J.I.A. 119, III, 457-543

STATISTICAL MOTOR RATING: MAKING EFFECTIVE USE OF YOUR DATA

BY M. J. BROCKMAN, B.Sc., F.I.A. AND T. S. WRIGHT, M.A., F.S.S., M.I.S.

[Presented to the Institute of Actuaries, 27 April 1992]

ABSTRACT

The paper gives details of statistical modelling techniques which can be used to estimate risk and office premiums from past claims data. The methods described allow premiums to be estimated for any combinaton of rating factors, and produce standard errors of the risk premium. The statistical package GLIM is used for analysing claims experience, and GLIM terminology is used and explained thoughout the paper.

Arguments are put forward for modelling frequency and severity separately for different claim types. Fitted values can be used to estimate risk premiums, and the incorporation of expenses allows for the estimation of office premiums. Particular attention is given to the treatment of no claim discount.

The paper also discusses possible uses of the modelled premiums. These include the construction of 'standardised' one way tables and the analysis of experience by postal code and model of vehicle. Also discussed is the possibility of using the results for assessing the impact of competition, and for finding 'niche' markets in which an insurer can operate both competitively and profitably.

KEYWORDS

General Insurance; Motor; Pricing; Statistical Analysis

1. INTRODUCTION

1.1.1 There have been a large number of papers written on motor insurance which make use of statistical modelling. Many of these papers were published well over twenty years ago (Almer, 1957; Bailey & Simon, 1960; Jung, 1966). Further progress was made during the 1970s with papers by Johnson & Hey (1971) and Bennett (1978) and more recently papers by Baxter, Coutts & Ross (1980) and Coutts (1984). These papers are reviewed in more detail in Section 1.7. It is rather surprising, therefore, that actuaries have not established themselves more firmly or been more widely accepted by the market as having an essential role to play in the pricing process of motor business. Surely motor insurance, the largest single class of personal lines business in the U.K., is one line of business where the talents of actuaries should be utilised to the full.

1.1.2 Current practice and methodology in the U.K. market can be considered, on the whole, statistically unsophisticated. The authors believe that a full statistical analysis of a company's data is essential in the pricing decision-making process, where it can be combined with careful underwriting control to help

improve profitability. Recent developments in micro computer technology have meant that detailed statistical analyses can now be carried out quickly and efficiently on a micro computer, and that there is no longer any need for extensive mainframe processing. For very large accounts, the mainframe need only be used for the initial data processing stages. A data transfer to a micro based database can provide the flexibility required for statistical analysis.

1.1.3 Profit margins in the U.K. motor market over the period 1984-1990 have been extremely small. During this period, the average return after investment income has amounted to little more than 2% of premiums before tax. It is doubtful whether this level of profitability can justify the substantial amount of capital which is required to support the business. In order to provide an acceptable return on capital, a company must now aim to be a consistent above-average performer.

1.1.4 This paper outlines a statistical approach which, we believe, can help a company achieve improved profitability. Our approach requires substantial data analysis, in which statistical modelling techniques help unravel the trends and patterns in the data. We have found that such an exercise can provide the underwriter with much additional information on all the relevant factors which are driving the claims experience of the account. It is essential to get the theory right, and, unavoidably, this paper is rather more theoretical than some might hope. However, it is important not to lose sight of reality and that the theoretical results are capable of practical application. Also, the methodology must be capable of adapting to the constant changes in market practice. These points were firmly in our minds when we developed the theory in this paper.

1.2 The Two Main Aspects of Premium Rating

1.2.1 There are two main aspects of premium rating. First, the relative premium levels need to be determined; for example, it is important to charge the correct premium for old drivers relative to young drivers, new cars relative to old cars, or high car groups relative to low car groups. Secondly, the overall premium levels must be adequate in order to meet particular profit objectives.

1.2.2 It is this first aspect which we address in this paper, indeed we do not believe that the second can be tackled effectively using a statistical analysis of past claims data. All that is required for the second aspect is to adjust the overall level of premiums to meet a particular profit objective, taking into account shortterm economic effects and other external factors. If the broad structure of rates is correct, then the second aspect, that of determining the overall level of premium rates, becomes much easier.

1.3 The Importance of Getting the Rating Structure Right

1.3.1 We emphasised in the previous section the importance of getting the broad rating structure right. There are two reasons for this. First, the motor market is highly competitive. If a company does not broadly charge the correct premium rate in a particular part of its portfolio, it will be selected against.

Secondly, it allows a company to operate profitably in particular 'niches' of the market.

1.3.2 It is essential that a company minimises the possibility of adverse selection, since it could otherwise lead to a spiral of unstable pricing action. Suppose a company charges too little for low car groups and too much for high car groups. Over the whole portfolio the premiums may be deemed to be at the correct level. However, eventually the company will tend to gain low car group policies and lose high car group policies. The overall profitability of the portfolio will begin to deteriorate, since low car groups are undercharged. If, in future, an overall premium increase is applied in order to address the deterioration in profitability the situation will be further aggravated, and the profitability will continue to deteriorate. Unless the root cause is corrected, and premiums are increased for low car groups relative to high car groups, the portfolio will not be profitable or produce stable results. If the broad structure of premium rates is right, then overall premium increases can be applied with more confidence and a more stable pattern of premium increase can be achieved. It is usually sufficient to concentrate on the primary rating factors if selection is to be minimised. These are usually considered to be cover, vehicle group, vehicle age, policyholder age, district, no claims discount (NCD), and vehicle use. However, we have found that the sex of policyholder and voluntary driving restrictions also have an important influence.

1.3.3 There has been much recent comment on the 'search for the niche' and 'niche underwriting'. Much of the comment has focused on hunches and market 'feel'. We believe that, if the techniques advocated in this paper are followed, it will help an underwriter to use his own data to identify the combination of rating factors which will enable him to operate both profitably and competitively. We have no doubt that an underwriter's 'gut feeling' will always be important in the pricing decision; however it is one thing to think you know, another to know you know. Even in the current extremely competitive market there are clear segments where companies could improve their profitability considerably.

1.4 The Multifactor Approach

1.4.1 This paper outlines a multifactor approach to premium rating. The historical claims experience is analysed on a multi-way rating factor basis. That is, for every possible combination of rating factors or 'cell', the historical claims experience is used to calculate a claim frequency and an average cost. Statistical models are fitted to the raw claim frequency and raw average costs, and the results are used to calculate an expected frequency and average cost for each cell. Separate models are chosen for each type of claim: for example, windscreen, accidental damage, third party property damage, third party bodily injury, fire, theft and zero claims. The expected claim frequencies and average costs for each type of claim are combined to produce theoretical premium rates for each cell.

1.4.2 It is worth considering at this point the size of the estimation problem. Suppose we limit the analysis to the seven principal rating factors mentioned in

Section 1.3. The cells are constructed by defining a number of levels for each of the rating factors. One possibility is shown in Table 1.4.2.

Cover	Vehicle	Vehicle	Policyholder	District	NCD	Vehicle
	group	age	age			use
Comprehensive	1	0 1	17-19	Α	0	SDP
Non-comprehensive	2	2-3	20-22	В	30	Class 1
	3	46	23-25	С	40	Class 2/3
	4	7-9	26-30	D	50	,
	5	10+	31-35	Е	60	
	6		36-45	F	Protected	
	7+		46-60			
			61+			

Table 1.4.2. Multifactor breakdown of data

This produces over 60,000 cells of information. The statistical estimation problem appears immense. One solution would be to reduce the number of factors analysed: a second to introduce wider bands within each factor. However, this seriously reduces the effectiveness of the statistical modelling approach. The underwriter needs to know the shape of the rating structure in increasingly fine detail. Recently the ABI has proposed the introduction of 20 vehicle groups. This trend is likely to continue. We therefore believe it is essential to provide the underwriter with the relative premium rates in as much detail as possible. In order to achieve this it is essential that statistical models are chosen very carefully, so that sensible results can be estimated from the small exposures in each cell. In particular, it is important to take full advantage of possible relationships between adjacent cells: for example, the likely progression in experience from age to age of policyholder or vehicle.

1.4.3 We have included NCD level as one of the 'rating factors' in §1.4.2, because both claim frequency and severity differ significantly between NCD levels for most claim types. Therefore, the inclusion of NCD level as an explanatory variable in the statistical models will allow a closer fit to the data. However, the fitted models usually indicate that the true relativities of expected loss between levels of the NCD scale are not the same as those implied by the existing percentage discounts. Unless the percentage discounts for each NCD level are changed to correspond to these estimated relativities, the other parameters of the fitted model will not indicate the correct relativities for the other rating factors. The NCD relativities indicated by the model will not necessarily be round percentages and, although they will represent the 'true' relationship between NCD level and risk, it is often undesirable to adopt them. For this reason, we believe it is often better not to include NCD level as an explanatory variable in practice, especially in cases where the difference in the true and existing percentage discounts is not material. If these differences are material, a change to the scale and stepback rules may be more appropriate.

1.5 Why Fit Statistical Models?

There will be many patterns which will 'fit' the claims data. Some will fit better than others, although none of them will be 'right'. Amongst these will be the underwriter's current rating structure. The statistical model will attempt, in engineering terminology, to separate the signal from the noise: in other words, extract the underlying pattern from the actual claims experience. The parameters of the model can help in understanding the patterns in the data. The relative importance of the underwriting factors can be tested statistically, as can the existence of 'interaction' effects between underwriting factors. There will always be a trade off between finding the simplest model and the model which fits the data closest. The aim of statistical modelling is to find the best compromise between these two objectives. We believe that the methods outlined in this paper go a long way towards achieving this aim.

1.6 One-Way Tables

1.6.1 Most companies regularly produce underwriting factor summary tables from their statistical databases. These generally show the key underwriting statistics for each underwriting factor on a one-way basis. Hence statistics such as claim frequency, average cost, risk premium and loss ratio are monitored, for each underwriting factor and level within each underwriting factor.

1.6.2 These 'one-way' tables are a useful means of monitoring the relative profitability of the levels within each underwriting factor, but will not give an accurate indication of the relative profitability between levels. Indeed, they can often be misleading and result in incorrect rating action.

1.6.3 The first problem is that there may be important 'interactions' between different underwriting factors, which will produce misleading results if presented on a one-way basis. This can be overcome by analysing the underwriting statistics on a multi-dimensional basis. Provided that the dimensions do not exceed about three, this should be manageable. For example, if the relative claims experience of males and females varies by policyholder age, then separate one-way tables for policyholder age can be produced for males and females. However this assumes a priori knowledge that there is an 'interaction effect' between sex and policyholder age.

1.6.4 The second problem is that the distribution of business is not identical within each level of each underwriting factor. For instance, the claims experience on a one-way basis may show that young females are more profitable than young males. However, a further inspection of the data may reveal that young females tend to drive vehicles in lower classified groups. Hence the apparent improved profitability may well be intrinsically related to the relative profitability of vehicle groups. There may well be no difference in the relative profitability between young males and young females insuring the same type of vehicle.

1.6.5 This problem is more difficult to overcome. The traditional actuarial solution to this problem is to 'standardise' the claims experience in an attempt to

eliminate the distorting effects of the uneven distribution of business within each level of each underwriting factor, so that valid comparisons of the claims experience can be made on a one-way basis. The standardisation approach requires careful application; we discuss this issue further in Section 9.

1.6.6 The modelling approach, as advocated in this paper, deals with both these problems automatically. If the approach is integrated with a system of careful control using both traditional one-way tables and standardised one-way tables, an effective underwriting control system can be developed.

1.7 Review of Previous Papers

1.7.1 The early papers on statistical modelling in motor insurance focused on claim numbers with little consideration of the size of claims. The debate centred mainly on whether an additive or multiplicative model should be used in relating claim frequency to rating factors. Almer (1957) suggested a multiplicative model and this was taken up and developed by a series of subsequent authors, notably Jung (1968), whereas Johnson & Hey (1971) in an influential paper, advocated an additive model. These early papers were very well summarised by Bennett (1978), who compared the various fitting methods which had been proposed at that time for both the additive and multiplicative models. Bennett came down in favour of the additive model, largely on the grounds that it could be fitted relatively simply.

1.7.2 These early papers might now be considered statistically unsophisticated (which perhaps accounts for the plethora of estimation methods which evolved). Baxter, Coutts & Ross (1980) attempted to remedy this. They advocated a statistically more rigorous approach in which assumptions concerning the structure of the random variation are explicitly stated, fitting is carried out in accordance with those assumptions, and the assumptions are critically examined after fitting by means of a residual analysis. They used the Royal Statistical Society's computer package GLIM (first released in 1977) to implement this approach, which they applied to both claim numbers and claim sizes.

1.7.3 For claim numbers Baxter *et al.* carried out an empirical comparison of several models using a single data-set. In our opinion, such an investigation cannot be conclusive and should not lead to any general recommendations, because the findings may depend on peculiarities of the particular data-set used. We think that the use of a small number of levels for each rating factor in their analysis renders generalisations particularly unsafe. However, Baxter *et al.* finished by recommending a method called OWLS (for which GLIM was not used) on the grounds that, for their single data-set, it gave results in close agreement with those from other models fitted using GLIM, but was computationally simpler than the GLIM models. The matter of computational simplicity was important in 1980 largely because of the limits on computer memory space. Today, with increased memory space, the GLIM models (which Baxter *et al.* implicitly accepted as technically superior to the OWLS approach) are computationally very simple, as we will show. For claim sizes, Baxter *et al.* used

an additive model, but failed to find a satisfactory error structure (a fact clear from their paper, but only later made explicit by Coutts (1984)).

1.7.4 Coutts (1984) continued with the OWLS model recommended in Baxter et al. for claim frequency. For claim sizes he accepted that the model of Baxter et al. was not satisfactory, and tried to formulate a model based on a suggestion by Nelder. This was unsuccessful and he settled for an additive model, again fitted using his own OWLS technique. One area in which Coutts (1984) made significant progress was in the calculation of office premiums (from the results of the component frequency and severity models) for input to a final modelling stage to produce a 'smooth' structure, which can easily be compared with the existing rate book, or with theoretical rates obtained using different assumptions.

1.8 GLIM

1.8.1 We make extensive use of GLIM in motor-rating analyses in practice. The problems encountered in motor-rating are of precisely the type for which GLIM was designed. Models which would appear complex if tackled manually or using ad hoc computer programs can be fitted quickly and simply using GLIM. This allows models to be selected solely on the basis of their technical suitability. GLIM code is used throughout Sections 2-8 as a concise means of describing our models. A brief outline of GLIM, covering only what is necessary for an understanding of this paper, is given in Appendix A. Appendix A also shows how the claim frequency models described by Bennett (1978) can be fitted very easily using GLIM. We strongly recommend that readers unfamiliar with GLIM study Appendix A before proceeding to Section 2. Apart from the papers by Baxter et al. and Coutts mentioned above, there have been several other papers advocating the use of GLIM in actuarial work, for example Haberman & Renshaw (1988) and Renshaw (1991). The theory behind GLIM is very general, and the program is used widely outside the actuarial world. A good example is Little (1987): this paper includes a lucid account of the underlying ideas. A more comprehensive account of the theory, with example applications, is given by McCullagh & Nelder (1989). For the details of how to use the software, reference should be made obviously to the GLIM manual (NAG, 1985), but we also recommend Aitkin et al. (1989) and Healy (1988).

1.9 Summary of Later Sections

1.9.1 The particular models which we advocate in this paper have been selected mainly on theoretical grounds, but are supported by empirical experience. For claim frequency we advocate the use of Bennett's model C, that is, the multiplicative model with a Poisson error structure (See Appendix A). This is also one of the models considered by Baxter *et al.* (1980): they used it to illustrate GLIM in their appendix. Our reasons for favouring this model are given in Section 2 and Appendices B and C. For claim sizes, we favour a multiplicative model with a Gamma error structure (i.e. $V(\mu) = \mu^2$: see Appendix

A). This is the error structure suggested by Nelder (see McCullagh & Nelder (1983, 1989)), which Coutts (1984) tried and found to be unsatisfactory. Full details on this model, including prior reasons for favouring it, are given in Section 3, whilst Appendices E and F give details of techniques which can be used to check its validity.

1.9.2 By using these models for frequency and severity we have been able to introduce a number of innovations to the modelling process, which increase the reliability of estimated risk and office premiums, and which give significant practical advantages. These are dealt with in Sections 4, 5 and 6. Having obtained estimated office premiums (Section 7), we follow Coutts (1984) in carrying out a final stage of modelling to smooth these into a simple structure (Section 8). Our methodology for this stage is, however, very different from that used by Coutts, because we allow for the fact that the distribution of policies over NCD levels may be different for each cell defined by the other rating factors. We consider two possibilities:

- (i) the structure of the NCD scale is to remain unchanged, but the percentage discounts for each level of the scale are to be changed to reflect the true relationship with loss as indicated by the data, and
- (ii) both the structure of the NCD system and the percentage discounts for each level of the scale are to remain unchanged.

The former possibility is the simpler. It can be tackled by having NCD level as an explanatory variable in both the frequency and severity models. The second option, which is often required in practice (for the reasons outlined in § 1.4.3), cannot be handled in the same way (this is proved in Appendix J). Rather than including NCD level as an explanatory variable, we propose a separate model to estimate the average NCD within each cell (Section 7.2 and Appendix H). The numerical examples in the paper all relate to an analysis of type (ii).

1.9.3 In Section 9 we show how the results from the modelling can be used to build an effective standard table, which can also be used to investigate the importance of the underwriting factors not included in the modelling process. We also discuss how the standard table can be used to investigate the appropriateness of postal code to district classifications, and make/models to vehicle group classifications.

1.9.4 Finally, in Section 10 we show how the results of the modelling can be used in conjunction with an analysis of competitors' rates to test how far a company can move its existing rating structure to that indicated by the theoretical analysis. The objective is to improve profitability by identifying areas of the market where profit margins are likely to be larger, and by increasing exposure in these more profitable areas.

1.9.5 Although the paper deals specifically with motor insurance rating, the methods may be applied more generally, for example to health and household insurance, where the premium rates depend on several factors.

2. BASIC CLAIM FREQUENCY MODEL

2.1 The Case for a Multiplicative Model

2.1.1 In the past there has been much debate over whether an additive or multiplicative model should be used in relating claim frequency to rating factors: a brief history is given in Section 1.7. We believe that there are clear and substantive reasons for preferring a multiplicative model. First, it should be noted that an additive model can give rise to negative fitted values for the claim frequency, whereas a multiplicative model cannot. However, this is merely a symptom of the inappropriateness of the additive model: we attempt to explain our reasoning in the remainder of this sub-section.

2.1.2 Car age is now widely recognised as an important risk factor: risk generally decreases quite substantially with car age. The reason is probably a negative association between car age and annual mileage. (Clearly it would be preferable to use annual mileage directly as the rating factor, but this is not common practice, owing to the difficulty in obtaining a reliable and objective forecast. If annual mileage were used as the rating factor, it would strengthen the argument given here.) Focus attention on car group and car age: suppose all other factors are held constant. Suppose that for car group 1 the claim frequency is 0.05 for cars aged 10 years, and 0.10 for cars aged 1 year. Now consider car group 6 and suppose that the claim frequency is 0.10 for cars aged 10 years. What should we expect the frequency to be for cars aged 1 year? Under the additive assumption the figure would be 0.15, whereas under the multiplicative assumption it would be 0.20. Given the belief that car age is a surrogate for annual mileage, the multiplicative model is clearly more plausible (the explanation is that 10-year-old cars cover about half the mileage of 1-year-old cars, for each car group).

2.1.3 Of course car age is a special case (in that it is believed to be highly associated with the exposure in terms of annual mileage), so the above argument in favour of the multiplicative model does not generalise directly to other rating factors. However, consider policyholder age and car group. For policyholder age 30, suppose we have claim frequencies of 0.05 for car group 1 and 0.10 for car group 6, and suppose that for policyholder age 17 we have a claim frequency of 0.20 for car group 1. For car group 6 the additive assumption gives a frequency of 0.25, whilst the multiplicative assumption gives 0.40. We find it much more plausible a priori that 'a car in group 6 has twice the risk of a car in group 1 whatever the drivers age' than that 'a car in group 6 is likely to have one more accident every 20 years than a car in group 1 regardless of the safety of the driver'. The same argument applies to any other combination of rating factors.

2.1.4 On these grounds it should be expected that an additive model may need interaction terms to provide a good fit (in the example, we need a policyholder age $17 \times \text{car}$ group 6 interaction of 0.15 in order to give the figure expected under the multiplicative model). An empirical investigation of these points is described in Appendix B. As mentioned in Section 1.5, modelling is always a matter of

finding the best compromise between accuracy and simplicity. In the present context, the multiplicative model is superior to the additive model in this respect: simplicity can be achieved (fewer interaction terms) without a major sacrifice in accuracy.

2.2 The Case for a Poisson Error Structure

2.2.1 The Poisson model for the incidence of claims (in any type of insurance) is so well known that there is no need to discuss it in full detail here. Chapter 2 of Beard *et al.* (1984) gives a comprehensive account. Given the suitability and general acceptance of the Poisson model, it is somewhat puzzling that it has not often been invoked when attempting to relate claim frequency to rating factors in motor insurance. Johnson & Hey (1971) is a case in point: in most sections of the paper they implicitly assume that the claims arising from each policy follow a Poisson process, but in the section which deals with a model relating claim frequency to rating factors, the Poisson assumption is dropped, and the model fitted on the basis that the number of claims on each policy has constant variance rather than a variance equal to the mean.

2.2.2 Of course, there are problems with the Poisson assumption, and these are discussed below, but initially an idealised situation will be considered. We suppose, for the purpose of this discussion, that there are two rating factors and their levels are indexed by i and j. Each particular combination of i and j is referred to as a rating cell (for example, all those policyholders aged 20-25 and driving cars in group 4). We use the following notation:

- x_{ij} = exposure in rating cell (*i*, *j*) (this could be the number of vehicle years for example), and
- n_{ij} = number of claims arising from the x_{ij} units of exposure in cell (i, j).

2.2.3 In an idealised situation, the rating factors are selected so that we have perfect homogeneity throughout each cell: that is, each unit of exposure in cell (i,j) has the same chance of yielding 1, 2, 3, etc. claims as any other unit of exposure in the same cell. Further, the distribution of the number of claims from a single unit of exposure is an exact Poisson distribution, and so has a variance equal to its mean. Thus, if f_{ij} denotes the mean of this Poisson distribution for cell (i,j), and m_{ijk} denotes the number of claims arising from the kth unit of exposure in this cell then we have:

$$E(m_{ijk}) = f_{ij}$$
 and $Var(m_{ijk}) = f_{ij}$.

Now, n_{ij} is the sum of the m_{ijk} for all x_{ij} units in cell (i, j), so, if we further assume that these units are mutually independent, we can apply elementary results on the mean and variance of a sum of independent random variables to obtain:

$$E(n_{ij}) = x_{ij} \cdot f_{ij}$$
 and $Var(n_{ij}) = x_{ij} \cdot f_{ij}$

(and in fact, n_{ij} is exactly Poisson distributed).

Dividing by x_{ij} and writing r_{ij} for the observed claim frequency in cell (i,j) $(r_{ij}=n_{ij}/x_{ij})$ we have:

$$\mathbf{E}(r_{ij}) = f_{ij}$$
 and $\operatorname{Var}(r_{ij}) = f_{ij}/x_{ij}$.

Thus, apart from the divisor x_{ij} , which is a known quantity, the variance of the data r_{ij} is equal to the mean. In the terminology of generalised linear models, r_{ij} has a Poisson error structure with prior weights x_{ij} and scale parameter $\phi = 1$ (see Appendix A).

2.2.4 We aim to discover the relationship between the mean claim frequency and the levels i and j of the two rating factors. For the intuitive reasons given in Section 2.1, this is usually best done using a multiplicative model, which can be expressed as:

$$f_{ij} = \exp(\alpha_i + \beta_j).$$

If the data r_{ij} and x_{ij} have been read into vectors R and X, and corresponding levels of the rating-factors have been named A and B, then this model can be fitted in GLIM using:

\$YVAR R	!	claim-frequency data R to be the y-variate
SERROR P	!	Poisson error structure
\$LINK L	!	log link function gives multiplicative model
\$WEIGHTS X	!	exposure data X to be prior weights
\$FIT A+B	!	estimate main effect of each rating factor

(Note that when **SERROR** P is declared, GLIM automatically assumes the value $\phi = 1$ for the scale parameter.)

2.2.5 In practice, none of the assumptions made above is likely to hold exactly. Most importantly, we will not be able to find rating factors such that we have perfect homogeneity of risk within each cell. It is well known that if the heterogeneity of risks follows a Gamma distribution (also known as Pearson type III), then the total number of claims $(n_{ij}$ above) is negative-binomially distributed (see for example Johnson & Hey (1971) or Beard *et al.* (1984)). This result is used in Appendix C to show that within-cell heterogeneity can be largely taken into account simply by allowing for over dispersion (that is by allowing $\phi > 1$). This is achieved in GLIM by inserting the command SCALE 0 before the fit command. This makes no difference to the parameter estimates given by GLIM, but it increases their standard errors. The GLIM code given above remains valid for investigating the relationship between the mean claim frequency and the rating factors. Note, however, that it is desirable to select rating factors which minimise within-cell heterogeneity, as this will maximise the explanatory power of the model. Further discussion on this point is given in Appendix D.

2.2.6 The assumption that each unit of exposure generates claims according to a Poisson process is violated in reality, because the risk intensity tends to decrease for a period after an accident: the vehicle may be out of service for repairs, or the driver, sobered by the experience, may drive more carefully for a while. It is

shown in Appendix C that these effects will merely tend to decrease the scale parameter, and that the decrease cannot possibly be substantial.

2.2.7 The assumption that the incidence of claims is mutually independent for the units within a cell is violated, because each vehicle covered is at risk from other vehicles covered by the same insurer and in the same rating cell: that is two or more of the vehicles may be involved in the same accident. It is shown in Appendix C that this tends to increase the scale parameter. However, the data will usually cover only a small proportion of the vehicles on the road, and the effect will be extremely slight. Note that effects such as weather conditions, which tend to increase claim frequencies for all units simultaneously, are not relevant here: such effects increase the mean claim frequencies f_{ij} , but we are concerned here with the question of mutual independence of the random variations about these means.

2.3 Model Testing in GLIM

2.3.1 The GLIM commands given in Section 2.2.4 will cause the main effects of the factors A and B to be estimated. In general, there may be more than two rating factors, and we may be interested in testing the significance of interaction terms, or in testing the suitability of other models such as those described in Section 6. For all these alternatives, it is the \$FIT command which varies: the arguments given in Sections 2.1 and 2.2 continue to apply, so the other parts of the model specification should remain unchanged.

GLIM estimates the parameters of the linear predictor specified in the \$FIT command by minimising an objective function known as the deviance. It then displays the minimised value of the deviance, together with the number of degrees of freedom (this is the number of data points less the number of parameters in the linear predictor). These quantities can be used to compare the quality of fit obtained using different linear predictors and hence to find the model which is in closest agreement with the data.

2.3.2 In this paper, the minimised deviance is denoted Q and the corresponding number of degrees of freedom D. In the idealised situation described in Section 2.2, if the linear predictor includes all those terms which genuinely influence claim frequency, then Q is approximately from the chi-squared distribution with D degrees of freedom. The quality of this approximation depends on the number of claims in each rating cell: if these are all sufficiently large, the approximation is good. In such a case, the value of Q could be used to test whether all terms which genuinely influence claim frequency have been included in the linear predictor: a value which is in the extreme right tail of the chi-squared distribution is evidence against this hypothesis. In other words, a large value of the deviance Q indicates that a significant amount of variation in the data is not explained by the current model. However, this test is not valid in practice for two reasons:

(1) We may not have a large enough number of claims in each cell for the chi-

squared approximation to be valid. This is particularly likely if we do not use broad categories for levels of the rating factors, but instead have a larger number of smaller categories (we advocate this in Section 6). It is well known that in such cases the deviance tends to be reduced. Thus, a value which is small compared to the chi-squared distribution would not necessarily imply a good fit, but a large value could still be taken as indicative of a poor fit.

(2) For the reasons given in Section 2.2. (the main one being within-cell heterogeneity) we are likely to have random errors which are over-dispersed relative to the idealised situation. This implies that the deviance will be increased by some unknown scale factor $\phi > 1$.

These two effects work in opposite directions, so when they both apply (as is likely in practice), nothing can be deduced about the quality of fit of a single model regardless of whether the deviance is large or small compared to the chi-squared distribution.

2.3.3 However, there is an alternative to the chi-squared test, namely the Ftest. This is much more robust, and will be reasonably reliable despite the effects mentioned in the previous paragraph. It cannot be used to test objectively whether a particular model taken in isolation provides a good fit, but only to compare two models, one of which is a simplified version of the other. This is not a problem in practice, for we can usually start with a model which contains so many terms in the linear predictor that we are quite confident that all those genuinely affecting claim frequency have been included, so the fit will be good. Let Q_1 and D_1 be the minimised deviance and the degrees of freedom for such a model, and let Q_2 and D_2 be the values obtained by fitting a simplified version of the model (terms may be removed from the linear predictor, or levels of one of the rating factors may be combined, for example). Note that we necessarily have $Q_2 > Q_1$ and $D_2 > D_1$. The F-statistic is given by:

$$F = \frac{(Q_2 - Q_1)/(D_2 - D_1)}{Q_1/D_1}.$$

This should be compared to values from the theoretical F-distribution with $(D_2 - D_1)$ and D_1 degrees of freedom. If the F-statistic appears unreasonably large in this comparison, then we would conclude that the simplified model is not valid: the increase in the minimised deviance is so large that it must partly be attributed to a significant loss of fit (i.e. the terms removed from the linear predictor are, in fact, significant in explaining the variation in claim frequency). On the other hand, if the F-statistic appears to be consistent with the theoretical F-distribution, then the simplified model can be accepted. Q_2 and D_2 can then be renamed as Q_1 and D_1 and further simplifications to the model investigated. Note that the quantity Q_1/D_1 in the denominator of the F-statistic is an estimate of the scale parameter ϕ , but a more reliable estimate when some claim numbers are small is the mean square of the standardised residuals (defined below).

2.3.4 When a model has been selected using a sequence of F-tests as described above, the assumptions it relies on can be checked by examining the standardised residuals. This is a standard procedure in good statistical practice, and is stressed by Baxter, Coutts & Ross (1980). They state that the standardised residuals should be approximately normally distributed. However, we believe that this is not necessarily true if many cells have a small number of claims. When allowance is made for over-dispersion (Appendix C) the model (§ 2.2.3) becomes:

 $E(r_{ii}) = f_{ii}$ and $Var(r_{ii}) = \phi f_{ii}/x_{ii}$

This implies that the quantity e_{ij} defined by:

$$e_{ij} = (r_{ij} - f_{ij}) / \sqrt{(f_{ij} / x_{ij})}$$

has

$$\mathbf{E}(e_{ij}) = 0$$
 and $\operatorname{Var}(e_{ij}) = \phi$.

After fitting, these quantities e_{ij} can be estimated by using the estimated mean claim frequencies obtained from the model in place of the unknown true values f_{ij} . The resulting values are known as standardised residuals and, if the model is good, their mean and variance are approximately 0 and ϕ . Since the variance does not depend on the levels *i* and *j* of the rating factors, we should see an evenly spread scatter of values around the mean value of zero in any plot of the standardised residuals (but the distribution is not necessarily Normal, and with small claim numbers, we would expect it to be skewed). We have always found these residual plots to be satisfactory in applications of this model.

2.4 Numerical Example

2.4.1 The data used for this and later examples are for own-damage claims arising from about 17,000 comprehensive policies over the three-year period 1986 to 1988, and have been chosen for illustrative purposes only. The rating factors are: district (DI), policyholder age (PA), car age (CA) and car group (CG). The levels of these factors are given in Table 2.4.1.

Level	DI	PA	CA	CG
1	Α	17-18	0-3	1
2	В	19-21	45	2
3	С	22-24	6.7	3
4	D	25-29	8-9	4
5	E	30-34	10+	5
6	F	35-44		6
7	G	45-54		7
8	N	55+		8

Table 2.4.	l. Rating f	actor i	levels	5
------------	-------------	---------	--------	---

Districts A to G are areas of the mainland U.K. categorised in order of increasing risk, as assessed by the insurer. Similarly, car-groups 1 to 8 contain models of car sorted in order of increasing risk. District N is Northern Ireland.

The total number of rating cells is $8 \times 8 \times 5 \times 8 = 2,560$, but there is no exposure in 670 of these cells over the period concerned, so there are only 1,890 data points for the claim frequency analysis. Each data point consists of six items: the level of each of the four rating factors, the exposure x (the number of policy years) and the number n of own-damage claims.

2.4.2 Table 2.4.2 gives some results of fitting multiplicative Poisson models to the observed claims frequency r=n/x, using the GLIM commands given in §2.2.4. Several models have been fitted by including different terms in the \$FIT command: first the main effects only (\$FIT DI + PA + CA + CG), then the main effects plus each two-factor interaction in turn.

Table 2.4.2. Analysis of deviance

Model	D	Q	F
DI+PA+CA+CG	1864	1892-1	
$DI + PA + CA + CG + DI \cdot CG$	1815	1828.8	1.28
$DI + PA + CA + CG + PA \cdot CG$	1815	1848·2	0.88
$DI + PA + CA + CG + CA \cdot CG$	1836	1861.8	1.07
$DI + PA + CA + CG + DI \cdot CA$	1836	1857-3	1.23
$DI + PA + CA + CG + PA \cdot CA$	1836	1850-5	1.46
$DI + PA + CA + CG + DI \cdot PA$	1815	1834-3	1.17

Using the minimised deviance Q and the residual degrees of freedom D, the F-statistic has been calculated to compare the model with only main effects, to each of the other models. For example:

$$1 \cdot 28 = \frac{(1892 \cdot 1 - 1828 \cdot 8)/(1864 - 1815)}{1828 \cdot 8/1815}.$$

The most significant interaction term on this basis is $PA \cdot CA$: tables show that the F-value of 1.46 corresponds to a probability of just over 5%. Because interaction terms make a fitted model considerably more difficult to interpret, we prefer not to include them unless the probability is less than 5%. None of the twofactor interactions is significant on this basis.

It would be more satisfactory to start by fitting a model with all two-factor interactions present simultaneously (and perhaps some three-factor interactions), and to test the significance of the interaction terms by removing each one in turn and carrying out an F-test as described in §2.3.3. However, this is sometimes impractical when there is a large number of factors.

2.4.3 The standardised residuals of the model with main effects only are plotted against fitted claim frequencies in Figure 2.4a, and against PA in Figure 2.4b. These plots show no evidence of heteroscedasticity, nor do similar plots of the residuals against the other rating-factors. The plots do show evidence of



Figure 2.4b.

positive skewness in the data: this is quite consistent with the modelling assumptions.

N.B. These plots and all other residual plots in this paper show the residuals for only one of the 8 districts. This is because of technical problems in reproducing the complete plots. We have tried to select a typical district for each residual plot.

The parameter estimates from the model with main effects only are given in Table 2.4.3.

Factor	Estimate	Standard error
1	-1.816	0.1216
DI(2)	0.0812	0.0424
DI(3)	0.1515	0.0443
DI(4)	0.2102	0.0437
DI(5)	0.0836	0.2274
DI(6)	0.4095	0.0721
DI(7)	0.4869	0.0701
DI(8)	0.2041	0.0892
PA(2)	- 0·395 7	0.1308
PA(3)	-0.4808	0.1207
PA(4)	-0.3587	0.1148
PA(5)	-0.4566	0.1214
PA(6)	-0·5678	0.1162
PA(7)	-0.6068	0-1161
PA(8)	0.8150	0.1148
CA(2)	-0·1090	0.0335
CA(3)	-0.2862	0.0439
CA(4)	0 ·3287	0.0609
CA(5)	0.5391	0.0753
CG(2)	0.0433	0.0580
CG(3)	0.0204	0.0546
CG(4)	0.0229	0.0564
CG(5)	0.2096	0.0571
CG(6)	0.1252	0.0606
CG(7)	0.2050	0.0617
CG(8)	0.5737	0.0811

Table 2.4.3. Parameter estimates

The first parameter is for the cell with each factor at level 1, that is, DI = A, PA = 17-18, CA = 0-3, CG = 1. The estimated claim frequency for each policy year in this cell is exp(-1.816) = 0.163. The remaining parameters represent the claim frequency in other cells relative to this first cell: the parameters are the natural logarithms of the multiplicative relativities. For example, claim frequency is higher in a cell with DI at level 2 than in a cell with DI at level 1 by a factor exp(0.0812) = 1.085. Similarly, claim frequency is lower for PA level 2 (19-21) than PA level 1 by a factor exp(-0.3957) = 0.673.

2.4.4 Levels 1 to 7 of DI represent areas of the mainland U.K. in order of ascending risk, as assessed by the insurer. However, the results indicate a decrease in claim frequency between DI levels 4 and 5, by a factor $\exp(0.0836-0.2102)=0.881$. This unexpected decrease is explained by the large standard error of the estimate for level 5 of DI. The estimate 0.0836 is very unreliable: the reason is that the data contain only a small sample of policies for level 5 of DI. It may be better to pool the data for levels 5 and 6 of DI, on the grounds of a prior belief that the claim frequency should not vary much between these two levels. This can easily be done in GLIM as follows:

ASSXDI = 1, 2, 3, 4, 5, 5, 6, 7	!	new levels of the district factor
		corresponding to previous levels 1 to 8
\$FAC DI7 7	!	new factor DI7 to have 7 levels
\$CAL DI7=XDI(DI)	!	calculation of new factor DI7

Similarly, the results indicate a decrease in claim frequency between car groups 3 to 4 and 5 to 6. These decreases appear to be insignificant (they are small compared to the standard errors) and, as they contradict prior beliefs, it may be better to pool the data for car groups 3 and 4 and for car groups 5 and 6 instead of attempting to estimate a separate parameter for each car group. This can be done

Table 2.4.5. Parameter estimates

Factor	Estimate	Standard error
1	-1.819	0.1217
DI7(2)	0.0814	0.0424
DI7(3)	0.1521	0.0443
D17(4)	0.2114	0.0437
DI7(5)	0.3814	0.0700
DI7(6)	0.4866	0.0702
DI7(7)	-0.2035	0.0892
PA(2)	-0.3906	0.1308
PA(3)	- 0.4760	0.1207
PA(4)	-0.3557	0.1149
PA(5)	-0.4557	0.1215
PA(6)	- 0.5673	0.1162
PA(7)	0.6065	0.1162
PA(8)	- 0.8128	0.1149
CA(2)	-0.1100	0.0335
CA(3)	-0.2882	0.0438
CA(4)	-0.3305	0.0608
CA(5)	0.5401	0.0753
CG6(2)	0.0433	0.0580
CG6(3)	0.0628	0.0203
CG6(4)	0.1731	0.0522
CG6(5)	0.2057	0.0617
CG6(6)	0.5751	0.0812

by creating a new car group factor with only 6 levels to replace the original 8 level factor:

\$ASS XCG = 1, 2, 3, 3, 4, 4, 5, 6 \$FAC CG6 6 \$CAL CG6 = XCG(CG)

2.4.5 The new main effects model FIT DI7+PA+CA+CG6, gives a minimum deviance of 1896.8: an increase of 4.7. There are 3 fewer parameters than in the original main effects model, so the F-statistic for comparing the present model with the original is 1.54, on 3 and 1864 degrees of freedom. This is insignificant, confirming that the pooling of districts and car groups is consistent with the data. The parameter estimates are given in Table 2.4.5.

There remains an unexpected increase in estimated claim frequency between PA levels 3 and 4. The construction of factor levels such that the parameter estimates increase or decrease in a manner consistent with prior beliefs is somewhat arbitrary: more satisfactory models are introduced in Section 6.

3. BASIC CLAIM SEVERITY MODEL

3.1 The Model: Multiplicative with Gamma Error Structure

3.1.1 The reasons for preferring a multiplicative model to an additive model for claim numbers (given in Section 2.1) apply equally to claim severity. For example, if own damage claims made by young policyholders tend to be z-times as expensive for cars in group 8 as for cars in group 1, it seems plausible that the same car group factor will apply for older drivers also. In the same way as for claim frequency, empirical comfirmation that the multiplicative model is suitable is provided by the lack of significance of interactions between rating factors.

3.1.2 The question of what error structure to assume for the claim severity model needs more care. Baxter, Coutts & Ross (1980) assumed that the distribution of individual claim sizes had the same variance for all rating cells. They fitted a model on the basis of this assumption, and then examined residuals in the usual way to check the assumption. They concluded that the residuals did not conflict with the assumption. However, Coutts (1984) refers to the analysis in Baxter *et al.* saying that, in fact, there was clear evidence that the variance of the residuals increased with the mean claim size. Thus, it appears that in cells where the mean claim size is large the variance of the individual claim sizes is correspondingly large.

We think that this makes intuitive sense, and that one can reasonably go further on intuitive grounds to postulate the nature of the relationship between mean and variance of the claim sizes. Suppose we know that, in a certain rating cell, the mean claim size is $\pounds1,000$ and the standard deviation of the claim sizes is $\pounds800$. If, in another cell, claims tend to be twice as severe so that the mean size is $\pounds2,000$, then we would expect to find a proportionate standard deviation of about $\pounds1,600$. Thus, it seems plausible to assume that the variance is proportional to the

mean squared, in other words that the coefficient of variation rather than the variance is constant across cells. Nelder has also advocated this assumption for the size of motor claims (McCullagh & Nelder 1989). We have empirically investigated the assumption and found that it provides a good starting point for severity models (see Section 3.3 and Appendices E and F). In GLIM, such an assumption is implemented by specifying a Gamma error structure, using the command \$ERROR G. As mentioned in Appendix A, this does not necessarily imply a belief that claim sizes follow a Gamma distribution.

3.1.3 We are now in a position to formulate our claim severity model in detail. As in Section 2.2, we suppose for simplicity that we have only two rating factors, and these are indexed by i and j. The data consist of:

 n_{ij} = number of claims arising from cell (*i*, *j*), and

 Y_{ij} = total amount of the n_{ij} claims in cell (i, j).

If Z_{ijk} denotes the size of the kth claim in cell (i, j) and m_{ij} denotes the mean claim size, then by the assumption of constant coefficient of variation we have:

$$E(Z_{ijk}) = m_{ij}$$
 and $Var(Z_{ijk}) = \sigma^2 \cdot m_{ij}^2$

for some constant σ (the coefficient of the variation).

Now Y_{ij} is the sum of the Z_{ijk} for all n_{ij} claims in cell (i, j), so if we further assume that these claim sizes are mutually independent, we can apply elementary results on the mean and variance of a sum of independent random variables to obtain:

$$\mathbf{E}(Y_{ij}) = n_{ij} m_{ij} \quad \text{and} \quad \operatorname{Var}(Y_{ij}) = n_{ij} \sigma^2 m_{ij}^2.$$

Dividing by n_{ij} and writing S_{ij} for the observed mean claim size in cell (i, j):

 $E(S_{ij}) = m_{ij}$ and $Var(S_{ij}) = \sigma^2 m_{ij}^2 / n_{ij}$.

Thus the data S_{ij} have a Gamma error structure with weights n_{ij} and scale parameter $\phi = \sigma^2$ (see Appendix A).

3.1.4 We aim to discover the relationship between the mean claim size and the levels *i* and *j* of the two rating factors. As in Section 2.2, this is best done using a multiplicative model, which can be expressed as:

$$m_{ij} = \exp(\alpha_i + \beta_j).$$

If the data S_{ij} and n_{ij} have been read into vectors S and N, and corresponding levels of the rating factors have been named A and B, then this model can be fitted in GLIM using:

\$YVAR S \$ERROR G \$LINK L \$WEIGHTS N \$FIT A+B GLIM will give both the parameter estimates (with standard errors) and an estimate of the scale parameter, which, for this model, is the square of the withincell coefficient of variation of individual claim sizes. In any application, the best form for the linear predictor can be found by using a sequence of F-tests in the standard way as described in Section 2.3 for the claim frequency model. This is considered further in Section 6.

3.2 Testing the Assumptions

3.2.1 Coutts (1984) pointed out that the claim severity model of Baxter *et al.* (1980) was suspect, because the scale parameter given by GLIM (which for their model was an estimate of the within-cell variance of claim sizes, assumed constant) was about twice as large as a direct estimate of the within-cell variance obtained from data on individual claim sizes. This is not surprising in view of the intuitive argument given in Section 3.1 that the within-cell variance cannot, in fact, be constant (thus the two problems mentioned by Coutts in relation to the model of Baxter *et al.* are intimately related).

However, for our model, we should attempt to verify that the scale parameter given by GLIM agrees with a direct estimate of the within-cell coefficient of variation obtained from data of individual claim sizes. A direct estimate can be obtained as described in Appendix E. We have never found a substantial discrepancy between these two estimates of the within-cell coefficient of variation. See § 3.3.2 for example.

3.2.2 The only assumption of our claim severity model, other than that claims are mutually independent, is that the coefficient of variation σ is the same for all cells. In any application of the model, this assumption can and should be checked by examining plots of the standardised residuals. From the model of Section 3.1:

$$E(S_{ij}) = m_{ij}$$
 and $Var(S_{ij}) = \sigma^2 \cdot m_{ij}^2/n_{ij}$

we define

$$e_{ij} = (S_{ij} - m_{ij}) \cdot \sqrt{n_{ij}/m_{ij}}.$$

The standardised residuals are obtained by replacing the unknowns m_{ij} by their fitted values. We should then find $E(e_{ij}) = 0$ and $Var(e_{ij}) = \sigma^2$ approximately. Thus, the variance should not depend on the values of the rating factors, and should appear constant in any plot. Any significant variation in the variance (heteroscedasticity) implies that the assumption of constant coefficient of variation is false. In such a case, it is necessary to model the variation of σ^2 across cells in order that the assumption can be modified. This is not as straightforward as it may appear at first sight. We could use the individual claim size data to calculate for each cell the observed mean and sample standard deviation, and hence an estimate of the within-cell coefficient of variation. However, in view of the small number of claims likely in some cells and the large amount of variation in individual claim sizes, a great deal of between-cell random variation in this estimate should be expected, and it would be extremely difficult to judge whether

or not the observed variation were significant. Thus such an exercise would not be conclusive.

3.2.3 In Appendix F we give a formal statistical test of the hypothesis that the within-cell coefficient of variation is constant. The techniques described there also allow the coefficient of variation to be modelled, giving estimates $\hat{\sigma}_{ij}^2$. If these are read into a GLIM variate V the statement \$WEIGHTS N should be replaced by \$WEIGHTS N/V, when fitting the severity model as in § 3.1.4.

There is a general technique for fitting generalised linear models whilst allowing for the possibility that the scale parameter may depend on the explanatory variables. Details are given in Chapter 10 of McCullagh & Nelder (1989). This general technique can always be applied in fitting claim severity models; the method given in Appendix F is a relatively simple alternative applicable in cases where claim sizes are approximately log-Normally distributed.

3.3 Numerical Example

3.3.1 The data are the own-damage claim amounts corresponding to the claim numbers used in Section 2.4. Out of the 1,890 cells with non-zero exposure, 736 yielded no claims over the three-year period 1986 to 1988. There are, therefore, 1,154 data points for the claim severity analysis. Each data point consists of six items: the level of each of the four rating factors, the number n of own-damage claims, and the total amount Y of these claims. Estimates of amounts outstanding were included in the individual claim amounts. These, along with all part payments, were totalled to give a single figure for each claim. For each claim originating in 1986 or 1987, this figure was inflated using factors 1.10 and 1.21 respectively (to put the amounts in the same terms as the 1988 claims), before calculating the total Y for each cell.

3.3.2 Multiplicative models with a Gamma error structure have been fitted to the observed mean claim sizes S = Y/n using the GLIM commands given in §3.1.4. As for claim frequency, none of the two-factor interactions are significant. The main effects model gives a minimised deviance of 1314.1 on 1128 degrees of freedom. The scale parameter (estimated by GLIM as the deviance per degree of freedom) is therefore 1.165. Taking the square-root gives an estimate for the coefficient of variation of individual claim amounts within each cell of 1.08. This is reasonably close to the direct estimate of 1.19 obtained using the method described in Appendix E. (The discrepancy is much smaller than that noted by Coutts, discussed in §3.2.1.)

3.3.3 Table 3.3.3 gives the parameter estimates for the main effects model (FIT DI + PA + CA + CG).

These results suggest that there is no significant difference in own-damage claim severity between levels 2 to 5 of DI, or between levels 1 to 3 of CG. This could be tested by defining a new DI factor with only 5 levels (level 2 corresponding to districts B to E), and a new CG factor with 6 levels (level 1 being for car groups 1 to 3 combined) and refitting the model in the same way as was done for claim frequency (Section 2.4).

Factor	Estimate	Standard error
1	6.33	0.1301
DI(2)	-0.0315	0.0452
DI(3)	-0.0258	0.0472
DI(4)	-0.0305	0.0465
DI(5)	-0.0469	0.2445
DI(6)	0 0453	0.0770
DI(7)	0.1635	0.0746
DI(8)	0.6521	0.0946
PA(2)	0.3123	0.1395
PA(3)	0.2497	0.1289
PA(4)	0 1693	0.1227
PA(5)	0.1542	0.1296
PA(6)	0.1181	0.1241
PA(7)	0.0868	0.1240
PA(8)	-0.0015	0.1228
CA(2)	-0.0392	0.0357
CA(3)	-0.0617	0.0467
CA(4)	-0.1890	0.0643
CA(5)	- 0 ·2928	0.0796
CG(2)	-0.0095	0.0616
CG(3)	-0.0041	0.0281
CG(4)	0.0873	0.0600
CG(5)	0.2262	0.0608
CG(6)	0.3543	0.0644
CG(7)	0.5679	0.0658
CG(8)	0.7812	0.0867

Table 3.3.3. Parameter estimates

Own-damage severity appears to increase monotonically with car group (from group 3 onwards) and decreases steadily with car age. The estimated severity is higher for level 2 than for level 1 of PA by a factor of exp(0.3123) = 1.37, but for higher values of PA the mean severity decreases monotonically. These patterns could be investigated further using the methods of Section 6. The most striking feature of the DI relativities is that the severity is much higher in Northern Ireland (level 8) than elsewhere.

3.3.4 Figure 3.3a shows the standardised residuals plotted against the estimated severity for each cell. This shows no clear evidence of heteroscedasticity: the coefficient of variation does not appear to depend on the mean claim size. This confirms that the Gamma variance function $V(\mu) \alpha \mu^2$ is correct for this data set. However, a plot of the same residuals against car age (Figure 3.3b) suggests that the variance decreases with car age. This implies that the coefficient of variation of individual claims is not, in fact, constant over all cells, but decreases with car age. A detailed analysis confirms this. Full details, including the effect on the parameter estimates of allowing for this variation, are given in Appendix F.



Figure 3.3b.

4. MODELLING THE DEPENDENCE ON TIME

4.1 Theory

4.1.1 In selecting the data to be used for a premium rating analysis, there are two conflicting criteria: the volume of data should be large to reduce random variation in the estimates, but the data should all relate to exposure periods that are sufficiently recent that future claims experience can be expected to be similar. Usually a reasonable compromise is achieved by using data covering the most recent three- or four-year period. It is often argued that very recent claim severity data should not be used in premium calculations, because many claims will not be finalised. However, this is not a problem when the methods proposed in this paper are used, provided the severity data comprise both amounts paid and case estimates of amounts outstanding. The theory of Section 3 is hardly affected if Z_{ijk} denotes the incurred amount (i.e. paid plus estimated outstanding), rather than the fully developed paid amount, for an individual claim.

4.1.2 It is desirable to allow for time as an explanatory variable when modelling the data for two reasons:

- (1) As premiums should be based on a forecast of future levels of claims, it is desirable to identify and quantify any current trends. A trend will usually exist in the size of claims because of inflation, but the magnitude of this cannot be assumed the same as for inflation in the RPI, earnings, or claims in some other class of insurance. If the latest claim-size data contain substantial estimated components on claims outstanding, then the trend may not give a good indication of true claims inflation: it will indicate the combined effect of inflation and any bias in case estimates (which will affect the most recent data most). There may also be a trend or seasonal variation in claim frequency.
- (2) If a time effect has been accompanied by a change in the mix of business, the rating-cell relativities could be distorted. For example, suppose we have data for three consecutive past years, the last of which had many more claims than the other two owing to bad weather. Consider two rating-cells C1 and C2, and suppose that the claim frequency has remained constantly twice as high for C2 as for C1 (it is this relativity which we aim to estimate). Suppose that over the three years concerned, exposure (number of policies) increased in cell C2 but decreased in cell C1. Thus, perhaps 40% of the total exposure in C2 relates to the most recent year, whereas the figure for C1 may be only 30%. If the dependence of the claim frequency on time is not taken into account, the relativity of cell C2 to cell C1 will appear to be greater than the true value of two.

The basic models proposed in Sections 2 and 3 can easily be enhanced to allow for dependence on time. This eliminates the problems mentioned above. In particular, the level of claims inflation can be estimated directly from the claimsize data at the same time as estimating the relativities: there is no need to attempt to remove inflation from the claim-size data by making prior adjustments.

4.1.3 Suppose we have data for three years, indexed by t = 1, 2, 3. Each claim is assigned to the year in which the accident occurred. Claim frequency will be considered first. The basic model (Section 2) is:

$$\mathbf{E}(r_{ij}) = f_{ij}$$
 and $\operatorname{Var}(r_{ij}) = \phi \cdot f_{ij} / x_{ij}$.

The main effects $f_{ij} = \exp(\alpha_i + \beta_j)$ can be fitted in GLIM using: \$YVAR R \$ERROR P \$LINK L \$WEIGHTS X \$FIT A + B. The possibility that claim frequency may depend on time in a multiplicative manner across all cells (a reasonable assumption for the effects of weather, petrol prices, road conditions, economic activity, etc.) can be expressed as $f_{ijl} = \exp(\alpha_i + \beta_j + \gamma_l)$. To fit this model we need data r_{ijl} and x_{ijl} for each (i, j, l) combination: it is assumed that these are read into vectors R and X as before. We also need an additional vector T holding the corresponding values of t (T must be a 'factor' in GLIM terminology). The GLIM commands are as above, except \$FIT A + B is replaced by \$FIT A + B + T. This will give estimates of the parameters $(\alpha_i, \beta_j, \gamma_l)$. Of these, the α_i and β_j give the required rating-cell relativities.

The γ_i describe the relative experience of past years. Usually there is no clear trend in claim frequency and these parameter estimates are of little value in themselves: the underwriter's judgement on the overall level of claim incidence in future years cannot be dispensed with. The primary reason for including these time parameters in the frequency model is to ensure that the rating factor estimates α_i and β_j are not distorted by any change in the mix of business as described in §4.1.2 (see also §7.1.2).

4.1.4 Now, consider claim severity. The basic model (Section 3) is:

$$E(S_{ij}) = m_{ij}$$
 and $Var(S_{ij}) = \sigma^2 \cdot m_{ij}^2 / n_{ij}$.

The main effects $m_{ii} = \exp(\alpha_i + \beta_i)$ can be fitted in GLIM using: **\$YVAR S SERROR G SLINK L SWEIGHTS N SFIT A + B.** A time factor can be introduced exactly as described above for claim frequency. For each claim, the total of paid amounts plus estimates of amounts outstanding is assigned to the year of accident. However, in the case of severity we may expect to find a trend (corresponding to inflation) in the estimated parameters γ_1 , γ_2 , γ_3 . For example, the values 0.07, 0.15, 0.21, would suggest that the force of claims inflation had been about 0.07 p.a. This hypothesis can be tested and, if true, a more accurate estimate of the constant force of inflation obtained by redeclaring T to be a variate (rather than a factor). This has the effect of fitting the model $m_{ii} = \exp (m_{ij})$ $(\alpha_i + \beta_i + t \cdot \gamma)$ instead of $m_{ij} = \exp(\alpha_i + \beta_i + \gamma_i)$. That is, instead of estimating three separate parameters γ_i we estimate a single parameter γ_i , representing the average force of inflation per annum. The latter model is a restricted case of the former, so can be tested using an F-test in the usual way. Whichever model is accepted, the estimates of α_i and β_i give the required rating cell relativities free from any distortion which might have been caused by a change in the mix of business with time.

4.1.5 In the above, it was assumed that time was indexed by the calendar year

of accidents. The same modelling techniques could, of course, be used with a finer time scale, using quarters perhaps. However, any changes in the mix of business over periods of less than a year will usually be so slight that they can reasonably be ignored. Also, we have found that inflation can be reliably estimated using claim severity data spanning three consecutive years indexed by the year only. Every further subdivision of the time scale increases the volume of data substantially, and computer-memory limitations may be exceeded. For these reasons, we recommend the use of a yearly scale.

4.1.6 If very recent data are included in the analysis, the coefficient of variation σ for the claim severities may vary with the time index: later claim amounts may be subject to greater random variation, because there are larger estimated components on average than the more fully developed claim amounts for earlier exposure intervals. This should be monitored by plotting the standardised residuals against the time index. If this plot suggests that the variance increases with time, this can be formally tested and allowed for using the methods of Appendix F.

5. SELECTION OF CLAIM SETS

5.1 Practical Aspects

5.1.1 A motor claim can give rise to multiple claim types such as fire, theft, vehicle own damage, third party vehicle damage and third party bodily injury. Fundamental to the successful application of our methodology is the separate treatment of each identifiable claim type. The theory discussed in Sections 2 and 3 applies equally to each claim type. Not only are there strong statistical reasons for following this approach, but there are strong practical reasons.

5.1.2 The parameters from the models applied to each claim type provide valuable help in the interpretation of the patterns and trends within the data. This gives the underwriter a much deeper insight into the factors driving the claims experience. In particular, the time parameter, as discussed in Section 4, will provide useful information on the claims inflation for each claim type and any trends in claim frequency for each claim type. Recently the industry has seen a rapid rise in theft frequency. The modelling approach will accurately reflect this trend, independent of any changes in mix of business during the investigation period.

5.1.3 The handling of large claims is simplified since these are mainly for bodily injury. We have found that the average cost of bodily injury claims does not usually vary significantly with the various underwriting factors. Indeed, we have found generally that it is only policyholder age and vehicle group which have an important influence. In practice, however, it is usually sensible to cap bodily injury claims at an appropriate level. The capping level will depend on the size of portfolio, although for a medium sized portfolio in the U.K. (excluding Northern Ireland), a capping level of about £25,000 in 1990 money is likely to be reasonable. The cost of the excess claims will be important only in the assessment

of the overall level of premiums, and will not affect the relative premiums between cells.

5.1.4 Changing 'knock-for-knock' agreements can significantly distort the relative claim costs between different cells. However, the distortion will be limited to vehicle damage costs which are relatively short tailed. Hence the statistical analysis can be linked to the period since the last major change in agreements.

5.1.5 Finally, a thorough understanding of the factors which are driving the claims experience of each claim type is invaluable to the actuary/statistician when carrying out a statistical review of the company's claim reserves.

5.2 Statistical Aspects

5.2.1 Consider the following categories of claims, all of which are usually covered by a comprehensive policy:

- (i) fire and theft,
- (ii) damage to the windscreen of the policyholder's vehicle not caused by a third party,
- (iii) other damage to the policyholder's own vehicle arising from incidents not involving a third party,
- (iv) damage to property of both the policyholder and the third party, arising from incidents involving a third party, and
- (v) bodily injury.

5.2.2 In general, the relativities may differ for each of these types of claim. For example, claims of types (ii) and (iii) may be more frequent in rural districts than urban districts (because speeds are often greater), whereas claims of type (iv) may be more frequent in urban areas than rural areas (because traffic is denser). If the average claim size were the same for all these types of claim (but varying across rating cells), this would not be important: risk premium relativities could be obtained by multiplying the average claim size for each cell by the total frequency for all types of claim combined, with no regard to how this total frequency breaks down.

5.2.3 However, the average claim size generally differs for each of these types of incident. Furthermore, the between-cell relativities for claim size may also differ. For example, the severity of bodily injury claims does not depend greatly on car age, whereas the severity of material damage claims does. Therefore, to obtain reliable risk premium relativities, it is advisable to apply the basic models for claim frequency and claim severity to each identifiable class of claims separately. The claim frequency model should also be applied to nil claims (i.e. claims for which the severity is zero), because these involve an expense to the insurer.

5.2.4 If there is any doubt a priori about whether either the frequency or severity for two types of claim is likely to differ, it is advisable to assume initially that they may differ, and to obtain separate data sets for the two claim types. It is then possible to test formally the hypotheses that the relativities and the overall

level are the same for both types (this can be done for both frequency and severity separately).

5.3 Implementation in GLIM

5.3.1 Subscript t will be used to denote the claim type, t=1, 2, etc. For simplicity of presentation, it is assumed that the data cover a single time period or (equivalently) that the data for all time periods have been aggregated. In principle, there is no reason why the methods described here should not be applied to data indexed also by time, as described in Section 4, but in practice, computing limitations are likely to militate against this approach.

5.3.2 Initially the frequency and severity models are fitted separately for each type of claim. For frequency the following GLIM commands are used:

\$YVAR R \$ERROR P \$LINK L \$WEIGHTS X \$FIT A+B

and for severity:

\$YVAR S \$ERROR G \$LINK L \$WEIGHTS N \$FIT A+B.

5.3.3 Suppose that, for two claim types, the frequency relativities for one of the rating factors, factor A say, appear to be similar. That is the estimates α_{i1} are close to the estimates α_{i2} (the second subscript here is for claim type t = 1, 2). We wish to test the hypothesis that these two sets of relativities are in fact identical, and if so, to obtain a single set of estimates from both data sets. The first step is to concatenate the data vectors for the two claim types to give a single large data set, R, X, A and B. An extra data-vector T must then be created holding values 1 and 2 to indicate the claim type of each entry in R, X, A and B (and T must be declared to be a factor in GLIM). The separate frequency models for each claim type can then be fitted simultaneously for both claim types using:

\$YVAR R \$ERROR P \$LINK L \$WEIGHTS X \$FIT T*(A+B).

The command \$FIT T*(A+B) allows interactions between the claim type T and the factors A and B: this is essentially the same as allowing the A and B parameters complete freedom to differ for each level of T. This command is equivalent to \$FIT T+A+B+T·A+T·B. The hypothesis that the factor A relativities are the same for both claim types is represented by \$FIT T+A+B+T·B (the interaction T·A has been removed). The hypothesis can therefore be tested using an F-test in the usual way, after fitting both these models. If the hypothesis is accepted, the parameter estimates given by \$FIT T+A+B+T·B will include the best estimates for α_i obtained from both data sets simultaneously. One can then go on to test whether the factor B relativities are also the same for both claim types by using \$FIT T+A+B.

5.3.4 In § 5.3.3 we considered testing the equality of frequency relativities for two claim types. The same techniques can be used for testing equalities of severity relativities, and also for comparing more than two claim types simultaneously, computer memory space permitting. Particularly when looking at severity, the

hypothesis that the overall level of the dependent variable (claim size) is the same for two claim types may also be of interest. This can be tested by using the usual F-test to compare FIT A + B with FIT T + A + B.

5.3.5 It might be asked why we are interested in testing the equality of parameters for different claim types. Why not be satisfied with separate models for each type of claim, as this would give reasonable estimates whether or not two of the claim types are similar in some way? The reason is the principle of parsimony: unnecessary parameters should not be included in a model. The more parameters in a model, the less reliably each of them is estimated. For each genuinely different set of relativities we should seek estimates based on the largest possible amount of data, by combining data for different claim types if possible.

6. CURVE FITTING

6.1 Practical Aspects

6.1.1 To be useful to an underwriter the statistical analysis should aim to provide premium relativities in as much detail as possible.

An underwriter is generally not interested to know the relative premiums between a group of policyholders aged 17–20 compared to a group aged 21–24. The underwriter requires the premium relativities for individual ages, at least over the age ranges where the risk is likely to vary most. Similar considerations will apply to vehicle age, as indeed they will to all the underwriting factors.

6.1.2 It may be thought, in the first instance, that this not unreasonable request will produce immense statistical difficulties, owing to the limited exposures which arise in many of the cells.

We define below the statistical problem and discuss how this apparent conflict can be resolved routinely within the analysis.

6.2 Statistical Aspects

6.2.1 In previous sections, the rating factors have been treated, in GLIM terms, as factors rather than variates. This means that separate parameters are estimated for each of a number of levels of each rating factor. This approach has also been adopted in all previous papers on quantitative motor rating methods known to us. For example, Bennett (1978) treats policyholder age as having 4 levels: 17–22, 23–26, 27–65 and 66–80, and estimates a separate frequency level for each of these. The other rating factors used by Bennett: car age and level of NCD, have 4 and 3 levels respectively, giving a total of 48 rating cells.

6.2.2 There are difficulties with this approach, as can be appreciated by considering the following points:

- (i) Each level of each factor should cover a range sufficiently narrow that we have approximate homogeneity of risk among the units of each cell.
- (ii) The ranges for each level of each factor should be narrow also, because information is lost by aggregation of data.





Figure 6.1. Claim frequency results from Baxter et al. (y-scale is log of multiplicative relativity).

- (iii) The number of exposure units and the number of claims in the data at each level of each rating factor should be sufficiently large that reasonably reliable estimation of the parameters is possible.
- (iv) The total number of rating cells (hence the volume of data) must not exceed computer memory limitations.

Clearly, objectives (i) and (ii) are in conflict with objectives (iii) and (iv) and, unless we have a vast volume of data and very powerful computing facilities, some sort of compromise is necessary if we use the conventional factor approach.

6.2.3 However there is an alternative, and to illustrate this we consider the analysis of Baxter *et al.* (1980). Among other models for claim frequency Baxter *et al.* used the model advocated in this paper (Section 2.1). They fitted the main effects model (i.e. no interactions) using the rating factors: policyholder age, car group and district, each with 4 levels. Although they do not give their parameter estimates of the paper, we have repeated the analysis using their data. Each of the graphs of Figure 6.1 shows a set of rating factor parameters (i.e. the α_i or β_j , etc.) plotted against the corresponding rating factor.

6.2.4 The claim frequency plots clearly suggest a functional relationship in the case of both policyholder age and car group. A similar analysis of their data for claim severity shows that a relationship is apparent for policyholder age and car

age. It should be noted that, in common with all models in which the explanatory variables are factors, no assumption was made in either case of any relationship between the parameters; they were free to take any values relative to each other, and the obvious patterns displayed are purely reflections of the data.

6.2.5 These relationships should not be surprising. The very fact that contiguous ages are grouped to form the levels of factors: policyholder age and car age, implies a prior belief of continuous relationships with risk. The age bands provide a discrete approximation to what is really a continuous relationship. The situation is much the same for district and car group; in reality there is almost a continuous spectrum of neighbourhoods/car models, from those with the lowest risk to those with the highest risk. This spectrum is approximated by placing each neighbourhood/car model into one of a small number of categories.

6.2.6 Given these reasons for expecting continuous relationships between the rating factors and risk, it seems reasonable to model the relativities using continuous functions. This approach has substantial advantages:

point (i) of §6.2.2 no longer applies: we propose fitting a continuous curve rather than approximating a continuous curve with a step function,

point (iii) no longer applies: we no longer wish to estimate a separate parameter for each level of each rating factor, so there is no problem in having a small volume of data for some levels, and

we may be able to fit a more parsimonious model: if the number of parameters needed to specify a continuous function is less than the number using a conventional model, then the parameters will be more reliably estimated.

Points (ii) and (iv) of § 6.2.2 continue to apply, so there is still a conflict, but this is relatively easy to resolve: we should simply choose the number of rating cells to make full use of the available memory space.

Section 6.3 describes how the fitting of continuous relationships between risk and rating-factors can be implemented in GLIM, and how to test formally whether such models do indeed fit the data better than the conventional type of model.

6.3 Implementation in GLIM

6.3.1 To simplify the presentation, attention is restricted to the case of a single claim type, with only two rating factors, and no time index. The techniques can easily be used more generally.

As usual, the levels of the two rating factors are assumed to be held in GLIM factors A and B. If there is no significant interaction between the factors (as is usually the case) the basic frequency model is:

 $E(r_{ij}) = f_{ij}$ and $Var(r_{ij}) = \phi \cdot f_{ij}/x_{ij}$

where $f_{ij} = \exp(\alpha_i + \beta_j)$. This is fitted in GLIM using the commands: \$YVAR R \$ERROR P \$LINK L \$WEIGHTS X \$FIT A+B.

6.3.2 Suppose factor A is policyholder age, and that the parameter estimates for the different levels of A when plotted against policyholder age give a plot similar to that in Figure 6.1. In such a case we would wish to try a straight line relationship:

$$\alpha_i = \mu + \alpha \cdot a$$

where a is the mean policyholder age (in years) corresponding to the *i*th level of factor A (for example, for i = 1 the age range may be 17–20, in which case the first value of a would be about 19), and μ and α are the unknown parameters to be estimated from the data: α is the slope of the straight line.

6.3.3 To fit this model in GLIM, the *factor* A must be replaced by a *variate* VA say, holding the values for a (e.g. 19, 23.2, 32.7, 58.8....), instead of the factor levels 1, 2, 3, 4.... This is achieved using the following sequence of commands:

ASS XA = 19, 23.2, 32.7, 58.8CAL VA = XA(A) (see GLIM manual for explanation)

The straight line model for factor A can then be fitted using \$FIT VA + B (the other commands remaining as before). Since the straight line model is a restricted case of the original model (the α_i parameters are restricted to lie on a straight line) the quality of fit can be tested using the usual F-test: if the F-statistic is not too large the straight line model should be accepted. Note that, since the linear predictor is the log of the mean claim frequency (because we are using a multiplicative model), the straight line model corresponds to an exponential model for the claim frequencies. For example, if GLIM gives an estimate of -0.02 for the slope parameter α of the straight line, this implies that the mean claim frequency decreases by a factor of just over 0.98 for each 1 year increase in policyholder age.

6.3.4 Although a straight line model fits well in some cases, much more flexibility is introduced by using a quadratic relationship. This allows for a fitted curve which is either strictly increasing or decreasing (with varying slope), or which has a minimum or a maximum at some value of the rating factor. To fit a quadratic in GLIM, it is necessary first to calculate a new variate holding the squares of the rating factor values a^2 :

CAL VA2 = VA*VA

The quadratic model can then be fitted using FIT VA + VA2 + B (the other commands remaining as before). Since the straight line model is a restricted version of the quadratic model, the quality of fit of the straight line model can be compared to that of the quadratic model using the usual F-test: if the F-statistic is not too large, then the quadratic term VA2 is not needed. Thus, the best strategy is first to fit the model with A as a factor, then fit the quadratic model, and if this is acceptable, try the straight line model. In some cases, a cubic and perhaps higher order terms may be needed.

6.3.5 In general, the effect of any number of factors may be modelled in this

way, so we may have models such as FIT VA + VA2 + VB + VB2 + VC for example. Interactions between rating factors can be included in such models, if necessary, by defining variates such as VAB = VA*VB for inclusion in the linear predictor. In a case where one of the rating factors cannot be modelled using a curve, so we have a factor such as D in the linear predictor, an interaction between this and another rating factor which is modelled using a curve is achieved by including a term such as $VA \cdot D$ in the linear predictor. This allows the straight line parameter of rating-factor A to take a different value for each level of rating factor D.

6.3.6 If the ranges of values for the original rating factors are broad (e.g. 17-24, 25-44, 45+ for policyholder age) then the mean values for use in the corresponding variate may be very approximate (values of about 21, 35, 58 may be appropriate). However, if continuous relationships are to be used, point (iii) of § 6.1.2 does not apply, so, subject to computer memory limitations, the width of the ranges can be reduced. As well as using more of the available information, the use of a larger number of narrower ranges allows more accurate estimation of the mean value for each range (needed to create the variates such as VA).

6.3.7 Although we have considered only the claim frequency model in this section, the same modelling techniques can be used just as easily for claim severity: we simply have to replace R by S, X by N, and SERROR P by SERROR G.

7. CALCULATION OF PREMIUMS

7.1 Calculation of Risk and Office Premiums

7.1.1 After fitting a model as described in Section 6 to each of frequency and severity for each claim type as described in Section 5, it is a simple matter to calculate the risk premium for any combination of rating factors. The risk premium for rating cell (i, j) is given simply by:

$$P_{ij} = \sum_{t} f_{ijt} \cdot m_{ijt}$$

where f_{ijt} is the fitted value obtained from the claim frequency model for claim type *t*, and m_{ijt} is the fitted value obtained from the claim severity model for claim type *t*.

For each rating cell (i, j), the estimates f_{ijt} and m_{ijt} are all asymptotically unbiased, and mutually independent (this is shown in Appendix G). Since each of the component models invariably has many degrees of freedom per estimated parameter, asymptotic results can be expected to give good approximations. Therefore, the estimate P_{ij} obtained in this way is unbiased, that is, it is the expected total loss per policy in cell (i, j) (or 'risk premium' in other words).

7.1.2 In using P_{ij} to find the relativities for future premiums, the following assumptions are made:

the relativities of both frequency and severity will be similar in the near future to their values in the recent past, for each claim type, and the relative volumes of the different claim types will be similar in the future and the past.

The plausibility of this second assumption can be checked by examining trends in the time parameter of the frequency models (see §4.1.3). If trends seem to differ between claim types, a suitable adjustment could be made in calculating P_{ij} from the component models.

7.1.3 The risk premiums P_{ij} must next be converted to office premiums by including expenses. As described by Coutts (1984, Section 4), some expenses relate to each claim made, while others relate to each policy written. The latter type of expense can be either additive (new business, lapse, renewal and endorsement expenses) or multiplicative (commission). If ex_1 denotes the total per claim expense, ex_2 denotes the total additive per policy expense, and w denotes the commission, then the net office premium is given by:

$$O_{ij} = (P_{ij} + ex_2 + ex_1 \cdot \sum_{i} f_{iji})/(1 - w).$$

Values for ex_1 , ex_2 and w have to be obtained from supplementary data: as indicated by the absence of subscripts, they do not usually depend on the rating cell.

Other formulae for O_{ij} may sometimes be appropriate. For example, expenses might depend on claim size also.

7.1.4 In addition to giving estimates of the parameters for each component model, GLIM also gives standard errors for the estimates. Just as the fitted values f_{iji} and m_{iji} can be calculated for any rating cell using the parameter estimates, the standard errors of these fitted values can be calculated using the standard errors supplied by GLIM. These, in turn, can be used to calculate the standard errors of P_{ij} and O_{ij} : it can be very useful in interpreting these values to know how reliably they have been estimated. The standard error could also be used as the basis of safety loadings: the pure risk premium could be augmented by some multiple of its standard error. The necessary calculations are outlined in Appendix G.

7.1.5 The remaining steps of the analysis depend on which of the two options for the treatment of NCD is being used (see 1.9.2).

If the percentage discounts are being reviewed (option (i) of § 1.9.2), then NCD level *I* will be an explanatory variable for each of the estimates f_{ijlt} , m_{ijlt} , so estimates P_{ijl} and O_{ijl} can be calculated for each NCD level *I* within each combination of the other rating factors *i*, *j*. It is highly unlikely that the dependence of O_{ijl} on NCD level will be purely multiplicative, because the dependence on NCD level will be different for each component f_{ijl} and m_{ijl} . Therefore, in order to obtain new percentage discounts for the NCD scale, the best estimates of the office premiums O_{ijl} obtained above must be approximated in such a way that the dependence on NCD level is multiplicative, and has no interactions with the other rating factors. This can be achieved using a final stage of modelling as described in Section 8.

If the percentage discounts are to remain unchanged (option (ii) of §1.9.2), then the rating cells used in each component model should not be broken down by NCD level. (It is shown in Appendix J that this would not give the correct relativities unless the existing NCD percentages are exactly right, or the distribution of policies over NCD levels is the same for all rating factor combinations.) In this case, each estimate P_{ij} (or O_{ij}) is the average over all NCD levels of the expected loss per policy in rating cell (*i, j*). In other words, they are average 'net premiums', that is, average premiums after deduction of NCDs. To convert these to gross premiums (to which the existing percentage NCD can be applied for each individual policy), it is necessary to estimate the average NCD in each rating cell. This is described in Section 7.2.

7.2 Calculation of Gross Office Premium from Average Net Office Premium

7.2.1 By 'gross office premium' we mean the rate book premium applicable to all policyholders with identical rating factor values before deduction of individual NCDs: this is denoted G_{ij} for rating cell (i,j). The remainder of this section describes how to convert the net office premium O_{ij} (given in §7.1.2) to G_{ij} , in cases where NCD level is not an explanatory variable in the component models, because the existing NCD percentages are to remain unchanged.

7.2.2 As in earlier sections, x_{ij} denotes the number of policy years in rating cell (i, j). Thus, the total premium necessary to cover pure risk and expenses for the cell is $x_{ij} \cdot O_{ij}$, where O_{ij} is the average net office premium calculated as described in § 7.1.2.

If C_{ijk} denotes the NCD for the kth policyholder in cell (i, j) and x_{ijk} denotes the exposure for this policyholder, then the net premium for this individual is $(1 - C_{ijk}) \cdot x_{ijk} \cdot G_{ij}$. (C_{ijk} may be 20% or 30% for example, $x_{ijk} = 1$ in most cases, but is often less than 1.)

 G_{ij} is given by equating premium income to expected loss (pure risk plus expenses) for each cell:

$$\sum_{k} (1 - C_{ijk}) x_{ijk} G_{ij} = x_{lj} O_{ij}$$

since

$$\sum_{k} x_{ijk} = x_{ij}$$

G

This gives

$$y = \frac{O_{ij}}{1 - C_{ij}}$$

 $C_{ij} = \left(\sum_{k} x_{ijk} C_{ijk}\right) / \sum_{k} x_{ijk}$

where

= exposure weighted mean NCD, in cell i, j.

However, if cells are small (in terms of total exposure) the observed means C_{ij} will be very variable, both between cells and over time, so for the relativities of G_{ij}
to be suitable for the future, it is better to use expected values of the C_{ij} rather than the observed exposure-weighted means, in converting O_{ij} to G_{ij} . Appendix H describes a method for estimating these expected values from the observed value C_{ijk} .

7.2.3 Coutts (1984), does not include NCD level as an explanatory variable in frequency and severity models, and does not attempt any conversion of net office premium to gross office premium: he suggests basing the rating structure on relativities of the quantities we have denoted O_{ij} . Thus, he makes an implicit assumption that the mean NCD C_{ij} does not vary between rating cells. We have found in practice, using the model of Appendix H, that there is significant variation in the C_{ij} between cells, and that this variation is quite substantial, so it should not be ignored.

7.3 Numerical Example: Standard Errors of Risk Premiums

7.3.1 To illustrate orders of magnitude of the standard errors in risk premiums and office premiums, the calculations of Appendix G have been carried out using the 17,000 comprehensive policies over 1986–1988 used in previous examples. Some of the results are given in Table 7.3.1. For the previous examples, only owndamage claims data were used: the figures given here incorporate the results of

				Risk		Net office
DI	PA	CA	CG	premium	SE	premium
1	18.0	1.5	1	351-39	30.39	487.73
1	18.0	1.5	2	360.22	30.21	501.45
1	18.0	1.5	3	382.69	31.41	532.75
1	18.0	1.5	4	402.97	33-31	557.10
1	18.0	1.5	5	452.46	38.12	618-85
1	18·0	1.5	6	514-90	44.32	695-99
1	18.0	1.5	7	638.66	59.05	844.58
1	18.0	1.5	8	959-59	100.62	1242.72
1	27.5	1.5	1	206-66	9.73	299-36
1	27.5	1.5	2	212.05	9.11	308.02
1	27.5	1.5	3	225.40	8.83	327.23
1	27.5	1.5	4	237.69	9.12	341.98
1	27.5	1.5	5	270.38	9.99	382.75
1	27.5	1.5	6	309.50	11.26	431.61
1	27.5	1.5	7	382.97	17.62	519·87
1	27.5	1.5	8	594.47	38.51	783-26
1	27.5	5.0	1	180.07	8.55	264.49
1	27.5	5.0	2	185-29	8.06	272.73
1	27.5	5.0	3	197-31	7.86	289.99
1	27.5	5.0	4	207.54	8.09	302-29
1	27.5	5.0	5	234.62	8.76	336-12
1	27.5	5.0	6	267.00	9.78	376.68
1	27.5	5.0	7	329.59	15.44	451-89
1	27.5	5.0	8	508·92	33.69	675-42

Table 7.3.1. Estimated theoretical premiums

modelling all claim types covered by the comprehensive policy. The units of the risk premium, its standard error, and the net office premium are 1988 pounds.

The standard error is quite high for the combination PA = 18 and CA = 1.5, particularly for the higher car groups. This is the combined effect of low exposure in this region (a small volume of data) and high variability of the claim amounts for some types of claim (see Appendix F).

8. SMOOTHING OF PREMIUMS

8.1 Theory

8.1.1 It is possible, using the techniques described in this paper, to calculate premiums for any combination of rating factors, with each rating factor being measured on a continuous scale. We could, for example, calculate the gross office premium for a driver aged exactly 34 in a two-and-a-half-year-old car in group 3, and also the standard error of this premium (Appendix G). In principle, there is no need to proceed any further. When a judgement has been made on the overall levels of claim frequency and severity likely for the coming year, values of the rating factors could be typed into a computer program holding the estimated relativity parameters for all the component models, to calculate the apprepriate premium. However, there are four reasons why a further stage of modelling (to approximate the 'best estimates' obtained in Section 7 by some 'smoothed' structure) may be necessary in practice:

If the NCD percentages are to be reviewed (case (i) of § 1.9.2), then NCD level will have been included as an explanatory variable in each of the component models. The resulting office premiums O_{ijl} (obtained from the component models as described in Section 7.1) will depend on NCD level *l* in a complex way. The dependence is very unlikely to be purely multiplicative: that is, there will be interactions with other rating factors. To obtain new percentage discounts, a purely multiplicative dependence on NCD level is required.

A simplified rating structure may be necessary in order to simplify the rate book for brokers. This requirement is likely to diminish with the increasing use of quotation systems and the trend towards direct selling.

To facilitate comparison of the new estimated relativities with the existing relativities of the current rate book, the new estimates can be forced into the same overall structure, which may be, for example, a purely multiplicative structure (i.e. multiplicative with no interactions).

To build a standard table which, under certain circumstances, can be used to further refine the relativities produced by the modelling process. This is discussed further in Section 9.

Coutts (1984, Section 4) also advocated a final stage of modelling in order to approximate the 'best estimates' using a simple structure.

8.1.2 The 'data' are either estimated office premiums O_{ijl} for each NCD level l

(see §7.1.3) or estimated gross office premiums G_{ij} obtained as described in Section 7.2. Y_{ijl} will be used for both possibilities. If the rating factors are treated as continuous variates as proposed in Section 6, we can generate as many 'data points' as we wish. Intuitively, it seems that a good approximation will be obtained by having the density of these data points proportional to actual exposure: in this way, the discrepancy between the smoothed surface and the best estimates Y_{ijl} will be smallest where it matters most, i.e. where exposure is high. Equivalently, and more simply, one 'data point' can be used to represent each of the existing rating cells, and these weighted by exposure in fitting the simplified structure. It is not correct to use the inverse of the squared standard errors (Appendix G) for weights, because the Y_{ijl} are not mutually independent.

8.1.3 Each data point for this final stage of modelling consists of:

the dependent variable Y_{ijl} , and

independent variables, being the values of the rating factors used to calculate the Y_{ijl} .

If the independent variables are denoted A, B, etc., as in previous sections, the smoothing can be carried out in GLIM using:

\$YVAR Y \$ERROR G \$LINK L \$WEIGHTS X \$FIT A+B.

The Gamma error structure is appropriate if we wish the percentage discrepancy between best estimates Y_{ijl} and the fitted surface to depend only on exposure (as specified by the \$WEIGHTS command) and not on the magnitude of Y_{ijl} (all else being equal). The use of a log-link function gives a multiplicative approximating structure. An identity link would given an additive structure: this is unlikely to give such a good approximation to the best estimates Y_{ijl} because the component models are multiplicative.

8.1.4 Conventional F-tests are not valid for determining which terms to include in the linear predictor: the 'data' are actually values obtained from the fitted values of other models so contain no mutually independent random components. The terms to be included, along with the link function, should depend mainly on the purpose of the model: if we wish to compare estimated values to rate book premiums which have an additive structure and no interactions, than we should use \$LINK I and \$FIT A+B+C+D. For the purpose of obtaining a multiplicative approximation to the dependence on NCD level, we should use \$LINK L \$FIT NCD+XXX where XXX represents any combination of terms involving the rating-factors other than NCD.

9. BUILDING AN EFFECTIVE STANDARD TABLE

9.1 Standardisation Techniques

9.1.1 We briefly discussed in Section 1.6 how 'one-way' tables are generally used to summarise the relative claims experience of the levels within each underwriting factor. We commented that these tables can be misleading, since the

distribution of business is not identical within each level of each underwriting factor.

9.1.2 A number of standardisation techniques have been developed in the past to cater for this specific problem. Most of these techniques are based around a standard table, which sets out the perceived relativities between the various underwriting factors and levels within underwriting factors. If there is no prior knowledge of an appropriate initial standard table, the starting point is often taken to be the relativities derived from the current underwriting guide or that of a competitor.

9.1.3 The standard table is updated to take into account the emerging claims experience. The updated standard table can then be used to obtain the relative claims experience of the levels within each underwriting factor, independent of the distribution of business within each level.

9.1.4 The structure of the standard table will determine the approach which can be taken in order to update the table.

In principle, the standard table could be multiplicative, additive or perhaps mixed, with or without interactions. Assuming that the structure is multiplicative, the risk premiums in each cell can be multiplied by appropriate scaling factors (or relativities) which represent the relative difference in risk premium to a defined standard risk. The scaling factors are derived from the standard table. Standardised one-way tables can then be constructed for each underwriting factor, which will then provide an updated set of relativities. This procedure can be carried out iteratively until the relativities converge. The updated relativities can then be used to define a revised standard table. If the standard table allows for per policy expenses, then the above approach would be applied to the office premiums for each cell.

9.1.5 One of the problems with the standardisation approach is that a structure is being imposed on the data. The statistical modelling approach, as described in this paper, attempts to extract the structure from the data. This is important, since the failure to identify an important interaction effect can result in an underwriting opportunity missed. Further, if the underlying structure of the standard table is wrong, for example an important interaction effect is omitted, then the relativities derived are likely to be unstable and they may not converge. It may also be difficult to identify and understand the cause of sudden changes in claims experience. A good example is the recent rapid increase in the frequency of theft claims, which will affect some parts of the portfolio to a greater extent than others. Furthermore, the standard table approach will not provide any measure of the reliability of the estimates produced or quantify the relative importance of each rating factor.

9.1.6 However, the standard table approach does have some useful merits, since the relativities can be continually updated without the need for extensive statistical analysis. This will provide management with valuable additional information.

We have found in practice that, if the standard table approach is combined

with the results of the statistical modelling approach, then the above criticisms can be overcome. Such a standard table can then form part of an effective pricing and underwriting control system.

9.2 Developing A Standard Table

9.2.1 The results of smoothing the estimated office premiums, as described in Section 8, can be used to build an effective standard table. The parameters derived from the modelling of the office premiums can be used as the starting relativities for an initial standard table. If important interaction effects had been identified in the data, these would be included in the standard table. The iterative procedure as described in §9.1.4, can be carried out in order to 'refine' the initial relativities produced from the modelling of the office premiums.

9.2.2 For a large account, where all the major rating factors had been included in the modelling process, it would be expected that very few, if any, iterations would be required for convergence. This would be particularly true if the procedures of Section 6 had been carried out, where the initial standard table would include relativities for individual policyholder ages and vehicle ages. The standard table derived can then be used to update the relativities as the new claims experience emerges over time. This will also allow the consistency of the relativities to be monitored.

9.2.3 For a smaller account, where it has not been possible to include all the underwriting factors in the modelling, or where the data have been grouped in some way before the modelling has been carried out, a number of iterations may be required before the relativities converge. However, the initial standard table, being based on the results of the modelling, will provide reasonably accurate relativities of the factors included in the modelling. This will be particularly true if the mix of business of the factors not included in the modelling is evenly spread throughout the portfolio. It would, therefore, be expected that few iterations would be needed for the relativities to converge. The final standard table derived provides the 'refined' relativities of the underwriting factors which were not included in the modelling can be estimated.

9.2.4 Whilst the standardisation approach provides a useful means of monitoring and updating the relativities as new claims experience emerges, we do not believe the approach should be used as a substitute for the more rigorous statistical modelling approach. The statistical approach entails detailed data analysis, which provides invaluable information on the factors which are driving the claims experience. In particular, any changing trends in the claims experience which can potentially affect the rating structure can be identified as quickly as possible, to enable appropriate rating action to be taken.

9.3 Postal Code and Make/Model Analyses

9.3.1 The analyses of the underwriting factors, district and vehicle group, give rise to special problems, since the district or vehicle group classifications have

been compiled from the allocation of individual postal codes or make/models. The objective is to classify individual postal codes or make/models to a manageable number of reasonably homogeneous risk groups. An underwriter may be satisfied with the existing classifications, and will be interested only in the adjustments required to the relativities for district or vehicle group. Alternatively, the underwriter may require to appraise the reasonableness of the existing classifications, indeed he may believe that further risk groups or classifications should be created.

9.3.2 The standard table, as derived in Section 9.2, can be used for an analysis of the claims experience of individual postal code or make/models. The analysis requires particular care since there are approximately 1,300,000 separate postal codes (e.g. KT19 8HB), 2,700 postal districts (e.g. KT19), and on many companies' systems over 7,000 different make/models. In the case of make/ models however, the claims experience tends to be heavily concentrated in the popular vehicles.

9.3.3 Standardised office premiums can be calculated for each postal code (at postal district level) or make/model. This enables a valid comparison of the claims experience of each postal code or make/model, allowing for any differences in mix of business between risks. Ideally, these should be calculated for each of the last three years together with the average standardised office premium over the period. It is also advisable to calculate the vehicle year exposure, standardised claim frequency and loss ratio for each year, to be used as supporting underwriting information. For most companies there will be little exposure in many of the risk groups, and this will result in a high degree of statistical variation between years. It is, therefore, sensible to apply a fairly low capping level to the bodily injury claims for the calculation of the office premiums. We have found, in practice, that for a medium sized insurer a capping level of about £5,000 is reasonable. It is possible that a higher capping level for the make/model analysis can be justified, since the average bodily cost tends to increase with vehicle group.

9.3.4 If the underwriter wishes to assess the reasonableness of the existing classifications, then the average standardised office premiums can be sorted by size within each of the existing classifications. Those postal codes or make/ models with relatively good or bad experience during the investigation period can be identified. It is important in the selection of the good or bad risks to choose those risks which have been consistently good or bad during the period. The underwriter can then decide to which new classification to allocate the risk.

In reaching his decision the underwriter can also take into account the supporting underwriting information described above.

9.3.5 The statistical analysis of postal codes and make/models is not an exact science, and much underwriting judgement is required. There will always be the problem of where to allocate new make/models or postal codes where there is little or no exposure. For the postal code analysis, we have found that the results of the statistical analysis is most usefully presented on mapping software which

can be purchased very cheaply. The good or bad postal codes can be colour coded, so that geographical trends can be easily identified. Local knowledge can normally be an invaluable aid to interpreting the trends.

9.3.6 Since the standardised office premiums of different risks are all comparable, the analysis can be used to assess the impact of large scale reclassifications, such as those currently being recommended for vehicle groups by the ABI.

10. USING THE RESULTS IN A COMPETITIVE MARKET

10.1 Comparing with the Existing Rate Book

10.1.1 The object of this paper has been to develop an approach to motor premium rating which can extract as much information as possible from the data, such that underwriters can clearly and easily assimilate the patterns and trends in the data, thereby helping them make better decisions. This is particularly important for companies with small portfolios, and companies which are planning to expand the size of their account.

10.1.2 Table 10.1.2 shows how the results of the modelling can be used to assess the reasonableness of a company's existing rating structure. We have taken as an example the analysis of premiums by vehicle age.

Table 10.1.2 gives for each vehicle age band the vehicle year exposure, the theoretical relative premiums based on the company's own data, the relative premiums currently being charged, and the theoretical premium adjustments necessary to move the existing structure to the rating structure implied by the theory. Although in this example we have shown the results for small bands of vehicle ages, our methodology produces relative premiums for each individual vehicle age, if required.

For simplicity, the above example gives the premium adjustments which are required to produce an unchanged amount of premium income assuming that the mix of business by vehicle age remains unaltered. Alternative calculations could be made incorporating further assumptions or alternative mix of business assumptions.

10.1.3 Table 10.1.2 shows that, on the basis of the company's past claims

Vehicle	Exposure	Relative pr	Adjustment	
age	(%)	Theoretical	Actual	(%)
0-1	20.1	1.060	1.000	7.4
2-3	27.5	1.000	1.000	1.3
4-5	25.1	0.934	0.950	-0.4
6-7	13.5	0.863	0.900	- 2.9
89	7∙0	0.789	0.850	- 6.0
10+	6.8	0.641	0.800	- 18.9
Total:	100.0			

Table 10.1.2. Analysis of vehicle age loadings/discounts

experience, its current rating structure is unbalanced. There are substantial cross subsidies between vehicles of different ages. In general older vehicles, and in particular 10 + year-old vehicles, are subsidising the newer vehicles.

Profitability could be significantly improved if the company could increase premium rates for new vehicles, and in particular 0-1 year-old vehicles, without significant loss of business volume. Similarly, profitability could be improved if an increase in business volume were obtainable for older vehicles following a small reduction in premium rates for these risks.

10.1.4 These decisions cannot be made without knowledge of the competitiveness of premium rates within the market for a variety of different risks. If it were possible to obtain a sample of premium quotations for different combinations of risks, then the financial implications of the theoretical recommendations above could be assessed.

10.2 Assessing the Effect of Competition

10.2.1 There are a number of premium quotation services available in the market, and these services can be used by companies to great advantage. Table 10.2.1 shows how the results of the statistical analysis can be used in conjunction with the output from one of the many quotation services available. In this example we have compared the premium rates from 35 different companies for a typical comprehensive policy. In each case we have selected a quotation for six different vehicle ages, all other rating factors remaining unchanged. In practice, a much larger number of quotations would be required to produce an effective analysis.

Vehicle age	Current premium (£)	Ranking	New premium (£)	New ranking
0	181	26	195	29
2	181	26	183	27
4	171	28	170	27
6	164	28	159	26
8	156	29	147	23
11	149	26	121	12

Table 10.2.1. Assessing the impact of competition

Table 10.2.1 shows the current premium for each of the six quotations. We also show the competitive ranking for each quotation. For example, the premium for a 0-year-old car is £181, and this was the 26th cheapest quotation out of the 35 companies. We also show the theoretical premium, calculated by applying the recommended premium increases of Table 10.1.2 to the current premiums, together with the revised competitive ranking.

10.2.2 Clearly, this company is currently very uncompetitive for each quotation. Furthermore, for the newer cars a move towards the theoretical premiums will not significantly affect its competitive position, although for older

cars a significant improvement in competitiveness is achievable. In practice, one would also look at the size of the proposed premium adjustment compared with the degree of change in competitive position.

10.3 The Contribution towards Profit

10.3.1 It is worth considering, at this stage, the current levels of profitability achievable. Table 10.3.1 shows the contribution towards profits of the company, assuming that its portfolio consists of the above policies with a mix of business similar to Table 10.1.2.

Vehicle age	Exposure	Current premium (£)	Expected cost (£)	Profit (£)
0	201	181	195	2,734
2	275	181	183	- 666
4	251	171	170	174
6	135	164	159	648
8	70	156	147	658
ļ 1	68	149	121	1,918
Overall:	1,000	172	172	0

Table 10.3.1. Analysis of profit

10.3.2 Significant losses are being made on 0-year-old vehicles, with profits being made on the older vehicles, in particular for 11-year-old vehicles. The company's profitability is, therefore, vulnerable to a change in mix of business towards new vehicles. Given the company's present competitive position, this would appear unlikely, although the situation should be closely monitored.

10.3.3 To improve profitability immediately, the premiums for 0-year-old vehicles could be increased by up to 7.4%. We have seen in Table 10.2.1 that this does not seriously affect the competitive position, although it is likely that a sizeable amount of business will be lost. In addition, sizeable reductions in premiums can be made for the older vehicles, in particular for vehicles aged 11 vears. This would significantly improve the company's competitive position for these risks, with the opportunity to increase business volume. It should be remembered that, provided the theoretical premiums make some allowance for per policy expenses, it is more important to ensure that business volume is maintained than it is to ensure that premium income is maintained.

10.3.4 A similar analysis can be carried out for each rating factor included in the analysis. The object will be to identify areas of the portfolio where both profitability and competitiveness can be improved. By considering various combinations of rating factors, it is quite possible to identify potentially profitable 'niches'. We have found, in practice, that the motor market is not uniformly competitive, some areas of the market being much more sensitive to rate changes than others. The competitiveness for different areas of the market is constantly changing, hence close monitoring is essential. It is also important that

profits are not thrown away unnecessarily, and a balance must be struck between the business volumes achievable in different sectors of market and the achievable profit per policy.

10.4 The Practical Advantages of the Approach

10.4.1 Having discussed how the results of the detailed statistical approach which we have advocated in this paper can be used to help the underwriter in practice, it is worth reconsidering some of the advantages of our proposed methodology.

Our analysis, being based on the company's own data, will:

- (1) allow for the company's individual underwriting standards,
- (2) allow for the company's method of distribution,
- (3) allow for the 'knock-for-knock' agreements in operation,
- (4) allow for the treatment of per policy expenses, and
- (5) enable the company to use a flexible design of rating structure.

10.4.2 The method of distribution has a significant effect on the amount of operational expenses. It is important that expenses are correctly allowed for in the analysis. For example, home service offices tend to pay lower rates of commission to their agents than those paid to brokers, but also tend to have higher overhead costs than the offices which obtain their business through brokers. Direct selling companies will have no commission, but relatively high fixed costs. It is important to build these expenses into the premium structure in a sensible manner, since it can affect the relative structure of premium rates.

10.4.3 For all the reasons mentioned above, a 'niche' for one company may not necessarily be a 'niche' for another. We believe that a true 'niche' can be only identified with confidence by applying a sophisticated statistical analysis to a company's own data.

11. ACKNOWLEDGEMENTS

Peter Lee prepared the data and carried out some of the GLIM runs for the numerical examples. Ian Cook wrote computer programs for the calculations described in Section 7. Andrew Smith made an important contribution to Appendix G.

We are also grateful to the many people who commented on earlier drafts of this paper, in particular, Professor John Nelder, Dr Pat Altham, Graham Ross and Kjell Andersson.

REFERENCES

- AITKIN, M., ANDERSON, D., FRANCIS, B. & HIND, J. (1989). Statistical Modelling in GLIM, Oxford Science Publications.
- ALMER, B. (1957). Risk Analysis in Theory and Practical Statistics. 15th International Congress of Actuaries, 2, 314.

- BAILEY, R. A. & SIMON, L. (1960). Two Studies in Automobile Insurance Ratemaking. ASTIN Bulletin, 1, 192.
- BAXTER, L. A., COUTTS, S. M. & ROSS, G. A. F. (1979). Applications of Linear Models in Motor Insurance. 21st International Congress of Actuaries, 2, 11.
- BEARD, R. E., PENTIKAINEN, T. & PESONEN, E. (1984). Risk Theory, the Stochastic Basis of Insurance (3rd edition). Chapman and Hall.
- BENNETT, M. C. (1978). Models in Motor Insurance. J.S.S. 22, 134.
- COUTTS, S. M. (1984). Motor Insurance Rating: an Actuarial Approach. J.I.A. 111, 87.
- HABERMAN, S. & RENSHAW, A. E. (1988). Generalised Linear Models in Actuarial Work, Staple Inn Actuarial Society paper.
- HEALY, M. J. R. (1988). GLIM: An Introduction. Oxford Clarendon Press.
- JOHNSON, P. D. & HEY, G. B. (1971). Statistical Studies in Motor Insurance. J.I.A. 97, 199.
- JUNG, J. (1968). On Automobile Insurance Ratemaking. ASTIN Bulletin, 5, 41.
- LITTLE, R. J. A. (1978). Generalised Linear Models for Cross Classified Data from the WFS, World Fertility Survey, technical bulletin.
- MCCULLAGH, P. & NELDER, J. A. (1983) and (1989). Generalised Linear Models, 1st and 3rd editions. Chapman and Hall.
- NUMERICAL ALGORITHMS GROUP. (1985). Generalised Linear Interactive Modelling.
- RENSHAW, A. E. (1991). Actuarial Graduation Practice and Generalised Linear and Non-Linear Models. J.I.A. 118, 295.
- WEDDERBURN, R. W. M. (1974). Quasi-Likelihood Functions, Generalised Linear Models and the Gauss-Newton Method. Biometrika, 61, 439-47.

APPENDIX A

GLIM

A.1 Theory

A.1.1. Suppose we have a set of data values Y_i for i = 1, 2, ..., n (in the present context Y_i would be either the actual number of claims per unit of exposure, or the actual mean claim size, for the *i*th combination of rating factors). GLIM deals with models of the form:

$$E(Y_i) = \mu_i$$
 and $Var(Y_i) = \phi \cdot V(\mu_i)/w_i$

where:

the mean μ_i has the form $\mu_i = h(\eta_i)$ where $h(\cdot)$ is some known monotonic function and η_i is a linear function of the unknown parameters of the model (η_i is called the linear predictor),

V() is a known function (called the variance function),

the quantities w_i (called the prior weights) are known, and

the constant factor ϕ is not necessarily known (it is called the scale parameter).

The random components of the data points Y_i for $i = 1 \dots n$ must be mutually independent.

A.1.2 In the motor rating context, there is generally one data point for each combination of several rating factors: if we have three rating factors (A, B and C say) the data may be represented as y_{ijk} . The linear predictor can then be expressed as:

$$\eta_{ijk} = \alpha_i + \beta_j + \gamma_k.$$

This is a linear function of the unknown parameters α_i , β_j and γ_k .

(In general there may also be interaction terms, but this is not important at this stage.) The parameter α_i is regarded as the effect of having rating factor A at its *i*th level, and similarly for the other two rating factors B and C. When the model has been specified in GLIM, the programme will use the data to estimate values of the parameters α_i , β_i , γ_k .

A.1.3 All the user has to do to fit his chosen model is to specify the following:

the data vector to be used as the y-variate, the form of the variance function V(), the form of the function h(), the data-vector to be used as the weights w, and the terms to be included in the linear predictor.

The above five stages of model specification are done respectively using the following GLIM commands: \$YVAR, \$ERROR, \$LINK, \$WEIGHTS, \$FIT

(all GLIM commands begin with a \$). Some examples are given below, but first some additional comments are necessary regarding specification of the functions V() and h().

A.1.4 The inverse of h() is known as the link function, and it is this which is specified using the \$LINK command. For example, if we wish to have $h(\eta) = \exp(\eta)$ we must specify the link function to be the log to base e. It is simply a matter of convention that h() is specified by way of its inverse and is of no consequence. Note, that since the mean μ of the data is given by $\mu = h(\eta)$ and η is a linear function, a log-link function corresponds to a multiplicative model for μ .

A.1.5 The variance function $V(\mu)$ is specified in GLIM by naming an error distribution, e.g. Normal, Poisson or Gamma. These mean respectively: $V(\mu) = 1$, $V(\mu) = \mu$, and $V(\mu) = \mu^2$. It is for technical and historical reasons that the variance function is specified in this way, and is of no real consequence. In particular, it is not necessary for the random variation in the data actually to follow the type of distribution specified in the \$ERROR command: GLIM will give conventional maximum-likelihood estimates if this is the case, but in other cases the parameter estimates remain optimal in a sense (see Wedderburn (1974) for more details on this point).

A.2 Examples

A.2.1 To illustrate the simplicity of model fitting using GLIM, we show below the GLIM commands needed to fit some of the claim frequency models described by Bennett (1978). These should be compared to the mass of equations given in Bennett's paper (as these would form the basis of an *ad hoc* computer program). It is assumed that the data have previously been read into *variates* named R and X. These are respectively: the actual number of claims per unit of exposure for each rating factor combination, and the exposure for each rating factor combination. The corresponding levels of the rating factors (policyholder age, car age and level of NCD) are assumed to have been read into *factors* named PA, CA and NC respectively. Explanatory comments are given to the right of the GLIM commands after the ! sign.

A.2.2 Bennett's model A is the additive model used by Johnson & Hey (1971) and can be fitted in GLIM as follows:

\$YVAR R	!	claim frequency data R to be the y-variate Y_{lik}
\$ERROR N	!	Normal $\Rightarrow V(\mu_{ijk}) = 1$ so $Var(Y_{ijk}) = \phi/W_{ijk}$
\$LINK I	!	identity link function gives additive model
\$WEIGHTS X	!	exposure data X to be the weights w_{ijk}
FIT PA + CA + NC	1	linear predictor to include one parameter for each
		level of each of the three rating factors

A.2.3 Bennett's model D differs from this only in that the mean claim frequency is multiplicative. This is achieved in GLIM by specifying a log link function; thus the code is as above, but with \$LINK I replaced by \$LINK L.

A.2.4 Bennett's model B is the additive model, but with the weights inversely

proportional to fitted values as advocated by Bailey & Simon (1960). In other words, the variances are assumed to be proportional to the fitted values. This is achieved in GLIM by specifying a Poisson error structure, so the code is as for model A, but with \$ERROR N replaced by \$ERROR P. GLIM will give parameter estimates slightly different from those obtained by Bennett, because he used minimum chi-squared estimation, whereas GLIM uses minimum deviance estimation, and these differ for non-Normal error structures.

A.2.5 Bennett's model C is the multiplicative model of Bailey & Simon (1960), but with the chi-squared function (to be minimised) adjusted as described by Jung (1968), so that the one-way marginal totals of claim numbers are equal for both actual and fitted. This is equivalent to minimum deviance estimation, so GLIM will reproduce Bennett's results exactly using the above code with \$ERROR P \$LINK L in place of \$ERROR N \$LINK I.

A.2.6 The distinction between a *variate* and a *factor* is important in GLIM. Sections 4 and 6 of the paper give examples of this distinction: the GLIM manual (NAG 1985) gives full details.

APPENDIX B

EMPIRICAL COMPARISON OF ADDITIVE AND MULTIPLICATIVE MODELS FOR CLAIM FREQUENCY

B.1 We wish to compare the following models for claim frequency:

1. Multiplicative model	SYVAR R SERROR P SLINK L! log link function SWEIGHTS X
2. Additive model	\$YVAR R \$ERROR P \$LINK I! identity link function \$WEIGHTS X

The Poisson error structure, with the exposure X as weights, is used in both models for the reasons given in Section 2.2 and Appendix C.

B.2 As explained in Section 2.1, we expect interaction terms to be more significant under the additive model than under the multiplicative model in general. To investigate this, we could first include all two-factor interactions:

 $FIT A+B+C+B\cdot C+C\cdot A+A\cdot B$

and then include the main effects only:

FIT A+B+C

The conventional F-statistic could then be calculated as described in §2.3.3 for both the additive and multiplicative models. This should indicate whether the interactions are more significant under the additive model or the multiplicative model. Unfortunately, this type of analysis cannot be carried out using many datasets, because the additive model usually gives some negative fitted values unless several two-factor interactions are included. (This is itself a point in favour of the multiplicative model.) However, the relatively small claim frequency datasets of Baxter *et al.* (1980) and of Coutts (1984) do allow the models to be compared in this way.

B.3 There are three rating factors in the data from Baxter (1980), each with 4 levels:

district	(DI)
car group	(CG)
policyholder age	(PA).

The minimised deviances for:

- (1) the multiplicative model, and
- (2) the additive model,

are given here:

	Residual	Deviance		
Model terms	degrees of freedom	1	2	
DI+CG+PA+CG·PA+PA·DI·CG	27	27.3	28·0	
DI+CG+PA	54	51.4	51·8	

The deviance is slightly lower for the multiplicative model than for the additive model (whether or not the two factor interactions are included), indicating that it gives a slightly better fit. However, the two-factor interactions are very insignificant under both models, the F-statistic on 27 and 27 degrees of freedom being 0.88 for the multiplicative model and 0.85 for the additive model.

B.4 There are four rating factors in the data from Coutts (1984):

cover	(CO) with 2 levels
car age	(CA) with 3 levels
car group	(CG) with 4 levels
policyholder age	(PA) with 5 levels.

Fitting main effects only, the multiplicative model gives a much better fit than the additive model, the minimised deviances being 115 and 135 respectively. However, under both models, some interaction terms are significant. Starting with all two-factor interactions, and removing insignificant ones in a step wise fashion, leaves three two-factor interactions in each case. These are shown in lines B and C of the next table.

	Residual	Devi	ance
Model terms	degrees of freedom	1	2
A Main effects + all 2-factor interactions	74	48	53
B Main effects + $CA \cdot CO + PA \cdot CA + CA \cdot CG$	93	68	76
C Main effects + $CA \cdot CO + PA \cdot CA + CO \cdot CG$	96	74	69
D Main effects + $CA \cdot CO + PA \cdot CA$	99	78	79
E Main effects only	109	115	135

Model C appears to be best under the additive assumption, and the fit is better than under the multiplicative assumption (69 < 74). However, model B, which is best under the multiplicative assumption, fits much better than under the additive assumption (68 < 76). Models with both CA \cdot CG and CO \cdot CG included or both excluded, all fit better under the multiplicative rather than the additive assumption.

B.5 Although these results are not conclusive, they support the belief that the multiplicative assumption is generally better.

APPENDIX C

ERROR STRUCTURE OF CLAIM FREQUENCY DATA

C.1 In Section 2.2 the case of just two rating factors indexed by i and i is considered, with data denoted x_{ii} , n_{ii} .

In this appendix we first relax the assumption (made in $\S 2.2.3$) of perfect homogeneity of risks within each cell, so we now have $m_{iik} \sim \text{Poisson}(f_{iik})$ (i.e. the Poisson rate depends on k).

If the f_{ik} for $k = 1, ..., x_{ii}$ are from a Gamma distribution with:

$$\mathbf{E}(f_{ijk}) = f_{ij} \quad \text{and} \quad \operatorname{Var}(f_{ijk}) = f_{ij}^{2}/h_{ij} \quad (*)$$

for some parameters f_{ij} and h_{ij} , then the number of claims m_{ijk} of an individual chosen at random from cell (i, j) is Negative Binomial with:

$$\mathbf{E}(m_{ijk}) = f_{ij}$$
 and $\operatorname{Var}(m_{ijk}) = (1 + f_{ij}/h_{ij}) f_{ij}$.

That is (dropping subscripts i and j) the same distribution as the number of failures before success number h in independent trials each with probability p of success where p = h/(h+f) (Beard et al., 38-40).

Hence the data n_{ij} are Negative Binomial with:

$$\mathbf{E}(n_{ij}) = x_{ij}f_{ij}$$
 and $\operatorname{Var}(n_{ij}) = (1 + f_{ij}/h_{ij}) x_{ij}f_{ij}$

and if $r_{ii} = n_{ii}/x_{ii}$ (as in §2.2.3):

$$\mathbf{E}(\mathbf{r}_{ij}) = f_{ij} \quad \text{and} \quad \mathbf{Var}(\mathbf{r}_{ij}) = (1 + f_{ij}/h_{ij}) f_{ij}/x_{ij} = \phi_{ij} f_{ij}/x_{ij} \quad \text{say.}$$

From (*) we see that $1/\sqrt{h}$ is the within-cell coefficient of variation of the Poisson rates. We advocate (Section 6) the use of as many rating cells as possible in modelling. One effect of this is that h is unlikely to be less than 1, and, in fact, a typical value is about 2. We also advocate (Section 5) the fitting of separate frequency models for different types of claim. This tends to keep the f_{ij} low, typically varying across cells in the range 0.02 to 0.16. Using these figures for fand h, we have ϕ varying across cells in the range 1.01 to 1.08: a difference of less than 7%. This is dwarfed by the between-cell variation in the other factors of Var(r_{ii}). Thus we can reasonably work on the basis that ϕ is constant, that is, we can use a Poisson error structure with weights x_{ij} .

C.2 We next relax the assumption of a Poisson claim process for each unit of risk: as mentioned in §2.2.5, the risk intensity is not constant as required for a Poisson process, but tends to decrease for a period after each claim. The nature and magnitude of the consequences of this can be determined by considering the extreme case in which the risk reduces to zero after the first claim and remains at zero for the remainder of the year. In this case m_{ijk} is either 0 or 1, and so $n_{ii} \sim \text{Binomial}(x_{ij}, f_{ij}).$

509

Hence:

$$E(r_{ij})=f_{ij}$$
 and $Var(r_{ij})=(1-f_{ij})f_{ij}/x_{ij}$.

So, in this extreme case the scale parameter is decreased by a factor $(1-f_{ij})$. In reality the factor will be much closer to 1 (it would be $(1-f_{ij}/12)$ if it were possible to make no more than one claim each month), and, of course, the variation of such a factor with *i* and *j* is very slight, so again the assumption of a constant scale parameter ϕ remains reasonable.

C.3 Finally, we relax the assumption of mutual independence of the risk-units within each cell. The possibility of two policyholders in the same cell being involved in the same accident means:

$$cov(m_{ijk}, m_{ijl}) > 0$$
 for all k and l.

Hence:

$$\operatorname{Var}(n_{ij}) > \sum_{k} \operatorname{Var}(m_{ijk})$$

so:

$$\operatorname{Var}(m_{ijk}) = \phi f_{ij}$$
 implies $\operatorname{Var}(n_{ij}) > \phi x_{ij} f_{ij}$

so the scale parameter for r_{ij} (= n_{ij}/x_{ij}) is increased, but the effect is obviously very slight.

APPENDIX D

WITHIN-CELL HETEROGENEITY OF CLAIM FREOUENCY

D.1 Theory

D.1.1 Suppose we have mutually independent Poisson processes for the claim incidence on each policy (so of the three assumptions considered in Section 2.2 and Appendix C, the last two hold). It is shown in Appendix C that if the Poisson rates are Gamma distributed within each cell, with coefficient of variation $1/\sqrt{h_{ij}}$, then the observed mean claim frequency r_{ii} has:

$$E(r_{ij}) = f_{ij}$$
 and $Var(r_{ij}) = \phi_{ij} f_{ij} / x_{ij}$

where:

 f_{ii} is the mean Poisson-rate (i.e. the mean of the Gamma distribution) and

$$\phi_{ij} = 1 + f_{ij}/h_{ij}.$$

In this appendix we consider the question of how much of the heterogeneity in claim frequency is accounted for by the model. Johnson & Hey (1971) also investigated this question: we base our analysis on the same equation as they used.

D.1.2 Although the range of Poisson parameters is continuous, it can be approximated arbitrarily closely by a finite sequence of values, λ_k say. Let p_{ijk} = proportion of all exposure which is in cell (*i*, *j*) and has Poisson rate λ_k for claim frequency. Then:

$$\sum_{k} p_{ijk} = \text{proportion of exposure in cell } (i, j)$$
$$= x_{ij} / \sum_{ij} x_{ij}.$$

This is denoted p_{ii} .

Since p_{iik}/p_{ii} for $k = 1, 2 \dots$ represents the distribution of Poisson rates within cell *i*, *j* (it is a discrete approximation to the Gamma distribution) we have:

$$\sum_{k} (p_{ijk}/p_{ij})\lambda_k = f_{ij} \qquad \text{i.e.} \sum_{k} p_{ijk}\lambda_k = p_{ij}f_{ij}. \tag{1}$$

The mean Poisson-rate over all cells is denoted μ , that is:

$$\mu = \sum_{ijk} p_{ijk} \lambda_k$$

512 Statistical Motor Rating: Making Effective Use of Your Data From (1) we have

$$\mu = \sum_{ij} p_{ij} f_{ij}.$$
 (2)

Now consider the overall variance of the Poisson-rate:

$$\mathbf{V} = \sum_{ijk} p_{ijk} (\lambda_k - \mu)^2. \tag{3}$$

It is straightforward to show that this can be expressed as:

$$\mathbf{V} = \sum_{ijk} p_{ijk} (\lambda_k - f_{ij})^2 + \sum_{ij} p_{ij} (f_{ij} - \mu)^2.$$
(4)

These two terms can be regarded as the 'within-cell variance' and 'between-cell variance' respectively, for the Poisson-rates.

D.1.3 Johnson & Hey (1971) used equation (4) as follows:

- (i) they calculated the between cell variance (directly from its definition) using fitted values from their model for f_{ij} ,
- (ii) they calculated the total variance V, and
- (iii) the within-cell variance can be calculated as the difference, using equation (4).

They carried out step (ii) by assuming that the distribution of Poisson-rates for all cells combined followed a Gamma distribution. Hence, by fitting a Negative Binomial distribution to data on numbers of claims per policy (for all cells combined) they were able to estimate the parameters of the Gamma distribution and hence its variance V.

This approach suffers from difficulties over the handling of incomplete policy years (as Johnson & Hey acknowledged). For our model, there is an alternative approximate method for estimating the within-cell variance, which does not suffer from these difficulties. This is to use the residual variation in the observed mean frequencies r_{ii} .

D.1.4 From the assumption of a Gamma distribution of Poisson-rates within each cell we have:

$$\sum_{k} (p_{ijk}/p_{ij}) (\lambda_{k} - f_{ij})^{2} = f_{ij}^{2}/h_{ij}$$

(each side is the variance of the Poisson-rate within cell i, j).

Therefore, 'within-cell variance' can be expressed:

$$\sum_{ijk} p_{ijk} (\lambda_k - f_{ij})^2 = \sum_{ij} p_{ij} f_{ij}^2 / h_{ij}$$
$$= \sum_{ij} p_{ij} f_{ij} (\phi_{ij} - 1).$$

The estimate of the scale parameter given by GLIM when frequency is modelled as proposed in this paper is an estimate of ϕ_{ij} made on the assumption that ϕ_{ij} is the same for all cells. There may in fact be some slight variation in ϕ_{ij} (see Appendix C) but the GLIM estimate $\hat{\phi}$ can be regarded as an average value. Approximately, we have therefore:

within-cell variance =
$$\sum_{ij} p_{ij} f_{ij} (\hat{\phi} - 1)$$
$$= \mu (\phi - 1).$$
(5)

A value for μ can, of course, be obtained using the fitted values f_{ij} from the frequency model. Note that the approximation is likely to be poor if $\hat{\phi}$ is close to 1 as is usual.

D.2 Example

٩

D.2.1 These calculations have been carried out after fitting a model of the type described in Section 6 to the own-damage data used in Section 2.4. The mean own-damage claim frequency over all cells, calculated from equation (2) using the fitted values \hat{f} , is 0.1043. The between cell variance, the second term of equation (4) calculated using the fitted values \hat{f} , is 0.001005. The estimate $\hat{\phi}$ calculated as the mean deviance, is $1911 \cdot 7/1879 = 1.0174$. The approximate within-cell variance, calculated from equation (5), is therefore 0.0018.

D.2.2 Although this result is very approximate, it suggests that a substantial proportion of the variation in accident proneness between individuals is not explained by the available information on the rating factors DI, PA, CA and CG. Possible explanations are:

- (i) The bands of the available rating factors were too broad: the use of smaller cells would increase the between-cell variance and decrease the within-cell variance.
- (ii) Other rating factors are also important, for example, sex of policyholder and class of vehicle use.
- (iii) The accident proneness of individuals, and their readiness to claim, depends on other factors which cannot be measured by an insurer, e.g. drinking habits, and the psychological make-up of the policyholder.

D.2.3 NCD systems attempt to take account of the factors mentioned under (iii). Johnson & Hey (1971) included NCD level as an explanatory variable in their model for claim frequency for all claim types combined, but still found:

> between-cell variance $\simeq 0.0025$ within-cell variance $\simeq 0.0055$.

However, it would be wrong to conclude that the NCD system is poor at allowing for unmeasurable risk factors, because points (i) and (ii) apply to Johnson & Hey's analysis.

APPENDIX E

DIRECT ESTIMATION OF WITHIN CELL COEFFICIENT OF VARIATION FOR CLAIM SEVERITIES

E.1 In this appendix we consider the question of whether the scale parameter given by GLIM, when the model of Section 3.1 is used, agrees with the true within cell coefficient of variation.

The model is:

$$E(Z_{ijk}) = m_{ij} \quad \text{and} \quad Var(Z_{ijk}) = \sigma^2 m_{ij}^2.$$

If $\dot{Z}_{ijk} = Z_{ijk}/m_{ij}$ we have $E(\dot{Z}_{ijk}) = 1$ and $Var(\dot{Z}_{ijk}) = \sigma^2.$

Values for \dot{Z}_{ijk} can be obtained by dividing individual claim data Z_{ijk} by the fitted values m_{ij} obtained using the GLIM model. A direct estimate of σ^2 (for comparison with the GLIM estimate) can then be obtained as the sample variance of the \dot{Z}_{ijk} .

That is:

$$\hat{\sigma}^{2} = \left(\sum_{ijk} \dot{Z}_{ijk}^{2}\right)/n - \left(\sum_{ijk} \dot{Z}_{ijk}\right)^{2}/n^{2}$$

where n = total number of claims in all cells,

$$n=\sum_{ij}n_{ij}.$$

The fitted values m_{ij} necessarily satisfy:

$$\sum_{ij} n_{ij} S_{ij} / m_{ij} = \sum_{ij} n_{ij}$$

so we have:

$$\sum_{ijk} \dot{Z}_{ijk} = n$$

and hence:

$$\hat{\sigma}^2 = \left(\sum_{ijk} \dot{Z}_{ijk}^2\right)/n - 1.$$

If q_{ij} denotes the sum of squares of claims in cell i, j:

$$q_{ij} = \sum_k Z_{ijk}^2$$

Statistical Motor Rating: Making Effective Use of Your Data 515 then this can be expressed:

$$\hat{\sigma}^2 = \left(\sum_{ij} q_{ij}/m_{ij}^2\right)/n - 1$$

so the only additional data required for this direct estimate of σ^2 are the quantities q_{ij} .

APPENDIX F

TESTING THE ASSUMPTION OF THE CLAIM SEVERITY MODEL

F.1 Theory

F.1.1 In the notation of Section 3 we have:

 $E(Z_{ijk}) = m_{ij}$ and $Var(Z_{ijk}) = \sigma_{ij}^2 m_{ij}^2$.

The only assumption of the basic claim severity model, other than that the claim sizes Z_{ijk} are mutually independent, is that the coefficient of variation σ_{ij}^2 is the same for all cells. A formal test of this hypothesis, based on data for individual claims, can be formulated by assuming that claim sizes are log-Normally distributed. This is a common assumption for claim sizes, and is probably a good enough approximation for the present purpose. Thus, if $L_{ijk} = \log_e (Z_{ijk})$ we have:

$$L_{ijk} \sim N(\lambda_{ij}, T_{ij}^2)$$
 for some λ_{ij}, T_{ij}^2 .

Using standard results for the log-Normal distribution:

$$m_{ij} = \exp(\lambda_{ij} + \frac{1}{2}T_{ij}^2)$$

$$\sigma_{ij}^2 = \exp(T_{ij}^2) - 1$$

so we wish to test the hypothesis that T_{ij}^2 is the same for all *i*, *j*.

F.1.2 Using individual claim size data, the sample variance of the log of the claim size can be calculated for each cell:

$$t_{ij} = \sum_{k} (L_{ijk} - L_{ij})^2 / (n_{ij} - 1)$$
 where L_{ij} is the mean of the L_{ijk} .

This has a scaled chi-squared distribution:

$$t_{ij} \sim (T_{ij}^2/(n_{ij}-1))\chi_{n_{ij}-1}^2$$

from which we have:

$$E(t_{ij}) = T_{ij}^2$$
 and $Var(t_{ij}) = 2T_{ij}^4/(n_{ij}-1)$

that is the sample variances t_{ij} have a Gamma error structure with weights $(n_{ij}-1)$ and scale parameter $\phi = 2$ (Appendix A).

Therefore, the dependence of T_{ij}^2 on the levels *i*, *j* of the rating factors can be investigated in GLIM using:

where T is a vector containing the sample variances t_{ij} and A, B contain the corresponding levels of the rating factors.

F.2 Example

F.2.1 In the example of Section 3.3, plots of residuals against rating-factors appeared to show some heteroscedasticity. The above analysis was carried out using all four rating factors:

district with 8 levels	(DI)
p/h age with 8 levels	(PA)
car age with 5 levels	(CA)
car group with 8 levels	(CG).

F.2.2 In recognition of the fact that the log-Normal assumption is only an approximation, sample variances t were calculated only for those cells with $n \ge 6$ (299 cells out of the 1890 cells with non-zero exposure). A sequence of F-tests was used in the usual way to examine the significance of the main effects and two factor interactions leading to the model: \$FIT DI+PA+CA+CG+PA·CG +DI·PA. The scale-parameter corresponding to this model was estimated as $\phi = 2.07$. The closeness of this to the theoretical value of 2 suggests that the log-Normal approximation is adequate for the present purpose. The significant variation in T^2 is summarised below:

higher by a factor of about 2.5 for NI compared to other districts,

decreases with car age, by factor of 0.6 for cars aged 10 + compared to cars aged 0-3 years, and

increases with car group by factor of 1.4 for cars in group 7–8 compared to cars in group 1.

F.2.3 Fitted values T_{ij}^2 were calculated using this final model and transformed to estimates of the within-cell coefficient of variation using $\sigma_{ij}^2 = \exp(T_{ij}^2) - 1$. These were used to adjust the weights of the severity model: \$WEIGHTS N replaced by \$WEIGHTS N/V where V is a vector holding the estimates σ_{ij}^2 . The residual plots were then satisfactory. Some of the severity relativities were significantly affected by correcting the weights in this way. For example, the model with \$WEIGHTS N gave car group relativities (by fitting a quadratic curve) of:

1.00 0.99 1.02 1.09 1.22 1.43 1.74 2.20

After correcting the weights these became:

1.00 1.03 1.08 1.16 1.27 1.43 1.64 1.93.

APPENDIX G

STANDARD ERRORS OF RISK PREMIUMS

G.1 In §7.1.1 we propose calculating risk premiums P_{ij} from fitted values of the component frequency and severity models as follows:

$$P_{ij} = \sum_{t} \hat{f}_{ijt} \, \hat{m}_{ijt}$$

(the symbol indicating an estimate was omitted in $\S7.1.1$).

By the asymptotic theory of generalised linear models, each item on the right of this equation is approximately unbiased. It is shown in Section G.2 that within each rating cell (i, j) these items are also asymptotically independent. These results imply that each P_{ij} is approximately unbiased.

Asymptotic approximations will be good, because there are typically very many individual claims per estimated parameter for each of the component models.

G.2 In this section, we show that for each rating cell (i, j fixed), the estimates \hat{f}_{ijt} and \hat{m}_{ijt} are asymptotically mutually independent. Mutual independence is obvious between different claim types *t*, because the models for different claim types are based on mutually disjoint data sets, each consisting of a sample of independent claims.

We need to consider each pair \hat{f}_{ii} and \hat{m}_{ii} in the same cell *i* and claim type *t*. (To simplify the presentation, rating cells are indexed using a single subscript instead of the pair *i*, *j* used in §7.1.1 and §G.1.)

For each (i,t): \hat{f}_{it} is estimated from the data $r_{jt} = n_{jt}/x_{jt}$; and \hat{m}_{it} is estimated from the data $s_{it} = y_{it}/n_{it}$ (where j varies over all rating cells).

The appearance of the same random numbers n_{jt} in both data sets leads to stochastic dependence between \hat{f}_{it} and \hat{m}_{it} . However, \hat{f}_{it} and \hat{m}_{it} are asymptotically independent (for each fixed *i* and *t*) as the following argument shows.

In Sections 2 to 6 we propose the following steps for each claim type t:

- (i) use assumptions about the distribution of the data r_j to estimate \hat{f}_i (where both *i* and *j* vary over all rating cells), and
- (ii) use assumptions about the conditional distribution of $s_j|r_j$ to estimate \hat{m}_i .

The assumptions for step (i) are given in Section 2. Estimation is carried out by maximising the likelihood of r_j , corresponding to a Poisson distribution for claim numbers. This likelihood is denoted $P_{\alpha}(\mathbf{r})$, where α represents the parameters of the model f_i . If the distribution of claim number is not exactly Poisson, then $P_{\alpha}(\mathbf{r})$ is a quasi-likelihood, and the estimates $\hat{\alpha}$ obtained in this way retain all the desirable properties of maximum likelihood estimates, see Wedderburn (1974).

The assumptions for step (ii) are given in Section 3. Estimation is carried out by

maximising the likelihood of $s_j|r_j$ corresponding to a Gamma distribution for claim sizes. If β represents the parameters of the model for m_i , this likelihood can be denoted $P_{\beta}(s|r)$. The same comments apply on the efficacy of estimates β obtained in this way when the claim size distribution is not, in fact, a Gamma distribution (Wedderburn, 1974).

An alternative to this two stage procedure would be to estimate α and β simultaneously by maximising the joint likelihood of r and s. By definition of the conditional distribution of s|r, the joint likelihood P(r,s) is given by:

$$\mathbf{P}(\mathbf{r},\mathbf{s}) = \mathbf{P}_{\alpha}(\mathbf{r})\mathbf{P}_{\beta}(\mathbf{s}|\mathbf{r}). \tag{1}$$

As α appears only in the first factor, and β only the second factor on the right, maximisation of P(r,s) is equivalent to the two stage procedure proposed in the paper.

Applying the general theory of maximum likelihood estimation to the joint estimation procedure: the estimates $(\hat{\alpha}, \hat{\beta})$ are asymptotically normal with variance-covariance matrix V given by:

$$\mathbf{V} = \begin{pmatrix} \frac{\partial^2 \mathbf{D}}{\partial \alpha^2} & \frac{\partial^2 \mathbf{D}}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \mathbf{D}}{\partial \alpha \partial \beta} & \frac{\partial^2 \mathbf{D}}{\partial \beta^2} \end{pmatrix}^{-1}$$

where D is the deviance, which is defined as $-2\ln$ (likelihood).

From (1) we have $D = -2\ln(P_{\alpha}(r)) - 2\ln(P_{\beta}(s|r))$, from which $\partial^2 D/\partial \alpha \partial \beta$ is trivially zero.

Hence V is block diagonal, that is $\hat{\alpha}$ and $\hat{\beta}$ are asymptotically uncorrelated, and by asymptotic normality they are asymptotically independent. Since \hat{f}_i is a function of $\hat{\alpha}$ and \hat{m}_i is a function of $\hat{\beta}$, \hat{f}_i and \hat{m}_i are asymptotically independent, for each cell *i*.

G.3 Since, for each cell (i, j), the quantities \hat{f}_{ijt} , \hat{m}_{ijt} (t = 1, 2...) are all mutually independent (to a good approximation) the variance of the risk premium P is given by:

$$\operatorname{Var}(P) = \sum_{i} \operatorname{Var}(\hat{f}_{i} \hat{m}_{i})$$
$$= \sum_{i} \{\operatorname{Var}(\hat{f}_{i}) m_{i}^{2} + \operatorname{Var}(\hat{m}_{i}) f_{i}^{2} + \operatorname{Var}(\hat{f}_{i}) \operatorname{Var}(\hat{m}_{i})\}$$
(2)

and a standard error for P can be obtained by replacing m_t by \hat{m}_t and f_t by \hat{f}_t in this formula.

(Subscripts *i*, *j* have been dropped for simplicity.)

Since we have a multiplicative model for both frequency f_t and severity m_t , we can use μ as a generic symbol for any of these, and we have a model of the form:

$$\mu = \exp(\mathbf{u}^T \cdot b)$$

where:

u = known vector of rating factor levels, and

b = model parameters.

GLIM gives estimates \hat{b} together with a standard error matrix \hat{V} satisfying approximately $\hat{b} \sim \text{Normal } (b, V)$.

Hence:

$$\boldsymbol{u}^T \hat{\boldsymbol{b}} \sim N(\boldsymbol{u}^T \cdot \boldsymbol{b}, \, \boldsymbol{u}^T \cdot \hat{\boldsymbol{\nabla}} \cdot \boldsymbol{u}) = N(\lambda, \, \hat{\sigma}^2) \text{ say.}$$

Consider:

$$\hat{\mu} = \exp(\boldsymbol{u}^T \cdot \hat{\boldsymbol{b}} - \frac{1}{2} \sigma^2). \tag{3}$$

Using standard results for the log-Normal distribution, this has:

$$E(\hat{\mu}) = e^{\lambda} = \mu \quad \text{and} \quad Var(\hat{\mu}) = e^{2\lambda} (e^{\sigma 2} - 1)$$
$$\simeq \hat{\mu}^2 (e^{\sigma 2} - 1).$$

Replacing σ^2 by $\hat{\sigma}^2$ in expression (3) for $\hat{\mu}$, allows $\hat{\mu}$ and $Var(\hat{\mu})$ to be evaluated for substitution in formula (2).

In practice $\hat{\sigma}^2$ is invariably very small compared to $\hat{\lambda}$, so the 'bias correction' at (3) makes virtually no difference: fitted values can be calculated using $\hat{\mu} = \exp(\mathbf{u}^T \hat{b})$.

APPENDIX H

MODELLING THE MEAN NO CLAIM DISCOUNT

H.1 In this appendix, the notation of Section 7.2 is simplified, by using the single subscript *i* to index rating cells (regardless of how many rating factors there may be). We assume for simplicity that only complete policy years are used for this analysis, that is $x_{ik} = 1$ for all policyholders k. The model easily generalises to include fractional policy years.

Thus:

 $C_{ik} = \text{NCD}$ of policyholder k in rating cell i, and $x_{ik} = \text{exposure of policyholder } k$ in rating cell i.

Within each rating cell the C_{lk} can be regarded as independent identically distributed random variables taking values in the set {0, 20%, 30%, 40%, 50%, 60%} or similar.

We wish to estimate the mean NCD, C_i of this distribution for use in converting net office premium to gross.

H.2 If C_{max} denotes the maximum NCD allowed (e.g. $C_{\text{max}} = 60\%$) and we define $C_{ik} = C_{ik}/C_{\text{max}}$, then, for fixed *i*, the C_{ik} are independent and identically distributed, taking values {0,1, some intermediate values}.

Thus, the distribution of the (\dot{C}_{ik}) can reasonably be modelled as a scaled binomial distribution:

$$C_{ik} \sim 1/m_i \cdot \mathbf{B}(m_i, p_i)$$

for some parameters m_i, p_i .

This implies:

$$E(\dot{C}_{ik}) = p_i$$
 and $Var(\dot{C}_{ik}) = p_i(1-p_i)/m_i$.

Hence, if y_i denotes the total of the \dot{C}_{ik} :

$$y_i = \left(\sum_k C_{ik}\right) / C_{\max}$$

we have:

$$E(y_i) = p_i n_i$$
 and $Var(y_i) = p_i n_i (1-p_i)/m_i$

where n_i is the number of policies observed in cell *i*, that is:

$$n_i = \sum_k x_{ik}.$$

This is a scaled binomial error structure for y_i , the scale parameter being $1/m_i$. Note that $y_i = n_i \bar{C}_i / C_{\text{max}}$ where $\bar{C}_i =$ sample mean NCD for the *i*th cell.

It seems reasonable to assume that m_i does not depend on the rating cell *i*,

because m_i will be related to the number of levels of the NCD scale, and this is usually the same for all cells.

H.3 The data y_i can thus be modelled in GLIM using:

\$YVAR Y \$ERROR B N \$SCALE \$

where N holds the data n_i (this is the binomial denominator: see GLIM manual for details). The command SCALE allows a scale parameter $\phi \neq 1$ which is necessary because the model has $\phi = 1/m$. Various linear predictors can be compared using FITA + B, etc., followed by F-tests, where A, B hold the rating factor levels as usual. When the best model has been found, fitted values of the mean NCD are given by:

$$\hat{C}_i = C_{\max} \cdot \hat{y}_i / n_i.$$

H.4 Although no direct investigations have been carried out as to whether the distribution of within-cell NCD levels can be adequately approximated by a binomial distribution (or some other distribution with a similar variance function), the model described here seems to work:

residual plots are generally satisfactory, the scale parameter given by GLIM is consistent with $\phi = 1/m$, and the fitted values look plausible.

Also, the binomial distribution can approximate quite closely the steady-state distribution of a Markov process, which is the model used by Johnson & Hey.

APPENDIX J

INCLUSION OF NO CLAIM DISCOUNT LEVEL AS AN EXPLANATORY VARIABLE

J.1 In this appendix, it is shown that if the NCD percentages are to remain unchanged (option (ii), of §1.9.2) then the inclusion of NCD level as an explanatory variable in the component models does not yield correct relativities for the other rating factors in general. (The 'correct' relativities are taken to be those which give no cross-subsidisation between the rating cells defined by the other rating factors.)

Without loss of generality, the cells defined by the other rating factors are indexed using the single subscript i.

J.2 Initially, to clarify presentation, a simple case is considered. Later it is shown that the same conclusions apply in more realistic scenarios. The simplifying assumptions are:

- (i) NCD scale has just two levels j=0, corresponding to no discount on gross premium, and j=1 for which a proportion c of gross premium applies (e.g. c=70%, discount = 1-c=30%).
- (ii) Only one claim type, that is a single frequency model and a single severity model.
- (iii) No interaction between NCD and other rating factors in determining either frequency or severity. That is, the effect of NCD level is purely multiplicative in both cases, and we have:

Expected claim frequency $f_{ij} = f_i \alpha_j$ Expected claim size $m_{ij} = m_i \beta_j$

where $\alpha_0 = \beta_0 = 1$, so α_1 , β_1 are the 'discount factors' for policies at NCD level 1, for each of frequency and severity.

(iv) No expenses.

Since we have only one claim type, the risk premium for policies in cell i at NCD level j is given by:

$$P_{ij} = f_{ij} m_{ij} = f_i m_i d_j$$
 where $d_j = \alpha_i \beta_j$

hence $d_0 = 1$ and d_1 (denoted d below) is the 'true' discount factor.

Consider the quantities P_{i0} . They are the true risk premiums for policies with no no-claims-discount (NCD level j=0), for each rating cell *i*. However, they do not in general give the correct relativities for the gross risk premiums. This can be seen as follows:

Suppose we have x_i policies in cell *i*, of which a proportion q_i are on NCD level j=1, so have discount factor *c*.

524 Statistical Motor Rating: Making Effective Use of Your Data If the quantities P_{i0} are used for the gross premium then:

total premium for cell $i = x_i P_{io}[(1 - q_i) + q_i c]$

however, the true expected loss for cell $i = x_i P_{io}[(1 - q_i) + q_i d]$.

The ratio of these is:

$$\phi_i = \frac{1 - q_i + dq_i}{1 - q_i + cq_i}$$

For the P_{i0} to give correct relativities, we must have ϕ_i constant. This is true only if either:

c = d i.e. existing NCD scale is correct,

or:

 q_i is constant i.e. the same distribution of policies over NCD levels for all rating cells.

J.3 More generally, one can relax assumptions (i) and (ii):

(i) several NCD levels $j = 0, 1, 2 \dots$

with c_j = proportion of gross premium applicable to NCD level j, $c_0 = 1$, and (ii) several claim types $t = 1, 2, 3 \dots$

We have separate models for each claim type t:

$$f_{ijl} = f_{il} \alpha_{jl}$$
 and $m_{ijl} = m_{il} \beta_{jl}$

for some α_{jt} , β_{jt} with $\alpha_{ot} = \beta_{ot} = 1$.

The true risk premium is given by:

$$P_{ij} = \sum_{t} f_{ijt} m_{ijt}$$
$$= \sum_{t} f_{it} m_{it} d_{jt}$$

where $d_{jt} = \alpha_{jt} \beta_{jt}$ == true discount factor for NCD level *j* for claims of type *t*. The 'true' discount factor for NCD level *j* in cell *i* is defined by:

$$D_{ij} = P_{ij}/P_{io}$$

hence:

$$D_{ij} = \frac{\sum_{t} p_{it} d_{jt}}{\sum_{t} p_{it}}$$

where: $p_{ii} = f_{ii} m_{ii}$

= risk premium for policies on NCD level 0, cell i, type t.

Now suppose q_{ij} represents the distribution of policies in cell *i* over the NCD levels:

$$\sum_{j} q_{ij} = 1 \text{ for all } i.$$

Consider the use of the quantities P_{io} as gross premiums:

total premium from cell *i* (given
$$q_{ij}$$
) = $x_i P_{io}\left(\sum_j q_{ij} c_j\right)$
expected loss from cell *i* (given q_{ij}) = $x_i P_{io}\left(\sum_j q_{ij} D_{ij}\right)$

Hence the correction factor to be applied to the P_{lo} in order to get correct relativities for gross premiums is:

$$\phi_i(q_{ij}) = \frac{\sum\limits_j q_{ij} D_{ij}}{\sum\limits_j q_{ij} c_j}.$$

Note that this is a function of the distribution q_{ij} .

If the factors ϕ_i could be estimated, estimates \hat{P}_{io} obtained by fitting the models could be corrected to give the required relativities.

Estimates \hat{D}_{ij} could be obtained from the fitted models, and the distribution q_{ij} of policies over NCD levels in each cell is known. However, for small cells, this distribution will be subject to a great deal of random variation both between cells, and over time. Therefore, for useful relativities, we require $E_{q_{ij}}(\phi_i(q_{ij}))$. This would be extremely difficult to estimate reliably. A first approximation

This would be extremely difficult to estimate reliably. A first approximation could be obtained by using the model of Appendix H to find the expected values of the numerator and denominator of ϕ_i in turn, but the reliability would be very uncertain.

In practice, the situation would be further complicated because assumptions (iii) and (iv) might not apply.

J.4 Our method overcomes these problems as follows:

we use data for all NCD levels combined, for each claim type. This has:

expected frequency
$$f_{it} = \sum_{j} q_{ij} f_{ijt}$$

expected severity $m_{it} = \sum_{j} q_{ij} m_{ijt}$.

Note that the q_{ij} on the right side of these equations are expected values, so there is no problem with random variation.

Hence risk premium for cell i:

$$P_i = \sum_{t} f_{it} m_{it}.$$

We then use the model of appendix H to estimate the expected value of the within-cell mean NCD level:

$$C_i = \sum_j q_{ij} c_j.$$

For gross premium we use:

$$P_{io} = P_i / \hat{C}_i$$

We then have:

expected total premium received from cell $i = x_i P_{io}\left(\sum_j q_{ij} c_j\right)$

 $= x_i P_i$

= expected total loss for cell i.

ABSTRACT OF THE DISCUSSION

Mr M. C. Bennett (opening the discussion): If we wish to maintain and develop a profitable motor insurance account, then it is at our peril if we fail to take adequate account in the rating structure of the variation in risk within the portfolio. The paper addresses this topic of variation in risk, which clearly is far from trivial in view of the large number of factors which are taken into account in determining the premium to be charged. Anyone seeking quotations for a particular motor risk will know that premiums for identical cover vary widely from one insurer to another. This variation seems at least as great today as at any time since the early 1970s, the last time motor insurance was the subject of an Institute paper. Much of this variation is because the structure of the premiums differs so much among the companies writing motor business in the United Kingdom, rather than because of different levels of generosity in claim settlement or different profit margins. A paper which draws attention, once again, to ways to avoid writing business which will produce a loss is to be welcomed as a step to improving the profitability of individual motor accounts and, more generally, to developing a somewhat more orderly market for motor insurance. While a free market and healthy competition are, in themselves, to be applauded, the remarkably wide variations in premiums between companies, which are to a major extent a consequence of inadequate risk assessment, lead to undue lapsing of policies and undue competitive pressures in the motor insurance market.

The authors refer to past papers on modelling a motor insurance portfolio, including those which advocate simple additive models with parameter values which are relatively easy to determine with limited computer capacity. Such additive models have given a reasonable fit of actual to expected values for many, but not all, sections of a private motor portfolio, in particular excluding the more extreme risks for which there is always the greatest uncertainty in risk assessment. Additive models have been successful as a means of standardising results, that is, making adjustments in the risk relativities to allow for the varying mix of business by the rating factors not currently under investigation. By using any simple model, additive or multiplicative, one can keep under review the ratios of claims plus expenses to earned premiums in one-way and broad two-way cells, and make small adjustments to the rating structure from time to time, based on the standardisations which the authors describe, to bring the premiums more closely into line with the experience. Furthermore, a simple additive model will permit the monitoring and development of a premium scale of multiplicative form and contains some interaction terms involving more than one rating factor. Given that so much can be achieved by simple models, it remains a matter for consideration as to the extent to which more elaborate modelling, as described by the authors, can be expected to produce significantly more satisfactory premium relativities than are possible using a simple model, given the important practical uncertainties to which I refer later.

The authors make use of the statistical package, GLIM, a remarkable 'black box', which is referred to in various places, but whose practical scope has been set out more clearly in this paper than some of us have seen before. Given the much larger capacity computers and the much faster processing speeds that are available today, there seems to be no strong argument against using GLIM in view of its power and flexibility.

The authors advocate the use of a multiplicative form of the basic model, and give examples in Sections 2 and 3 to support this form of model. I am not convinced by their argument regarding, for example, the multiplicative form of risk relativities for different combinations of age of policyholder and car group, and I grow much more doubtful when all the other rating factors, such as age of car and geographical area are added. In Section 3 the authors use a similar illustration to try to justify the use of a multiplicative form of model to represent average claim amount. Again I do not find the theoretical argument particularly convincing. In the absence of any other sufficiently simple and reasonably intuitive form of model, we may indeed find ourselves using a simple multiplicative model supplemented by such interaction terms as we judge to be appropriate, although I suspect that the best model may well be of some intermediate form between additive and multiplicative. All this said, does the precise form of model later in the paper. What does matter crucially is that we have *some* model, and that we understand the strengths and limitations of whatever model we are using.

In §4.1.1 the authors refer to the conflict between using sufficiently recent data and using a sufficient volume of data. This conflict has always been a serious one in rating motor insurance. For a reasonably sized account it is true that, for most claim types, the claims analysis can be confined to the most recent few years of claims. For bodily injury claims a much longer time span than the 3 or 4 years indicated by the authors is, I suggest, preferable: because the data are sparse and of exceptionally high variance; because there are intrinsic uncertainties in claim amount at the early stages of development of a claim; and because the magnitude of such claims can tend to be underestimated at early durations, and such understatement could easily distort the assessment of risk relativities.

The authors propose that certain claim types, such as those in § 5.2.1, be analysed separately in the modelling, as regards both frequency and amount of claim. This appears to be a sound approach, since trends can vary from one type of claim to another. The authors remark that judgement is needed regarding future trends in frequency of each claim type, but once those judgements have been made the modelling can be carried out. It would be particularly useful to do separate modelling for windscreen and theft claims and claims involving bodily injury. Having examined various claim types separately, it makes good sense to use the principle of parsimony, as the authors suggest, to combine data for different claim types where possible, and consequently avoid unnecessary parameters in the model. I suspect, incidentally, that the traditional forms of simple model could readily be extended to treat the main claim types.

The authors refer, in § 5.1.4, to distortions from changing knock-for-knock agreements. This is on the assumption that we are seeking to rate the business in line with the claims experience observed with knock-for-knock agreements in force. This will, for example, understate the claim costs properly attributable to business with non-comprehensive cover and high-risk business with comprehensive cover. This approach to rating is fine, so long as all companies adopt the same approach. However, there are now some insurers writing mainly low-risk business who are understood not to have knockfor-knock agreements, and for whom such agreements would be disadvantageous. In order to compete with them, it may be necessary for other insurers to rate on the basis of the relativities which would apply in the absence of knock-for-knock. There is clearly some considerable uncertainty as to what these relativities would be, and this provides one illustration of why we cannot hope to achieve precision in rating, however sophisticated our analysis of the data.

In Section 5, the authors appear rather dismissive of the problem of very large claims. They suggest a capping level of £25,000 for bodily injury claims in a medium-sized portfolio. This may be necessary to increase the stability of the model, but a substantial proportion of claim cost relates to amounts in excess of £25,000. I believe that the underlying risk relativities by age of policyholder, in particular, may be somewhat different if one caps at £25,000 rather than taking the total claim cost. A large volume of data is needed to investigate, but some research has been presented on this topic [Bennett, M.C. & Johnson, P.D. (1984). The treatment of large claims when deciding on a premium structure and on the relationships between the premiums for different groups. *Proceedings of the Four Countries ASTIN Symposium, Akerslool*], and there is scope for some updating of the results.

Section 6 relates to curve fitting, and I am uneasy about some aspects of this. There is a danger that a curve fitted to blocks of ages of policyholder, say, will give an inappropriate result at the extremes, for example at ages 17 and 18. I can imagine circumstances in which curve fitting gave a risk relativity for age 17 which pointed to a reduction in premiums at that age, whereas age 17 policyholders have produced consistently high claim ratios in recent years as a result of the age rating being too low. I suggest that we should not regard curve fitting as a panacea for dealing with grouped data, and this points us once again in the direction of a standardisation approach.

Sections 7 and 8 include a detailed analysis of the allowance to be made for No Claims Discount (NCD). Premium discount has tended to be earned in a true sense at a slower rate by younger and less experienced policyholders than by others. As a result, no simple NCD system can be right for the whole portfolio. If the proportion of young policyholders has changed or will change, as is often the case, the relative profitability of certain sections of the portfolio will also change, added to which the overall profitability of the portfolio will be changed also.

I suggest that we need to acknowledge the existence and significance of the various elements of uncertainty, including those to which I have referred. Their combined effect may well outweigh the statistical advantages of using one form of model rather than another.
Section 9 considers the use of a standard table. An insurer which does not include some process of standardisation as part of its data analysis will be very poorly placed to improve its rating structure. In Section 9.3 the authors very reasonably propose a standardisation approach for analysing by postcode grouping and individual car models. As they remark in §9.3.5, with some understatement, such analysis is not an exact science.

Section 10 considers how, after the modelling work is done, the premium structure might be modified. In Section 10.2 the effect of competition is assessed. The competitive position is expressed in terms of ranking, although, depending partly on the means by which the insurer obtains business, ranking may be an insufficient expression of competitiveness. Some concise measures of differences in, and the general spread of, premiums across the market may be particularly called for. The last sentence of § 10.3.3 indicates that there are circumstances in which, after making necessary changes in premium relativities, maintaining volume of business is more important than maintaining premium income. Observation of claim cost trends and insurers' rating structures suggest that many insurers are still liable to be writing some significant sections of their motor business at such unprofitable rates that they would be best not written at all at these rates, whatever assumptions are made about the spreading of fixed expenses.

The authors refer to the scope for discovering what they describe as profitable 'niches', although I prefer to avoid the term 'niche underwriting' when referring to the continuing process of premium adjustment which the authors describe. In particular, we need to be cautious in believing that we have identified a section of the business where there is scope for substantial profitable growth, since it can often be the case that a large volume of new business that we have encouraged turns out to be of poorer quality than its earlier counterpart.

In the U.K. motor market it is certainly the case, as the authors indicate, that insurers need to model their portfolio and to look at how the premium relativities should be changed. They also need to judge how changes in premiums might best be achieved, given the pattern of competitors' rates, how advantage might be taken if there is to some extent an imperfect market as regards premium relativities, and how a particular insurer might justify a shading of premium relativities to take account of features of that insurer's operations.

However, I suggest that the authors overstate their case in the very last sentence of Section 10 where, in spite of the practical uncertainties that abound in motor insurance rating, they seem to imply that a particular modelling technique applied to an insurer's own data is a necessary and sufficient condition for stealing a march on the competition.

Mr P. S. Carroll: About 10 years ago some actuaries discovered GLIM and one or two were enthusiastic about it. However, it seemed that the people working on motor insurance came to a negative conclusion about it. It is, perhaps, interesting to look back and think why they did so, and why the situation is somewhat different now:

- (1) Computers, 10 years ago, were of limited space and capacity and computational power; GLIM was a large program and the data sets for motor insurance, the policy files, were enormous. So when you put this large program and the large data file on your computer you did not have much space left. Now there is much more computer capacity, so the picture is different.
- (2) The GLIM literature was hard to read, hard to find, and dispersed. In § 1.8, the paper does a useful job in summarising some of the literature. It does not mention *GLIM Newsletters*, which were relatively easy to read, but quite hard to find, and not many people knew about them or were aware that these were worth discovering.
- (3) There was disagreement about the presentation of the results. One leading professor said it was important to interpret the parameters; another that there were so many parameters in these models that it was impossible to interpret them. Some experts said it was important to plot many graphs and look at residuals; others said it was not. The authors have given a better perspective on this.

I have tried to teach this subject, and it is difficult to give a clear, concise, intellectually satisfactory explanation of what these not very obvious formulae and distributions are. Whey are these GLIM functions recommended? Why are these peculiar error functions recommended? I found sufficient

statistics were a useful concept. It is because of the mathematics producing different sufficient statistics in different cases that we have these different GLIM functions and error distributions, as recommended.

I have some minor criticisms of the paper in the area of overfitting. When you have this nice program the temptation is to overfit, to exploit the chance variations in the data, and get a good fit and a nice model. However, the next set of data will not fit nearly as well. The authors are well aware that we are not measuring some of the things that we might measure in motor insurance. We do not know how many miles per year cars are travelling; we do not know how the teenage and young adult children are using the motor car. To try to explain everything with what we do measure in terms of the measured variables is leading to over-fitting. In § B.3 the authors err in that direction when they say, "The deviance is slightly lower for the multiplicative model than for the additive model". In fact, the deviance is much the same, and I support the opener's opinion that considerations of simplicity tend to favour additive models rather than multiplicative ones.

Mr P. D. England: I read the paper as a user of GLIM in other areas of interest to actuaries. The methodology is of practical benefit to the actuarial profession, and has wider application in other areas of general insurance-for example, household insurance. My comments are of a technical nature and may be helpful to the authors and others intending to use the techniques. The plots shown in the paper look highly unsatisfactory, because investigation has revealed that the plots shown use standardised residuals. In contrast to the authors' statement in §2.3.4 that "We have always found these residual plots to be satisfactory", I have never found these residual plots to be satisfactory for anything other than Normal models. The authors quite correctly state that we would expect the distribution of standardised residuals to be skewed. For this reason, standardised residuals are not usually used for non-Normal models, and deviance residuals are used instead, since they give more informative plots. The deviance residual is the signed square root of the contribution which each unit makes to the deviance, and is explained within the first 40 pages of the text by McCullagh & Nelder (1989), referenced in the paper. I believe that if the authors use deviance residuals for both the Poisson model for frequency and the Gamma model for severity, the plots will look far more satisfactory. It is usual to plot deviance residuals against the linear predictor rather than fitted values. Furthermore, a histogram of deviance residuals should show a distribution whose mean and variance are approximately zero and ϕ (using the notation of §2.3.4).

The authors make several references to limitations on computer memory space available. The solution to this problem is to buy more memory and, if need be, a better computer. For a DOS-based computer, the user is still limited by the amount of RAM and the version of GLIM supplied. For a computer running under the UNIX operating system, GLIM is supplied together with the FORTRAN source code and instructions on how to modify the source code to increase the allocated memory space. It is therefore possible to adapt GLIM to fit models using very large data sets.

The authors mention briefly the difficulties of fitting Poisson models with the identity link function (in other words the additive model). The problems can be alleviated to some extent by specifying the same model using the OWN directives, as explained in the GLIM manual. This procedure is more flexible than directly specifying a Poisson error with the identity link. The use of the identity link with a Poisson error structure is uncommon and is intuitively unappealing. I agree with the authors that it is better to use a multiplicative model, purely from a model fitting point of view. The additive model should only be used if there is good empirical evidence for doing so; for example, if it produces a much lower deviance than a model using the logarithmic link, implying a much better fit. It is also possible to use what is called the power link, whereby it is possible to fit a model which is in between an additive and a multiplicative model, which may be, say, on a square root scale of the linear predictor.

Professor J. A. Nelder (a visitor): I hope that the paper will encourage others to look at the possibility of analysis of their own data. I have the following comments:

Residuals

Generalised Linear Models (GLMs) give rise to two definitions of residuals: one is the signed square root of the contribution to the Pearson χ^2 statistic; and the other is the deviance residual which uses

the signed square root of the contribution to the deviance. For Normal errors these definitions are identical, but for other distributions of errors they are not. Pierce & Shafer [Residuals in generalized linear models (1986) J. Am. Statistical Association, 81, 977] have shown that the deviance transformation is close to the optimum Normalising transformation for the GLM distributions, and thus we may expect that deviance residuals will look very much like a set of residuals from a Normal model, even when the error distribution of the model is not Normal. By contrast, Pearson residuals have a skew distribution for non-Normal errors, as is clearly shown in Figure 3. I confidently predict that the skewness would effectively disappear had deviance residuals been used in place of the Pearson kind. If deviance residuals are used, you have access to the Normal and the half-Normal plot for looking at ordered residuals; it is also easier to see if the variance function is correctly chosen if the distribution of residuals is symmetrical. I strongly recommend the use of deviance residuals as a standard technique.

Transformation to additivity

The link function in a GLM defines the scale on which the effects of the terms in the linear predictor are assumed to be additive. For their models with gamma errors, the authors have assumed a log-link function. An alternative, which has some statistical advantage, is to use the inverse scale. This is equivalent to modelling the number of claims that can be covered by a fixed sum rather than analysing the average sum. In the example on motor insurance in our book (McCullagh & Nelder, 1989) we found that the inverse scale gave a slightly better fit than the log. It would be interesting to know whether this is true more generally.

Model checking

Recent advances in model checking have fundamentally changed the way that we model data. Instead of choosing a model class a priori, fitting the model, summarising in terms of parameters and their standard errors and then stopping, we now introduce a loop into the process, whereby, after fitting the current model we check for its internal consistency. If these checks fail, we go back and have further thoughts about the model class, and then try again. As always, the introduction of a feedback loop into a system fundamentally changes the behaviour of that system. No analysis should be accepted today if standard model-checking techniques have not been applied.

Prediction

The calculations of quantities like risk premiums constitute the prediction phase that follows the analysis; and Lane & I in a paper in 1982 (Analysis of covariance and standardization as instances of prediction, *Biometrics*, **38**, 613) showed that many of the techniques of standardisation, the analysis of covariance and calibration can best be thought of as instances of prediction, which we defined as the calculation of derived quantities that answer questions of the what-if kind. A typical question from demography would be 'what would the incidence of the disease in city X be if its population structure were that of the country as a whole?'.

Software

Two established software packages have GLMs built in: GLIM and Genstat. Their capabilities for fitting GLMs are virtually the same. However, Genstat has some facilities that GLIM lacks, which might be useful in the analysis of motor insurance data. These include the manipulation of multi-way tables as a built-in data structure, matrix arithmetic for model extensions, colour bit-mapped graphics for displaying the results pictorially, and a very powerful storage and retrieval system for complex data structures. It also includes the PREDICT directive, which implements the prediction methods which I have described above.

Dr S. M. Coutts: This is a very important time for the motor insurance industry. Underwriting is under pressure, and so is the actuarial profession in so far that it can assist the underwriter. I have worked with GLMs for over 15 years, and find myself in agreement with most of the technical arguments for the use of GLMs. However, there are some significant omissions in the paper on the practical application of the statistical theory and on other matters.

From practical experience, the data base to perform a GLIM analysis is not always available in companies. This is normally because the data bases constructed are transaction based, and do not lend themselves to statistical analysis. Many companies contribute to the ABI motor statistical data base, and this data base is usually sufficient. If companies do not supply information to the ABI it could take up to 6 months to obtain sufficient control data for any statistical analysis.

A significant problem mentioned briefly before is that GLIM is not a user-friendly language. The Royal Statistical Society, through NAG, is trying to address this problem, and certainly the latest version is more friendly. However, the unfriendliness must be considered a drawback. Also, the authors do not address the interface problem of a large data base with GLIM. Professor Nelder has mentioned it, and it is a very important problem. A third alternative to Genstat and GLIM is SAS.

The authors managed to criticise work published by Baxter et al. (1979) and Coutts (1984). They were able to do so because these papers included the data. However, it is impossible to comment on the models or analysis made by the authors, since they did not supply their data. Baxter et al. commented on the fact that the profession does not help itself by hiding behind non-disclosure. I find that such a large data base is unnecessary. It would have been better to publish a smaller one, so that professional criticism of the analysis could have been performed, and I believe that all the conclusions made by the authors could have been made using a small data base. For example, in Section 3.3 the authors claim that the within cell variance in their data set is reasonably close to the model's estimate, and they claim that this is contrary to Coutts (1984). Without access to the data I am not in a position to verify or otherwise their views.

The most startling omission from the paper is the role of the underwriter, who is mentioned as a floating technician; but it is not clear how he interrelates with the actuary. The authors say that their models work in practice, so what is the underwriter's role? In my experience, where the actuary takes over the role of the underwriter the account usually makes a loss. The profession has to come to terms with the different roles of the actuary and the underwriter. My experience is that the actuary is part of a team, with the underwriter as the decision maker. For example, the actuary has to explain the GLIM model as part of explaining technically difficult statistics to non-statisticians. The actuary needs to market himself or herself.

On the marketing side, I do not believe that Section 10.2 is helpful. One-way tables are not the way forward. There is no mention of lapses and new business analysis by multiway factors. The way forward is to use a detailed multiway business plan which brings together marketing and profitability.

Professor S. Haberman: My colleagues and I at City University would like to think that we have pioneered the use of GLIM in solving real actuarial problems. We are currently engaged on a project investigating a statistical approach to motor rating for a very large portfolio—considerably larger than the case studies given in this paper, and it is comforting to find support for the approaches that we are currently adopting, for example, Poisson modelling of the claim frequency and Gamma modelling of claim severity. We are using GLIM in a UNIX context on a powerful SUN microcomputer, and we have thus circumvented the difficulties with memory limitations to which the authors allude.

In Section 4, I would like to hear the authors' views on forecasting the calendar year effect—just as a forward estimate of future inflation would need to be incorporated in a final premium formula.

Section 6 looks at curve fitting, and here I believe the approach is too restrictive. It would be useful to investigate: a transformation of the age scale, for example, prior to curve fitting; and perhaps more promisingly, the use of splines or break-point predictors should be considered, rather than just polynomials, straight lines, quadratics and cubics. I think this latter point would deal with the opener's argument about extreme ages.

I found Section 10 disappointing. I was expecting to see an attempt to present a model of the market itself, along the lines of the recent paper by Daykin & Hey (J.I.A. 117, 173). I suspect that using rankings, as is advocated in Section 10.2, would be too crude a device in reality, since both the company's relative position and its absolute position in the market are important, and the use of ranks and one-way tables throw away too much valuable information.

I was surprised at the literature review, in Section 1, which only goes as far as 1984. There are more

recent papers that have been published in actuarial journals; for instance by Stroinski (*I.M.E.* 8, 35) and by Taylor (*ASTIN Bulletin*, 19, 91), on the subject of motor premium rating.

On Appendix C, I have a technical comment. It is possible to use exactly the negative Binomial distribution in GLIM without resort to approximations by the Poisson (Users' Guide, 112).

Mr N. Shah: I believe that the amount of space devoted to NCD is out of proportion to its importance in rating because:

- For a mature portfolio with protected NCD, I would expect well over 80% of policyholders to be on a maximum rate, and of the rest 50% have not been driving long enough to be on maximum NCD.
- (2) The interrelationship between NCD and excess is an important point which needs to be examined.
- (3) Lapse and new business rates are important to assess the population. There is some evidence to suggest that the policyholder who leaves had an accident.

In Section 6.3 the authors advocate a method for obtaining more detailed results for the underwriter. The method shown is appropriate for interpolating, but in practice, the requirement for detail is at the ends of the factor ranges, and thus we need to extrapolate. Curve fitting methods do not generally help, as it is in these areas that the trend levels change. Therefore in practice, more analysis and discussion with the underwriter is needed.

Apart from a simple example, the paper does not discuss the possibility of parameters not behaving in a logical way, and I would have liked to see some more discussion on the interpretation of such estimates. In practice with small data sets, and in particular for claim cost-models, there are usually some outliers which need to be looked at. This aspect, again, was not fully discussed. There is also much work to be done on choosing the number of cells to use for a model.

It is not clear to me what is achieved by Appendix E. Surely, all that is required is the calculation of the sample variance for each of the cells. The total claim amount, the number of observations and the sum of the squares of the individual claim amounts are sufficient for this.

Mutual independence between different claim types is stated as being obvious in Appendix G.2. To my mind this is an assumption, since in practice there is some relationship between claim types; for example, a bodily injury claim normally occurs in conjunction with accidental damage or third party property damage claims.

Mr H. E. Clarke: The paper describes fully the mathematical approach to the detailed analysis of motor insurance data. I do not wish to discuss the actual statistical analysis, but how you would present the results of any analysis to the underwriter.

If you talk to underwriters, you find that they do not know how it is possible to take all the millions of cells and carry out an analysis that makes sense of them. The paper describes how to carry out that analysis; that is, how to extract the information from the data. The actuary will be part of a team, probably also consisting of the underwriter, claims person and a marketing person. These people will be relying on the actuary to analyse the data and extract as much information as possible from them. They will then be relying on the actuary to present the results of his or her analysis in such a way that the rest of the team can use them in the remainder of their analysis. In our experience, most companies have the relevant data somewhere in their computer systems. The difficulty that a number of them have is summarising them into a form in which they can be analysed, and then analysing them and presenting the results in a way which is intelligible to a non-actuary. I have found that the best way to present the results of analyses is to produce tables for each rating factor showing the following items:

---summary of the data,

-premium relativities calculated from the one-way data without any statistical analysis, and

-premium relativities from the full statistical analysis.

In the calculation of the premiums I would allow for NCD and expenses.

Dr Coutts stated earlier that he thinks that the final part of the paper is incomplete as, in his opinion, a detailed multiway model incorporating lapse and new business assumptions, etc., is required in order to decide on the final level of premiums to be charged. In the current state of development of the U.K. motor market this is oversophisticated. The problem that most underwriters have is getting the structure of the rates right, rather than deciding on the increase to apply once they have achieved this. The paper describes how to get the structure right, with the starting point being the production of a structure that is equally profitable in each cell. Against such a background, the overall level of rates can then be determined without the need to worry about being selected against. Thus, I do not consider that currently detailed models are required to decide on the overall rate increases to be applied.

A number of speakers have commented that the solution to some of the computational problems described in the paper is just to obtain one of the more powerful computers that are now available. This merely reflects the speed at which computers are developing. When the paper was being written the most powerful PCs then available were being used for the analysis.

Professor A. D. Wilkic: I have no direct experience of statistical motor rating, nor have I ever used GLIM. I have, however, had many years' experience in calculating credit scoring systems for banks and other institutions that do personal lending. Credit scoring has many similarities with motor insurance premium rating.

When you apply for a credit card or a bank personal loan or a hire purchase contract, you are probably asked to fill in a form which has quite a lot of straightforward questions in it: name, address, telephone number, etc. Also included are questions like: 'How long have you lived at your address?' 'How many years have you been in your present job?' 'Are you paid weekly or monthly?' 'How many dependent children do you have?' There are also some boxes down the right hand side of the form, which are used by the lender to give points for particular factors: for example, four points for monthly paid; no points for weekly paid. Ten or a dozen factors may be used. The points scored for all the factors are added up, and if the score is high enough, you pass and are granted the loan. If the score is too low, your application will be rejected. The bank may seek additional information for marginal cases. There are plenty of features of credit scoring which make it different from motor rating, but the important point is the similarity. In each case, there are a number of potential scoring factors, and within each factor a number of levels, which in some cases are ordered and in others are not. The response variable may be simpler than in motor insurance: it is usually just a question of whether the loan at some time runs into arrears, perhaps 2 or 3 months. This is equivalent to a claim. The amount of claim could be taken into account, but I do not think that it usually is.

The purpose of motor rating is to obtain a numerical premium. The purpose of credit scoring is simply to classify cases into accept or reject; banks usually do not vary the interest rate in this class of business according to the perceived riskiness of the borrower.

Although the points obtained are just added, the statistical model used is not necessarily an additive one in the authors' sense; it could just as well be multiplicative, where the logarithms of the relevant factors are being added. Indeed, since the response is strictly a Binomial variable, as the loan either defaults in the defined sense or it does not, a logistic transformation is logically the most appropriate, since this transforms the probability of response into something between 0 and 1. Indeed, since it is conventional to give a high score to good features and a low score to bad features, and the points are usually just small integers, the statistical scores have to be transformed by changing their sign and rescaling them, so that the range of the total score is something like 0-100.

Like motor rating, credit scoring is entirely suitable for applying GLIM, or rather it would be if there were not a problem of scale. In the authors' example, they use 4 rating factors, with a total of 21 levels and 2,560 potential rating cells. This provides GLIM with no problems. However, it would not be difficult for a motor rating system or a credit scoring system to use, say, 12 factors, averaging perhaps 10 levels in each, or 120 levels in all, and 10^{12} individual cells. These, of course, would mostly be empty, and so it has to be considered whether the data should be turned round and each loan or each policy be used as an individual cell. In the authors' example, they have 17,000 policies, which again is not too many, but some of the files I have dealt with have over one million loans, and I imagine that the larger motor insurance companies would have over one million policies.

One way of dealing with a large number of policies is to take a sample. All the loans that default, or all the policies on which there is a claim are included, and then an equal-sized sample from the successful loans or non-claiming policies is chosen. This sample can just be chosen randomly. I am not sure about the precise statistical effect of sampling the non-claims. For some models at least, it would give unbiased estimates of the factors, but the standard errors of the parameter estimates would be increased. If the sample of non-claimants is the same size as the population of claimants, then the standard errors of the estimates are increased roughly by the factor $\sqrt{2}$.

Another approach is to model using only first order effects. GLIM wants to know the contents of every possible cell, because it is able to calculate all possible interaction terms. For some purposes this is interesting, but it is not of practical use in our two cases. It is sufficient for a straightforward additive or multiplicative model to use only first order effects. In order to do this for an additive model, all that is needed is the square table showing the number of cases that combine each level of each factor with each other level of each factor, only taking pairs of factors. In my example, with 12 factors and 10 levels each, or 120 levels, this means a square table of 120×120 or 14,400 cells, which is quite manageable. Even in some of my preliminary analyses I would use 30 factors, with about 400 levels in all, which requires a square table with 160,000 cells, again quite manageable.

This approach leads one to the method described by Grimes in 'Claim Frequency Analysis in Motor Insurance' (J.S.S., 19, 147). Grimes's approach allows the constant μ to be the overall mean, rather than the value in the cell with levels 1, 1, 1, etc. The scores found for each cell are then relative to the overall mean. Grimes' includes an extra constraint for each factor to ensure that this is so. Whether Grimes's techniques can be used for a multiplicative model, I am not sure; but it should be possible to use a correspondingly simple method for a multiplicative model.

In some cases interaction between two factors is important, and needs to be brought into the model. Here is a simple example, though perhaps not one that can be readily used in practice. Sex and marital status are two single factors: sex with two levels (male and female) and marital status with two main levels (single and married) and possibly others that usually have quite small numbers in them and often have to be combined, either with single or married. Both for personal loans and, I expect, for motor insurance, there is likely to be a quite strong interaction term: single females have a good experience, single males a bad one; and married people of either sex much the same. This can be dealt with by combining the two factors into a single factor with more levels: single males, single females, married males, married males, married males, but I have seldom found that this is a serious problem.

A problem relevant both to motor insurance rating and to credit scoring is in the sub-division of variables like address and type of car on the one hand, or purpose of loan on the other. The authors mention this in §9.3.2. One approach to rating districts used by at least some lenders is to use one of the commercial systems such as ACORN or MOSAIC, which classify individual post codes into a district type. This is certainly useful for credit scoring; it might be useful for motor rating.

Those who are experts in credit scoring have much to learn from those who know about motor rating, and vice-versa. I recommend *Credit Scoring and Credit Control*, edited by Thomas, Crook & Edelman, published by Clarendon Press, Oxford, in 1992, which gives the proceedings of a conference on credit scoring held in Edinburgh in 1989.

Mr C. G. Lewin: A great deal of work is clearly going into looking at the experience of an individual office. What the office is really interested in is the statistics of the market as a whole, and whether it can carve out niches in that particular market which will be profitable. Looking only at your own experience may tell you a certain amount about whether you are making profits or losses on particular classes of business, but does not tell you very much about what you might be able to do. I was wondering, therefore, what the present state is of the ABI data that Dr Coutts made reference to, and whether it is going to become as useful as the Continuous Mortality Investigation data is for long-term business.

Mr M. J. Brockman: I do not hold much faith in the usefulness of the ABI Motor Risk Statistics Bureau analysis: there are serious weaknesses in the methodology; the data are always at least 2 years out of date when received; and it is not easy to fit in the correct expense loadings, which we advocated

you need to do if you are to get the rating structure right. Perhaps if the companies could get together, and bring the methodology much more up to date, they will find they have a much greater use of the analyses.

If you change your mix of business you do have to keep close control over the expenses, and that is why in the paper we commented that, in some respects, it is more important to make sure that the business volume is maintained rather than premium income, provided that you make sensible allowance for fixed costs.

Miss S. L. Dixon: I would like to pick up the point of heterogeneity of the within-cell coefficient of variation. I have been doing some research on implementing Appendix F, which addresses this question. I am currently in the middle of this research, and so do not have any final results. However, so far, using my data set, I have been unable to obtain a scale paramater close to the theoretical value of 2 mentioned in §F.1.1, the nearest being approximately 3.5. This throws serious doubt on the validity of the log-normal assumption for claim sizes within the cells of this data set; an assumption on which Appendix F relies. Instead, for such data sets, if 'fine-tuning' of the method of the paper is desired—although I am not entirely convinced that the improvement is worth the added complications and computer time—then I would recommend the method suggested by Professor Nelder, of fitting a second linear model for some measure of dispersion. Basically, a model for the mean and a model for dispersion are fitted iteratively, the dispersion model being used to calculate weights for the mean model, and the next dispersion statistic being calculated using fitted values from that mean model. I have found that this refinement produces better residual plots—although the extra fits require substantial computer time, which has been one of my main problems and which may render this approach impractical.

Dr A. E. Renshaw (a visitor): I wish to express much interest in Miss Dixon's comments suggesting the use of joint modelling techniques for mean and dispersion in order to generate weights for use in fitting claims data. I have recently applied these techniques to allow for the effects of duplicate policies in the graduation of mortality data based on policy counts, the results of which appear in *J.I.A.* **119**, 69. The joint modelling technique is readily facilitated in GLIM, through the aid of a user-defined switching macro which allows the modelling process to alternate between the two stages. One particularly interesting feature to be addressed in applying such techniques concerned the selection of the predictor-link for the dispersion GLM. It is suggested that the authors will need to search for relevant patterns in the data to establish these. It should, perhaps, be emphasised that the application of such joint modelling techniques would represent a refinement of the primary mean claims modelling process.

Ms G. Vera (a visitor): Conceptually, one of the objectives of premium rating is to define the underlying risk structure of past claims data. The foundations of such a structure are each and every one of the individual risk factors. Hence, adequate definition and analysis of all rating factors are fundamental to the soundness of the final risk premium structure.

To some extent, the analysis of individual rating factors is outside the scope of GLIM. The assumption of continuity between rating factor relativities is quite reasonable, but the approach for dealing with discontinuities, suggested in the paper, is only adequate for measurable rating factors, such as car or property age, policyholder age, mileage, etc. Parameters such as district, NCD, car group and type of property have features that lend themselves to a fuller analysis if defined differently. NCD is basically a stochastic problem, and the spacial relationship between postal codes and the availability of demographic statistics and/or geographical data bases, provide an excellent basis for full cluster analysis.

Interaction between rating factors is normally significant, particularly between all the incomerelated parameters and sex and age. Hence, the models used to develop the rating scales ought to include interaction terms, to remove the effect of other factors from the scale.

Mr T. G. Clarke (closing the discussion): Unfortunately, in the U.K., the involvement of actuaries in

the rating or premium setting process is not as great as it is for claims reserving, nor for that matter as great as the involvement in premium rating by members of the actuarial profession in the U.S.A. I hope that this paper, together with the discussion, has demonstrated that the profession has much to offer in this area, but is extremely important that we present our capabilities in a clear and practical way, under-pinned by sound statistical theory.

With the general insurance market currently unprofitable, with the trends of frequency and levels of court awards making it difficult to take the appropriate action on the level of premium rates, members of the profession will not have a better opportunity to display their capabilities to those responsible for the underwriting accounts, especially the personal lines of motor and household insurance. I was pleased to hear that Mr England feels that the methodology can be used in household insurance. This is very important, given the differing levels of premium for a category of risk as the opener mentioned, which can lead to an increase in lapse rates and thus expenses.

Some speakers have indicated an omission, in that the paper has concentrated on the use of the data rather than on the quality of the data. For those insurance companies who are members of the ABI Motor Risk Statistics Bureau, the quality of the data is likely to be good, but the experience of that organisation during the early years of development highlights the difficulties of obtaining reliable statistical data on which to carry out the statistical techniques discussed in the paper. Dr Coutts emphasised this point. Actuaries have the appropriate skills in this area to help companies improve the statistical data base as well as perform the analysis. Historically the companies were not good at using the data.

We have had some discussion regarding the types of mathematical models—the opener queried whether it was really necessary to use multiplicative models. He was not convinced by the example. It would appear that there is no real consensus of opinion on which form of model is preferred.

The problems of bodily injury were highlighted by the opener, including the capping of claims. The problem of large claims related to age was raised.

As the opener stated, whichever model is appropriate, it is important that the underwriters or management to whom actuaries are communicating their results understand the limitations of the analysis and the conclusions drawn therefrom. It is unusual for the actuary to be responsible for setting the rates—certainly, if I understood Dr Coutts, he would not allow it. The actuary is an adviser, and the recipient of the advice should understand the limitations if he is to make the appropriate decisions and, in the longer term, have confidence in the actuarial advice, especially when the future experience does not conform exactly with 'expected values'. As Mr Clarke said, the important fact is that the structure of rating is required. The fact that there is much more computer capacity now has certainly been helpful in the development and use of the methods, as Mr Carroll stated.

The subject of whether to include NCD as a parameter or not was raised. It is clear that it is important to know the exact purpose of the investigation, as was well stated in the paper. Mr Shah also mentioned the relationship between excess and NCD. The opener correctly raised the whole subject of knock-for-knock, and the type of analysis which was required.

The sections on calculation of premiums are important, as this is the stage where the results of statistical analysis are put into practice. Given the purpose of the paper, it is understandable that the subject of expenses was only briefly referred to. I believe that this is an important area, and probably requires further analysis and refinement. I hope it will not be too long before we have a further opportunity to discuss this important subject.

Section 10, on the competitive market, is important. It makes the statement that a 'niche' market for one company may not necessarily be a 'niche' for another. This is an important message, and should not be lost on those companies that actuaries are advising, although the opener queried this, especially if the expanded portfolio actually displays a worse experience.

One message should come out of this meeting loud and clear, and that is that actuaries do have something to offer to underwriters in premium rating, especially personal lines. It is important that there is good statistical information, but, from the profession's point of view, we need to explain clearly what we can offer, and then provide practical solutions which the underwriters and management of general insurers fully understand. Mr Clarke and Dr Coutts also made the point. We should be team members. Time and time again we come back to the subject of communication in

every section of actuarial work. For general insurance it is doubly important, because we frequently talk to people who have not been exposed to actuarial thinking before, and understandably are reticent about involving actuaries, based on the perceived image they have of us.

The President (Mr H. H. Scurfield): I am glad that at this, my last sessional meeting as President, there is a paper on general insurance. My earliest involvement as a Fellow with the Institute related to motor insurance rating, and in 1968 I wrote a paper on the subject for the then Students' Society (J.S.S. 18, 207). There has been increasing activity amongst the profession on general insurance, as the numbers of those who have attended the annual General Insurance Study Group Conventions attest: 34 in 1974 increasing to 140 in 1991.

However, I share the authors' disappointment that actuaries have not established themselves more firmly in the market place. In some other countries non-life reserves have to be certified by an actuary—and in Canada they are moving ahead with a scheme to have an Appointed Actuary for each general insurance company—a theme which needs to be followed up here.

I believe that we have been too shy in this country in putting forward the very significant benefits which we can offer general insurance. Of all the professionals involved we alone can offer expertise in all of:

probability, statistics, predictions, analysis of risk, finance, economics, and investment.

When I began to be involved in general insurance there were two omissions and both these have now been corrected:

- (1) General insurance is now one of our examination subjects, and we have a variety of published material available—both theoretical and practical.
- (2) Very importantly, we now have a sufficient supply of experienced actuaries to be able to certify the reserves of all general business companies in this country.

Other professions can offer some of what we can, and collectively they could provide it all, but none can provide the complete service which actuaries can.

I suggest that both the complexities and the speed of change of general business are such that the disciplines which would come from actuarial certification of reserves could be of significant benefit. Furthermore, with the growing length of the tail of the claims reserves and the greater importance of investment, the value of an Appointed Actuary system becomes evident. The value of the disciplines of analysis and the balancing of risks on the two sides of the balance sheet are becoming increasingly apparent.

We, alone of the professions, can bring all these issues together in a single report which would, I suggest, be of great value to the directors of the insurance company and to the supervisor. Such an objective report could bring out very much better information than pages of detailed statistics, such as are contained in the current DTI returns. I hope that the profession will be pressing the case in the months ahead.

As a profession we must, however, remember the need to communicate really well; that has not been one of our strongest points. We must do better. This paper communicates well, and we should congratulate the authors on their thorough explanation—and indeed on bringing the whole subject forward. It moves the profession ahead.

I ask you to show your thanks in the usual way to Mr Brockman and to Mr Wright.

Mr T. S. Wright (replying): Dr Coutts suggested that our approach to the role of the statistical analysis in pricing sought to exclude the underwriter from the process. The closer stressed that the actuary is one part of a team, and should ensure that the limitations of any statistical analysis are

understood by the others involved. We recognise that there will always be other considerations in the pricing than the risk premiums estimated from recent past claims data. Dr Coutts' perception of our position is, therefore, mistaken. The methods we have described allow standard errors of the estimated risk premiums to be calculated. The calculations are based on assumptions which can be checked from the data themselves by using diagnostic techniques such as residual plots. If the assumptions are validated in this way, the standard errors give a realistic and objective assessment of the reliability of the estimated risk premiums. The standard errors reflect both the volume and variability of data used in the analysis. They therefore allow the statistical analysis to be given its due weight in relation to other considerations in the pricing process.

Mr Shah commented that, as a single incident may give rise to more than one type of claim, the estimates for the different claim types are not mutually independent, as stated in Section G.2. I think that this comment is the result of a misunderstanding. We should, perhaps, have used the term 'incident types' instead of 'claim types' throughout Section 5. The incidents are classified into a number of mutually exclusive and exhaustive types, and a separate frequency and severity model is fitted for each type. For example, those incidents giving rise to both an own damage and a third party claim are treated quite separately from those giving rise to an own damage claim only.

Miss Dixon said that she has applied the methods we propose in Appendix F to analyse the variation of claim amounts, and for one data set has found that the scale parameter for the variance of the logs was about 3.5. This casts doubt on the adequacy of the log-Normal assumption for this purpose, so that the alternative approach mentioned in §3.2.2 might be preferable. She has been investigating this alternative, but omitted to mention that it suffers from considerable difficulties in testing which explanatory variables should be included in the model for claim severity. However, it may well prove to be a more appropriate method for some data sets.

On SAS, it is important to recognise the distinction between 'Generalised Linear Models' and 'General Linear Models' which are the special case of Generalised Linear Models with an identity link function. When I last looked at SAS, about a year ago, it was very good on General Linear Models, but to fit a Generalised Linear Model it would have been necessary to write special code in much the same way as if one were using a general programming language such as FORTRAN.

WRITTEN CONTRIBUTION

The authors subsequently wrote: The opener's main thesis was that, in view of many 'practical uncertainties' that exist in motor-rating, it is difficult to justify the use of statistical models as elaborate as those suggested in the paper. First it must be asked: in what sense are the models we suggest 'elaborate'? It is not that they are more difficult to fit. The basic model we use for claim frequency is in fact identical to model C of Bennett (1978), and we show in the paper (Appendix A, and § 2.2.4) how easily this can be fitted using GLIM. Our basic claim severity model is just as easy to fit (see §3.1.4). The only sense in which our models are more elaborate is in their justification (Appendices B to F). The objective function to be minimised (the deviance) is, in each case, derived from verifiable assumptions about the random component of the data. This approach has the additional benefits of allowing the influence of rating factors to be objectively assessed using formal statistical tests, and of allowing standard errors of the final results to be calculated. The assumptions are verified through residual analysis. In some earlier papers, models were fitted by minimising an objective function chosen purely on intuitive grounds. The authors were apparently unaware of the implicit assumptions involved, and so did not carry out any checks of these assumptions. The fact that there may be 'practical uncertainties' which have to be considered in premium rating, does not seem to us to justify a lack of rigour in the statistical modelling.

One of the practical uncertainties to which the opener referred concerned NCD. In his words 'no simple NCD system can be right for the whole portfolio'. We agree with this statement, but our treatment does not depend on the NCD system being right. The likelihood that it is not right is precisely why we treat NCD in the unconventional way that we do. We assume only that either:

- (1) the structure of the system is to remain unchanged, but discounts within the structure are to be updated, or
- (2) both the structure and the discounts are to remain unchanged.

As Ms Vera implied, investigating the correctness or efficiency of NCD systems is a separate problem on which much work could be done; we did not address that problem in the paper. Mr Shah expressed his belief that there would normally be no need to allow for variation in average NCD levels between cells, because most drivers would be on the maximum discount. The model described in Appendix H indicates whether or not there is any significant variation between cells, so renders judgement on the matter unnecessary. We have invariably found that there is significant and material variation between cells.

The opener also drew attention to the problem of large claims, and we accept that this is a difficult problem, about which much more could be said. Simple capping at a constant level could affect the relativities. More refined methods with which we are experimenting are briefly:

- (1) to cap at a constant level, but use a separate frequency model of capped claims to apportion the excess, and
- (2) to make the capping level a multiple of the fitted mean (this involves iterative fitting).

Another of the opener's 'practical uncertainties' concerned the severity of bodily injury claims. He gave two reasons for collecting data over a longer time span than for other claim types:

- (1) that they are sparse and have a high variance, and
- (2) that they tend to take longer to settle, so the total amounts of recent claims will contain large estimated components.

The first we accept, but the second is only a problem if case estimates are biased. If they are unbiased, the estimation will merely increase the magnitude of random variation, and this is allowed for in the calculations of standard errors. It should also be borne in mind that the arguments for using recent data are stronger for bodily injury than for other claim types: changes in levels of compensation awarded by the courts can quickly decrease the relevance of past data.

The debate over whether additive or multiplicative models should be used continues! In reply to the opener's point that the best model may often be some intermediate form between a purely additive and a purely multiplicative model, we would point out that this possibility is accommodated by allowing interaction terms in either an additive or a multiplicative model. We prefer the multiplicative form at present, because we suspect it will usually need fewer interaction terms. Mr Carroll cites simplicity in favour of the additive model. It would be slightly simpler to calculate standard errors of risk premiums from an additive model, but otherwise we cannot see any simplification. We agree with Mr England's comment that the additive model should only be used if there is good empirical evidence supporting it. Mr England also suggested the use of a power link function, whilst Professor Nelder suggested the inverse link function for claim severity models. Although such link functions might improve the fit, this would be unfortunate, because it would complicate both the calculation of standard errors for the fitted values and the interpretation of the models. This is discussed further below.

More generally, Professor Nelder stressed the importance of model checking. The model checking described in the paper is confined to residual plots, and the methods described in Appendices E and F, to check the variance assumptions. On the subject of residual plots, several contributors commented that deviance residuals would be less skewed than the Pearson residuals we have used. We have no doubt that this is true, but think it is unimportant. The purpose of residual plots is to check the nodelling assumptions. Our assumptions concern only the mean and the variance of the data, not higher moments, so only the mean and the variance of the residuals are of interest. Standardised residuals (whether Pearson or deviance residuals) should have zero mean and constant variance, and this is all we are concerned about. So a Normal plot of the deviance residuals (for example), would tell us nothing about the validity or otherwise of our assumptions. [If we were making full distributional assumptions, then of course such techniques would be useful. For example, if we had assumed that claim sizes were Gamma distributed, then the result quoted by Professor Nelder (that the deviance

transformation is close to the optimum Normalising transformation) would be relevant: deviance residuals which were not consistent with a Normal sample would cast doubt on the Gamma assumption.] It is important to note that the use of a 'Gamma variance function' in GLIM, does not imply an assumption that the data are Gamma distributed. GLIM is unfortunately misleading in this respect. Wedderburn (1974) showed that second moment assumptions are sufficient for minimum deviance estimates to be asymptotically unbiased, efficient and Normally distributed. Having said all this, we do accept Professor Nelder's point that it is easier to judge whether a residual plot is homoscedastic if the distribution is symmetrical. However, making judgements by eye about residual plots becomes unnecessary if a second model for the squared residuals is used (see § 3.2.3, and the comments by Miss Dixon and Dr Renshaw). Such a model, for which the *y*-variate can be either the squared deviance residual or the squared Pearson residual, allows the use of formal statistical tests to determine whether or not there is any significant heteroscedasticity. Details are given in Chapter 10 of McCullagh & Nelder.

Other model checking techniques for generalised linear models are described in Chapter 12 of McCullagh & Nelder (1989). One can attempt to check the link function, the variance function, the scale of explanatory variables, and the influence of individual data points (outliers). However, the results often depend on the order in which such checks are carried out, so it is necessary to apply them in many different combinations and permutations before the results can be interpreted with much confidence (to quote from McCullagh & Nelder "model checking remains almost as much art as science"). In our view, it is impractical to attempt thorough checking for every model, bearing in mind that in each application we have two models for each claim type. A more realistic approach is to carry out thorough model checking for a few typical data sets, and then to use the same link and variance functions for others. In a thorough analysis of the claim severity data set from Baxter et al. (1980), McCullagh & Nelder (§ 12.8.3) found the Gamma variance function and the log-link to lie in a 95% confidence region for these two components of the model. However, this is only one data set; more work is needed in this area. We think that the investigation of link functions should be given the highest priority; small deviations from the assumed variance function can be accommodated by making adjustments to the prior weights, using either the method described in Appendix F, or a second model for the squared residuals. Two additional factors must be taken into account in choosing the link function:

- (1) its effect on the calculation of standard errors of the fitted values, and
- (2) the interpretability of the model.

The second factor is often crucial; many underwriters will be reluctant to accept findings from a statistical model which does not have a simple intuitive interpretation. For this reason, we think it is reasonable to use a relatively simple link function which lies in a 95% confidence region around the optimum, rather than the optimum itself. On the question of outliers, we think that the typically large number of data points makes a thorough analysis impractical, and (fortunately!), unnecessary (provided some sort of capping is used for claim severities). The residual plots will draw attention to any dangerously extreme points. High leverage may be a problem if curve-fitting is used when there are a few claims at very extreme values of a rating factor (e.g. policyholder age 80-85): this is dealt with under curve fitting below.

We were pleased with the interest shown in Section 6 on curve fitting. The opener expressed concern, citing the possibility of poor fitted values at the extreme values of rating factors. Mr Shah also drew attention to the danger of extrapolating a fitted curve beyond the range of the available data. We accept that great caution should be exercised with extrapolation, and we prefer to avoid it as far as possible. This is partly why we advocate the use of as many cells as computer memory will permit; by having separate data-points for age ranges 60–65, 65–70, 70–75, 75–80, rather than a single point for the age range 60–80, less extrapolation is necessary to obtain fitted values up to age 80. Note that some extrapolation is necessary whether curves or the more conventional 'steps' are fitted. One simple technique which we use to test for poor fitted values at an extreme (or for high leverage), is to test the significance of a 'dummy variable' (a 'factor' in GLIM terms), which introduces a separate parameter for age 17 only (for example) on top of the fitted curve. Professor Haberman made some other promising suggestions. Partial residual plots should help with transformation of the

measurement scales; see §12.6.4 of McCullagh & Nelder. However, such transformations might again lead to problems with interpreting the models to underwriters.

The criticism of Baxter *et al.* (1979) to which Dr Coutts referred, was taken from § 7.5.1 of his own paper (1984), which states that 'two major errors were made by Baxter *et al.*...'. Baxter *et al.* assumed that the within cell variance of claim severities was the same for all cells (i.e. SERROR N in GLIM), and the two problems referred to by Coutts (1984) were:

- (1) heteroscedasticity: the variance of standardised residuals was an increasing function of the fitted value, and
- (2) the scale parameter estimated as the mean residual deviance was about twice as large as a direct estimate of the within cell variance.

Coutts (1984) did not solve these problems, but said: "some work establishing the error distribution by cell is necessary so that reasonable models can be used". This is what we have done; the two problems mentioned do not exist for the severity model described in our paper. Appendix E deals with the second issue. Our modelling assumption is that the within cell coefficient of variation (not the variance) is the same for all cells. Appendix E describes how this coefficient of variation can be estimated directly for comparison with the scale parameter given by GLIM. Mr Shah suggested that the sample variance for each cell could be used for this purpose. We cannot see how, and in any case, with the small cells which we advocate, the sample variance for an individual cell is usually highly unreliable.

Both Dr Coutts and Professor Haberman expressed disappointment in Section 10 and the opener questioned the use of ranking to assess competitiveness. We accept that there are other measures of competitiveness, and agree that much more useful work could be done on modelling lapses, renewals, and the market as a whole. The analysis of ranks, described in Section 10.2, is useful, because it is easily carried out and easily understood.

Professor Haberman asked for our views on forecasting the calendar year effect. What we have to say here is also relevant in reply to Dr Coutts' query about the role of the underwriter. One reason for considering premium rating to consist of the two aspects described in Section 1.2 is that the statistical analysis of past data deals with the first aspect (the relativities) much more completely than the second (the absolute level of future premiums). This is because there are frequent influences which cause discontinuous or rapid changes in the absolute level, for example: changes in the law (e.g. for seatbelts, MOT standards); in road conditions (e.g. government decision to change the level of maintenance); in traffic volumes (caused by, for example, a change in oil prices, or rail fares); and in claim settlement procedures. Because of these influences, it would be foolish to use the statistical analysis alone to project the absolute level for next year. Underwriting judgement must be combined with the results of the statistical analysis. This could perhaps be formalised using a Bayesian procedure to combine the estimates and standard errors from the statistical analysis with the judgements of the underwriter on the likely effects of the various short-term influences.

On Appendix C, Professor Haberman commented that it is possible to use a negative Binomial variance function exactly in GLIM, rather than approximating with an over-dispersed Poisson structure. This would be achieved by using the \$OWN directive. In Section C.1 we have:

$$Var(r_{ij}) = (1 + f_{ij}/h_{ij}) f_{ij}/x_{ij}.$$

First, it is interesting to note that the negative Binomial distribution could have exactly an overdispersed Poisson error structure: this occurs if f_{ij}/h_{ij} is constant across cells. From equation (*) of Appendix C this is equivalent to having the variance proportional to the mean in the (Gamma) distributions of within-cell risk-proneness. However, this seems less appealing than the possibility that the coefficient of variation of these distributions is approximately constant: this is equivalent to h_{ij} approximately constant. If the values of h_{ij} were known, the **\$OWN** directive could be used to specify the above error structure in GLIM, as Professor Haberman suggested. However, these values are not known a priori, so it would be necessary to proceed iteratively; using the scale parameter and fitted values f_{ij} from one fit to estimate the h_{ij} for use in specifying the error structure in the next fit. The other major violation of the simple Poisson assumption is that the risk intensity for each policyholder is likely to decrease after a claim. It is interesting to examine whether this too could be accommodated more exactly. We argue, in Section C.2, that the effect is approximately to introduce a further factor of the form $(1 - f_{ij}/\alpha_{ij})$ into the variance function. The product of this and the factor $(1 + f_{ij}/h_{ij})$, considered above, is obviously of the same form (to a good approximation), so iterative fitting with the **S**OWN directive could be carried out in the same way. This would considerably complicate the fitting of frequency models, and, we suspect, would not greatly affect the results. It would be useful, however, to examine this empirically.

Mr Carroll expressed concern about the danger of over-fitting. It seems to us that this strikes at the heart of statistical theory and practice. The raison d'être of statistical tests is to prevent over-fitting. In fact, the accepted practice of including explanatory variables only if there is strong evidence (by using 5% or 10% significance levels) causes a tendency towards under-fitting. If one concentrates on checking the assumptions underlying the statistical tests (via residual analyses and so on), then over-fitting will not be a problem.

Mr Shah would have liked some comment on the possibility of 'parameters not behaving in a logical way'. We have not found this to be a serious problem using the models described in the paper. The estimated parameters can always be compared to their standard errors; this usually shows that any unexpected behaviour is insignificant (this is illustrated in Section 2.4). In other cases, discussion with the underwriter sheds light on the matter. If this did become a serious problem we would try to find better models.

Ms Vera drew attention to the problem of definition of the rating factors. Like NCD, this is a whole new can of worms, but we think it can be regarded as quite separate from the main problem addressed in our paper (although Section 9.3 does touch on it). The main purpose of the methods we have described is to find the best premium structure based on the existing rating factors, on the assumption that the definition of these factors will remain unchanged, regardless of whether the definition is good or bad. Having said that, the basic models could be used to compare a number of alternative definitions of a rating factor. For example, one could compare the ACORN categories suggested by Professor Wilkie to the insurers' existing geographical categories, by first fitting a model with both included as explanatory variables, and then excluding each in turn. The F-tests would indicate which of the two alternatives had greater explanatory power. Incidentally, we suspect that the ACORN categories would be more appropriate for household than motor insurance.

We think Ms Vera's comment about the interaction between rating factors is mistaken. The correlation between wealth and age of policyholders affects the exposure (e.g., the combination of young policyholder, new car and high car group, contains relatively little exposure), but the exposure is not what we are attempting to model; the actual values for each cell are known and are used in the modelling.

Professor Wilkie's discussion of credit scoring was interesting. Generalised linear models would probably be as useful there as they have proved to be in so many other areas of actuarial work.