Institute of Actuaries of Australia

XIth Accident Compensation Seminar 2007

Combining GLM and data-mining techniques for modelling accident compensation data

Peter Mulquiney

Introduction

- Accident compensation data exhibit features which complicate loss reserving and premium rate setting
  - Speeding up or slowing down of payment patterns
  - Abrupt changes in trends due to legislative changes
  - Changes in the profile of claims
  - Other changes which emerge as superimposed inflation
- Complicated structure can be modelled with GLMs
  - structure chosen in an *ad hoc* manner
  - process can be laborious and can be fallible

Introduction

- Alternative: Data mining techniques
  - Artificial Neural Networks
  - CART, MARS etc
- Advantages:
  - flexible architecture can fit almost any data structure
  - model fitting is largely automated

Overview

- Examine general form of model of claims data
- Examine the specific case of a **GLM** to represent the data
- Consider how the **GLM structure** is chosen
- Introduce and discuss **Artificial Neural Networks** (**ANNs**)
- Consider how these may assist in formulating a GLM

Model of claims data

- General form of claims data model

$$Y_i = f(X_i; \beta) + \varepsilon_i$$

  - Y_i = some observation on claims experience
  - β = vector of parameters that apply to all observations
  - X_i = vector of attributes (covariates) of i-th observation
  - ε_i = vector of centred stochastic error terms

Model of claims data

- General form of claims data model

$$Y_i = f(X_i; \beta) + \varepsilon_i$$

  - Y_i = some observation on claims experience
  - β = vector of parameters that apply to all observations
  - X_i = vector of attributes (covariates) of i-th observation
  - ε_i = vector of centred stochastic error terms
- Examples
  - $Y_i = Y_{ad}$ = paid losses in (a,d) cell
    - » a = accident period
    - » d = development period
  - Y_i = cost of i-th completed claim

Examples (cont)

- Y_{ad} = paid losses in (a,d) cell
  - $E[Y_{ad}] = \beta_d \Sigma_{r=1}^{d-1} Y_{ar}$ (chain ladder)

Examples (cont)

- Y_{ad} = paid losses in (a,d) cell
  - $E[Y_{ad}] = \beta_d \Sigma_{r=1}^{d-1} Y_{ar}$ (chain ladder)
- Y_i = cost of i-th completed claim
  - $Y_i \sim$ Gamma
  - $E[Y_i] = \exp [\alpha+\beta\ t_i]$

  where

  » a_i = accident period to which i-th claim belongs

  » t_i = operational time at completion of i-th claim
  = proportion of claims from the accident period a_i completed before i-th claim

Examples of individual claim models

- More generally

$E[Y_i] = \quad \exp\{\text{function of operational time}\}$

Examples of individual claim models

- More generally

E[Y_i] = exp {function of operational time}

+ function of accident period (legislative change)}

Examples of individual claim models

- More generally

$E[Y_i] =$ exp {function of operational time}

+ function of accident period (legislative change)}

+ function of completion period (superimposed inflation)}

Examples of individual claim models

- More generally

$E[Y_i]$ = exp {function of operational time}

+ function of accident period (legislative change)}

+ function of completion period (superimposed inflation)}

+ joint function (interaction) of operational time & accident period (change in payment pattern attributable to legislative change)}

Examples of individual claim models

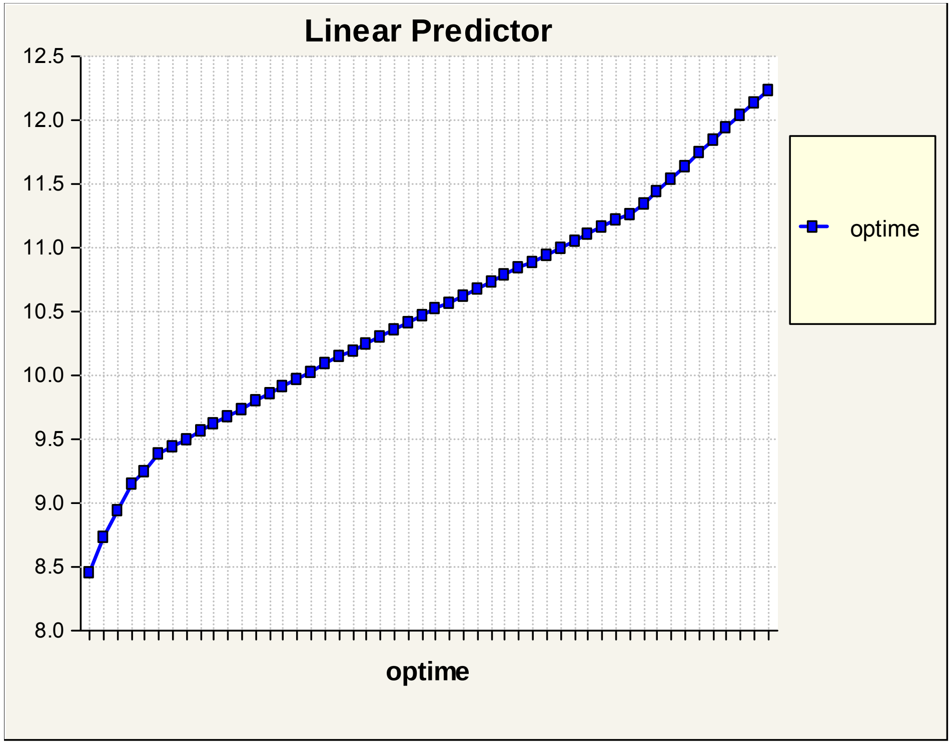
- Models of this type may be very detailed
- May include
  - Operational time effect (payment pattern)
  - Seasonality
  - Creeping change in payment pattern
  - Abrupt change in payment pattern
  - Accident period effect (legislative change)
  - Completion quarter effect (superimposed inflation)
  - Variations of superimposed inflation with operational time

Choosing GLM structure

- Typically largely *ad hoc*, using
  - Trial and error regressions
  - Diagnostics, e.g. residual plots
- Example:
  - Modelling 60,000 Auto Bodily Injury claims
  - Model of the cost of completed claims

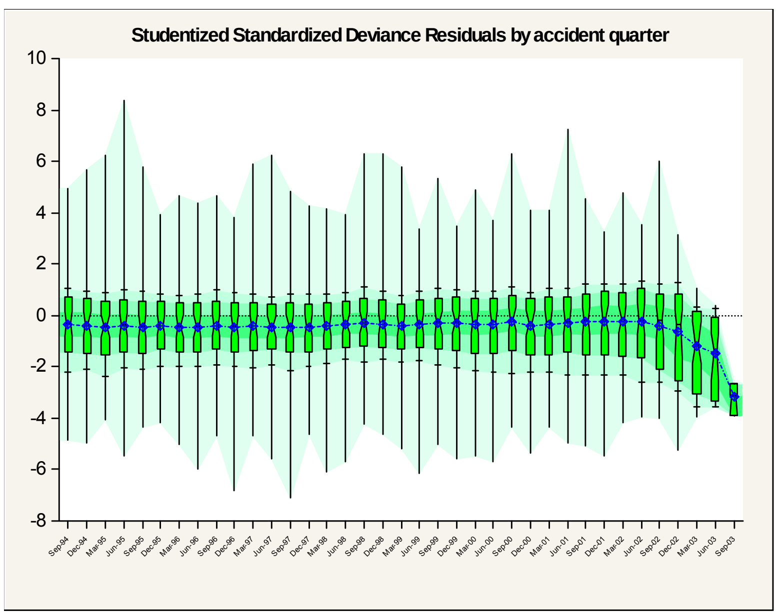
Choosing GLM structure

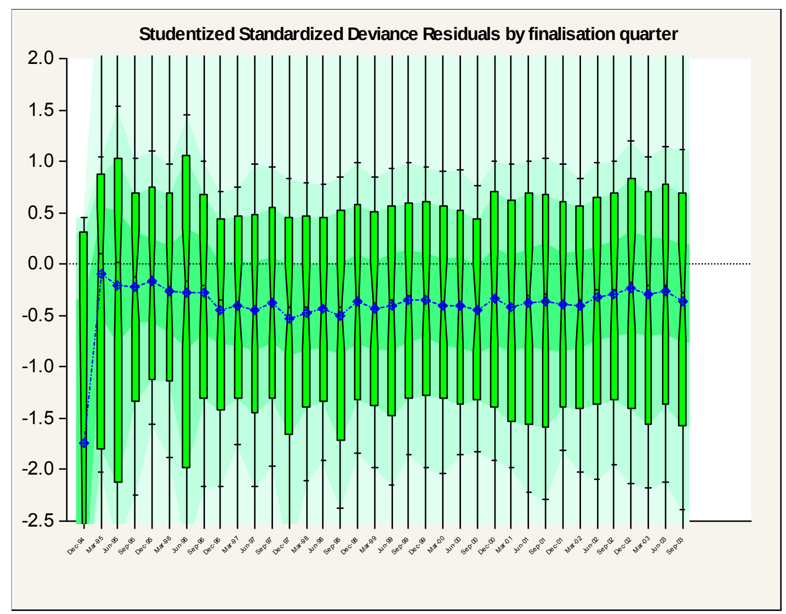
- First fit just an operational time effect



Choosing GLM structure

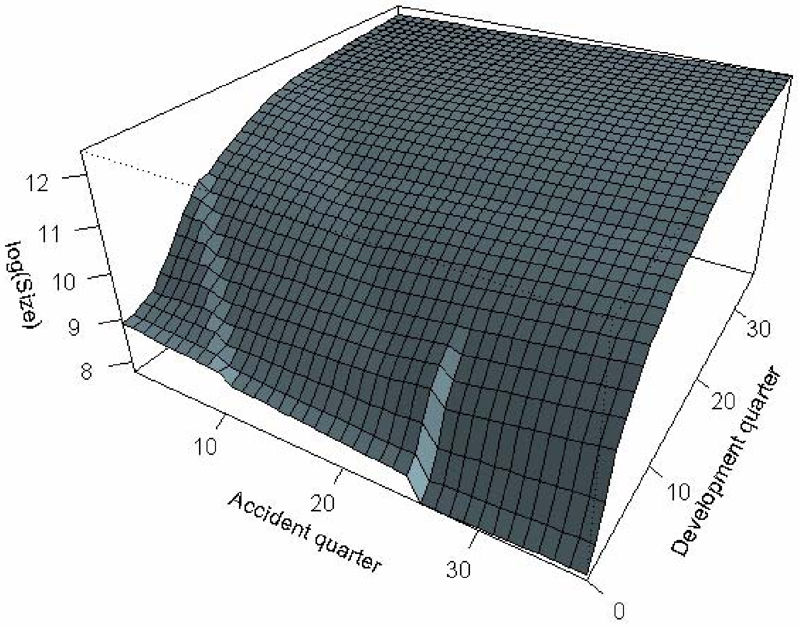
- But there appear to be unmodelled trends by
  - Accident quarter
  - Completion (finalisation) quarter





Choosing GLM structure

- Final model has terms for:
  - Age of claim
  - Seasonality
  - Accident quarter
    - Change in Scheme rules
  - Change in age of claim effect with change in Scheme rules
  - Superimposed inflation
    - Varying with age of claim



Choosing GLM structure

- Structure identified in *ad hoc* manner
  - Trial and error regressions
  - Diagnostics, e.g. residual plots
- More rigorous approach desirable
- **Can we use ANN to do it better?**

Introduction to ANN

- The ANN Regression Function
  - Start with vector of P inputs $X = \{x_p\}$
  - Create **hidden layer** with M **hidden units**
    - Make M linear combinations of inputs

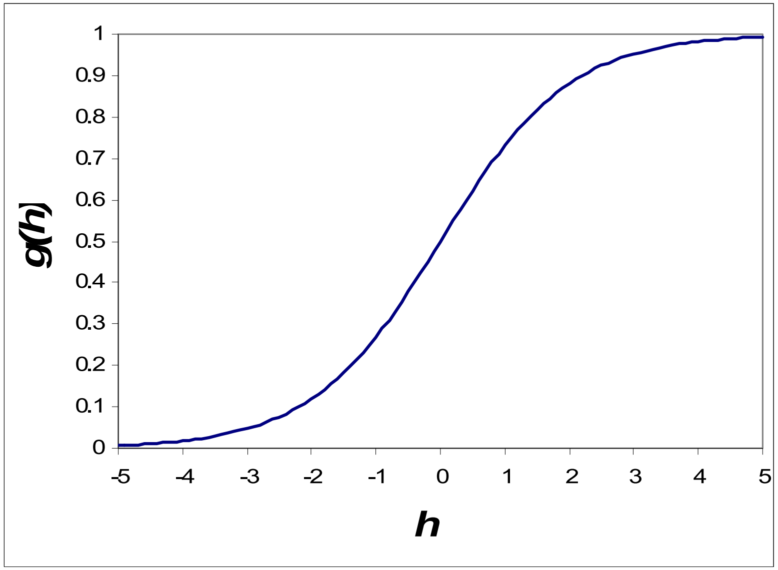
$$h_m = \sum_p w_{mp} x_p$$

    - Linear combinations then passed through layer of **activation functions** $g(h_m)$

$$Z_m = g(h_m) = g(\sum_p w_{mp} x_p)$$

Introduction to ANN

- Activation function
  - Usually a sigmoidal curve



$$g(h) = \frac{1}{1+e^{-h}}$$

  - Function $\Rightarrow$ introduces non-linearity to model

    $\Rightarrow$ keeps response bounded

Introduction to ANN

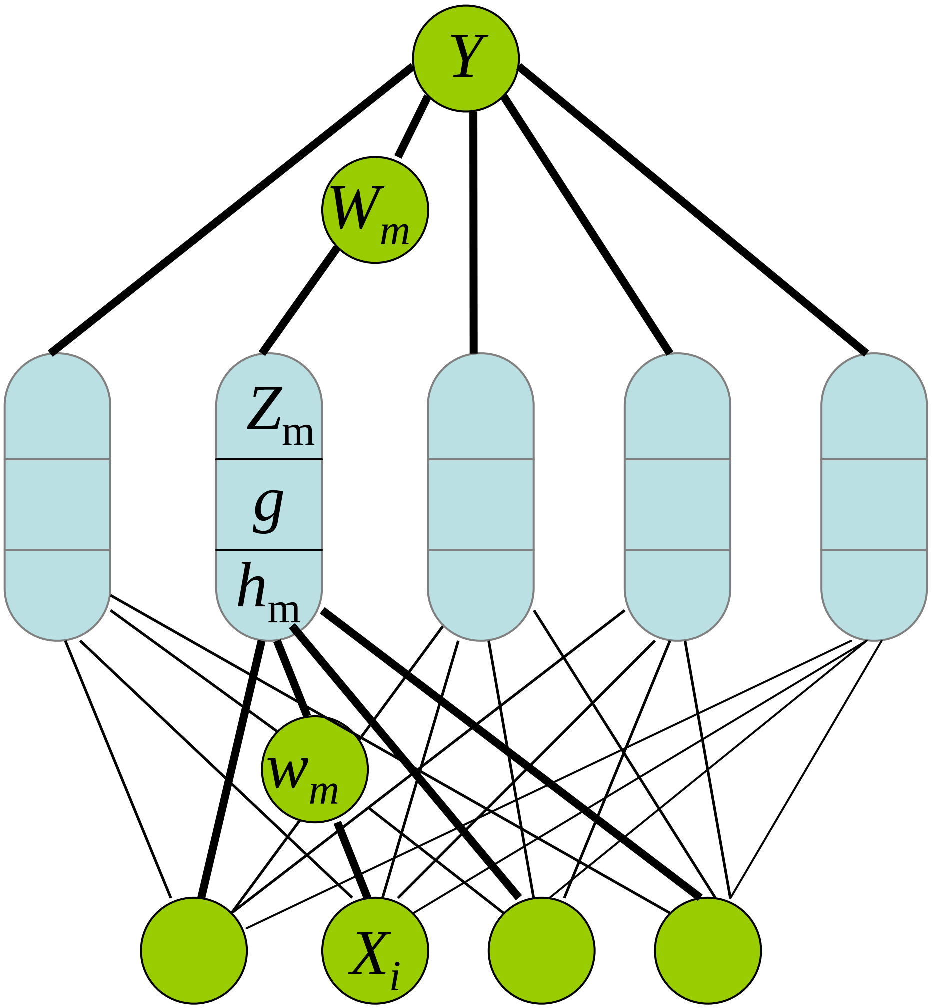
- *Y* is then given by a linear combination of the outputs from the hidden layer

$$Y = \sum_m W_m Z_m = \sum_m W_m g\left(\sum_p w_{mp} x_p\right)$$

- This function can describe any continuous function

- 2 hidden layers $\Rightarrow$ ANN can describe **any** function

Introduction to ANN



Training an ANN

- Weights are usually determined by minimising the least-squares error

$$Err = \frac{1}{2}\sum_{i=1}^{N}(y_i - f(X_i))^2$$

- **Weight decay** penalty function stops overfitting

$$Err + \lambda(\sum_m W_m{}^2 + \sum_m \sum_p w_{mp}{}^2)$$

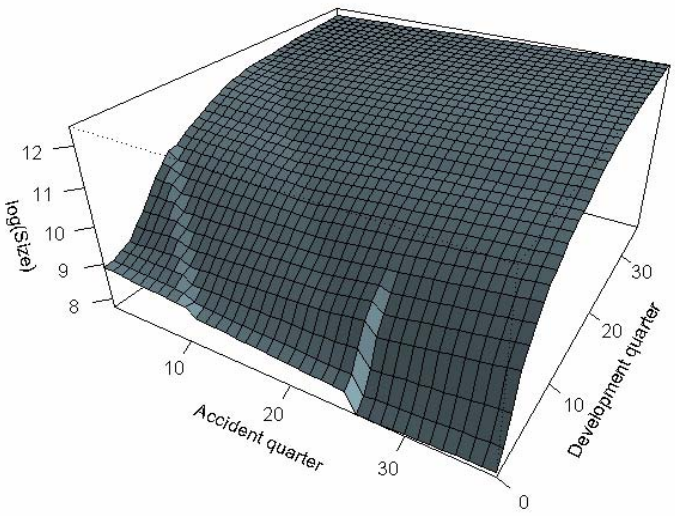
  - Larger λ ⇒ smaller weights
  - Smaller weights ⇒ smoother fit

Training of an ANN - Example

- Training data set: 70% of available data
- Test data set: 30% of available data
- Network structure:
  – Single hidden layer
  – 20 units
  – Weight decay λ=0.05
- These tuning parameters determined by **cross-validation**
  – Prediction error in test data set

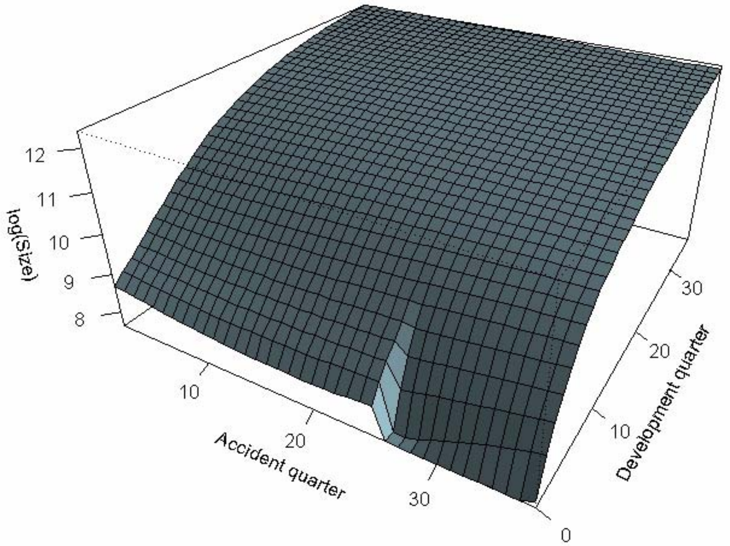
Comparison of GLM and ANN

- GLM



Average absolute error
= $33,777

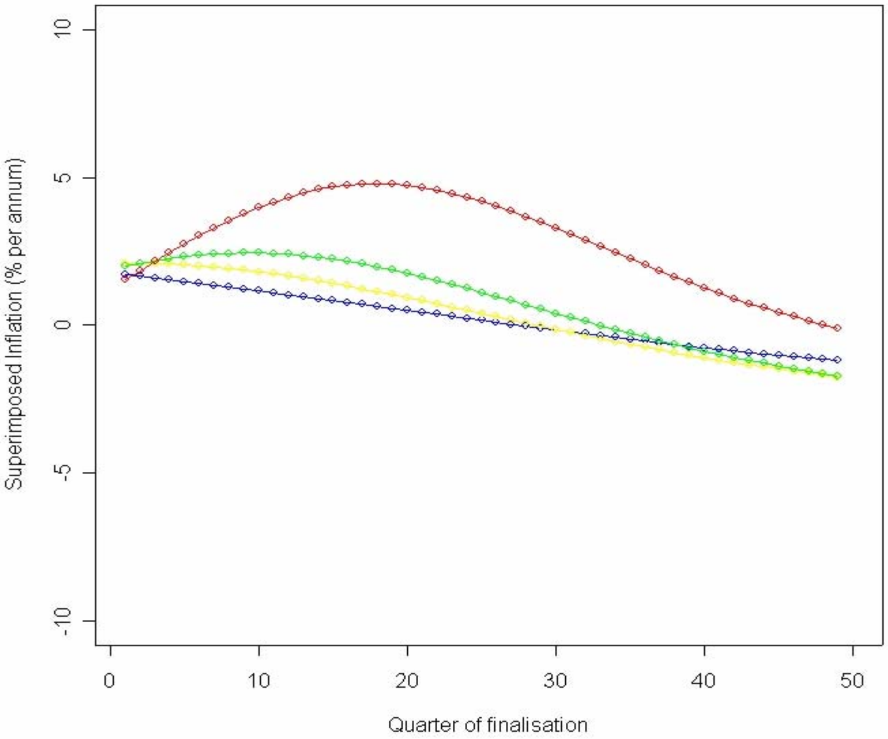
- ANN



Average absolute error
= $33,559

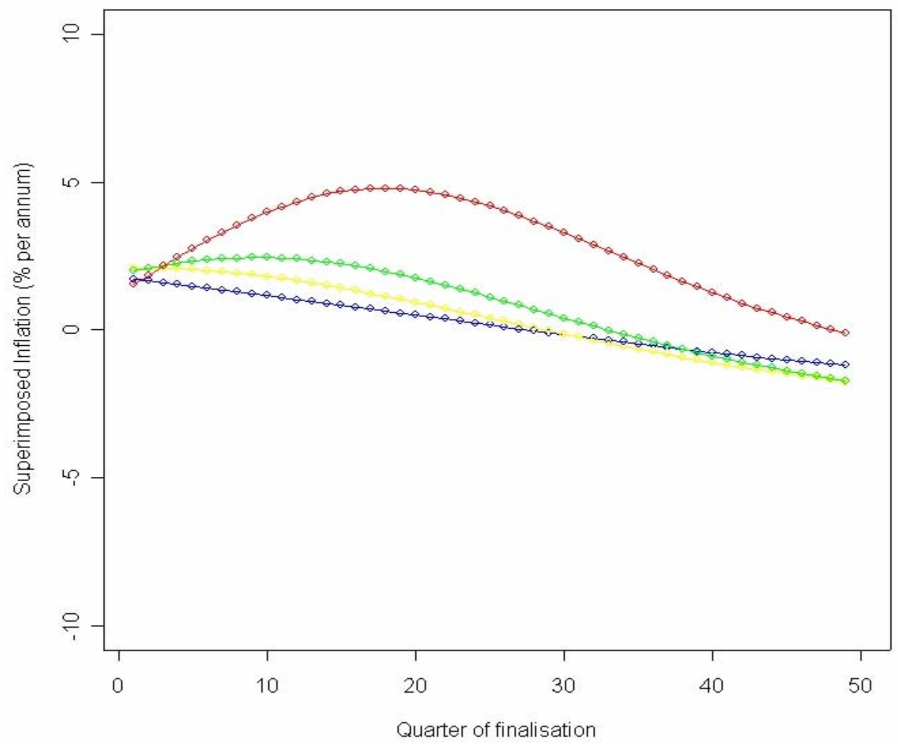
ANN forecasts

- 1D graphical plots to visualise data features
- e.g. historical and future superimposed inflation
    - Development quarter 10: red
    - Development quarter 20: green
    - Development quarter 30: yellow
    - Development quarter 40: blue
- ANN has searched out general form of past superimposed inflation (SI)
- Future SI determined by **simple extrapolation**



ANN forecasts

- Note forecast negative SI may be undesirable
- Need to consider expected claims environment in the future to determine appropriate SI forecast
- Because of problem of extrapolation with ANN usually necessary to supplement ANN forecast with a separate forecast of future SI.

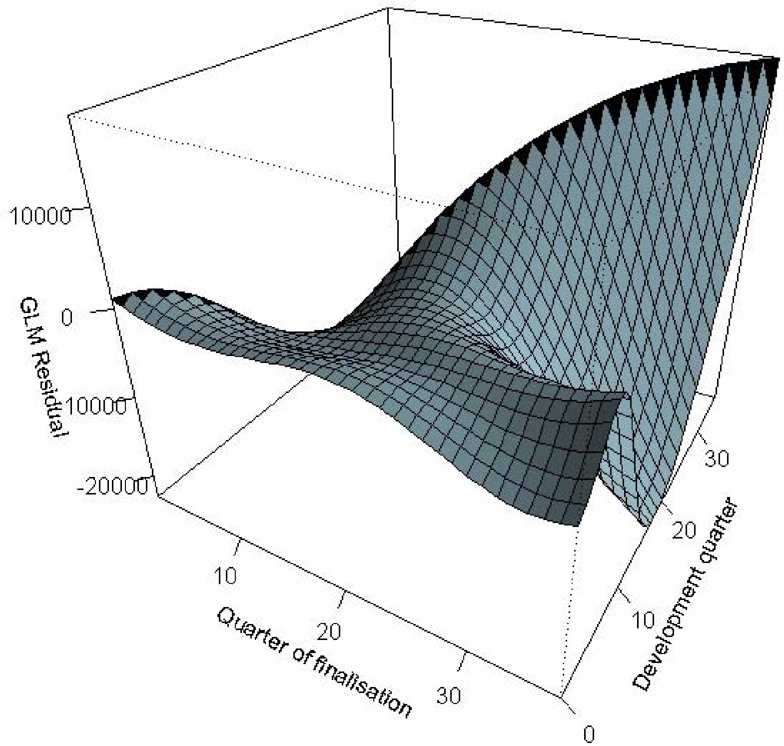


Combining ANN and GLM

- Often preferable to use a GLM over ANN due to model **simplicity** and **transparency**
  - ANN – 181 parameters
  - GLM – 13 pars
- May get best out of ANN and GLM if use in combination
- Use **ANN** as an automated tool to seeking out trends in data
  - Apply ANN to data set
  - Study trends in fitted model against a range of predictors or pairs of predictors using graphical means
- Use this knowledge to choose the functional forms to include in the **GLM** model

Combining ANN and GLM

- Ultimate test of the GLM is to apply ANN to its residuals, seeking structure
- There should be none
- The example indicates that there may the chosen GLM structure may:
- Over-estimate the more recent experience at the mid-ages of claim
- Under-estimate it at the older ages



Conclusions

- GLMs provide a powerful and flexible family of models for claims data
- Complex GLM structures may be required for adequate representation of the data
  - The identification of these may be difficult
  - The identification procedures are likely to be *ad hoc*
- ANNs provide an alternative form of non-linear regression
  - These are likely to involve their own shortcomings if left to stand on their own (e.g. reduced transparency)
  - They may, however, provide considerable assistance if used in parallel with GLMs to identify GLM structure