Of concern could be analytical approach inconsistencies behind reserves for delay periods before and after the last known claims, under reserving and unfair reserve allocation at underwriting year, array or contract levels.

As applications of Markov Chain Monte Carlo (MCMC) methods, the models proposed in this paper depart from the neat assumptions of quasi-likelihood and extended quasi-likelihood, and introduce random effects models. The primary focus is the close dependency of the *IBNR* on data variability structures and variance models, built with reference to the generic model derived in Vera (2003). The models have been implemented in BUGS (http://www.mrc-bsu.cam.ac.uk/bugs)

Keywords: Markov Chain Monte Carlo, Non-linear Random Effects and GLM, Reserving.

1. INTRODUCTION

Insurance data reflect and react to financial uncertainty associated with external events, quantifiable time varying factors such as inflation and interest rate fluctuations, and non-quantifiable factors such as variations in litigation practices and underwriting policy terms. In an interesting historical account of legislative changes introduced in Israel to deal with inflation, Kahane (1987) illustrates how external events can be given functional interpretation in a reserving model. Further examples can be found in Taylor

(2000). Data distortions due to external events could undermine all stochastic assumptions. Concerned with the analysis of claims data, from the simplest aggregation levels, such as class of business, to multiple-nested groups, this paper deals with the construction of claims reserving models capable of capturing variability structures in a claims portfolio.

Hierarchical or multi-level claims reserving models are potential source of wide-ranging contribution to claims portfolio analysis beyond reserving. Identification of the causes of data variability with reference to hierarchical model structures could provide a statistical framework for parametric analyses of claims across a number of underwriting years. This would enhance our ability to construct more discriminating models, set initial parameter values, review and update our assumptions on risk premium calculations, related management strategies for commutations, portfolio composition, analysis, etc.

1.1 Research Context

As one of the simplest claims reserving methods, the chain ladder has motivated an extensive body of work intended to establish statistical basis for the problem of reserving. Models that fall within the category of generalized linear models (GLM) (McCullagh and Nelder (1989)), such as Renshaw (1989), Renshaw and Verrall (1998), Verrall (1991), Wright (1990), Mack (1991) and many others, have extended the research beyond assumptions of lognormality and explored applications from exponential family distributions. Carroll (2003) remarks "there are many instances where understanding the structure of variability is just as central as understanding the mean structure". The *IBNR* definition given in this paper is integral to the definition of the model itself, and its value is highly sensitive to model specification. Hence, the emphasis of this research is in the identification of suitable representations for the mean and data variability structures beyond assumptions of known and specific mean-variance relationships.

Reserving model structures depend on the intended use of the predicted reserves and on the sector of interest in the claims portfolio, such as insurance class, contract, specific loss, etc. The data assessment should determine the selection of the analytical approach.

For instance, an insurance contract provides cover against the hazards listed in the contract. Premium calculations reflect policy management expenses, expected returns and risk premiums for all the perils covered by the contract. Risk premium analyses, in general, are carried out by peril, ignoring the fact that a particular event could simultaneously hit more than one kind of cover. When reserve analysis of all perils with a single model is viable, it could deliver, for example, relative cost measures capable of generating more competitive commercial premiums, hence allowing cover assessment on statistical basis, identification of cross-subsidies and unexplored niches, etc.

Within the context of hierarchical models, claims data can be differently interpreted depending on their levels of aggregation. For instance:

- Each underwriting year data set could be described as a set or cohort of longitudinal data.
- A claims array could be considered single-level longitudinal data for more than one subject.
- A book of business segmented by class, type of loss and underwriting year, could be treated as multilevel longitudinal data or as multiple nested groups of single level longitudinal data.

Davinian and Giltinan (1993 and 1996) provide an introduction to the theory of non-linear random effects models and an overview of various techniques for the analysis of non-linear models with repeated observations. More recently, Pinheiro and Bates (2000) reviews the theory and applications of linear and non-linear mixed effect models to the analysis of grouped data.

In this paper it is shown that the generic model in Vera (2003), briefly outlined below, is key to the extension of random effect models to the analysis of reserves. If the claims process for underwriting year w is reported at times $t_1, t_2, ..., t_e$, such that $0 < t_1 < t_2 < ... < t_e$, and t_e is the final settlement period, the generic model is given in terms of a percentage cash flow and a ultimate claim amount functions, denoted respectively by P_{w,t_j} and C_w. $P_{w,t_j} = \int_0^{t_j} \pi(w,z)dz$, where $\pi(w,t)$ is a probability density function taking values from positive real numbers. $S_{w,t_j} = 1 - P_{w,t_j} = \int_{t_j}^{t_e} \pi(w,z)dz$, $P_{w,t_j} \le 1$ for $j < e$ and $P_{w,t_j} = 1$ otherwise. Finally, h_{w,t_j} and H_{w,t_j} are the instant and cumulative hazard rate functions, defined for underwriting year w and payment year τ (τ = w + delay time - 1) by

$$h_{w,\tau-w+1} = -\left(\frac{\partial\left(\ln\left(1-P_{w,z}\right)\right)}{\partial z}\right)_{z=\tau-w+1} = \left(\frac{C_w}{IBNR_{(w,\tau-w+1)}}\right)\left(\frac{\partial P_{w,z}}{\partial z}\right)_{z=\tau-w+1} \tag{1.1}$$

$$H_{w,\tau-w+1} = -\ln\left(1-P_{w,\tau-w+1}\right)$$

Hence, the following are alternative representations of the claims process for cumulative data $Y_{w,\tau-w+1}$:

$$Y_{w,\tau-w+1} = C_w P_{w,\tau-w+1} \tag{1.2}$$

$$Y_{w,\tau-w+1} = C_w\left(1-\exp\left(-H_{w,\tau-w+1}\right)\right) \tag{1.3}$$

$$Y_{w,\tau-w+1} = C_w\left(1-S_{w,\tau-w+1}\right) \tag{1.4}$$

Equivalently, for incremental data $y_{w,\tau-w+1}$

$$y_{w,\tau-w+1} = C_w * \left(P_{w,\tau-w+1} - P_{w,\tau-w+1-1}\right) \tag{1.5}$$

$$y_{w,\tau-w+1} = C_w\left(\exp\left(-H_{w,\tau-w}\right) - \exp\left(-H_{w,\tau-w+1}\right)\right) \tag{1.6}$$

$$y_{w,\tau-w+1} = C_w * \left(S_{w,\tau-w} - S_{w,\tau-w+1}\right) \tag{1.7}$$

The underwriting year and array *IBNR* and reported *IBNR* projections are respectively

$$\begin{aligned} IBNR_{(w,\tau-w+1)} &= C_w S_{w,\tau-w+1} \\ IBNR(\tau) &= \sum_{w=1}^{u} IBNR_{(w,\tau-w+1)} \end{aligned} \tag{1.8}$$

$$\begin{aligned} RIBNR_{(w,\tau-w+1)} &= IBNR_{(w,\tau-w+1)} + \left(C_w S_{w,\tau-w+1} - Y_{w,\tau-w+1}\right) \\ RIBNR(\tau) &= \sum_{w=1}^{u} RIBNR_{(w,\tau-w+1)} \end{aligned} \tag{1.9}$$

where u is the number of underwriting years in the array. *RIBNR* links the reserving analysis to the accounting processes, by adjusting the *IBNR* by the difference between the total claim amount incurred to date and its estimate. Due to the additional noise induced by the adjustment, (1.9) is only applied in the final stages of the reserving analysis. In contrast to many published reserving methods, an important aspect of the models is the unrestricted *IBNR* projection periods, since the period before the last claim is generally unknown. The above equations could make explicit, and potentially highlight, the sources of data variability. Settlement speeds differences between underwriting years should be captured by $P_{w,r-w+1}$, $H_{w,r-w+1}$ or $S_{w,r-w+1}$. Although exposure levels are largely determined by underwriting volumes and contract terms, neither necessarily random, to accelerate convergence and formulate the final model variance function, random effects are introduced in C_w. When more than one claims array are analyzed, the additional aggregation level and source of variability is *array*, indexed by subscript *r*.

1.2 Objective

The examples' aim is to show that more than one model could fit historical data, but not all may reliably predict the reserves. The reliability of the *IBNR* and ultimate claim amount predictions depends on the models' capacity to extract from the data claims volume and settlement speeds measures. This is possible when the variability of both can be represented parametrically and formulated into the variance model. The scope of the models is made evident by their formulation and by the data. As the variability in settlement speeds and claims volumes increase the underlying assumptions of GLM are no longer sustainable, and more complex variance models and random effect parameters for the mean response become essential. To illustrate the process of constructing variance models two data sets are selected. One is a claims array simulated from a mixed portfolio, and the second consists of three arrays simulated from a marine hull, marine cargo and aviation hull portfolios. The second, selected to exacerbate the variability encountered in the first, in addition to large claims volume differences between underwriting years, contains also 20 negative incremental claims entries.

Since the concepts of population models (Zeger, Liang and Albert (1988)) are intended to average random variability between subjects, they are implemented around the percentage cash flow function. They can be used to obtain average (or array) *IBNR* predictions for a given ultimate loss. Other array or average results are the weighted average array or portfolio hazard rates. They provide thresholds, useful to quantify the impact on the claims portfolio of excluding from it underwriting contracts associated with particular underwriting years or arrays.

1.3 Outline

The paper structure is as follows. Section 2 introduces random effect models for one array with a general formulation of non-linear random effects models, and translated into a Bayesian framework in section 2.1.1. Noted in section 2.2 are amendments necessary to formulate multi-array models.

The models selected to analyze the two data sets are presented in sections 3 and 4 respectively. Denoted 1.0 and 2.0, in section 3.1 two preliminary models for one array are given, followed by numerical examples in section 3.3. The examples identify 2.0 as the basis for further analysis to construct the final models. In section 3.4.5 the results from two validation and two final models are discussed. Also in two stages, in section 4 multi-array models are constructed for two mean response functions denoted respectively 7.0 and 8.0. The preliminary models, used to establish data variability structures, are introduced in section 4.1, followed by numerical examples in section 4.2. For mean response functions 7.0 and 8.0, results for precision parameters σ^2, σ_r^2 and σ_{rw}^2 are obtained, identifying the three model versions by (a), (b) and (c). The final models, defined in section 4.3, are analyzed in section 4.5. They emphasise the contribution the generic model makes to the analysis of reserves, and to random effects models and variance models in general.

Section 4.4 extends the claims array average percentage cash flow definition given in section 3.2 to introduce portfolio model average for the percentage cash flow. As immediate by-products of the reserving analysis, hazard rates are discussed in section 4.6. The claims' hazard rate profile, essential for further portfolio analyses, can be used also as a portfolio management template. Discussion on the contribution made by the models proposed is given in section 5.

For the models in section 3, the results are fully reported in appendix A. Given the size of the data used in section 4, the reported results in this section are restricted to *IBNR* and ultimate claim amount projections for the selected preliminary and final models.

2. GENERAL FORMULATION OF NON-LINEAR RANDOM EFFECTS MODELS

In non-linear hierarchical models, inter and intra-underwriting year variations are analysed as a *two-stage process*. In the first, the intra-underwriting year variation is defined by a non-linear regression model for the underwriting year covariance structure. In the second stage, the inter-underwriting year variation is represented by both, systematic and random variability. The models can be constructed within a Bayesian hierarchical structure by noting that the intra-underwriting variation is associated with the sampling distribution, while the prior distribution is relevant to the inter-underwriting variation. Because the models' notation will depend on the number of aggregation levels, in sections 2.1 and 2.2 the array and multi-array analytical frameworks are respectively given.

2.1 Analytical Framework For a Claims Array

For the purpose of defining the general model, ignoring whether claims are cumulative or incremental, the observation at development time t of response vector for underwriting year w is simply denoted by $y_{w,t}$, and the model is defined as follows:

$$y_{w,t} = \mu_{w,t}(\phi_w) + \varepsilon_{w,t} \tag{2.1}$$

where $\mu_{w,t}$ is a non-linear function common to the entire array, while parameter vector ϕ_w is specific to underwriting year w. $t = t_1, \ldots, t_{n_w}$; with t_{n_w} representing the last period with known claims to date, $w = 1, \ldots, u$, and u is the number of cohorts or underwriting years in the claims array. Hence

$$y_w = \left[y_{w,t_1}, \ldots, y_{w,t_{n_w}} \right]^T$$
$$\mu_w = \left[\mu_{w,t_1}, \ldots, \mu_{w,t_{n_w}} \right]^T$$
$$\varepsilon_w = \left[\varepsilon_{w,t_1}, \ldots, \varepsilon_{w,t_{n_w}} \right]^T$$

and

$$\operatorname{cov}(\varepsilon_w) = \sigma^2 R_w \tag{2.2}$$

R_w is the intra-underwriting year covariance matrix for underwriting year w.