Statistical Premium Rating Working Party

Members

Tim Carter (Chairman)
Sanjiv Chandaria
Henry Johnson
Andrew Leifer
Helen Pitt
Andrew Turnbull

Summary

This paper considers how generalised linear models are used to assess the adequacy of premium rates for personal lines insurance business. It considers the practical steps involved and discusses the results for various models that were fitted to a sample set of data. The results of the models that were fitted illustrate the usefulness of these techniques when assessing the impact of several risk factors upon insurance claims experience. Being a practical paper it is hoped will make this an accessible introduction to the techniques involved.

1 Introduction

It is increasingly being recognised that insurers who can analyse their personal lines business effectively are obtaining a competitive advantage. Actuaries are increasingly involved in this area, in particular using statistical models to objectively assess the adequacy of premium rates.

The excellent paper by Brockman & Wright (JIA 119 part III) explained in some detail how generalised linear models have been fitted to individual company's motor insurance claims experience to produce theoretically correct premium rates. It is not suggested in a competitive market that premium rates are automatically changed to these theoretical values. However, understanding the profitability of different segments of a company's account enables informed management decisions to be made about premium rate adjustments or new business strategies. The aim is to identify niches of business which, because of a company's unique structure and distribution, are profitable. The statistical analysis can also be used to assess objectively the appropriateness of rating groups such as car groupings and to produce standardised tables against which the developing claims experience can be monitored to quickly identify any changes in claims experience from that predicted by the model.

The approach taken by Brockman & Wright is a thorough description of the methods that can be used. There remain many practical issues which, because of the comprehensive nature of the subject, are not discussed and which it is intended this working party will address. Being a practical approach to this subject should also make this an accessible introduction to the techniques involved.

A sample set of data was provided to all the members of the working party and they were invited to fit statistical models to this data. Models were fitted by a wide range of individuals including both actuaries and statisticians, although these are not necessarily mutually exclusive professions! This paper describes the practical steps required to fit these models and compares the approaches taken and the projected results of the different models.

The techniques used are not only applicable to premium rating motor insurance business, but also for any other classes of insurance where a number of rating factors are available which it is believed will influence the claims experience. The methods have been used successfully for household business, extended warranty cover, health insurance, credit scoring and automobile breakdown cover.

The paper's contents fall into three distinct sections. Section 2 briefly considers the underlying statistical theory and the data that is used in this investigation. The methods by which a suitable model is selected are discussed in Section 3. Section 4 then compares the approaches taken and the results of the different models that were fitted to sample data.

2 Background

Data

To perform a premium rating analysis information is required summarising the total exposure, the number of claims and the amount of claims for each permutation of the risk factors that are to be considered. Obtaining this information can often be the most time consuming and complex part of a premium rating investigation. This proved to be the case for this working party, although not for the usual reasons of poor data quality and IT resource problems.

Our initial intention was to use information collected by the Motor Risk Statistics Bureau ("MRSB") from a large number of UK insurers as a basis for this investigation. Analysis by the working party would have had the advantage of introducing the staff of the MRSB to the generalised linear modelling techniques without the substantial cost of consultancy fees. Also, by combining the information of different companies and because of the limited information that is available, it was not felt that the results of this information would be commercially sensitive. However, when the MRSB approached their members to see whether they could include the information from their company in the data extract a few companies objected so vehemently that the MRSB felt they could not provide any information at all.

We then had discussions with several insurers about whether they could provide information. They were understandably reluctant to do so if no other companies were involved and it was decided to generate a set of data using stochastic methods. This approach, although not ideal, does have the advantage that the "correct" model is known and can be compared to the results produced by the modelling exercise. To make this data as realistic as possible not all the risk factors that were used to generate the data were provided with the data set that was modelled. This causes the data to be overdispersed, in that there will be more variation present in the data than can be explained by the available risk factors, which is a typical feature in practice.

Basic Theory

Statistical models for a random variable are based upon the idea that the variable under investigation has a definite structure. It is assumed that this structure can be used to explain the values actually obtained and can be used to predict future values. In general insurance the random variable is often assumed to be the claim frequency or the average claim amount and a possible assumption would be that these are Normally distributed. It is also assumed that the variable of interest can be expressed in terms of other more basic variables. For example the frequency of motor insurance claims is in part explained by the age of the insured driver. The modelling process involves attempting to identify the structure of the random variable we are investigating and the relationship between this and the other available information.

The general approach when choosing statistical models is to explain as much of the variability of the data as possible whilst using as simple a structure as possible, a concept known as parsimony. In practice a sequence of models is often compared with a more complicated model being preferred if the additional parameters are justified by the increase in the variability of the data that is explained.

In this investigation we have restricted our attention to generalised linear models. This is a very flexible class of models which can be efficiently fitted using the GLIM statistical package. Generalised linear models are described in detail in many statistical texts. In particular "An Introduction to Generalised Linear Models" by Annette J Dobson is very accessible and "Generalised Linear Models" by McCullagh and Nelder provides a comprehensive treatment of the subject. They are defined by three features:

1 A random component which is the assumed underlying probability distribution of the variable of interest. This variable of interest may be a transformation of the raw data. The observed values of this random variable are assumed, in most cases, to be independently distributed.

2 A systematic component which is generally a linear combination of the explanatory variables.

3 A link between the mean of the variable of interest and the systematic components.

For example, the well known log-linear model shown below is a generalised linear model.

$$\ln\left(\mu_i\right) = \sum_{j=i}^{p} x_{ij}\beta_j \qquad\qquad y_i \sim N\left(\mu_i, \sigma^2\right)$$

where the random component is Normal,

the systematic component is $\sum_{j=i}^{p} x_{ij}\beta_j$,

and the link is a log.

To fit statistical models to a set of data involves two basic decisions.

1 The choice of the relationship between the response which is being modelled and the underlying parameters of the model. For generalised linear models this involves specifying the random component, systematic component and the link function discussed above.

2 The selection of a measure of fit which defines how closely a model represents the data. This is optimised to estimate the most suitable parameters for a particular model and used to assess the adequacy of a model's fit. The measure of fit used by GLIM is the deviance function, which is a measure of the lack of fit of the model. Parameters are estimated by minimising the deviance, which is equivalent to producing maximum likelihood estimates. A simple example in Appendix A illustrates how maximum likelihood estimates are obtained and that for a Normal distribution with identity link, maximum likelihood estimation is identical to estimation by least squares.

3 Model Selection

This section is intended as practical background to the modelling results in Section 4. It summarises the issues that are considered when fitting models to premium rating data in practice.

Available Software

Both the GLIM statistical package and SAS were used to fit models by members of the working party. The GLIM package is the most widely used for these investigations, it is inexpensive and provides a customised environment for fitting generalised linear models.

Cohort Definition

As has been discussed in previous papers on premium rating, cohorts are usually best defined by policies incepting in a given period. However accident year cohorts do not require a link to be made between the claim and premium computer files and may be the only type of information that is available. In particular where accurate historic claims exposure is not stored the accident year approach allows sampling to be used from snapshots of exposure that are available. Accident year models do not relate directly to a particular premium rate book but they are more easily adjusted for changes in claims handling experience, such as the removal of knock for knock agreements.

Subdivision of Data

The data should be subdivided as far as is possible into homogeneous groups. Models are separately fitted where possible to claim frequency and average claim amount, rather than to the average claim cost per policy. Generally fewer risk factors influence the claim amount than the claim frequency and this approach also allows perceived trends in average claim amount or claim frequency to be explicitly allowed for when projecting forward to determine premium rates in force in the future.

The data is also often split by type of claim, as this again produces more homogeneous groups. For example bodily injury claims are generally influenced by different and fewer factors than accidental damage claims. There are two ways in which the data can be subdivided; into the separate components of each type of claim such as the windscreen or accidental damage components, or by a claim event type. Claim event types are defined to represent independent events, such as separating accidental damage claims with no bodily injury component and accidental damage claims which include bodily injury claims. If independent events are not used the level of uncertainty in the premium is underestimated because several types of damage will be caused by one incident.

Large Claims

Large claims distort the experience in individual cells and so are usually capped at a relatively low level. Any excess over the limit is assumed to be randomly distributed across all cells. It is added back once risk premiums have been fitted.

Deductibles

As far as the average claim amount is concerned, deductibles are by definition additive. That is if a policyholder makes a claim the deductible is in the form of £x rather than y% of the claim. If a multiplicative model is being fitted to the average claim amount this can be allowed for by grossing up the average claim amount by the amount of the deductible prior to modelling.

When modelling claim frequency the level of deductible is included as a separate rating factor. Care needs to be taken when determining levels of this factor to allow for changes in the levels of deductible over time.

Selection of Link Function and Random Component

A logarithmic link function is almost always used in practice. As the simple example below shows this results in a model for the expected value with multiplicative factors, which is intuitively more obvious than an additive or more complex relationship between the factors.

$$\begin{aligned} \ln\left(\mu\right) &= \hat{\alpha} + \hat{\beta}x_1 + \hat{\gamma}x_2 \\ \Rightarrow \mu &= e^{\left(\hat{\alpha}+\hat{\beta}x_1+\hat{\gamma}x_2\right)} \\ &= e^{\hat{\alpha}} . e^{\hat{\beta}x_1} . e^{\hat{\gamma}x_2} \end{aligned}$$

The random components often used in practice are the Poisson distribution for claim frequency and the Gamma distribution for average claim amounts. (The random component for the average cost per policy is more difficult to determine if there are a large number of zero values.)

It is important that the random component is not blindly chosen. The relationship between the mean and the variance for different distributions will help in identifying the most appropriate model.

Distribution	Mean-Variance Relationship
Normal	Var(y) is constant
Poisson	Var(y) α mean
Gamma	Var(y) α (mean)^2
Inverse Gaussian	Var(y) α (mean)^3

By plotting the residuals (see tests for model adequacy below) against the fitted values it is possible to identify whether the random component selected for the current model has removed any relationship that exists between the mean and the variance.

Model Weights

When fitting models to the claims experience more credibility is given to cells which contain a large amount of information. This is usually obtained by weighting the models for claim frequency and average claim amount by the exposure and the number of claims respectively. An alternative approach when a logarithmic link function is specified is to use an offset. A generalised linear model with an offset vector of a is defined as

$$g\left(\mu_i\right) = a_i + \sum_{j=i}^{p} x_{ij}\beta_j$$

A possible approach for claim frequency is then to use a logarithmic link, an offset of the log of exposure and to model the number of claims.

An additional point of interest is how to adjust the weight to allow for the proposed spread of new business because these are the areas of financial importance to a company.

Tests for Model Adequacy

There are four main methods by which the adequacy of a model can be assessed which are considered below. Applying these in practice is more of an art than a science and involves a significant amount of judgement.

1 χ^2 Test

If the current model is an adequate representation of the data then the scaled deviance has an asymptotic χ^2 distribution with n-p degrees of freedom, where n is the size of the data sample and p is the number of parameters in the current model. In the special case of the Normal distribution the χ^2 approximation is exact. The χ^2 test may not be valid in practice because of either insufficient data in some cells, or heterogeneity in cells where the data has not been adequately subdivided.

Since the χ^2 test is additive, if a model C_2 is a valid simplification of C_1 then the change in scaled deviance will have an asymptotic χ^2 distribution with $(n-p_2) - (n-p_1) = p_1 - p_2$ degrees of freedom.

The GLIM package generates a value for the deviance rather than the scaled deviance. Since the scale parameter is unity for Poisson and Binomial distribution functions it is possible to apply the χ^2 test directly, but care must be taken if the data is overdispersed where the variation in the data exceeds that expected for a Poisson distribution and so the scale parameter is in fact greater than unity.

2 F Test

The F Test involves the ratio of two χ^2 distributions and so eliminates the unknown scale parameters. This test often proves to be more robust than the χ^2 approximation in practice when sparse or non-homogeneous data is being used.

In practice with the very large data sets that are used for insurance premium rating the total deviance is very large. Adding additional parameters to this model will generally reduce the deviance by a significant amount, even if this is only a small proportion of the total deviance. There is therefore a tendency for the F Test to suggest that most parameters should be included in the model.

3 t Test

When a model is fitted parameter estimates are produced together with the standard error of these estimates. The t test can be applied to assess whether these parameters are distinct. For example if the parameter estimate is less than twice the standard error then it is often assumed that this parameter is not distinct from zero. Care should be taken when applying the t test because it is not reliable when applying this test to several parameters at the same time. The F and t tests are in fact identical when a single hypothesis is being tested, but applying the t test avoids the need to fit additional models.

4 Residual Plots

A residual is the difference between the fitted value and the data, given some appropriate weight. It is generally accepted that the most appropriate type of residuals to use when fitting generalised linear models are deviance residuals, rather than the traditional standardised Pearson residuals, as deviance residuals are not expected to be skewed for non-Normal distributions.

Plots of the residuals are a very useful way of identifying departures from the fitted model. The most useful plots are

- against fitted values to investigate the appropriateness of the distribution of the response
- against each explanatory variate to identify in particular the requirement for higher order interaction
- plot of the leverage and Cooks statistic to assess the influence of outliers
- normal plot and histogram of residuals
- against time if this is available.

No Claims Discount

For marketing reasons it is generally not possible to change the scale of no claims discount offered. This is therefore not generally fitted as a factor in the models used. Once the theoretical risk premiums have been calculated from the models it is necessary to gross these up by the average level of no claims discount for each permutation of risk factors. These average discount values need to be calculated from the data. Once the premiums have been grossed up for no claims discount a final model is then fitted to these values to smooth the premium rates.

Expenses

Finally expenses are added to the smoothed premium rates. It is desirable to attribute these expenses as accurately as possible. For instance, the modelled claim frequency can be used to attribute a cost per claim for each permutation of rating factors. The process of accurately attributing expenses can itself provide a very useful insight into the adequacy of premium rates. For example, if a fixed percentage of premium has traditionally been used to allocate expenses this may have overestimated the profitability of low premium (low risk) policies and underestimated it for high premium (high risk) policies.

4 Practical Results

The data supplied to working party members consisted of the number of policies, number of claims and amount of claims for a hypothetical motor account. This information was subdivided by four rating factors: engine power, age of insured, profession and sex. Low, medium and high powered engines were considered and the professions used were Actuaries, Lawyers and Accountants. The data set was generated stochastically and as such included random variation in both the claim frequency and average claim amount. Additional factors, for geographical location and car type, were used to generate the data but were not available for modelling purposes.

The initial decision that needed to be made was the approach to be taken for the age factor. Nearly all models grouped this data rather than fitting curves to it. This is the approach often used in practice, although curves can be useful at extreme ages. Summarising the data into one-way tables of claim frequency and average claim amount against policyholder age enabled the most appropriate groupings to be identified. One-way tables are also an extremely useful way of identifying data problems and as an initial means of identifying significant risk factors.

A variety of different types of models were fitted to the data and the results of these are compared below. Where the same model was fitted by different individuals the results were consistent, which was encouraging! This was true, in particular, regardless of whether extreme data points had been removed from the fit. A common feeling was that the data was much "better behaved" than was the case in real life, which is interesting as it is both stochastically generated and overdispersed.

The models fitted were as follows.

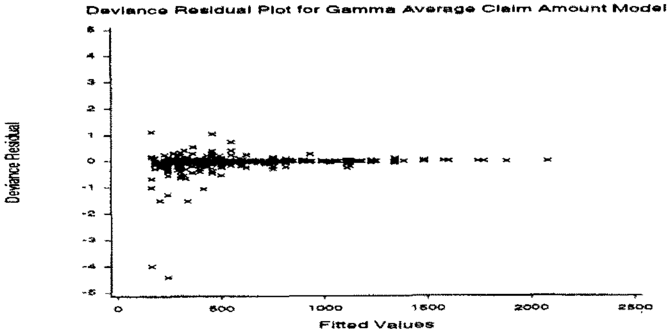
1 The average cost per policy with a log link and Gamma random component.

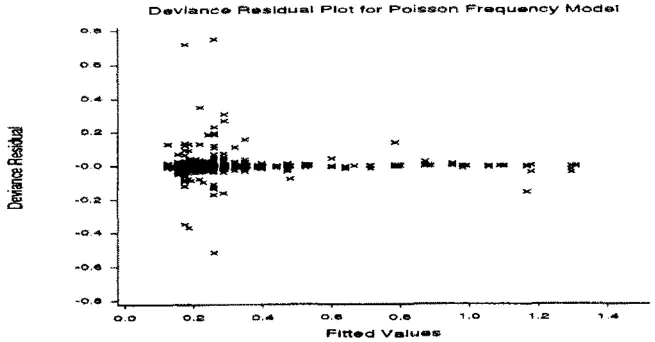
2 The frequency with a Poisson random component and average claim amount with a Normal random component, both with log links.

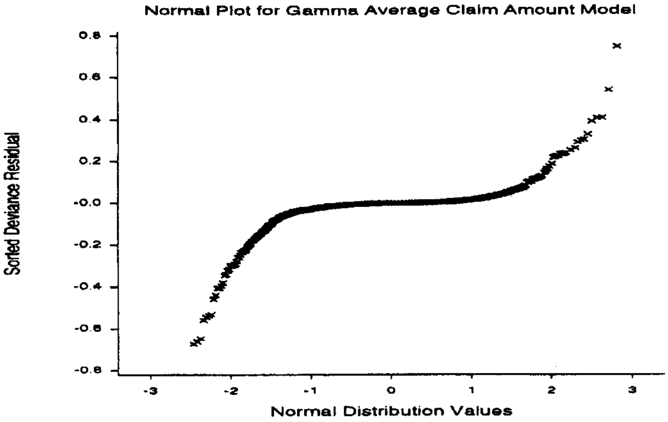
3 The frequency with a Poisson random component and average claim amount with a Gamma random component, both with log links.

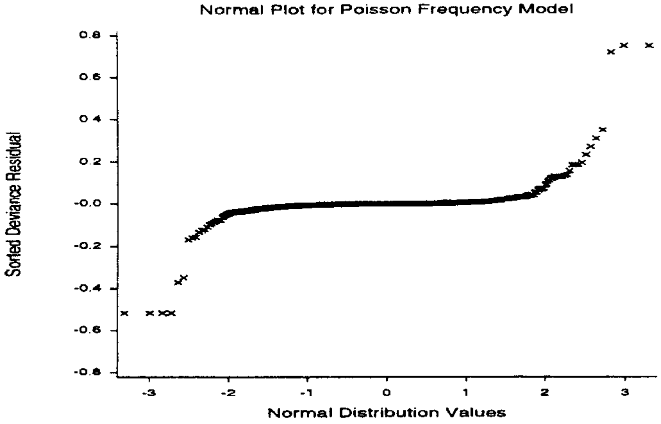
4 The average claim amount with a Gamma random component and a power link function.

Normal plots and plots of the deviance residuals against fitted values are illustrated below for model 3. These are particularly interesting because these distributions were used to generate the data. The scale parameter for the Poisson model was 1.1. The Normal plot is non-linear when deviance residuals are plotted, but is linear for standardised Pearson residuals.





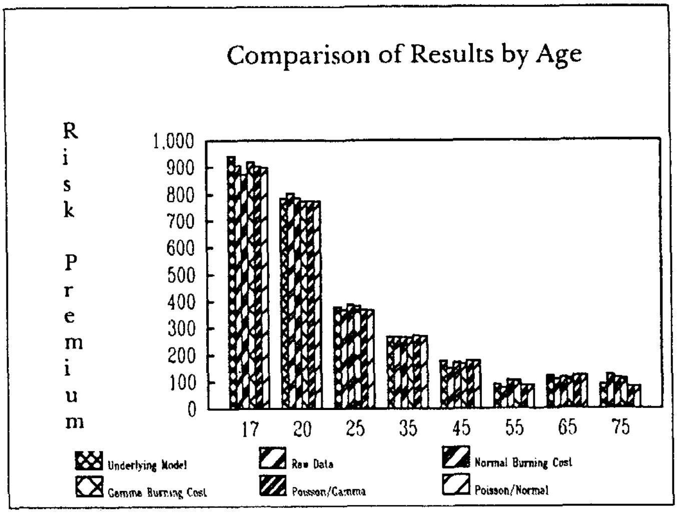

Normal Plot for Gamma Average Claim Amount Model
Sorted Deviance Residual
0.8
0.6
0.4
0.2
-0.0
-0.2
-0.4
-0.6
-0.8
-3
-2
-1
0
1
2
3
Normal Distribution Values

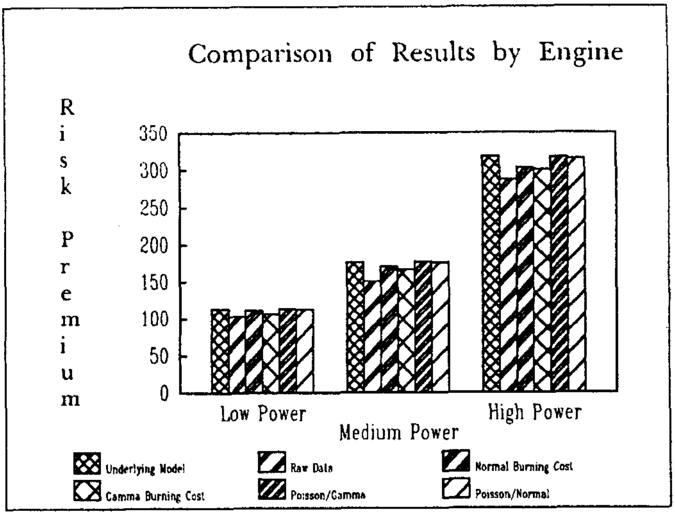

Normal Plot for Poisson Frequency Model
Sorted Deviance Residual
0.8
0.6
0.4
0.2
-0.0
-0.2
-0.4
-0.6
-0.8
-3
-2
-1
0
1
2
3
Normal Distribution Values

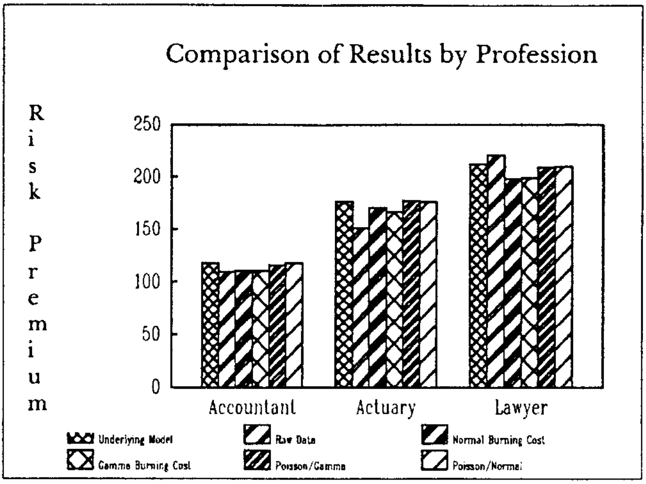
It is difficult to concisely compare the results of different models. To achieve this we have concentrated on one particular individual; a male accountant, aged 45, driving a medium powered car. The way in which the risk premium varies for this individual with changes in each of the four rating factors is compared in the graphs below.

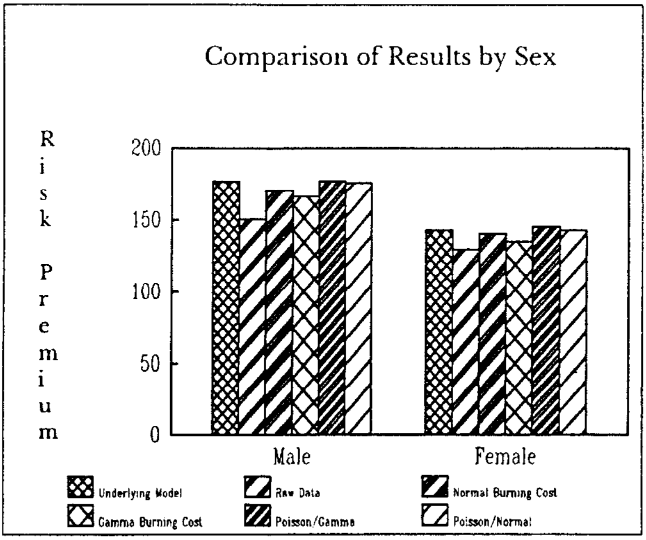
The key features are as follows.

- The power link function was used for claim severity by one working party member. This produced a relatively poor fit to the data, possibly because all the variables were treated as continuous rather than as factors with discrete levels.
- Estimates made using one-way tables from the raw data consistently overestimate the risk premium for this individual but provide a guide to the relativities for each factor.
- The data was summarised by each rating factor, for example the risk premium for sex is calculated from the data only for accountants aged 45 driving medium powered cars. This provides a more accurate indication of the true risk premium, but the relativities are distorted by random variation.
- Fitting a generalised linear model with a Gamma random component to the burning cost (the average claim amount per policy) generally provides a more accurate estimate of the true risk premium than summaries of the raw data.
- Separate models of frequency and amount give a consistently better estimate of the true risk premium. Selection of either a Gamma or Normal random function has only a marginal effect upon the estimated risk premium.
- Risk premiums generated by the models are nearly always less than that of the underlying model. This is thought to reflect the fact that the models produce maximum likelihood estimates, the distribution of which is skewed. Maximum likelihood estimates are only asymptotically unbiased, that is they are expected to equal the expected value of the underlying parameters with an infinite sample size. The practical effect of this is small, but it reflects an interesting area for possible future study.
- The poorest fit of the models is to the youngest ages where the experience is least stable.


Comparison of Results by Age
Risk Premium
1,000
900
800
700
600
500
400
300
200
100
0
17
20
25
35
45
55
65
75
Underlying Model
Raw Data
Normal Burning Cost
Gamma Burning Cost
Poisson/Gamma
Poisson/Normal


Comparison of Results by Engine
Risk Premium
350
300
250
200
150
100
50
0
Low Power
Medium Power
High Power
Underlying Model
Raw Data
Normal Burning Cost
Gamma Burning Cost
Poisson/Gamma
Poisson/Normal


Comparison of Results by Profession
Risk Premium
250
200
150
100
50
0
Accountant
Actuary
Lawyer
Underlying Model
Raw Data
Normal Burning Cost
Gamma Burning Cost
Poisson/Gamma
Poisson/Normal


Comparison of Results by Sex
Risk Premium
200
150
100
50
0
Male
Female
Underlying Model
Raw Data
Normal Burning Cost
Gamma Burning Cost
Poisson/Gamma
Poisson/Normal

Careful consideration needs to be made when applying any such models to future business. In particular the impact of competitiveness and hence new business volumes will be crucial. Model office projections provide a valuable insight when considering possible future scenarios. Also the effect policyholder selection can be significant. For example young drivers purchasing comprehensive cover may have been historically overcharged and so very profitable. This could reflect the fact that this cover was expensive and so was purchased by a particular type of young person. Reducing the cost of comprehensive cover for all young people may result in a different type of policyholder purchasing the cover.

Appendix A

Maximum Likelihood Estimation

Consider the following simple linear model, which has an identity link function and represents the normal straight line regression model,

$$y_j = \alpha + \beta\, x_j + \varepsilon_j \qquad j = 1,\ldots,n.$$

where $\varepsilon_j \sim N(0, \sigma^2)$

Then assuming the distribution function is correct, the likelihood of the values y_j and x_j being obtained is

$$\Pi\left(1/\sigma\,\sqrt{2\Pi}\right)\exp\left\{-1/2\sigma^2\left(y_j - \sigma - \beta x_j\right)^2\right\}$$

maximising this function is equivalent to minimising

$$\Sigma\left(y_j - \alpha - \beta x_j\right)^2$$

Taking partial derivatives with respect to α and β and equating to zero gives,

$$\hat{\alpha} = \bar{Y} - \beta\bar{X}$$

$$\hat{\beta} = \frac{\Sigma\, x_k y_k - 1/n\left(\Sigma\, x_k\right)\left(\Sigma\, y_k\right)}{\Sigma\, x_k^2 - 1/n\left(\Sigma\, x_k\right)^2}$$

These estimators are clearly identical to those produced using least squares estimation and are asymptotically unbiased. Unbiased estimators use a denominator of n-1 in the estimate for $\hat{\beta}$.