

Artificial Neural Networks in Insurance Loss Reserving

Peter Mulquiney¹

¹Taylor Fry Consulting Actuaries, Sydney, Australia

Abstract

In this paper we analyse **insurance data** using Artificial Neural Networks (ANN)[1]. In particular, we use ANN for the problem of **Loss Reserving**.

Loss reserving is the practice of estimating the future payments for the claims which have occurred on an insurance portfolio. A difficulty in forecasting future payments is that the time series of payments often depends on influences that are not observable in the historical data.

For example, claims cost inflation may depend on future events such as legislative change and changes in judicial attitudes. Because of this, it is often necessary to supplement ANNs with separate forecasts which account for the expected changes in the future claims environment.

Keywords: Insurance, Loss Reserving, Artificial Neural Networks.

1. Introduction

Certain classes of insurance involve substantial delay between occurrence of the event generating a claim and its settlement. During this interval, there may be considerable uncertainty as to the amount of the final settlement.

Loss reserving is the practice of estimating the future payments on the claims which have occurred on an insurance portfolio. The future payments that will be made on these claims are a liability to the insurer and most insurers are required by statute to estimate the size of these liabilities for inclusion in their financial statements.

Typically, the claims experience of an insurance portfolio has many features that result from events such as changes in claim management procedures, changes in legislation, seasonality and changes in the rates of claim cost inflation. We have found that ANN are useful modeling these features of an insurer's historical claims experience.

However, a difficulty in using ANN to forecast future claims experience results from the fact that the forecasts often depend on influences that are not

observable in the historical data. For example, future claims cost inflation may depend on future events such as legislative change and changes in judicial attitudes. Hence any influences not directly observable in the historical data need to be separately forecast to produce loss reserve estimates.

We have addressed this difficulty by supplementing our ANN with separate forecasts which accounted for the expected changes in the future claims environment.

In the following paper, the use of ANN for loss reserving is illustrated using data from a motor bodily injury portfolio. We also compare the results to those obtained using Generalized Linear Modelling (GLM) – a technique more often used for loss reserving than ANN.

2. Methodology

The insurance data we analyse relates to Motor Bodily Injury (CTP) insurance in one state of Australia. The payments for Motor Bodily Injury are usually dominated by a single lump sum near the date of claim finalisation. Hence a common approach to such payment types is to:

- Model the expected number of claim finalisations to be made at future dates; and
- Model the expected size of finalised claims at each future finalisation date.

In the following paper we restrict our attention to the model of expected claim sizes, however the general conclusions apply equally to the model of claim finalisations.

2.1. Data

The data set consists of a claim file with approximately 60,000 claims for a 9 year period up to 30 September 2003. For each claim various items are recorded, including the date of injury, date of notification, and histories of paid losses, case estimates and finalised/unfinalised status including dates of change of status.

For this analysis, all paid loss amounts have been converted to 30 September 2003 values in accordance with past wage inflation in the state concerned.

2.2. Regression Models

We fitted both an ANN and a GLM to the data. For both models we were interested in modelling the size of the r th finalised claim, Y_r in terms of:

- i_r = accident quarter = 1, 2, 3, ..., 37
- j_r = development quarter = 0, 1, 2, ..., 36
- k_r = calendar quarter of finalisation = $i_r + j_r$
- t_r = operational time = proportion of claims incurred in accident quarter i_r which have been finalised at the mid-point of development quarter j_r
- s_r = season of finalisation = March, June, September, and December

Hence both the GLM and the ANN have the general regression function:

$$Y_r = f(i_r, j_r, k_r, t_r, s_r) \quad [\text{Eqn 1}]$$

Note that calendar quarter is just the sum of accident quarter and development quarter. The dependency between these three predictors indicates that the model should be primarily based on 2 of these 3.

For both the ANN and GLM we found that a model based primarily on calendar quarter and development quarter was preferred. However in both cases an accident quarter binary variable was included to model the effect of a legislative change that came into effect in September 2000.

2.3. Software

All analysis was performed using the software “R” [2]. The algorithm package **nnet** was used for the neural network algorithm and the **glm** function was used for the GLM model. A random subset of 70% of the data was assigned to be the training data set, while the remaining 30% formed the test data set.

The tuning parameters were determined using cross-validation and the final neural network consisted of a single hidden layer with 20 units and a weight decay of 0.05.

3. Results

3.1. GLM

The procedures that were used to build the GLM model have been described previously [3] and here we

restrict ourselves to the main features of the model.

The GLM equation was:

$$E[Y_r] = \exp \left\{ \alpha + \beta_1^d t_r + \beta_2^d \max(0, 10 - t_r) + \beta_3^d \max(0, t_r - 80) + \beta_4^d I(t_r < 8) \right\} \quad [\text{Operational time effect}]$$

$$+ \beta^s I(k_r = \text{March quarter}) \quad [\text{Seasonal effect}]$$

$$+ \beta_1^f k_r + \beta_2^f \max(0, k_r - 2000Q3) + \beta_3^f I(k_r < 97Q1) \quad [\text{Finalisation quarter effect}]$$

$$+ k_r [\beta_1^{ft} t_r + \beta_2^{ft} \max(0, 10 - t_r)] \quad [\text{Operational time x finalisation quarter interaction}]$$

$$+ \max(0, 35 - t_r) [\beta_1^{at} + \beta_2^{at} I(i_r > 2000Q3)] \quad [\text{Operational time x accident quarter interaction}]$$

[Eqn 2]

with the response assumed to follow an exponential dispersion family distribution with a variance power of 2.3 (Taylor and McGuire, 2004). A plot of the log of the regression function (the linear predictor) is shown in Figure 1.

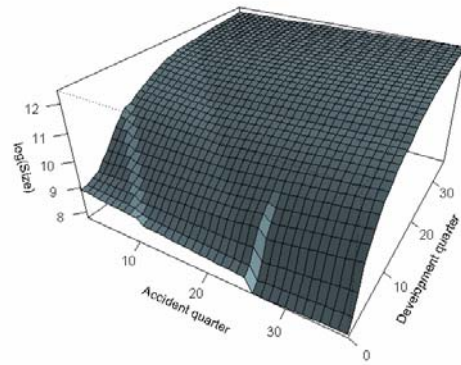


Fig. 1: Plot of the linear predictor of the GLM model. To smooth this plot I have assumed used the same rates of finalisation across each accident quarter, and I have ignored the effect of seasonality.

Eqn 2 and Fig. 1 illustrate the features that are present in the finalised claim data. There are 5 main features:

- Operational time effect: Because of changes in the rate of claims finalisation, the regression function includes an operational time effect rather than a development quarter effect. This effect shows that the average size of finalised claims increases with operational time.
- Seasonal effect: Claims finalised in the March quarter tend to be slightly lower than in other quarters.
- Finalisation quarter effect: This represents superimposed inflation. Because the historical payment data was adjusted to constant dollar values using a historical inflation index, any

additional inflation is termed superimposed inflation. The model indicates that there is a change in the rate of superimposed inflation before 1997 and at the end of the September 2000 quarter.

- Operational time and finalisation quarter interaction: This brings out the feature that smaller and larger finalised claims are subject to different rates of superimposed inflation.
- Operational time and accident quarter interaction: This feature resulted from legislative changes that came into effect in September 2000. This legislation placed limitations on the payment of plaintiff costs and effectively eliminated a certain proportion of smaller claims in the system in all subsequent accident quarters.

3.2. ANN

A plot of the log of the claim size for the ANN model is shown in Fig. 2.

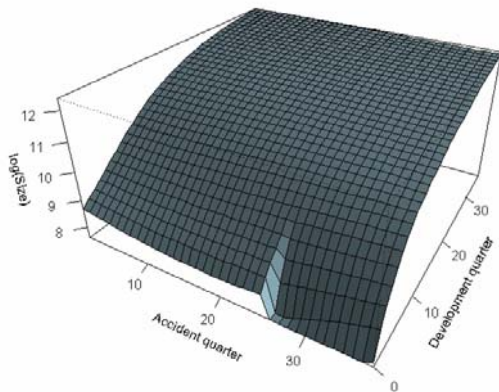


Fig. 2: Plot of log(size) for the ANN model. Smoothing as for Fig. 1.

The predictive accuracy of the ANN on the test data set compared favourably to the GLM for two different measures (Table 1).

Table 1 Test errors for the ANN and GLM models

Model	Average Sum of squares	Average Absolute Error
GLM	\$99,965 ²	\$33,777
ANN	\$99,843 ²	\$33,559

In addition, a variety of 1 dimensional residual plots showed that there appeared to be no systematic bias in the model fits

across the predictors. The quality of the residual plots was similar between the ANN and the GLM model.

3.3. Projection of future claim size

In Figs. 1 and 2 claim sizes have been projected into future quarters, that is quarters beyond the last historical data date of 30 Sep 2003. This is represented by the upper right hand triangles of these plots. In other words, the diagonal line joining the front corner (accident quarter = 37, development quarter = 0) to the back corner (accident quarter = 1, development quarter = 36) represents the latest quarter of finalisation in the historical data. Every point to the right of this line represents a future data point.

A particular concern with this data set when projecting future claim sizes is the assumed level of future superimposed inflation. Sources of superimposed inflation in a motor bodily injury portfolio such as the one under study include legislative changes and increasing generosity in court awards. Hence future projections of superimposed inflation should give consideration to the expected claims environment in the future - they will not usually be a simple extrapolation of past trends.

The past and future superimposed inflation that is predicted by the ANN and GLM models respectively (using simple extrapolation) are shown in Figs. 3 and 4.

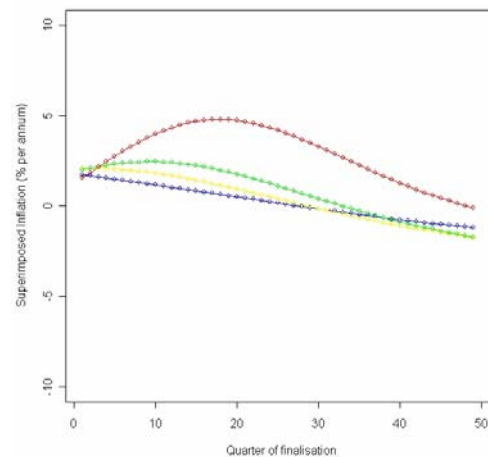


Fig. 3: Historical and projected superimposed inflation for the ANN model as a function of finalisation quarter and development quarter. Future superimposed inflation is from finalisation quarter 38. Development quarter was: red line, 10; green line, 20; yellow line, 30; blue line, 35. An Operational time appropriate for the development quarter was also chosen. All other predictors were constant.

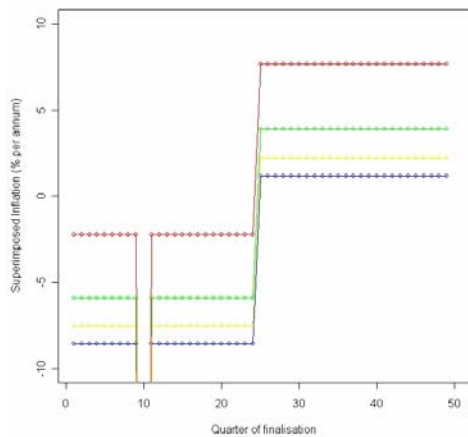


Fig. 4: Historical and projected superimposed inflation for GLM model. Future superimposed inflation is from finalisation quarter 38. Development quarter was: red line, 10; green line, 20; yellow line, 30; blue line, 35. Operational time appropriate for the development quarter was also chosen. All other predictors were constant.

Of interest is the significant difference in the estimated superimposed inflation. This results from the different architectures of the two models. In particular, while our ANN model included both development quarter and operational time as predictors, the GLM model only included operational time. If development quarter was excluded from the ANN model, the ANN model predicted negative superimposed inflation values at early quarters of finalisation also.

As discussed above it is often not appropriate to simply extrapolate past trends in superimposed inflation and it is usually necessary to make a separate forecast of the expected future values. We do not have sufficient space to discuss the considerations that are required when choosing future superimposed forecasts. We simply note that we have assumed that future superimposed inflation will be 0% in all future years and we have then forecast future claim sizes by supplementing our ANN with this assumption. Note that while a very simple model of future superimposed inflation has been chosen, the forecast could have easily taken a more complex form. The model of projected claim sizes made using the 0% future superimposed inflation forecast is illustrated in Fig. 5.

If the projections of the size of finalised claims are combined with projections of the number of finalised claims it is possible to estimate the total amount of future payments – the loss reserve. If we use simple extrapolation to project the model of finalised claims, it is found that the GLM produces a loss reserve 11% higher than the ANN. This is not surprising given the projected levels of superimposed inflation (Figs. 3 and 4). However, using the 0% future superimposed

inflation assumption for both the ANN and GLM models yields loss reserves that agree to within 0.1%.

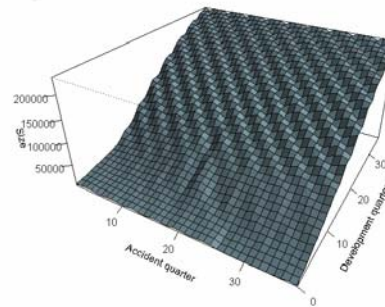


Fig. 5: Plot of ANN model of finalised claim size.

4. Discussion

The main points from the loss reserving exercise were:

- ANN were effective in modelling the complex features of the historical insurance data.
- The ANN model resulted in better predictive accuracy on the test data set compared to the GLM model.
- It took significantly less time to fit the ANN compared to the GLM model. The ANN algorithm was largely automated while fitting the GLM required significant input from the model builder.
- The functional form of the ANN was more complicated than the GLM having 181 parameters compared to the GLM's 13. It was necessary to use graphical techniques to understand the behaviour of the ANN.
- A difficulty in forecasting future payments is that the time series of payments often depends on influences that are not observable in the historical data. This was illustrated by superimposed inflation for which it was necessary to supplement the ANN with a separate forecast of future superimposed inflation.
- Further assessment of the ANN and GLM models is being undertaken.

5. References

- [1] C. Bishop, "Neural networks for pattern recognition," Clarendon Press, Oxford, 1995.
- [2] <http://www.r-project.org/foundation/>
- [3] G. Taylor, and G. McGuire, "Loss reserving with GLMs: A Case Study," *Casualty Actuarial Society Discussion Paper Program*, pp. 327-391, 2004.