JIA 111 (1984) 87-148

MOTOR INSURANCE RATING, AN ACTUARIAL APPROACH

By S. M. Coutts, M.Sc., Ph.D., F.I.A., F.I.S.

Section 1

Introduction

This paper is largely based on a Ph.D. thesis (Coutts (1983)) and provides an actuarial approach to the technical aspects of motor insurance premium rating, where equal weights are given to both the practical and statistical elements of the problem. The methods described are applicable to the competitive United Kingdom motor insurance premium market. However, it is also contended that in countries where motor rating tariffs are in operation, the analyses proposed are still necessary for management to judge where in the portfolio the business is potentially unprofitable.

Only broad outlines of motor premium rating are available in the Institute of Actuaries literature (e.g. Beard (1964); Scurfield (1968); Johnson and Hey (1971); Benjamin (1977)). Elsewhere there is no shortage of theoretical papers on premium rating models (e.g. Pitkanen (1974)). Bailey and Simon (1960) applied both a multiplicative and additive model to smooth claim ratios and applied their techniques to a set of Canadian motor insurance data. The first paper to analyse claim frequency and claims cost separately, together with expenses was that of Kahane and Levy (1975). The ASTIN Netherlands Group (1982) presented a practical report on the premium motor structure in the Netherlands. The study was commissioned by the large insurance companies in the wake of the collapse of the tariff. The data analysed represented a significant portion of the Netherlands' private car population and must be considered a major paper in reporting the results of a detailed premium analysis. The General Insurance Study Group (GISG), which is a forum in the U.K., for actuaries involved in general insurance, presented four case studies on the practical aspect of premium rating namely: on household insurance, Ajne (1982); motor no-claim discount systems, Christensen (1982); a motor insurance points table, Coutts (1982); reporting to the Insurance Commission in the United States of America, Grady (1982). Details of other important published work will be left to the relevant sections in the paper.

A central theme of this paper will be to emphasize a detailed within-portfolio analysis. The features taken into account will include:

(a) some of the important underwriting rating factors;
(b) statistical analysis of both claim frequency and claims cost;
(c) expenses and inflation.

These features will be combined to arrive at an office premium.

The statistical analysis will be applied to a set of data supplied by an insurance company. In some ways the data supplied does not take into account all the factors which, ideally, would be necessary to perform a full analysis. However, this paper attempts to give a framework for the whole analysis and, therefore, these deficiencies will not invalidate the approach.

By performing this detailed analysis it is believed that:

— the person responsible for the premium decision will be able to restrict his attention to the sensitive areas where judgements have to be exercised;
— as the underlying assumptions are explicitly stated a monitoring process can be set up to establish if these assumptions are reasonable;
— any change in portfolio can explicitly be taken into account;
— this will enable management to establish where in the motor portfolio the business is either profitable or unprofitable.

Section 2

Data breakdown

2.1 *Overview*

The first question to consider is whether the premium rates are to be reviewed and adjusted in an overall fashion, e.g. adding 10% over the whole portfolio, or whether to apply selective increases to different sections of the portfolio. It is argued that with the use of computers, a selective breakdown of important underwriting factors should be undertaken which, by aggregating the results, can automatically give the overall level of premium adjustment needed. If, however, the system of analysis is only geared to the overall review, it is much harder to obtain information about the selective parts of the portfolio.

If the data are subdivided by underwriting rating factors, this will lead to figures which are small both in exposure and claim numbers, so that simple averages will be suspect. Hence, it is suggested that simple statistical modelling be preferred. This has the following advantages:

1. Extension of actuarial principles to small databases so that the portfolios of small insurance companies can and should be analysed. This statement is strongly worded, but it is believed that it can be accomplished since the theory of statistics in the past 10 years has made 'space age' progress in the analysis of small databases in far more critical and sensitive areas than motor insurance, namely, medical and demographic statistics, Little (1978). This formal statistical approach to the problem of small databases can be used to offset uninformed comments such as 'the data are too scanty to support any meaningful analyses'.
2. When the data are analysed in sufficient detail, then the effects of portfolio changes are reduced and judgements on these effects can be made with confidence.

3. A detailed analysis reveals that the process of premium rating involves many different assumptions. Changes in some of these assumptions, for example different bodily injury assumptions (§ 12), can affect the premium rates significantly.
4. The establishment of statistical structures, however simple, provides 'bench marks' that can be used as a basis to monitor actual results as they emerge (§ 14).

2.2 *Advantages and disadvantages of an analysis*

Despite these points, it has been argued by non-actuaries that any actuarial input which might alter rates within the motor portfolio, is practically irrelevant when compared with overall marketing considerations, or in countries where the motor rates are fixed by a tariff. Hence, the statistical process is considered a mere theoretical exercise, the cost of which, given the personnel involved, is hard to justify. In addition, at the 1982 actuarial GISG Conference, a number of delegates supported this view. Their arguments were that an experienced actuary did not need to perform any detailed premium analysis, since the actuary should be aware of the premium situation and adjust the rates accordingly. It was also argued that past analyses should be a sufficient basis for changes in premium, if selected data were collected in order to monitor the process.

It is suggested that all these remarks are half-truths. Certainly, actuarial rates may differ considerably from market rates. However, it is the management's decision and they should be aware that the rates charged may in fact generate potential losses, the size of which should be quantified. If decisions are made without having all the relevant facts available, then this must be considered poor management. A topical example is that in the present market structure the underwriting rating factor 'car age' is generally ignored or given insufficient weight; however, from statistics available, it is evident that newer cars are being undercharged (§ 12.2). On the positive side, a detailed analysis may reveal unsuspected marketing opportunities within the present structure. In addition, in countries where there is fixed premium rating, set by a tariff, this may eventually break down, e.g. in 1982 in the Netherlands. Hence, it is advisable to be able to perform a detailed analysis to cover this eventuality.

As for the argument that the experienced actuary has little need of analysis, it is accepted that judgements could often be made without any in-depth investigation, but how is this to be achieved? The existing body of knowledge has been accumulated largely by trial and error. In fact, however, the actuarial profession ought to strive to establish sound scientific procedures. It is contended that both the present U.K. and international actuarial literature does not give sufficient information for this purpose.

By appealing to past analyses, this tacitly assumes that premium structures are stable over time. It is accepted that frequent changes to the system may be undesirable, but regular analyses are necessary to verify assumptions made in the original calculations. In particular, if the database is small, this will necessitate regular checks to judge whether the past statistical inferences were reasonable. Thus in the U.K., in the late 1970s, many companies reviewed their premium levels quarterly because of the rising rate of inflation. Furthermore, the underlying insurance risk pattern may also vary over time.

Finally, on the financial side, cost and time are put forward as reasons for not performing regular analyses. However, with the advent of microprocessor technology, it is believed that the cost has been cut to a minimum and time reduced to an irrelevant factor.

In summary, a breakdown of the data to take into account rating factors is fundamental to establishing premium bases. To analyse data in this way, modern statistical techniques must be understood and employed regularly since the systems are not necessarily stable. As judgements are required to be made, it is necessary to set up monitoring systems.

SECTION 3

General background

3.1. *Why project?*

Before discussing what background information is required in practice, to make judgements about premium levels, it would be useful to visualize the time span involved in the premium analysis. This is best explained by means of a simple example.

Consider a company that reviews its premium rates on 1 October 1980 and let us assume that these rates are expected to be in force for 1 year, Figure 1. New recommendations would have to be made in practice 3 months in advance. This time lag would be needed to alter computer output and prepare documentation for the rate book to be passed to the broker. The average policy will be effected halfway through the period over which the premium is expected to be in force, i.e. on 1 April 1981. This policy will be on risk for 1 year and, should it have a claim, the claim date will on average be 6 months after date on risk, i.e. 1 October 1981.

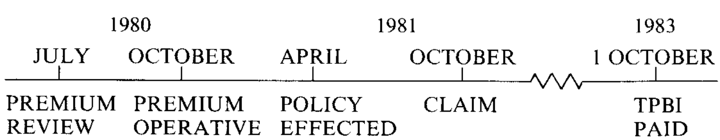

Figure 1. *Projection process*

The material damage costs will be expected to be settled within 3 months from the date of accident. However, the third-party bodily injury costs (TPBI) will take on average 2 years to settle. Ideally a stochastic approach would be preferred, but data were not available on this basis. Prospectively, the average future time span is $3\frac{1}{4}$ years from the date on which the premium decision has to be made, i.e. by 1 July 1980. Hence, data concerning claims and expenses have to be projected to 1 October 1983 and beyond. This will be referred to as an 'averaging process'.

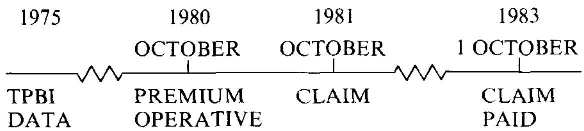

Figure 2. *Total time span to be projected*

Expense data will be based on information which is reasonably up to date at 1 July 1980 and is then projected to 1981. The claim data will be built up from claim numbers, material damage costs and TPBI costs, Figure 2. All but the latter costs will be based on recent data, say 1979/80. However, the most reliable TPBI data are likely to be 5 years old, i.e. claims occurring in 1975. The reason why the average settlement figure of 2 years is not appropriate is because, in practice, the larger and proportionately more important claims take in excess of 5 years from date of accident to settlement. Hence, the total time span to be projected is on average $8\frac{1}{4}$ years. This large time span necessitates a number of subjective decisions, the main ones being:

—Are third-party claims occurring in 1975 relevant in respect of liability claims expected to occur in 1981 and to be settled on average in 1983 or later? The main problem lies in the settlement figures, as court judgements change with changing social conditions. This has been commonly referred to as 'judgement drift'. If the Company has an individual liability claim estimation process (referred to as a manual basis) then it is possible, though not necessarily reliable, to use later liability information based on recent manual estimates as a substitute for settled data.

—Whatever method is employed as the base for projecting claim costs and expenses, a view of past and future inflation has to be taken. In particular, for liability claims, a view of the rate of inflation to bring 1975 values up to 1980 is needed and thereafter a future rate to project these costs into 1983 or later.

—Finally, a view of future levels of claim frequency has to be decided. Factors including future weather conditions, petrol prices and speed restrictions have to be considered. It will be shown later that these aspects may have a relatively large effect on the profitability of results (§ 14.2).

3.2. *The group*

In a Company, it is desirable to make premium decisions within a Group. The decision-maker is the person who ultimately says what the premium rates are going to be; he would usually be a General Manager or the Motor Underwriter.

The rest of the Group would act as advisers to the decision-maker, supplying information on various aspects of the business. The size of the Group might range from one person for each main function, to, in a small company, as few as two, the decision-maker and the underwriter.

The following are the main aspects of the business which have to be considered:

General underwriting principles, which vary from company to company and reflect, e.g. (i) prior views on occupations or socio-economic groups which the marketing should attempt to attract, e.g. teachers, civil servants; (ii) wording in the policy conditions to take into account the introduction of a new rating factor such as protected No Claim Discount.

The overall market position of the Company compared with its competitors, with regard to growth and pricing. Within the Company, production summaries will be available showing lapse and new business figures. The marketing department will also be arguing for sales increases over selected parts of the portfolio, by trying to keep any rate increases to a minimum.

Analyses from the claim personnel, who will report the latest manual estimates on present third-party liability claims.

Statistical analyses will produce information for various members of the Group; concerning in particular, past claims frequency, claims cost and production results.

General economic factors will be used by the Group to project into the future past claim costs and expenses. This will involve, *inter alia*, a view as to the effect of the Government's current economic policy on inflation rates, for salaries and prices.

3.3. *Definition of surplus*

Before the Group can judge the results of the latest premium analysis, it is necessary for them to define profitability. In the narrow sense, the profitability of an identical group of policies will depend on the balance between, on the one hand, premiums receivable and on the other, claims and expenses incurred by that business. In general, premiums and associated expenses and some elements of claims cost will be known very quickly. However, as discussed above, the costs of third-party liability claims may take some years to be accurately assessed. In practice, premium decisions have to be made before the ultimate claims costs are known; hence, the true profitability is not known when premium decisions are made.

At any point of time, before the ultimate claim result is known, an estimate of the claim cost is made and when this is used, the term surplus will be employed instead of profit.

This definition of surplus is too general to be of use to management. It is necessary to sharpen the definition to provide both a specific measure of surplus together with its associated time period. Various definitions of these two latter concepts have been used by companies and herein lies the problem of establishing a universally acceptable monitoring procedure.

The above argument is known to actuaries. It is well documented in life assurance (Fisher and Young (1965)).

The two concepts, surplus and time periods, have to be related to achieve any meaning. Section 3.3.1 discusses surplus and § 3.3.2 the time periods.

3.3.1. *Measurement of surplus*

In this section, several different definitions of surplus are discussed:

(i) The claim or loss ratio is defined as:

$$\frac{\text{Estimated Total Cost of Claims}}{\text{Earned Premium}}$$

where Earned Premium includes commission.

This ratio shows the proportion of earned premium which is used to pay claims cost. Typically, the ratio will lie between ·50 and ·75. The main criticism is that this measure does not take into account expenses. Users say that comparison between companies requires a measure which is independent of expenses. However, for one's own company, the expense ratio is usually fixed and can be notionally added on at the end of the calculation, i.e. plus ·25.

(ii) In the U.K., in recent years, a measure which explicitly takes into account expenses has been used, namely the operating ratio which is defined as:

$$\text{Claim Ratio} + \frac{\text{Premium Related Expenses}}{\text{Written Premium}} + \frac{\text{Other Expenses}}{\text{Earned Premium}}$$

An example of premium-related expenses is commission, and of other expenses, salaries. For a discussion of this ratio, see Scurfield (1968). Typically, the values will lie between ·95 and 1·10. The main criticism is the use of different denominators in the measurement which affect the sensitivity of the ratio, e.g. if written premium changes at a different rate to earned premium. This will happen when either the size of portfolio changes quickly or if there is a sudden increase in premium rates. This has the effect that underwriting surplus (see below) may not imply the same result as the operating ratio.

(iii) Sometimes it is useful to look at the actual surplus rather than at a ratio. Then the underwriting surplus (more usually known as underwriting profit) is defined as:

Earned Premiums—Estimated Claims Cost—Expenses

This measure will give the surplus in absolute terms and is usually used retrospectively. If used prospectively, it is very sensitive to portfolio change and requires some prediction as to the levels of classes of policies.

(iv) The following measure is not quoted in the literature. It compares a weighted average of two sets of alternative premiums, where e.g. one set could be based on the existing premium rates and the other on a new premium basis taking into account projected claim frequency, claims cost and expenses. The weights can either be the present or projected standing business. This measure will be referred to as the 'proposed to existing' basis:

$$\frac{\sum \text{Standing Business} \times \text{Proposed Premiums}}{\sum \text{Standing Business} \times \text{Existing Premiums}}$$

where the summation is over all policies.

As this paper is concerned with premium rates, the final measure 'proposed to existing' will be used to measure the effect of the office premium structure on the existing rates, and to compare the office premium to the fitted 'points table' (§ 11.6.7).

It is noticeable that none of the definitions of surplus take into account, explicitly, investment income. This omission will be briefly discussed later when the input assumptions are detailed. However, it is worth pointing out that, in general, the market ignores investment income for motor premium purposes. However, an implicit measure is used, where the assumed investment income, expressed as a percentage of premium, is added to the operating ratio.

For example, assume that the operating ratio = 1·03 and investment income is 10% of premium income. Then the trading result (underwriting surplus plus investment income) will give approximately a 7% return on premium income and the decision maker then has to decide if this is reasonable.

One reason why explicit account of investment income is not taken into account is the difficulty of deciding what the future return on premiums will be.

3.3.2. *Time periods associated with surplus*

As with the above measures, there are several associated accounting time periods used by companies to assess surplus. The problem is to relate premiums, claims cost and expenses to a defined period of risk.

(a) The revenue account relates to a specific time period and credit is taken for earned premium, against which is set claims and expenses received in that period. The problems which were discussed by Benjamin (1977) are, briefly, that it has no regard for the actual dates to which these values relate, e.g. a claim which occurred in 1976 could have a significant adjustment in 1980 and this adjustment would be in the 1980 revenue account. The advantages with this method are that it gives one figure for surplus quickly; it has been used by the insurance industry as the standard method of showing results and it seems to be reasonably well understood. The main disadvantage from a premium decision point of view is that it mixes up the experience of different premium decisions to a greater degree than the accident year method (see below). This means that any inadequacies of the present premiums will be hidden. The measures of surplus (i)–(iii) above can be used.

(b) The year of accident basis considers all premiums, claims cost and expenses generated by policies at risk in a fixed period. The two main advantages of this method are that these data are required by the Department of Trade returns under the Insurance Companies Regulations (1981), hence data are being collected on an overall level and it shows claims cost development for different accident years.

However, like the revenue year, but to a lesser extent, this method mixes up the claim experiences relating to the different premium bases. The measures of surplus (i)–(iii) above can be used.

(c) The policy year method considers the premium, claims cost and expenses generated from a cohort of policies which have been issued during a fixed period of time. Usually, the group of policies chosen relates to a specific premium basis. This method is satisfactory, since it allows the assumptions in the premium basis to be tested. However, to do this, it will be necessary to isolate the experience of this cohort separately. The main criticism is that the follow-up period would extend to 2 full calendar years and this might be considered too long a time to wait to test premium decisions. All measures of surplus can be used, but (iv) is preferred.

3.3.3. *Comparison of time periods*

These three time periods are very different and they can affect the premium rates in terms of equity from the points of view of the Insured, the Insurance Company and the Supervisory body.

From the Insured's point of view, the policy year method is the most equitable since it directly costs the premium prospectively. Hence, any shortfall in previous years' premium bases will not be brought into the calculation. Interestingly, the Labour Government's Price Commission Regulation (for Insurance Companies) (1977) looked at the prospective cost of a cohort of policies and explicitly stated that any shortfall in previous premium rates could not be taken into account. The price control regulations were repealed in the late 1970s and, since then, there has been no control on how to arrive at premium rates. Hence shortfalls of premium rates can be made up if the market allows. In addition, shortfalls are automatically made up under the year of accident method. Consider the following example.

Say, at 1 July 1980, we are interested in making an underwriting surplus equal to zero in 1981 (equivalent to making the operating ratio = 1·00) by introducing a premium increase on 1 October 1980. To apply the definition of underwriting surplus it is necessary to estimate the claims cost for claims in 1981. This will take into account policies effected between 1 January 1981 and 1 October 1981, but they will be subject to the present (i.e. pre-1 October 1980) rates. Hence, if the present rates are too low, this will affect underwriting surplus in 1981, as the earned premium generated by these policies will also be too low. Therefore, from the Insurers point of view, accident year is not very equitable. However, it is more equitable than the revenue year method.

As far as the Company is concerned, the year of accident is a sensible method. The Company has to charge premiums, not only matching the prospective claims, but also taking into account the most recent past experience. The Company would argue that, since the rates under the policy year method are only estimates and everyone admits that these assumptions can be incorrect, is it so wrong to 'claw back' some previous underestimates in the recent past? (see below.)

Also, if the previous inadequacies cannot be corrected explicitly within the rating structure, the Company will have to consider either raising more capital to finance this shortfall or making more conservative assumptions which would raise premium rates. This would also be in the interest of the Insured who would want the Company to safeguard itself against insolvency. The revenue account basis would be taking this argument to the extreme. It is believed that the year of accident is a reasonable compromise as far as the Company is concerned.

Finally, the supervisory body has to look at the situation from a solvency point of view, that is, it wants to be sure that the Company will pay retrospective and prospective claims (only IBNR) from the present and prospective value of assets.

However, the solvency regulations do not ask for any information concerning present premium bases and rely totally on past information to judge the solvency of that Company.

As a final comment, the above has only discussed the situation when previous premium rates are at best inadequate. Bohman (1979), in an interesting article, discussed methods where surplus is either repaid to the Insured by reducing premiums or accumulated to offset future inadequate rates. This is really a long-term view of rates and is outside the scope of this paper.

As a summary of this Section, it has been decided to consider the time period basis for calculation of premium rates to be related to a cohort of policies all effected during a particular calendar year. However, to obtain data on this basis would require an over-sophisticated database which is not available. Hence, a pragmatic approach has been taken, namely, as a compromise the averaging process outlined in § 3.1 will be applied to data based on year of accident. The measure of surplus will be method (iv) in § 3.3.1 namely a comparison of 'proposed to existing'.

SECTION 4

4.1 *Formulae*

It would be useful at this point to give the premium formula to be applied by combination of risk factors where appropriate.

The risk premium (RP), i.e. premium excluding expenses.

$$\text{RP} = fc$$

where $f = \text{claim proportion} = \dfrac{\text{number of claims}}{\text{exposure}}$

$$\begin{aligned} c &= \text{projected claims cost} \\ &= \text{AD}(1+i_1)t_1 + \text{TPPD}(1+i_2)t_2 + \text{TPBI}(1+i_3)t_3 + \text{M}(1+i_4)t_4 \end{aligned} \qquad (1)$$

where AD = average accident damage cost per claim if settled immediately.
TPPD = average third-party property damage cost per claim if settled immediately.
TPBI = average third-party bodily injury cost per claim if settled immediately.
M = average miscellaneous cost per claim if settled immediately.

and

—$i_1 i_2 i_3 i_4$ are inflation rates for respective types of claim cost.

—t_1, t_2, t_3, t_4 are average settlement periods for respective types of claim cost.

The following comments can be made about equation (1):

(a) An alternative suggested by Benktander (1982) is to replace $(1+i)^t$ by

$$\int_0^\infty e^{\delta u} dG(u) \simeq e^{\delta t}$$

where the inflation rate is e^δ, i.e. $(\delta—1)$ p.a. and $G(u)$ is unspecified with mean t the average time to settlement.

(b) Equation (1) divides claims cost, into AD, TPBI, TPPD and M, and so implicitly assumes that their respective proportion of the intimated claims remain stable.

The calculation of the office premium, OP, depends on how expenses are introduced. There are two main variants:

(i) Only fixed expenses:

$$\mathrm{OP} = \frac{\mathrm{RP}}{1-S} \tag{2}$$

where S is the $\dfrac{\text{total expenses}}{\text{total premium income}}$

i.e. a fixed percentage of ~~premium~~

(ii) This method entails variable plus fixed expense.

$$\mathrm{OP} = \frac{\mathrm{RP}+\left(cc.f+\dfrac{(nb+l)}{t_5}+r+ed\right)(1+j)^{\frac{1}{2}}}{1-w} \tag{3}$$

where
cc = claims cost expenses
f = claim proportion
nb = new business expense
l = lapse expense
t_5 = time period till lapse
r = renewal expense
ed = endorsement expense
w = commission plus other expenses related to premium
j = inflation applicable to expenses

and the costs per unit are inflated to the relevant date of premium. A short discussion about expenses appears in §9.1.

Two omissions will be noticed from the office premium:

4.2 *Contingency loading (or solvency loading)*

It is not certain whether companies explicitly take this factor into account. Although it has been omitted, its algebraic inclusion would be very easy. The problems are of estimating a value and its effect on the final rate. During the 1970s when inflation was high there was a great deal of discussion on how to maintain solvency levels. One method looked at was to include in the premium basis a solvency loading. It is believed that in practice no company did this.

4.3 *Investment income*

It is acknowledged that the market does not explicitly take this factor into account. Implicitly, income is taken into consideration when the whole motor account is scrutinized, in that the ultimate trading results can be compared with the underwriting results. Since equation (1) assumes the average date to settlement, i.e. payment; the adjustment to this equation to take into account investment income of $k\%$ *pa* is

$$\mathrm{AD}\left(\frac{1+i_1}{1+k}\right)^{t_1}+\mathrm{TPPD}\left(\frac{1+i_2}{1+k}\right)^{t_2}+\mathrm{TPBI}\left(\frac{1+i_3}{1+k}\right)^{t_3}+\mathrm{M}\left(\frac{1+i_4}{1+k}\right)^{t_4}$$

These expressions assume that there are no outstanding balances.

SECTION 5

5.1 *Data*

Private car motor insurance data were supplied by an Insurance Company. The Company also arranged the programming for all the grouped data by selected rating factors, showing vehicles exposed to risk, numbers of claims and the various average claims cost.

The list of information contained on the data file to arrive at these results are given in Table 1.

Table 1. *Data contained on computer file*

Exposure data	Claims data
Policy number	Claim number
Cover	Policy number
Policyholder age	Cover
Car age	Policyholder age
Vehicle group	Car age
	Vehicle group
Date of renewal	Date of accident
	Payment of AD
	Payment of TPPD
	Payment of TPBI
	Outstanding amount

The exposure figures were obtained by the census method. Number of claims (including zero claims), AD, TPPD, TPBI and M, payment costs, were all related to the date of accident and the rating factors at that date.

The database for exposure, claim numbers, AD, TPPD and M costs was from 1 April 1979 to 31 March 1980. TPBI costs were for years 1972 to 1977.

It will be assumed that all the existing market underwriting factors will continue in use since it is unlikely that any underwriter would consider altering them without other companies following suit. In order to illustrate the principles under discussion, the rest of the paper will be concerned with a worked example. The following rating factors will be used:

1. Type of cover—Comprehensive or Non-comprehensive
2. Policyholder's age (years)—17–20, 21–24, 25–29, 30–34, 35+

3. Car age (years)—0–3, 4–7, 8+
4. Vehicle group—A, B, C, D

The vehicle group code A represents very small cars e.g. Mini whilst D represents large or sports cars, e.g. BMW or Morgan sports.

Two comments are necessary:

(a) Some significant rating factors have been omitted from this list (e.g. district, no claim discount and use of car) since the data were not available. It is emphasized that these omissions may make the final results quoted in §§ 11 and 12 not totally practical, in so much as the premium rate is not in a suitable form for quoting to a client. However, the main aim of this paper is to present a framework for premium analysis rather than recommend a set of premiums for actual use by the company. Further, as it is straightforward to extend the analysis to include these factors it is considered that these omissions do not invalidate the work.
(b) Following discussion with the underwriter, it may be possible to aggregate some of the detailed data into relevant groupings, in order to make the analysis more manageable. In our example, as an illustration, policy-holder's age 35+ is grouped instead of subdividing into 35–50, 50–60 and 60+. This subdivision is in fact quite popular in the market since it can be used to select retired people, or parents whose children have their own car.

SECTION 6

The general principles of claims analysis

6.1. *Objective*

The principal objective of the analysis is to project past claims data for the relevant period. If the example in § 3.1 is taken, then on 1 July 1980 a premium for 1 October 1980 has to be decided. For convenience all claims cost data will be projected to 15 August 1980. The premium rates will be in force for 1 year and the average data of claims arising will be 1 October 1981. Hence, all claim information has to be projected to 1 October 1981 and onwards. There are two levels of decision to make, namely,

(a) the overall levels of frequency and claims cost, and thereafter
(b) the within-portfolio levels, i.e. the relationship between rating factors.

The first stage will be dealt with by the Group with a minimum statistical analysis, but the ultimate decision requires a great deal of judgement. The second stage is basically where the statistical modelling and actuarial judgement become important.

6.2. *Unbalanced designs*

Since the database of exposure and number of claims is relatively small the subdivision by rating factor produces some cells with a very small number of observed claims, making statistical modelling all the more necessary. The concepts of modelling are fundamental to a proper statistical analysis (Fisher, 1946). In particular, modelling techniques are used to make inference about the structure and reduce variation. Dawid (1980) discussed the advantages of modelling for this type of data.

The main statistical problem is not the size of the database but the estimation of the parameters and the subsequent analysis of variance table for the multiway table which has an unequal number of observations (i.e. exposures and claims) per cell. This is also referred to as an unbalanced design. By inspecting Table 5 (to be discussed in the next section) which shows, for the portfolio, the actual exposure column (1) and claims column (3), the largest numbers occur for comprehensive cover for policyholder's age 35+ and vehicle groups B and C, whilst for non-comprehensive, small numbers occur for newer cars and all policyholder ages.

Bailey and Simon (1960), Seal (1968) and Johnson and Hey (1971), all acknowledged the lack of balance in the design and restricted their analyses to estimation of the parameters (without interaction terms). However, the advantage of extending the analysis to consider the relative importance of rating factors was not discussed in depth. It is believed that part of the reason was the technical statistical difficulties.

Francis (1973) brought to the notice of statisticians that many of the standard computer packages produced inconsistent results for the unbalanced design. The reasons are discussed fully in Aitkin (1978)—but the principal reason is the number of parameters included in the model. For the simple additive model with one observation (or equal numbers of observations) per cell, the sum of squares in the analysis of variance table are partitioned, that is for each parameter the explained variation will not change if an additional parameter is included. This is known as an orthogonal model. The unbalanced design is non-orthogonal and the explained variation in the analysis of variance table will alter if a parameter is included or excluded from the model. Nelder and Wedderburn (1972) developed the general linear model and included an analysis of variance which, because it was generalized, they called an analysis of deviance. The computer package called GLIM (1975) (General Linear Interactive Model) solved all the programming problems. However, to run successfully the GLIM package on the size of data included in this paper a large amount of computer space is required (over 500 K). The limitations and lack of access to a large computer was one of the reasons for the development of Orthogonal Weighted Least Square (Coutts, 1975).

GLIM was introduced to actuaries by Baxter, Coutts and Ross (1979); however, it has not gained any formal recognition. Little's (1978) work is important, since it explains in very simple terms the underlying principles of

GLIM and leaves this reader convinced that is a solution for a large number of problems. Albrecht (1982) reviewed all the gėneral linear model literature and is theoretically a very important paper.

SECTION 7

The analysis of claim frequency and material damage claims cost

7.1. *Introduction*

This section describes the detailed analysis performed on both claims frequency and material damage claims cost. It also contains a critical review of past published work and concludes with the input values for the office premium calculation.

7.2. *Claims frequency: overall levels*

Data for past years and quarters, showing actual claim proportions would be shown to the Group. From experience it is only necessary to decide on the overall levels for both the comprehensive and non-comprehensive sections of the portfolio. Table 2 shows the data available for comprehensive.

Table 2. *Claim frequency overall: comprehensive*

Year of accident	Claim Proportions per 1000 vehicles Quarter of accident				Year
	1	2	3	4	
1977	142	125	115	152	138
1978	140	127	128	144	135
1979	184	130	131	154	150
1980	149	130*			

* includes an IBNR estimate

$$\text{where Claim Proportion} = \frac{\text{Number of claims in quarter}}{\text{Vehicles exposed to risk in quarter}}$$

Discussion on projecting the 1980 results into 1981 would be centred around such items as:

(i) weather conditions,
(ii) petrol prices,
(iii) road repairs,
(iv) general economic conditions, since these might affect the frequency with which policyholders have their cars serviced.

It is assumed that the Group decides to use the 1978 overall levels, i.e. approximately 135 per 1,000 vehicles as the basis for projection for 1981. The data indicate that apart from the first quarter 1979, the years 1978 and 1979 were very similar, and that there is no reason to think that 1981 would be any different.

A similar table could be prepared for non-comprehensive.

Finally the proportions of AD, TPPD, TPBI and zero claims were investigated for both comprehensive and non-comprehensive. From the data, the assumption that these remained reasonably constant over time were accepted.

7.3. *Claims frequency within portfolio*

7.3.1. *Review*

Statistical modelling techniques for claim frequency data are well documented, e.g. Ferrara (1971) and Bennett (1978). Many models have been used where the dependent variable was either the number of claims or claim proportions and the independent variables were the underwriter's rating factors. Interest has been centred on comparing additive and multiplicative models, and the different statistical procedures used to estimate the effect of the rating factor parameters.

Almer (1957) first suggested a multiplicative model, similar to

$$P_{ijk} = SR_iU_jV_k + E_{ijk} \tag{4}$$

where P_{ijk} is the claim proportion for cell ijk

S is overall mean

R_i is the effect of rating factor R at level i

U_j is the effect of rating factor U at level j

V_k is the effect of rating factor V at level k

E_{ijk} is the error term with mean zero and variance σ^2.

Bailey and Simon (1960) investigated claim ratios by comparing both Almer's multiplicative model and an additive model, using an additive model similar to

$$P_{ijk} = S + R_i + U_j + V_k + e_{ijk} \tag{5}$$

where e_{ijk} is the error term with zero mean and variance τ^2. The parameters were estimated by minimum χ^2 statistic.

Then followed a number of papers adapting the Bailey and Simon work on multiplicative models, e.g. Mehring (1964) Jung (1968) and Ajne (1974). It would appear that the multiplicative model became very popular in both the U.S.A. and Europe but not in the U.K. Seal (1968) reintroduced interest in the additive model and used as the dependent variable a standard statistical transformation of ln of the odds (or logit) namely

$$\ln\frac{P_{ijk}}{1-P_{ijk}} = S + R_i + U_j + V_k + E^1_{ijk} \tag{6}$$

The standard reason for the transformation is that $\ln P/(1-P)$ can be shown to be approximately normally distributed and the condition that the estimated P lies between 0 and 1 is always satisfied, Cox (1970). It has been found in practice, that if there are a number of actual claims near to zero, then the fitted estimates for that cell could be negative; hence the advantage of the logit transformation. The fitting criterion was weighted least squares where the weights were

$$\frac{P_{ijk}(1-P_{ijk})}{n_{ijk}}$$

and n_{ijk} are number of vehicles exposed to risk in cell ijk. As the weights depended on the parameter estimates, the solution required an iterative procedure. Seal's work did not seem to make an impact on the U.K. actuarial profession. It seems that general insurance practical problems only came of age after the major paper by Johnson and Hey (1971). They introduced a similar model to Seal (1968), but replaced the weights by simply the number of vehicles exposed to risk, i.e. they assumed that $P(1-P)$ was constant, for cells ijk and ignored the logit transformation. The fitted method was weighted least squares, i.e. minimize

$$\sum_{ijk} n_{ijk}(P_{ijk}-(\hat{S}+\hat{R}_i+\hat{U}_j+\hat{V}_k+\ldots))^2 \tag{7}$$

where is the least square estimate.

The analysis and results were based on a large data set and Johnson and Hey's conclusions were that the main effects explained the underlying structure.

Johnson and Hey (1971) did not give the mathematical analysis, but left it to Grimes (1971). Grimes' paper was difficult to follow, but the underlying theory is explained in Coutts (1975). In the early 1970s, computer time was relatively expensive, hence Bailey and Simon's iterative procedures were not encouraged and the very large matrix inversion potentially required in weighted least squares was not discussed in actuarial literature. Johnson and Hey acknowledged these practical problems and limited their analysis to the main effects.

The computer size restrictions and the problems concerning the analysis of variance table (§ 6.2) prompted Coutts (1975) to develop an approximation to the standard weighted least squares analysis called Orthogonal Weighted Least Squares (OWLS). The approximation depended on the factorization of the weights and was suggested by Please (1974). The method is briefly discussed in Appendix 1. Coutts (1975) showed that weighted least squares and OWLS with a logit transformation gave very similar results on a small data set. The main advantage of OWLS is that no computer inversion is necessary and a very large number of rating factors can easily be analysed and, in particular, the importance of the interaction terms can be investigated. (see § 7.3.2.)

A powerful statistical tool when fitting data to an additive model is the analysis of variance which has been mentioned in § 6. However, any inferences made depend on the assumption of the model being satisfied. Coutts (1975) and Baxter and Coutts (1977) and Baxter, Coutts and Ross (1979) discussed the variation explained by the rating factors, and brought evidence to show that (on a data set supplied by the same Insurance Company that supplied data for this paper) approximately 70% of the variation can be explained by the model of main effects. This was contrary to the result by Johnson and Hey (1971) who said that a very small amount of variation can be explained. The reason for this apparent difference of view lies in the definition of variation being explained. In Baxter *et al.* (1979), the variation being measured is for a group of policies falling into cell *ijk*, whilst in Johnson and Hey (1971) the variation being measured applies to an individual policyholder (Johnson, 1980).

The major criticism of work published in the 1970s was discussed by Baxter *et al.* (1979), i.e.:

> There is a tendency among many authors to advocate a particular model and proceed to estimate the parameters, often making no formal attempt to specify their assumptions concerning the error structure, and make no attempt to justify their choice in a statistical sense.

The paper went on to outline a statistical framework for modelling; namely, state the underlying assumptions, carry out the analysis and then attempt to verify the assumption by examination of the residuals. An analysis was performed on several data sets using GLIM. The paper demonstrated, on relatively small data sets,

(a) that the multiplicative and additive models gave similar results (confirming Bailey and Simon, 1960);
(b) that, for the error distributions, Poisson (Jung, 1968), binomial and normal assumption (Johnson and Hey, 1971) gave similar results;
(c) that OWLS gave similar results to the correct model even if the weights did not factorize. In particular, the approximate analysis of variance table produced by OWLS gave similar results to the correct analysis of variances produced by GLIM. The analysis of variance table showed that approximately 70% of the variation was explained by the main effects;
(d) that the standardized residual plots supported the model with main effects, hence the main effects were used to model claim proportions.

7.3.2. *The analysis*

In the spirit of the Baxter *et al.* (1979) paper and because the data analysed in this paper is very similar, the following model was used to smooth the actual claim proportions

$$\ln\frac{P_{ijk}}{1-P_{ijk}} = T+CO_i+\mathrm{PH}_j+\mathrm{VG}_k+\mathrm{CA}_l+(\mathrm{In})+E_{ijkl} \tag{8}$$

where P_{ijk} = claim proportion for cell *ijk*

T = overall mean

CO_i = rating factor cover (levels comprehensive and non-comprehensive)

PH_j = policyholder's age (17–20, 21–24, 25–29, 30–34, 35+)

VG_k = vehicle group (A, B, C, D)

CA_l = car age (0–3, 4–7, 8+)

(In) = interaction terms

E_{ijkl} = is assumed normal with mean zero and variance σ^2.

The fitting method was OWLS (see Appendix 1).

The data, which was collected from claims occurring from 1 April 1979 to 31 March 80, is Column (1) and (3) of Table 5 which shows for all combinations of rating factors the exposures and actual claims respectively. Table 3 shows the analysis of variance table.

Table 3. *Claim frequency analysis of variance fitted by OWLS: Dependent variable* $\ln P/(1-P)$

Factor	Sum of squares (1)	Degrees of freedom (2)	Mean squares (1)÷(2) (3)	Mean squares ratio (3)÷residual (4)
Cover	179·2	1	179·2	167·5
Car age	205·0	2	102.5	95·8
Vehicle group	187·6	3	62·0	58·4
Policyholder's age	260·3	4	65·0	60·8
Cover × Car age	25·2	2	12·6	11·7
Cover × Vehicle group	5·0	3	1·6	1·5
Cover × Policyholder's age	2·4	4	·6	·5
Car age × Vehicle group	2·2	6	·3	·3
Car age × Policyholder's age	19·3	8	2·4	2·2
Vehicle group × Policyholder's age	11·6	12	·9	·9
Residual	79·1	74	1·1	
Total	977·5	119		

By inspection of the mean square ratio column (4) the most important factor is cover, and the interaction terms are relatively unimportant. Hence the fitted proportions will be modelled using all the main effects. Also notice that for this example 85% of the variation has been explained by the main effects.

Table 4 shows the estimates of parameters.

To obtain an estimate of the proportion of claims for a particular cell *ijk*, with only main effects

$$\ln\frac{\hat{P}}{1-P} = \hat{T}+C\hat{O}+P\hat{H}+V\hat{G}+C\hat{A} \text{ where } \hat{} \text{ is at least squares estimate}$$

$$= \hat{X} \text{ (say)}$$

$$\hat{P} = \frac{e^{\hat{X}}}{1+e}\hat{X}$$

Table 4. *Claim frequency parameter estimates: Additive model (main effects only)*

	overall	=	−2·044
	cover comp	=	·111
	non-comp	=	−·244
Car age	0–3	=	·296
	4–7	=	·044
	8+	=	−·209
Vehicle group	A	=	−.220
	B	=	−·076
	C	=	·017
	D	=	·351
Policyholder's age	17–20	=	·687
	21–24	=	·487
	25–29	=	·180
	30–34	=	·042
	35+	=	−·100

For example, from Table 4 we have for cell definition non-comp., car age 8+, vehicle group D, and policyholder age 35+, the following calculations to estimate P:

Overall value	−2·044
Non-comprehensive	− ·244
Car age 8+	− ·209
Vehicle group D	·351
Policyholder age 35+	− ·100
total	−2·246

$$\therefore \hat{P} = \cdot 095$$

Finally Table 5 shows the actual verses fitted analysis where:

col (1) = Exposed to risk
col (2) = Fitted proportion of claims (main effects)
col (3) = Actual claims
col (4) = Expected number of claims
col (5) = Actual/Expected
col (6) = (Actual–Expected)^2/Expected = χ^2 statistic

and

df = degrees of freedom
= Number of non-empty cells—number of parameters + number of constraints on the parameters.

Table 5. *Claim frequency actual v. fitted analysis dependent variable* $ln\ P/(I-P)$, *OWLS, main effects*

Cover	Veh. age	Veh. gp	P/H age	Expos (1)	Fitted prop'ns (2)	Claims Actl (A) (3)	Exp (E) (1)×(2) (4)	A/E 100 (5)	$(A-E)^2/E$ (6)
Comp	0–3	A	17–20	36	·230	5	8	60	1·29
			21–24	68	·198	18	13	133	1·51
			25–29	123	·155	25	19	131	1·86
			30–34	179	·138	29	25	117	·74
			35+	1,609	·124	230	199	116	4·88
		B	17–20	49	·259	13	13	102	·01
			21–24	146	·225	36	33	110	·30
			25–29	349	·177	63	62	102	·03
			30–34	555	·158	93	88	106	·34
			35+	5,772	·140	825	810	102	·29
		C	17–20	49	·278	7	14	51	3·22
			21–24	176	·242	44	43	103	·04
			25–29	433	·191	79	83	96	·16
			30–34	863	·171	158	147	107	·77
			35+	8,165	·152	1,262	1,240	102	·40
		D	17–20	18	·347	7	6	112	·09
			21–24	119	·309	24	37	65	4·41
			25–29	289	·248	69	72	96	·10
			30–34	576	·223	135	129	105	·32
			35+	4,203	·200	818	840	97	·58
	4–7	A	17–20	47	·188	10	9	113	·15
			21–24	104	·162	23	17	137	2·26
			25–29	174	·125	15	22	69	2·08
			30–34	274	·111	25	30	82	·97
			35+	2,545	·099	243	252	97	·29
		B	17–20	135	·216	30	29	103	·02
			21–24	281	·185	61	52	117	1·58
			25–29	526	·143	82	75	109	·59
			30–34	1,005	·127	126	128	99	·03
			35+	10,117	·113	1,153	1,139	101	·17
		C	17–20	66	·230	19	15	125	·95
			21–24	211	·199	39	42	93	·21
			25–29	505	·155	72	78	92	·50
			30–34	950	·138	115	131	88	1·96
			35+	7,825	·122	1,034	956	108	6·35
		D	17–20	20	·289	9	6	156	1·78
			21–24	108	·257	28	28	101	·00
			25–29	241	·203	49	49	100	·00
			30–34	476	·182	95	87	109	·77
			35+	2,968	·163	515	483	107	2·17
	8+	A	17–20	18	·139	1	3	40	·90
			21–24	40	·124	3	5	61	·77
			25–29	94	·097	11	9	120	·38
			30–34	192	·087	17	17	101	·00
			35+	3,105	·078	240	244	98	·05
		B	17–20	35	·169	6	6	102	·00
			21–24	81	·146	8	12	68	1.25
			25–29	241	·114	24	27	87	·43
			30–34	525	·101	57	53	107	·28
			35+	6,607	·090	534	592	90	5·73
		C	17–20	20	·177	4	4	113	.06
			21–24	40	·154	6	6	97	·01
			25–29	130	·122	17	16	107	·08
			30–34	284	·109	34	31	109	·27
			35+	2,800	·097	262	273	96	·43
		D	17–20	2	·124	0	0	0	·25
			21–24	11	·187	2	2	97	·00
			25–29	51	·160	8	8	98	·00
			30–34	111	·145	21	16	130	1·50
			35+	722	·131	90	94	95	·20

Table 5 (cont.)

Cover	Veh. age	Veh. gp	P/H age	Expos (1)	Fitted prop'ns (2)	Claims Actl (A) (3)	Claims Exp (E) (1)×(2) (4)	A/E 100 (5)	$(A-E)^2/E$ (6)
N/Comp	0–3	A	17–20	8	·139	0	1	0	1·11
			21–24	9	·113	1	1	99	·00
			25–29	6	·052	0	0	0	·31
			30–34	11	·066	0	1	0	·73
			35+	67	·084	5	6	89	·07
		B	17–20	10	·171	1	2	58	·30
			21–24	29	·159	4	5	86	·08
			25–29	26	·117	6	3	196	2·84
			30–34	25	·101	4	3	158	·85
			35+	192	·101	17	19	88	·28
		C	17–20	12	·193	1	2	43	·75
			21–24	30	·174	6	5	115	·12
			25–29	35	·132	5	5	108	·03
			30–34	47	·119	5	6	90	·06
			35+	267	·110	27	29	92	·20
		D	17–20	6	·241	2	1	138	·21
			21–24	21	·227	5	5	105	·01
			25–29	15	·168	4	3	159	·88
			30–34	21	·152	1	3	31	1·51
			35+	93	·145	15	14	111	·16
	4–7	A	17–20	78	·140	11	11	100	·00
			21–24	90	·113	15	11	142	1·84
			25–29	97	·088	6	9	70	·77
			30–34	118	·078	11	9	119	·35
			35+	578	·071	43	41	117	1·22
		B	17–20	184	·162	25	30	84	·77
			21–24	252	·137	41	34	119	1·25
			25–29	315	·104	35	33	107	·14
			30–34	382	·092	37	35	105	·10
			35+	2,207	·082	167	180	93	·94
		C	17–20	114	·174	24	20	121	·88
			21–24	204	·148	27	30	90	·33
			25–29	271	·113	27	31	88	·43
			30–34	306	·100	22	31	72	2·40
			35+	1,585	·089	128	141	91	1·13
		D	17–20	45	·225	12	10	119	·35
			21–24	123	·195	20	24	84	·64
			25–29	125	·150	16	19	85	·41
			30–34	144	·133	15	19	78	·92
			35+	599	·119	59	71	83	2·18
	8+	A	17–20	386	·115	48	44	108	·28
			21–24	365	·096	40	35	114	·69
			25–29	488	·073	39	35	110	·37
			30–34	633	·064	35	40	87	·73
			35+	3,764	·056	223	212	105	·56
		B	17–20	512	·131	77	67	115	1·44
			21–24	630	·110	71	69	102	·04
			25–29	843	·083	81	70	115	1·64
			30–34	1,100	·073	84	81	104	·13
			35+	6,470	·065	507	418	121	19·00
		C	17–20	179	·141	29	25	115	·57
			21–24	272	·119	40	32	124	1·85
			25–29	357	·090	33	32	103	·02
			30–34	478	·080	41	38	108	·24
			35+	2,608	·070	205	183	112	2·54
		D	17–20	45	·182	9	8	110	·08
			21–24	77	·156	13	12	109	·09
			25–29	102	·119	14	12	115	·28
			30–34	145	·106	21	15	137	2·06
			35+	549	·095	55	52	105	·16
Total				1030678		11548	11310		117·11

Test statistics on 109 degrees of freedom (assuming main effects only used) total col (6) unweighted $\chi^2 = 117{\cdot}11$.

The Actual/Expected statistic is given since it is the standard actuarial test, but it is difficult to interpret for this data. It is found that the χ^2 statistic gives a better overall measure. Only one cell stands out as being very peculiar namely, non-comprehensive, vehicle age 8+, vehicle group B, and policyholder 35+. The χ^2 statistic = 117.11 with 109 degrees of freedom, assuming only main effects; this shows a good fit (compare with the Expected Values of a χ^2 statistic E $(\chi^2_{109}) = 109$).

The main effects estimated will be inputted as the frequency data for the premium calculation (Table 5 column (2)).

7.3.3. *Future work*

The main problem which has not been addressed, is, how well does the model predict numbers of claims: (a) overall; (b) within portfolio?

Because of the exogenous factors referred to in § 7.2 it is felt that overall levels cannot be statistically predicted.

The within-portfolio analysis was considered by Coutts (1975) and the additive model with main effects was found to predict reasonably well. However, no other published work has been found.

Three methods suggest themselves to judge forecasting methods:

(i) compare actual claims, with predicted claims, using past data to estimate parameters, and the present exposures;
(ii) compare parameter estimates of rating factors over time;
(iii) formally include in the model a time factor using some form of time series model.

7.4. *Material damage and miscellaneous: overall*

For Accidental Damage (or Fire and Theft for Non-comprehensive), TPPD and Miscellaneous Costs, the object is to obtain the average cost at the mid point of the third quarter 1980 if settled immediately. This value is then projected forward after taking into account the date at which the claim is expected to be notified and the average settlement date. In practice, inflation is assumed to be the main factor for projecting overall levels. This is supported by Ziai (1979). He fitted various probability distributions to several comprehensive AD claims cost data sets. Various techniques were used to estimate the parameters for each of the probability distributions investigated. After estimating the parameters, claim cost distributions were generated by simulation techniques and goodness of fit tests were applied. Next Ziai investigated the prediction ability of these distributions. The parameters of the original data were adjusted for 1 year's inflation and, conditional on the known numbers of claims 1 year later, a simulated predicted claim distribution was obtained. This was compared with the actual claims cost.

For distributions log normal, gamma and the inverse Gaussian, reasonable predicted results were obtained.

Ziai's work was performed on relatively small data sets. Since the same insurance company supplied both Ziai's data and the data analysed in this paper, Ziai's results are considered representative of the data used in this paper. Further, even though Ziai's work only considered comprehensive AD, it seems reasonable to use inflation as the basis for projection for all material damage costs. The actual rates of inflation are discussed in §9.2. Therefore it is assumed that the overall levels of material damage data costs (i.e. AD and TPPD) will reflect the actual average values for the period 1 April 1979–30 March 1980, adjusted for inflation to bring them up to the start of the projection period, i.e. 15 August 1980. For miscellaneous costs, nominal figures of £2.50 for comprehensive and £1.50 for non-comprehensive were assumed.

Finally, some analyses relating to average settlement have been performed and the following average dates to settlement are being used. For both comprehensive and non-comprehensive: AD 3 months; TPPD 6 months and M 3 months.

7.5. *Material damage: within portfolio*

7.5.1. *Review*

Very little work has been published analysing claims cost data sets. During the 1960s when claim frequency models were being developed and discussed, no published papers were found discussing in-depth analysis of claims cost. Bailey and Simon (1960) and Seal (1968) implicitly discussed claims cost since they analysed claim ratios. Why claims cost was ignored is not clear and, some 15 years on, it is only possible to hazard a guess. A plausible reason is that, as the 1960s inflation was less than 6%, it seemed likely that all the variation in claims experience was attributed to claim frequency. Another factor as far as the U.K. was concerned, was that until the end of the 1960s, there was a motor insurance tariff in operation. This was organized by the Accident Office Association (AOA) which collected statistics by one way rating factors, so that no tariff office needed to analyse their own data in depth. When the AOA tariff collapsed and the British Insurance Association (BIA) took over the data collecting through the Motor Risk Statistics Bureau (MRSB) in the earlier 1970s, companies which contributed to the MRSB were forced to set up data bases which allowed an in-depth claims cost analysis to be studied. During the 1970s, inflation became a very significant factor and though the MRSB analysed in-depth claims cost, their results were confidential and were not published. Johnson and Hey (1971) did not analyse claims cost explicitly; however, their model is being used by the MRSB to smooth claims cost data. Kahane and Levy (1975) analysed Israeli claims cost data, using an additive model but their analysis is only briefly explained.

Baxter and Coutts (1977) and Baxter *et al.* (1979) analysed the same set of comprehensive AD data using the model suggested by Johnson and Hey (1971) and the results were compared with OWLS. The dependent variable was the average AD cost per intimated claim, the independent variables were the rating factors, of policyholder age, car age and vehicle group and the weights were the number of intimated claims.

The results using both models were very similar. In addition, Baxter and Coutts (1977) quoted the analysis of variance table, which showed that the main effects explained about 70% of the variation and that the interaction terms could be ignored for fitting purposes. Baxter and Coutts (1977) started to consider the residual analysis and Baxter *et al.* (1979) continued this analysis and confirmed the earlier results and the validity of the underlying assumptions were satisfied, viz:

— the additive model with main effects gave a reasonable fit
— the variance was proportional to the inverse of the number of intimated claims.

However, two major errors were made by Baxter *et al.* (1979), and so both these conclusions have to be considered suspect. The errors were:

(a) If the model and variance assumptions were correct then a residual analysis showing fitted values against standardized residuals would give a plot of points which would look random, Anscombe and Tukey (1963). It was brought to the authors' notice that there was a trend in the residual plot, as the fitted values became larger, so did the standardized residuals. This implied that the underlying assumption of the variance being proportional to the inverse of the number of intimated claims was suspect. This was discussed with Nelder (1980) who suggested using a gamma-error distribution. However, this also failed to give a satisfactory result.
(b) The GLIM package produces an estimate of the residual means square, which has expected value σ^2 (the error variance) for an adequate model. In addition any statement concerning the percentage of variation explained by the analysis of variance depends on the estimate of σ^2 being reasonable. For the main effects model suggested by Baxter and Coutts (1977), the residual mean square was equal to 82,000 units. However, an independent estimate of σ^2 obtained from the data of individual claims, estimated σ^2 to be 43,000 units, approximately half that assumed by the model. Hence it appears that the main effects model is not entirely adequate. It was hoped that by taking logs of the individual claims (i.e. assume a log normal distribution) this might overcome this problem, but this proved unsatisfactory.

The ASTIN Netherland Group (1982) also analysed material damage claims cost and used an additive model, similar to the Johnson and Hey (1971) model. They did not test the underlying assumptions and in particular an independent estimate of residual sum of squares was not calculated. So it is felt that given the preceding comments these results must be suspect.

7.5.2. *The analysis*

The data collected for AD, TPPD was typically unbalanced, with a number of cells giving zero claims cost. Four separate analyses were performed on comprehensive, non-comprehensive AD and TPPD. For comprehensive AD the actual average costs and number of claims are listed in Table 6 columns (1) and (4) respectively.

For the limited objective of smoothing the data, OWLS was used even though it does not provide a close fit. It was considered better to use a smoothing technique rather than either the actual averages or one way tables. In Table 6 column (2) shows the comprehensive AD values, only main effects. Column (3) shows the difference between actual and smoothed values, whilst column (5) shows the count times column (3). By inspecting column (5), it is obvious that the fit is not especially good but, overall, the sum of column (5) = −£3,504. Details are only shown for comprehensive AD, the other material costs being smoothed in the same way.

The smoothed results input are shown in the premium calculation input Table 17. For convenience TPPD value will include the M cost.

7.5.3. *Future work*

The analysis of claims cost is still relatively new. In particular some work establishing the error distributions by cell is necessary so that reasonable models can be used. The GLIM package should be of great assistance for the researcher. Possible lines of research could be to work on using the gamma or inverse Gaussian error distribution.

The analyses so far have explicitly ignored the variable time, which can be interpreted as explaining inflation. Dawid (1981) suggested the following generalized model:

let $\ln \mathrm{AD} = N_t$
$+ Q_i + R_j + S_k$
+ interaction terms involving time
+ interaction terms not involving time
+ error not including time
+ error including time

where AD is the average AD cost

N_t is the main effect of time

and Q_i, R_j, S_k are the underwriting factor effects.

This model is known as a mixed model, where the mixture is between the fixed effects rating factors Q, R, S, (which are not random) and the random effect time N. Also there are two variances to estimate. The statistical problem lies in estimating these variances. Standard methods usually produce one negative variance, which is difficult to explain (Harville 1978).

Table 6. *AD comprehensive smoothed values, actual V. fitted*

Veh. age	Veh. gp	P/H age	Actual (1)	Expected (2)	Act-Exp (3)	Count (4)	Count × (3) (5)
0–3	A	17–20	455·37	416·70	30·67	5·00	193·34
		21–24	320·65	302·95	17·70	19·00	336·26
		25–29	255·86	302·13	−46·27	28·00	−1,295·59
		30–34	232·69	257·86	25·17	28·00	704·63
		35+	282·36	267·06	15·30	239·00	3,657·78
	B	17–20	423·69	421·11	2·58	15·00	38·76
		21–24	312·26	307·36	4·90	39·00	191·21
		25–29	307·88	306·54	1·34	68·00	91·40
		30–34	204·81	262·26	−57·45	96·00	−3,515·23
		35+	269·18	271·46	−2·28	859·00	−1,958·03
	C	17–20	472·01	468·10	3·91	9·00	35·19
		21–24	272·54	354·35	−81·81	18·00	−3,926·92
		25–29	366·93	353·53	— 13·40	83·00	1,112·23
		30–34	286·72	309·25	−22·53	170·00	−3,030·77
		35+	325·59	318·45	7·14	1,320·00	9,419·46
	D	17–20	1,638·30	568·71	1,069·59	9·00	9,626·35
		21–24	687·58	454·96	232·62	27·00	6,200·83
		25–29	463·63	454·14	9·49	78·00	740·50
		30–34	468·53	409·86	58·67	147·00	8,624·53
		35+	394·99	419·06	−24·07	872·00	−20,988·89
4–7	A	17–20	225·98	373·42	−147·44	12·00	−1,769·22
		21–24	175·20	259·67	−84·47	23·00	−1,942·72
		25–29	234·71	258·84	−24·13	19·00	−458·56
		30–34	167·72	214·57	−46·85	27·00	−1,264·93
		35+	237·56	223·77	13·79	266·00	3,668·37
	B	17–20	345·05	377·82	−32·77	38·00	−1,245·26
		21–24	232·19	264·07	−31·88	67·00	−2,136·02
		25–29	247·06	263·25	−16·19	92·00	−1,489·45
		30–34	244·14	218·97	25·17	135·00	3,397·41
		35+	231·62	228·17	3·45	1,254·00	4,321·19
	C	17–20	250·16	424·81	−174·65	21·00	−3,667·73
		21–24	346·63	311·06	35·57	41·00	1,438·18
		25–29	284·90	310·24	−25·34	85·00	−2,154·19
		30–34	327·23	265·97	61·26	129·00	7,902·84
		35+	270·77	275·17	−4·40	1,123·00	−4,938·69
	D	17–20	272·59	525·42	−252·83	9·00	−2,275·47
		21–24	330·07	411·67	−81·60	31·00	−2,529·61
		25–29	515·88	410·85	105·03	53·00	5,566·64
		30–34	298·31	366·57	−68·26	104·00	−7,099·40
		35+	374·87	375·77	−·90	586·00	−529·47
8+	A	17–20	308·37	307·20	1·17	2·00	2·33
		21–24	173·78	193·46	−19·68	6·00	−118·05
		25–29	191·68	192·63	−·95	14·00	−13·36
		30–34	69·06	148·36	−79·30	20·00	−1,505·97
		35+	132·80	157·56	−24·76	263·00	−6,511·55
	B	17–20	118·74	311·61	−192·87	9·00	−1,735·83
		21–24	294·81	197·86	96·95	12·00	1,163·39
		25–29	117·06	197·04	−79·98	25·00	−1,999·48
		30–34	172·46	152·76	19·70	66·00	1,299·96
		35+	181·25	161·96	19·29	619·00	11,938·25
	C	17–20	241·51	358·60	−117·09	4·00	−468·37
		21–24	267·59	244·85	22·74	6·00	136·41
		25–29	371·94	244·03	127·91	20·00	2,558·14
		30–34	246·63	199·76	46·87	38·00	1,781·16
		35+	186·35	208·96	−22·61	304·00	−6,872·63
	D	17–20	107·21	459·21	−352·00	1·00	−352·00
		21–24	284·34	345·46	−61·12	3·00	−183·36
		25–29	247·41	344·64	−97·23	9·00	−875·06
		30–34	414·82	300·36	114·46	27·00	3,090·34
		35+	312·60	309·56	3·04	99·00	300·65
					TOTAL		−3504·06

Interpretation of the estimate is:

(a) N_t is the effect of inflation and can be compared with the inflation used in the past;
(b) the interaction terms including time relate underwriting rating factors to time after eliminating inflation. If these interactions are important then it will be difficult to predict costs using this model.

SECTION 8

Third party bodily injury (TPBI) costs

8.1. *Introduction*

The hardest part of the statistical analysis is that of TPBI costs, since, on average, only about 5–7% of all claims per year involve such costs, yielding for the example portfolio less than 1,000 TPBI claims, but this represents about 20% of the total cost. However, before discussing details, it is necessary to look at the arguments against pooling data from different companies to arrive at an input to premium calculations.

Since the number of claims for each company is, in practice, small, it seems reasonable that all companies should pool their data. That would not, however, solve the problem, for a reason best illustrated by way of an example. Table 7 shows some typical average bodily injury costs per claim for an individual company. They have been adjusted for earnings inflation plus 'judgement drift' (§3.1) to bring them all up to 15 August 1980, the projected date.

Table 7. *Bodily injury costs by year*

Year of accident	Average bodily injury costs per claim inflated up to 15 August 1980 £
1972	73
1973	50
1974	70
1975	35
1976	65
1977	52

If inflation were the only factor operating on these averages, relatively constant values would be expected assuming that inflation and 'judgement drift' assumptions were correct. However, it is clear that there is wide variation. This is due to the extreme skewness of the underlying distribution of bodily injury claims; there is no reason to suppose that the distribution shifts from year to year.

Corresponding pooled data for all companies are available in the U.K. from the MRSB but are not available for publication. However, if they were, then the sample size would be far greater and the variability smaller than for any single company. Suppose an average claim of £58 (in 1980 values) resulted from pooled data for years 1972–77. If the Company based rates on this, a large positive surplus would be shown if the 1975 experience were repeated, or a large negative surplus if the 1972 figures occurred again. The real problem is that the Company experiences a small sample of the total market experience and, therefore, its premium rates should make an allowance for its own variability. How this is to be accomplished in the model is not obvious. Theoretically, a factor for the variance can be included (Kahane and Levy (1975)), but the rates obtained may be too uncompetitive.

Alternatively, it could be included explicitly in the claim-reserving philosophy. Additional reserves could be set up in 'good' years which will be released to supplement the claim experience in bad years. This reserve is often referred to as the equalization or catastrophe reserve, e.g. Trayhorn (1980). In Finland this reserve is allowed to be set up, and can offset positive surplus, so reducing the tax liability. However, in the U.K. it has to be met out of profits after tax, and so is not popular.

8.2. *Overall*

8.2.1 *Base Period*

The Group has to decide the basis of projecting bodily injury claims. Assuming that they have decided to use their company's data, they have to select the base period. The possibilities are:

(i) using the latest results (1978 and 1979), which, however, entail predominantly manual estimates;

(ii) combining claims experience from earlier years, which will reflect 'good' and 'bad' years experience. The results will contain some manual estimates, but these are expected to be realistic and more reliable than those estimates contained in method (i); as older estimates will be based on claims 3 or more years old and the claim assessor is expected to establish reliable liability costs.

There is no correct answer to the problem of prediction and the Group has to arrive at a view, and hold it until evidence is produced to indicate that the view has to be altered. A choice between methods (i) and (ii) has to be made bearing in mind that:

— method (i) contains the latest information, but has inherent subjective estimates.

— method (ii) is more objective than method (i), but it uses data which are 3 or more years old to predict claims cost 4 years into the future.

Due to the uncertainty about the most recent claim estimation procedures, method (ii) was accepted in preference to the more subjective approach.

The following data were collected by year of accident to be used as the base period. It was also decided on statistical evidence that comprehensive and non-comprehensive data were to be shown separately.

(a) The average TPBI claims cost per intimated claim (payments plus outstanding) for years of accident 1972 to 1977.
(b) All averages calculated in (a) are discounted to allow for the time to settlement, i.e. to estimate its value if the claim was settled immediately. Then this value is inflated to the start of the projection period, i.e. 15 August 1980.

The method of adjustment is best explained by an example. For the average TPBI cost for 1972, it is assumed it takes 2 years to average settlement (§ 3.1), therefore, the average TPBI is to be discounted to 1 July 1972. This value is then inflated from 1 July 1972 to the mid-point of 3rd quarter 1980. Since it is assumed that the rate of discount for settlement is the same as the inflated value, it is only necessary to inflate 1972 from 1 July 1974 to 15 August 1980 and 1973 from 1 July 1975, etc.

The inflation rates used are detailed in § 9 and are based on the BIA Economic Advisory Group (EAG) quarterly reports.

8.2.2 *Analysis*

Table 8 summarizes the data collected for comprehensive.

Table 8. *Motor comprehensive TPBI claims cost analysis*

Year of accident	(1) Inflation adjustment to 15 Aug. 80	(2) Actual average TPBI per intimated claim £	(3) Inflation adjusted TPBI col. (1) × (2) £
72	2·587	28·1	72·7
73	2·127	23·3	49·6
74	1·755	40·0	70·2
75	1·548	22·4	34·7
76	1·377	47·2	65·0
77	1·197	43·5	52·1
72–76			58·1
72–77			57·2

As has been pointed out above, the variation by year is due in part to:

(a) the underlying skew TPBI distribution;
(b) the small database;
(c) the use of past inflation rates based on Economic Advisory Group data which may not be appropriate, in particular the effect of 'judgement drift' is very difficult to estimate;

(d) the proportion of TPBI claims is very small, about 5–7%, and a small deviation in this proportion will affect the intimated average;
(e) The effect of large claims on the average (see § 8.3.2).

These problems make selection of the base period difficult. Certainly the use of 1 year is not advisable. If this is rejected the decision of what years to include has to be made.

From Table 8, it seems that the 'good' years are 1973, 1975 and 1977, whilst the bad years are 1972, 1974 and 1976. Combining 'good' and 'bad' years seems a reasonable approach (as with constructing a standard life table). It is suggested that two combinations of years are selected, 1972 to 1976 and 1972 to 1977. The values were calculated by weighting the inflated averages by the number of intimated claims. The respective values being £58.1 and £57.2. For comparison purposes, the actual average TPBI claim cost per intimated claim, based mainly on manual estimates and unadjusted for inflation for 1978 and 1979 at 1 July 1980, were both £47.00. A similar analysis could be done for non-comprehensive business.

To show consistency between comprehensive and non-comprehensive two combination of years are selected, 1972 to 1976 and 1972 to 1977 the respective values being £155.0 and £145.7. For comparison purposes, the actual average TPBI claims cost per intimated claim, based mainly on manual estimates, unadjusted for inflation for 1978 and 1979 at 1 July 1980 were £163 and £136 respectively.

Finally, after some data investigations the evidence suggests that on average the assumption of a 2-year period to settle was reasonable for both comprehensive and non-comprehensive.

8.3. *Within portfolio analysis*

8.3.1. *Review*

The small database, when subdivided even further will make any formal statistical analysis difficult. This is not to say it should not be undertaken. Little (1978) has used generalized linear models to analyse small databases, where the underlying distributions are not necessarily normally distributed.

Papers have been published in the past few years giving details of some analysis of TPBI claims outside the U.K., but all are based on large databases. Chang and Fairley (1980) investigated the relationship of TPBI costs from the State of Massachusetts using the rating factors, type of driver and district within State. The emphasis of this paper was on the difference in fit between the additive and multiplicative models. The ASTIN Netherlands Group (1982), performed an analysis of third-party data by using weighted least squares. The rating factors analysed were weight of vehicle and car type. They came to the conclusion that weight of vehicle was more important.

A general criticism of Chang and Fairley (1980) and ASTIN Netherlands

Group (1982) were that no investigation of the underlying distribution was undertaken and no residual tests comparing fitted against actual were published to support the assumptions of the models. In addition no independent estimate of the residual sum of squares from the data was calculated, §7.5.1, to judge whether the inferences made about the analysis of variance tables were reasonable.

Hallin and Ingenbleek's (1981) work was non-parametric. They analysed third-party costs collected by all the Swedish Insurance Companies. The emphasis of the work was to establish an ordering of several rating factors. Westenberg (1983) on the Swedish data has shown a contrary result, namely that the third-party costs are approximately proportionate to the exposure. This data were also analysed using GLIM (1975) and Shrewsbury (1983), supported Westenberg's results. There was evidence to suggest that the rating factors can be ordered but they do not explain a statistically significant part of the variations, hence inference concerning the ordering is not all that important.

8.3.2. *Large claims*

As the data are broken down into even smaller groups, the effect of large claims becomes very important, as a large claim in a small cell will disproportionately affect the results. Hence it seems reasonable to apply some method of smoothing. In the example, all claims over £10,000 were cut off at that value and the excess was respread over the whole portfolio.

To attempt to judge the effect of this crude truncation on the overall yearly results, Table 9 has been prepared for comprehensive.

Table 9. *Motor: comprehensive TPBI claim cost analysis: large claim adjustment*

	(1)	(2)	(3)
		Average TPBI per Intimated claim inflated	
Year	Number of large claims in excess of £10,000	large claims Unadjusted £	large claims Adjusted £
72	2	72·7	70·0
73	6	49·6	43·8
74	10	70·2	41·2
75	7	34·7	30·0
76	12	65·0	42·0
77	10	52·1	35·4

The effect of removing the claims in excess of £10,000 reduces the variation between years, except for 1972. A more reasonable method is to decide on the excess point for 1972 and inflate all subsequent years by the assumed inflation rate. Another approach would be by taking into account inflation and the underlying distribution, e.g. Ziai (1979). The truncation method is applied by the MRSB on its claim analysis as it is easy to apply.

The problems of the inadequacy of the models and the further problem of large claims lead one to reject any in-depth analysis on small data. Consequently an overall TPBI cost for comprehensive and non-comprehensive should be used in the office premium calculation and the resulting premium structure would reflect claim frequency and average damage costs.

However this defeatist approach was rejected and some attempts to analyse one-way rating factors were undertaken. If it had failed then it was always possible to revert to an overall value but the empirical approach did seem to give sensible results.

All discussion concerning large claims assume no reinsurance. For the insurance company the largest claim made was £100,000 which is well below its retention level. Hence the cost of reinsurance was ignored.

8.3.3. *A practical approach*

The following was the rationale for the proposed analysis.

Even though the overall levels of average TPBI costs vary by year, it would seem reasonable that when comparing several years the relative cost by policyholder's age to the overall value should not fluctuate greatly between years. In addition to help reduce the variation of the relativities, claims in excess of £10,000 were truncated at £10,000.

The relative value obtained would then be multiplied by the overall input value, as detailed in § 8.2.2. The following example in Table 10 for comprehensive, policyholder age 35+ give the details of the calculation.

Table 10. *Motor: comprehensive: TPBI within portfolio example: policyholders age 35+ —relativities*

Year of accident	(1) Claims cost after large claim adjustment No inflation	(2) Overall average after large claim adjustment No inflation	(3) (1)÷(2) Relativity
72	26·2	27·0	·97
73	20·7	20·6	1·01
74	22·4	23·5	·96
75	21·1	19·4	1·09
76	29·5	30·5	·97
77	31·4	29·6	1·06
72–76			1·00
72–77			1·01

The view was taken that the variation between years was reasonable. The input value for 35+ for 72–76 would be $1{\cdot}00 \times 58{\cdot}1 = 58{\cdot}1$ and for 72–77 would be $1{\cdot}01 \times 57{\cdot}2 = 57{\cdot}8$ (from Table 8).

Table 11 summarizes all the relativities:

Table 11. *Motor comprehensive TPBI within portfolio: summary and relativities*

Year of accident	Policyholder's age 17–20	21–24	25–29	30–35	35+
72	1·29	3·06	1·38	·60	·97
73	·24	3·31	·84	·76	1·01
74	1·55	1·99	1·18	·94	·96
75	1·13	·67	·64	1·60	1·09
76	·18	·45	2·04	·82	·97
77	5·59	·98	1·38	·63	1·06
72–76	·92	2·14	1·17	·94	1·00
72–77	1·60	2·03	1·19	·90	1·01

The small number of claims generate a wide variation of relativity by year in the younger ages. Apart from 17–20 which is based on less than 200 intimated claims per year the trend is that the relative costs for TPBI reduces as the policyholder gets older. Table 12 gives the input values, i.e. relativities times values in Table 8.

Table 12. *Input values: TPBI comprehensive*

Base	Age 17–20	21–24	25–29	30–35	35+
72–76	53·63	124·45	67·81	54·90	58·14
72–77	91·50	116·16	68·03	51·63	57·81

A similar analysis could be done for non-comprehensive. This completes the statistical analysis of TPBI costs.

Table 13 summarizes the analysis performed for the claims proportions and claims cost.

Table 13. *Summary of all claim experience analyses*

Type of analysis	Rating factors	Whether modelling technique used	Overall level
Claim proportion	All	Yes	1978
Accidental damage	All	Yes	Inflation
Property damage	All	Yes	Inflation
Miscellaneous	Comp. & non-comp.	No	Inflation
Bodily injury	Comp., non-comp. Policyholder age	Yes	1972–76 1972–77 +Inflation

This ends the claim experience analyses.

SECTION 9

Economic factors

9.1. *Expenses*

It is not proposed to go into detail about the collection of data to obtain a breakdown of expenses. A paper by Rushton (1977) has described the process very well. As far as the Group is concerned, it will obtain this information directly from the accounting department within the Company, which projects future expenses. The only factor it would consider is the future rate of inflation related to expenses (see § 9.2).

As far as the premium formula is concerned, two approaches are available.

The first is to consider all expenses as fixed in the short term—equation (2). The total value of expenses is obtained from the accountants and then expressed as a percentage of premium income. Commission and other premium related expenses are added to this percentage. In practice, the value should be between 25–45%, depending on type of company. The rationale behind this is that in the short term (say 2 years) staff levels (which make up 80% of expenses) are virtually fixed. Thus, it is argued, further sophistication is needless. The fixed expense level was calculated to be 45% of premium for the example under consideration.

The second method, equation (3), is to divide up costs into those which are fixed and expressed as a percentage of premium, and those which are identifiable as separate totals such as claim cost expense, new business expense, lapse expense and endorsement expenses.

For example, the following were used and then inflated to 1980 levels:

Fixed as a percentage of premium (incl. commission)	17%
Claim cost	£14 per claim
New business	£6 per policy
Lapse	£1.50 per policy
Renewal	£1.50 per policy
Endorsement	£2.25 per policy
Time period to lapse	3 years

Both methods will give the same overall expense allocation, if the portfolio is similar to that of the base period. However, if the portfolio changes, the allocations will change. The choice between the two methods depends on the accounting methods of the company. Section 12.2 will show that the different allocations affect the premium structure significantly.

9.2. *Inflation*

The Group will have to form some overall views on inflation. As mentioned earlier, general economic factors and government policies will dominate. Usually, several scenarios will be followed. Different rates of inflation will be applied to different parts of the analysis. In this respect, in the U.K., the B.I.A.

Economic Advisory Group (EAG) is helpful in reporting and projecting inflation separately for each type of claim. Another problem is to establish which index will be an appropriate indicator for each type of inflation risk. For purposes of this example the EAG results were followed.

Table 14 is a summary of the inflation rates and inflation indices used to project costs.

Table 14. *Inflation summary*

	Inflation index	Future inflation: 1980–84 High (%)	Low (%)
Material damage claims	Earnings + material goods	11	7
Bodily injury claims	Earnings + judgement drift	17	13
Expenses	Internal, based on salary increase	10	10

Table 15 shows the assumed previous inflation rates to adjust past claims up to the projection date, 15 August 1980.

Table 15. *Past inflation rates AD, TPPD, M (based on Company data)*

Year	Quarter	Inflation per annum (%)
1979	2	14
1979	3	14
1979	4	17
1980	1	17
1980	2	15

TPBI (Based on EAG figures)

Year	Inflation per annum (%)
1974	17·3
1975	26·1
1976	16·5
1977	10·3
1978	14·6
1979	15·5
1980	18·8

Notice in Table 15, that the past inflation adjustments for material costs are significantly higher than the forecast values. The reason is that the future rates of inflation are expected to fall dramatically in 1980–81 and beyond.

The following method was applied to the TPBI rates in Table 15 to obtain the past inflation adjustments used in Table 8.

Table 17. *Office premium calculation*
Expense ratio 45% TPBI 72–76: Low inflation

Cover	Veh. age	Veh. grp	P/H age	Claim prop'n	AD cost	TPBI cost	(TPPD+M) cost	Office premium
Comp	0–3	A	17–20	·2300	416·70	53·63	27·54	244·83
			21–24	·1980	302·95	124·45	23·82	202·12
			25–29	·1550	302·13	67·81	18·52	131·63
			30–34	·1380	257·86	54·90	23·81	101·36
			35+	·1240	267·06	58·14	31·90	96·80
		B	17–20	·2590	421·11	53·63	30·80	279·94
			21–24	·2250	307·36	124·45	27·08	233·36
			25–29	·1770	306·54	67·81	21·78	153·21
			30–34	·1580	262·26	54·90	27·07	118·64
			35+	·1400	271·46	58·14	35·16	111·57
		C	17–20	·2780	468·10	53·63	26·15	324·30
			21–24	·2420	354·35	124·45	22·43	271·73
			25–29	·1910	353·53	67·81	17·13	181·69
			30–34	·1710	309·25	54·90	22·42	143·05
			35+	·1520	318·45	58·14	30·51	134·16
		D	17–20	·3470	568·71	53·63	40·60	487·54
			21–24	·3090	454·96	124·45	36·88	420·65
			25–29	·2480	454·14	67·81	31·58	295·00
			30–34	·2230	409·86	54·90	36·87	239·73
			35+	·2000	419·06	58·14	44·96	224·23
	4–7	A	17–20	·1880	373·42	53·63	31·40	185·10
			21–24	·1620	259·67	124·45	27·68	152·43
			25–29	·1250	258·84	67·81	22·38	96·18
			30–34	·1110	214·57	54·90	27·67	72·66
			35+	·0990	223·77	58·14	35·76	69·37
		B	17–20	·2160	377·82	53·63	34·66	216·20
			21–24	·1850	264·07	124·45	30·94	177·09
			25–29	·1430	263·25	67·81	25·64	112·35
			30–34	·1270	218·97	54·90	30·93	85·21
			35+	·1130	228·17	58·14	39·02	81·03
		C	17–20	·2300	424·81	53·63	30·01	249·93
			21–24	·1990	311·06	124·45	26·28	207·54
			25–29	·1550	310·24	67·81	20·99	135·06
			30–34	·1380	265·97	54·90	26·28	104·42
			35+	·1220	275·17	58·14	34·37	97·94
		D	17–20	·2890	525·42	53·63	44·46	382·96
			21–24	·2570	411·67	124·45	40·73	329·32
			25–29	·2030	410·85	67·81	35·44	225·30
			30–34	·1820	366·57	54·90	40·73	181·11
			35+	·1630	375·77	58·14	48·81	169·72
	8+	A	17–20	·1390	307·20	53·63	30·85	117·85
			21–24	·1240	193·46	124·45	27·13	99·71
			25–29	·0970	192·63	67·81	21·84	61·36
			30–34	·0870	148·36	54·90	27·13	45·06
			35+	·0780	157·56	58·14	35·21	43·99
		B	17–20	·1690	311·61	53·63	34·11	146·03
			21–24	·1460	197·86	124·45	30·39	119·79
			25–29	·1140	197·04	67·81	25·09	73·98
			30–34	·1010	152·76	54·90	30·38	53·96
			35+	·0900	161·96	58·14	38·47	52·23
		C	17–20	·1770	358·60	53·63	29·46	168·13
			21–24	·1540	244·05	124·45	25·74	139·55
			25–29	·1220	244·03	67·81	20·44	89·62
			30–34	·1090	199·76	54·90	25·73	67·57
			35+	·0970	208·96	58·14	33·82	64·60
		D	17–20	·1240	459·21	53·63	43·91	147·36
			21–24	·1870	345·46	124·45	40·19	214·05
			25–29	·1600	344·64	67·81	34·89	155·70
			30–34	·1450	300·36	54·90	40·18	124·47
			35+	·1310	309·56	58·14	48·27	118·49

Table 17 (*cont.*) *Office premium calculation*
Expense ratio 45% TPBI 72–76: Low inflation

Cover	Veh. age	Veh. grp	P/H age	Claim Prop'n	AD cost	TPBI cost	(TPPD+M) cost	Office premium
Non-comp	0–3	A	17–20	·1390	31·35	245·26	37·35	115·69
			21–24	·1130	39·92	181·13	42·39	77·10
			25–29	·0520	50·30	150·66	22·75	29·87
			30–34	·0660	54·55	106·79	17·35	29·60
			35+	·0840	44·74	142·58	31·22	47·00
		B	17–20	·1710	21·69	245·26	37·85	139·13
			21–24	·1590	30·26	181·13	42·89	105·53
			25–29	·1170	40·64	150·66	23·26	65·02
			30–34	·1010	44·89	106·79	17·86	43·42
			35+	·1010	35·08	142·58	31·73	54·63
		C	17–20	·1930	36·57	245·26	28·12	158·68
			21–24	·1740	45·14	181·13	33·16	116·97
			25–29	·1320	55·52	150·66	13·53	74·48
			30–34	·1190	59·77	106·79	8·12	52·17
			35+	·1100	49·96	142·58	22·00	60·44
		D	17–20	·2410	63·10	245·26	52·95	224·71
			21–24	·2270	71·67	181·13	57·99	177·62
			25–29	·1680	82·05	150·66	38·35	113·31
			30–34	·1520	86·30	106·79	32·95	83·40
			35+	·1450	76·49	142·58	46·82	95·65
	4–7	A	17–20	·1400	35·03	245·26	62·08	125·38
			21–24	·1180	43·61	181·13	67·12	87·98
			25–29	·0880	53·98	150·66	47·48	56·11
			30–34	·0780	58·23	106·79	42·08	39·92
			35+	·0710	48·42	142·58	55·95	44·22
		B	17–20	·1620	23·37	245·26	62·58	142·06
			21–24	·1370	33·95	181·13	67·62	99·59
			25–29	·1040	44·32	150·66	47·99	64·37
			30–34	·0920	48·57	106·79	42·59	45·37
			35+	·0820	38·76	142·58	56·46	49·54
		C	17–20	·1740	40·25	245·26	52·85	154·06
			21–24	·1480	48·82	181·13	57·89	108·85
			25–29	·1130	59·20	150·66	38·26	70·91
			30–34	·1000	63·45	106·79	32·85	50·17
			35+	·0890	53·64	142·58	46·73	54·53
		D	17–20	·2250	66·78	245·26	77·68	224·03
			21–24	·1950	75·35	181·13	82·72	164·92
			25–29	·1500	85·73	150·66	63·08	110·66
			30–34	·1330	89·98	106·79	57·68	81·39
			35+	·1190	80·17	142·58	71·55	86·03
	8+	A	17–20	·1150	15·36	245·26	63·32	98·68
			21–24	·0960	23·94	181·13	68·36	67·98
			25–29	·0730	34·31	150·66	48·72	43·81
			30–34	·0640	38·57	106·79	43·32	30·36
			35+	·0560	28·76	142·58	57·19	32·78
		B	17–20	·1310	5·70	245·26	63·82	109·97
			21–24	·1100	14·28	181·13	68·86	75·85
			25–29	·0830	24·65	150·66	49·23	48·27
			30–34	·0730	28·91	106·79	43·83	33·27
			35+	·0650	19·10	142·58	57·70	36·84
		C	17–20	·1410	20·58	245·26	54·09	119·56
			21–24	·1190	29·16	181·13	59·13	83·07
			25–29	·0900	39·53	150·66	39·50	53·11
			30–34	·0800	43·79	106·79	34·09	37·14
			35+	·0700	33·98	142·58	47·97	40·27
		D	17–20	·1820	47·11	245·26	78·92	174·39
			21–24	·1560	55·68	181·13	83·96	126·09
			25–29	·1190	63·06	150·66	64·32	83·33
			30–34	·1060	70·31	106·79	58·92	60·89
			35+	·0950	60·50	142·58	72·79	65·12

Assume that 1972 TPBI claims occurred on 1 July 1972. This value has to be inflated to the start of the projection period, i.e. 15 August 1980. As equation (1) takes into account the 2-year average date to settlement, the 1972 projected value has to be discounted to allow for this. Therefore the inflation adjustment is from 1 July 1974 to 15 August 1980, that is in terms of inflation:

$$(1{\cdot}173)^{\frac{1}{2}}\,(1{\cdot}261)\,(1{\cdot}165)\,(1{\cdot}103)\,(1{\cdot}146)\,(1{\cdot}155)\,(1{\cdot}188)=2{\cdot}587$$

Table 16 gives the full set of TPBI adjustments.

Table 16. *TPBI inflation adjustments for years*

Year	Adjustment
1972	2·587
1973	2·127
1974	1·755
1975	1·548
1976	1·377
1977	1·197

This ends the economic assumption discussion.

Section 10

The calculation of premium rates

The calculation to arrive at an office premium was performed on a microprocessor by applying equations (2) and (3) and the input data described in §§ 7, 8 and 9.

All the claims cost data were adjusted to bring them up to the inflation levels assumed at the start of the projection period, i.e. 15 August 1980. Table 17 shows for all rating factor combinations, the input data namely:

col (1) Claim proportion
col (2) AD cost
col (3) TPBI cost
col (4) (TPPD + M) cost

These values then have to be adjusted by inflation according to the rates applicable in Table 14 to the premium date 1 October 1981 as per § 6.1.

For example Table 17 shows the office premium column (5) applicable on 1 October 1981 for the following sets of assumptions:

TPBI: 1972–76
Inflation: LOW: AD, TPPD, M 7%; TPBI 13%
Expense: Fixed at 45% of premium.

In § 12, the effects of different sets of assumptions on office premium rates are discussed.

SECTION 11

Presentation of results

11.1. *Introduction*

The traditional way of presenting rates is via a rate book, where every combination of rating factors is defined with its respective rates. In the early 1970s, the Co-operative Insurance Company introduced a 'points table' and several companies are now using similar systems. In the following sections, a method of deriving a premium for a 'points table' is developed. First, a 'points table' is defined. Then follows a discussion as to why a Company would want to determine its premiums in this way. Then the mathematical background is outlined, and an algorithm for deriving a 'points table' is listed. Finally, an example of the algorithm is discussed.

However, it is emphasized that the 'points table' is only an alternative to the rate book and is sometimes used as a selling point. In addition many companies still use the traditional rating book; however, the premiums are based on a 'points table'.

11.2. *What is a points table?*

The workings of a 'points table' is probably best explained by way of an example. Consider Table 18.

Table 18. *Points table*

	Comp	Non-Comp			
Cover	19	0			
Policyholder's	17–20	21–24	25–29	30–34	35+
age	30	12	6	2	0
Car age	0–3	4–7	8+		
	8	4	0		
Vehicle	A	B	C		D
group	0	5	10		20

There are four rating factors each with its associated scale. For each point on the scale there is a value, expressed in points.

The procedure used is to record the points score for each rating factor and then aggregate them. Then use a points conversion table to arrive at the premium to be charged. The conversion table is simply a list of points scores with associated monetary values—see Table 19. For example, consider Comprehensive, Policyholder age 35+, Car age 4–7 and Vehicle group A:

19+0+4+0 pts. = 23 pts: equivalent to £12.35 from Table 19.

Table 19. *Points conversion table*

Points	Premium £	Points	Premium £	Points	Premium £
0	5·90	30	15·45	60	40·25
1	6·10	31	15·95	61	41·60
2	6·30	32	16·45	62	42·95
3	6·50	33	17·00	63	44·35
4	6·70	34	17·55	64	45·75
5	6·95	35	18·10	65	47·25
6	7·15	36	18·70	66	48·80
7	7·40	37	19·30	67	50·40
8	7·65	38	19·95	68	52·00
9	7·90	39	20·55	69	53·70
10	8·15	40	21·25	70	55·45
11	8·40	41	21·95	71	57·25
12	8·70	42	22·65	72	59·10
13	8·95	43	23·40	73	61·05
14	9·25	44	24·15	74	63·00
15	9·55	45	24·95	75	65·05
16	9·85	46	25·75	76	67·20
17	10·20	47	26·55	77	69·35
18	10·50	48	27·45	78	71·60
19	10·85	49	28·35	79	73·95
20	11·20	50	29·25	80	76·35
21	11·55	51	30·20	81	78·85
22	11·95	52	31·20	82	81·40
23	12·35	53	32·20	83	84·05
24	12·75	54	33·25	84	86·75
25	13·15	55	34·30	85	89·60
26	13·60	56	35·45	86	92·50
27	14·00	57	36·60		
28	14·45	58	37·80		
29	14·95	59	39·00		

11.3. *Why use a points table*?

The workings of a 'points table' are mainly practical and are listed below, not necessarily in order of importance:

—It is cheaper to produce than the rate book.

—It is easy to revise rates.

—The calculation is relatively straightforward.
—The method is easy to understand.

There are, however, problems associated with such a system. Some are listed below:

—It may be regarded by brokers and underwriters as over-simplified.
—Fine tuning of the rate structure is no longer possible for marketing purposes. In the traditional rating book, particular rates could be altered without affecting any other rates; in a 'points table' this is not possible.
—Since the 'points table' is easy to interpret, any adjustments to the points which reflect statistical experience, are thought by brokers and underwriters to be errors.
—The resulting premium structure does not necessarily reflect exactly the underlying office premium.

11.4. *Theory*

The problem is to convert the office premium into a 'points table'. Then, it is necessary to compare fitted premium (estimated from our model) with the office premium. That is:

$$(\mathrm{OP}-\widehat{\mathrm{OP}})$$

for each combination of rating factors, where

$$\widehat{\mathrm{OP}} = \text{fitted premium}$$
$$\mathrm{OP} = \text{office premium}$$

Proceed as follows. Let

$$\mathrm{OP}_{ijkl} = (1{\cdot}0325)^{Y} \tag{9}$$

where OP_{ijkl} = office premium associated with a given rating factor
i = level of cover (CO)
j = level of policyholder's age (PH)
k = level of vehicle group (VG)
l = level of car age (CA)

Then define

$$Y = T+\mathrm{CO}_i+\mathrm{PH}_j+\mathrm{VG}_k+\mathrm{CA}_l+(\mathrm{In})+e_{ijkl} \tag{10}$$

where the definitions are similar to § 7.3.2, equation (8).

The choice of the figure 1·0325 is quite arbitrary. In equation (9), take logs to the base 1·0325, whence

$$\log_{1\cdot0325}\mathrm{OP}_{ijkl} = Y_{ijkl}$$

This relationship is fitted by weighted least squares:

$$\min\sum_{ijkl} W_{ijkl}\,(\log \mathrm{OP}_{ijkl}-\hat{Y}_{ijkl})^2 \tag{11}$$

where ^ denotes the least squares estimator and

W_{ijkl} is some set of weights to be selected and CO_i PH_j VG_k and CA_l, etc. are the estimated points for the table.

The method of Johnson and Hey (1971) or GLIM (1975) can be used to arrive at a 'points table' minimizing equation (11). However, for simple assumptions, OWLS can be used.

OWLS relies on the fact that it is possible to factorize the weights in equation (11), that is

$$W_{ijkl} \alpha (co)_i (ph)_j (vg)_k (ca)_l \tag{12}$$

This is demonstrated in Appendix 1.

11.5. *An algorithm*

An algorithm follows for the computation of a 'points table'.

1. The first step is to choose a set of suitable weights. A suitable choice would be standing business, since, as will later be demonstrated, the method is reasonably robust to the choice of weights selected.
2. A standard analysis of variance is performed, using

$$\log W_{ijkl} = \log\,(\text{of standing business})_{ijkl}$$

3. The factors $(\hat{co})_i\,(\hat{ph})_j\,(\hat{vg})_k\,(\hat{ca})_l$ are selected, after checking for interactions from the analysis of variance.
4. The W_{ijkl} is replaced by the product of the estimates of co, ph, vg, ca from the previous step, i.e. then find $\hat{Y}_{ijkl}$ from the expression

$$\min_{ijkl} \sum (\hat{co})_i (\hat{ph})_j (\hat{ca})_k (\hat{vg})_1 \,(\log \mathrm{OP}_{ijkl} - \hat{Y}_{ijkl})^2 \tag{13}$$

5. The weighted analysis of variance of log OP is then checked to ensure that there are no significant interactions.
6. Assuming that no significant interactions appear, then the estimates of the main effects serve as the points within the table.
7. The fitted values for premiums computed through the model are compared with the office premiums, in order to assess the goodness of fit.
8. Then an overall adjustment is made to make sure that the sum of the office premium is equal to the sum of fitted premiums.

11.6. *Results*

11.6.1. *Step 1*

An example will now be presented. Table 25 shows the basic data. The rating factors can be seen on the left. Column (1) shows the Standing Business, which is being used for the weights, i.e. W_{ijkl}.

11.6.2. *Step 2*

Analysis of variance on log W_{ijkl} gives the following results in Table 20:

Table 20. *Analysis of variance of log W*

Factor	Sum of squares	Degrees of freedom	Mean square ratio
Cover	1·5	1	258
Car age	5·9	2	487
Vehicle group	4·6	3	253
Policyholder's age	33·0	4	1,360
Cover × Car age	12·4	2	1,028
Cover × Vehicle group	0·0	3	1
Cover × Policyholder's age	3·9	4	161
Car age × Vehicle group	2·3	6	62
Car age × Policyholder's age	0·2	8	3
Vehicle group × Policyholder's age	0·5	12	7
Residual	0·5	74	
Total	64·8	119	

From this, it can be seen that, in addition to the main effects, there is a significant association between cover and car age. This point will be returned to later (§ 12.2).

11.6.3. *Step 3*

The standing business distribution yields weights for the factors:

Table 21. *Weights for standing business*

Cover		*Car age*	
Comp	4·8	0–3	1·8
Non-comp	2·8	4–7	5·3
		8+	5·3
Policyholder's age		*Vehicle group*	
17–20	·9	A	2·3
21–24	1·9	B	6·1
25–30	2·8	C	5·5
30–34	4·3	D	2·4
35+	32·3		

The values represent the distribution of standing business within the portfolio. For example, there are 32 times as many policies in age 35+ as in age 17–20.

11.6.4. *Step 4*

Using these estimates W_{ijkl}, $\hat{Y}_{ijkl}$ can be estimated from equation (13).

11.6.5. *Step 5*

See Table 22 for the analysis yields of variance on log OP_{ijkl}.

Table 22. *Analysis of variance log OP*

Factor	Sum of squares 1,000 (1)	Degrees of freedom (2)	Mean squares 1,000 (1)÷(2) (3)	Mean square ratio (3)÷Residual (4)
Cover	4,351	1	4,351	7,515
Car age	3,178	2	1,589	2,745
Vehicle group	3,931	3	1,310	2,262
Policyholder's age	2,746	4	687	1,185
Cover × Car age	226	2	113	195
Cover × Vehicle Group	81	3	27	47
Cover × Policyholder's age	43	4	11	18
Car age × Vehicle group	5	6	1·0	2
Car age × Policyholder's age	5	8	0·7	1
Vehicle group × Policyholder's age	21	12	2·7	3
Residual	43	74	·6	
Total	14,632	119		

It is important to note that this analysis does not allow us to make an statement about the goodness of fit of the model, since there is not a independent estimate of the residual variance. All that the analysis shows is that relatively, the main effects are far more important than the interaction terms which can therefore be neglected.

11.6.6. *Step 6*

From these results, we estimate the main effects as follows:

Table 23. *Estimates of points for office premium*

Weighted mean	134·1	*Car age*	
Cover		0–3	10·9
Comp.	6·3	4–7	4·0
Non-comp.	−10·7	8+	−7·7
Policyholder's age		*Vehicle Group*	
17–20	27·6	A	−8·9
21–24	20·6	B	−4·2
25–29	6·8	C	0·8
30–34	−2·5	D	17·0
35+	−2·2		

It would be easier to make comparisons if there were no negative values. This is simply accomplished, by transforming the smallest value in each line to zero, making the same addition to the other values in the line, and adjusting the overall mean subsequently.

For example, with Car age, we can alter the value for '8+' to zero by adding 7.7, whence 18·6 for '0–3' and 11·7 for '4–7'. The overall mean is later reduced by 7.7, as in Table 24.

Table 24. *Altering the base to facilitate comparison*

	Unadjusted	*Adjusted*
Overall	134·1	104·3
Cover		
Comp.	6·3	17·0
Non-comp.	−10·7	0
Policyholder's age		
17–20	27·6	30·1
21–24	20·6	23·1
25–29	6·8	9·3
30–34	−2·5	·0
35+	−2·2	·3
Car age		
0–3	10·9	18·6
4–7	4·0	11·7
8+	−7·7	·0
Vehicle group		
A	−8·9	·0
B	−4·2	4·7
C	0·8	9·7
D	17·0	25·9

11.6.7. *Step 7*

The goodness of fit of the new premium structure can now be examined. There are three possible tests:

1. The fitted premiums can be compared with the office premiums for each rating factor (i.e. Expected/Actual).
2. Using 'proposed to existing' method, i.e. comparing premium incomes.
3. The expected premium income from the fitted premiums can also be compared with that of the present premium structure, using the 'proposed to existing' method.

Column (1) of Table 25 shows the standing business as projected to 1 October 1980, and column (2) shows the equivalent points applicable to the fitted premium (column (3)).

Column (6) of Table 25 shows the relationship between the fitted premiums and the office premiums, expressed as a percentage. It can be seen that, generally,

Table 25. *Office premium results comparison of actual v. fitted expense ratio 45% TPBI 72–76 low inflation*

Cover	Veh. age	Veh. grp	P/H age	Standing business (1)	Points (2)	Fitted premium (3)	Office premium (4)	Diff (3)–(4) (5)	Ratio (3)/(4) (6)	SB diff (1)×(5) (7)
Comp	0–3	A	17–20	32	170	230·10	244·83	−14·73	93	−471·43
			21–24	62	163	184·18	202·12	−17·94	91	−1,112·22
			25–29	103	149	118·50	131·63	−13·13	90	−1,351·97
			30–34	152	140	87·92	101·36	−13·44	86	−2,043·48
			35+	1,446	140	88·61	96·80	−8·19	91	−11,836·68
		B	17–20	52	175	267·48	279·94	−12·46	95	−648·10
			21–24	146	168	214·10	233·36	−19·26	91	−2,812·65
			25–29	287	154	137·75	153·21	−15·46	89	−4,435·84
			30–34	472	145	102·20	118·64	−16·44	86	−7,757·51
			35+	5,063	145	103·00	111·57	−8·57	92	−43,390·91
		C	17–20	60	180	313·55	324·30	−10·75	96	−644·92
			21–24	188	173	250·97	271·73	−20·76	92	−3,902·07
			25–29	444	159	161·48	181·69	−20·21	88	−8,974·85
			30–34	820	150	119·81	143·05	−23·25	83	−19,061·82
			35+	8,142	150	120·75	134·16	−13·42	89	−109,238·11
		D	17–20	21	196	527·02	487·54	39·48	108	829·05
			21–24	153	189	421·84	420·65	1·19	100	182·19
			25–29	291	175	271·42	295·06	−23·64	91	−6,879·41
			30–34	630	166	201·37	239·73	−38·36	83	−24,167·99
			35+	4,498	166	202·95	224·23	−21·27	90	−95,686·62
	4–7	A	17–20	35	163	184·74	185·10	−·36	99	−12·62
			21–24	93	156	147·87	152·43	−4·55	97	−423·38
			25–29	126	142	95·14	96·16	−1·02	98	−128·44
			30–34	181	133	70·59	72·66	−2·07	97	−375·10
			35+	2,081	133	71·14	69·37	1·77	102	3,691·94
		B	17–20	4	168	214·75	216·20	−1·45	99	−136·29
			21–24	229	161	171·89	177·09	−5·20	97	−1,189·99
			25–29	308	147	110·60	112·35	−1·75	98	−679·81
			30–34	744	138	82·06	85·21	−3·16	96	−2,347·75
			35+	7,911	138	82·70	81·03	1·67	102	13,241·29
		C	17–20	65	173	251·75	249·93	1·82	100	118·26
			21–24	199	166	201·50	207·54	−6·04	97	−1,201·41
			25–29	437	152	129·65	135·06	−5·41	95	−2,364·95
			30–34	801	143	96·19	104·42	−8·23	92	−6,592·59
			35+	6,877	143	96·95	97·94	−·99	98	−6,818·66
		D	17–20	21	189	423·14	382·96	40·18	110	843·80
			21–24	113	182	338·69	329·32	9·37	102	1,059·02
			25–29	215	168	217·92	225·30	−7·38	96	−1,586·78
			30–34	427	159	161·68	181·11	−19·44	89	−8,298·80
			35+	2,818	159	162·95	169·72	−6·77	96	−19,070·35
	8+	A	17–20	17	151	126·94	117·85	9·10	107	154·65
			21–24	39	144	101·61	99·71	1·90	101	73·91
			25–29	76	131	65·38	61·36	4·02	106	305·30
			30–34	170	121	48·50	45·06	3·45	107	586·35
			35+	2,826	122	48·89	43·99	4·90	111	13,840·15
		B	17–20	43	156	147·57	146·05	1·52	101	65·25
			21–24	108	149	118·12	119·79	−1·68	98	−181·13
			25–29	219	135	76·00	73·98	2·02	102	442·37
			30–34	487	126	56·38	53·96	2·43	104	1,182·98
			35+	6,992	126	56·83	52·23	4·60	108	32,161·91
		C	17–20	27	161	172·99	168·13	4·85	102	131·07
			21–24	60	154	138·46	139·55	−1·09	99	−65·54
			25–29	158	140	89·09	89·62	−·54	99	−84·61
			30–34	310	131	66·10	67·57	−1·48	97	−457·35
			35+	3,346	131	66·62	64·60	2·01	103	6,731·67
		D	17–20	3	177	290·76	147·36	143·40	197	430·20
			21–24	19	170	232·73	214·05	18·67	108	354·82
			25–29	61	157	149·74	155·70	−5·95	96	−363·29
			30–34	126	147	111·10	124·47	−13·37	89	−1,684·53
			35+	966	148	111·97	118·49	−6·52	94	−6,296·00

Where col. (1) = standing business at 1 October 80. (2) = points applicable to fitted premium. (3) = fitted premium. (4) = office premium.

Table 25 *(cont). Office premium results comparison of actual v. fitted G.B. private car premium analysis expense ratio 45% TPBI 72–76 low inflation*

Cover	Veh. age	Veh. grp	P/H age	Standing business (1)	Points (2)	Fitted premium (3)	Office premium (4)	Diff (3)–(4) (5)	Ratio (3)/(4) (6)	SB diff (1)×(5) (7)
Non-comp	0–3	A	17–20	6	153	133·42	115·69	17·74	115	106·42
			21–24	6	146	106·80	77·10	29·69	138	178·15
			25–29	6	132	68·71	29·87	38·85	230	233·08
			30–34	12	123	50·98	29·60	21·38	172	256·52
			35+	52	123	51·38	47·00	4·38	109	227·77
		B	17–20	15	158	155·10	139·13	15·97	111	239·48
			21–24	29	151	124·14	105·53	18·62	117	539·90
			25–29	30	137	79·88	65·02	14·85	122	445·64
			30–34	23	128	59·26	43·42	15·04	136	364·31
			35+	181	128	59·73	54·63	5·10	109	922·34
		C	17–20	17	163	181·82	158·68	23·14	114	393·30
			21–24	34	156	145·53	116·97	28·56	124	971·10
			25–29	37	142	93·64	74·48	19·15	125	708·62
			30–34	42	133	69·47	52·17	17·30	133	726·45
			35+	291	133	70·02	60·44	9·58	115	2,787·07
		D	17–20	12	179	305·60	224·71	80·89	135	970·62
			21–24	25	172	244·61	177·62	66·99	137	1,674·68
			25–29	23	158	157·38	113·31	44·07	138	1,013·60
			30–34	22	149	116·77	83·40	33·37	140	734·09
			35+	121	149	117·68	95·65	22·03	123	2,666·02
	4–7	A	17–20	85	146	107·12	125·38	–18·25	85	–1,551·31
			21–24	78	139	85·75	87·98	–2·24	97	–174·47
			25–29	77	125	55·17	56·11	–·94	98	–72,39
			30–34	83	116	40·93	39·92	1·01	102	83·94
			35+	482	116	41·25	44·22	–2·96	93	–1,428·81
		B	17–20	147	151	124·53	142·06	–17·53	87	–2,576·89
			21–24	194	144	99·67	99·59	·08	100	15·53
			25–29	211	130	64·13	64·37	–·24	99	–51·29
			30–34	293	121	47·58	45·37	2·21	104	647·12
			35+	1,723	121	47·95	49·54	–1·59	96	–2,733·44
		C	17–20	140	156	145·98	154·06	–8·09	94	–1,132·11
			21–24	220	149	116·84	108·85	7·99	107	1,758·53
			25–29	253	135	75·18	70·91	4·27	106	1,080·07
			30–34	300	126	55·78	50·17	5·61	111	1,682·52
			35+	1,636	126	56·22	54·53	1·69	103	2,757·69
		D	17–20	67	172	245·36	224·03	21·34	109	1,429·50
			21–24	122	165	196·39	164·92	31·48	119	3,840.27
			25–29	130	151	126·36	110·66	15·70	114	2,041.08
			30–34	174	142	93·75	81·39	12·36	115	2,151·39
			35+	695	142	94·49	86·03	8·46	109	5,880·24
	8+	A	17–20	346	134	73·61	98·68	–25·07	74	–8,673·84
			21–24	303	127	58·92	67·98	–9·06	86	–2,746·13
			25–29	392	114	37·91	43·81	–5·90	86	–2,313.40
			30–34	504	104	28·13	30·36	–2·23	92	–1,123·44
			35+	3,217	105	28·35	32·78	–4·43	86	–14,258·56
		B	17–20	528	139	85·57	109·97	–24·40	77	–12,882·09
			21–24	609	132	68·49	75·85	–7·36	90	–4,479·21
			25–29	774	118	44·07	48·27	–4·20	91	–3,250·09
			30–34	989	109	32·69	33·27	–·57	98	–567·51
			35+	6,036	109	32·95	36·84	–3·88	89	–23,449·80
		C	17–20	209	144	100·31	119·56	–19·26	83	–4,024·43
			21–24	284	137	80·29	83·07	–2·78	96	–788·94
			25–29	346	123	51·66	53·11	–1·45	97	–500·57
			30–34	558	114	38·33	37·14	1·19	103	662·43
			35+	2,835	114	38·63	40·27	–1·64	95	–4,648·54
		D	17–20	56	160	168·60	174·39	–5·80	96	–324·57
			21–24	136	153	134·95	126·09	8·86	107	1,203·14
			25–29	151	140	86·83	83·33	3·50	104	527·77
			30–34	167	130	64·42	60·89	3·53	105	508·94
			35+	858	130	64·93	65·12	–·19	99	–164·40

Where col. (1) = standing business at 1 October 1980. (2) = points applicable to fitted premium. (3) = fitted premium. (4) = office premium.

the divergence is only four percentage points, though in a few cases, e.g. Comprehensive cover for an 8-year-old vehicle in group D driven by a teenager, it is very large. Note also that for Comprehensive vehicle age 0–7, the ratio is in general less than 100, whilst for the corresponding non-comprehensive values, it exceeds 100. This suggests that comprehensive and non-comprehensive require separate tables, as discussed below.

Column (7) of Table 25 shows the difference between the fitted premiums and office premiums, weighted for the standing business. Column (7) reveals that the actual difference by cell in required premium income is in some cases considerable. It would seem that comprehensive is being undercharged and non-comprehensive overcharged. This will be returned to in § 12.2. However, overall the total premium income is reasonable, as shown by the following:

A. Total standing business	90,362
B. Points premiums (fitted using main effects)	£8,165,000
C. Office premiums	£8,545,000
D. Difference (B)–(C)	–£380,000
E. Ratio (B)/(C) (%)	96

Hence the fitted premium would have to be increased by 4% to match the office premium.

A similar comparison may be made with the present premium structure (Table 26, column (6)).

A. Total standing business	90,362
B. Points premiums (fitted using main effects)	£8,165,000
C. Actual premiums (charged in the existing rate book)	£6,538,000
D. Difference (B)–(C)	£1,627,000
E. Ratio (B)/(C) (%)	125

This indicates that a premium increase of about 25% is required to break even (see § 12.2).

11.6.8. *Problems*

There are a number of problems which such a method might encounter in practice, such as:

1. The effect on the results if the weights alter? (§ 12.2).
2. What happens if we later change the assumptions, e.g. as to expenses or the impact of inflation on claim costs? (§ 12.2).
3. If there is association between the main effects at the first stage? (§ 12.2).

These problems are discussed in the next section.

SECTION 12

Comparison of different sets of assumptions

12.1. *Comparisons*

One of the great advantages of this method is the relative ease with which calculations may be performed. Hence the 'points table' can help examine the effect various input assumptions have on the premium-structure. The following changes will be investigated:

(a) Inflation: Two scenarios will be tested—High, with inflation at 11% for material damage and 17% for liability settlements, and Low, with inflation at 7% for material damage and 13% for liability settlements.
(b) Base periods: TPBI claims based on the period 1972–76 will be compared with data based on the period 1972–77.
(c) Expenses: Treating all expenses as fixed, will be compared with the effect of distinguishing between fixed and variable expenses.
(d) The effect of altering the weights.
(e) For the main assumptions, comparison with the existing points.
(f) Separate 'point tables' for comprehensive and non-comprehensive.

12.2. *Results*

A summary of results are shown in Table 26.

The effect of varying the inflation assumption can be seen by comparing columns (1) and (2). As expected, the overall level has risen, but the relativities are virtually unchanged.

Changes in the base year assumptions for TPBI are reflected in columns (1) and (3). The lower overall level suggests that the 1972–77 basis yields a lower overall average. This time there has been a change in the policyholder age relativities. The factors affected are cover and policyholder age, reflecting the bases for TPBI input values (§ 8.3.3).

Altering the expense assumptions from Fixed (column (1)) to Fixed plus Variable (column (4)), gives results as expected, namely, an increase in overall level combined with a narrowing of the relativities for the fixed and variable basis.

To investigate the effect of changing the weights, an experiment was undertaken. The weights used in the previous example—column (1) of Table 25—form a series, starting 32, 62, 103 . . . and ending . . . 151, 167, 858. This series was reversed, i.e. the weights used on the second occasion began 858, 167, 151 . . . and ended 103, 62, 32.

Column (5) shows the results obtained when the weights were reversed. It can be seen that the overall level does not change but the relativities do. Given the drastic nature of the change, the fitting procedure seems reasonably robust to portfolio changes.

Table 26. *Summary of comparison results*

Assumptions	(1)	(2)	(3)	(4)	(5)	(6)
Inflation	Low	High	Low	Low	Low	Present points table
TPBI	72–76	72–76	72–77	72–76	72–76	
Expenses	Fixed	Fixed	Fixed	Fixed & Var.	Fixed	
Weights	Present	Present	Present	Present	Reversed	
Overall	104	106	100	109	104	104
Cover						
Comp.	17	16	19	15	19	18
Non-Comp.	0	0	0	0	0	0
Policyholder age						
17–20	30	30	33	27	32	29
21–24	23	23	23	20	23	23
25–29	9	9	10	8	9	8
30–34	0	0	0	0	0	0
35+	0	0	1	0	2	0
Car age						
0–3	19	19	19	16	13	7
4–7	12	12	12	10	10	4
8+	0	0	0	0	0	0
Vehicle group						
A	0	0	0	0	0	0
B	5	5	5	4	5	5
C	10	10	10	8	9	12
D	26	26	26	23	21	23

Comparing column (6) which represents the existing premium table, with say column (1) the major difference lies in the car age. Namely that newer cars are being significantly undercharged. Any changes to the existing table will depend on market considerations.

In § 11.6.2 brief mention was made of the question of what is to be done if associations between main effects are observed at the first stage of the OWLS analysis. In Table 20, an association was noticed between cover and age of car. One possible solution is to have separate rating structures for the two different types of cover. The results are summarized in Table 27.

These results suggest that there may be valid theoretical grounds for having separate premium structures. In particular, notice that car age and vehicle group are less important under the non-comprehensive points. This seems reasonable, since there is no AD cover. As decisions are taken on practical as well as theoretical grounds it is highly probable that despite the above results, a single premium structure for comprehensive and non-comprehensive will be selected, because of marketing considerations.

This now concludes the technical analysis of the premium bases.

Table 27. *Separate tables for comprehensive and non-comprehensive*

	Comp.	Non-Comp.
Overall	118	105
Policyholder's age		
17–20	28	36
21–24	23	25
25–29	10	11
30–34	2	0
35+	0	3
Car age		
0–3	23	11
4–7	13	9
8+	0	0
Vehicle group		
A	0	0
B	5	4
C	11	7
D	28	21

SECTION 13

Marketing aspects

The Group will now be in a position to discuss the premium recommendation. Factors they will have to consider will include *inter alia*:

(i) what other companies have done since the last meeting;
(ii) whether the competitive position will allow all or some of the recommendations to be implemented. For example, in Table 26, compare the actuarial premium recommendations column (1) with the existing structure column (6). The outstanding differences occur when comparing the points for car age. Hence the actuarial recommendation might indicate an increase in rates for new cars. However, this is unlikely to be acceptable because of the market conditions, which give little weight to this rating factor;
(iii) whether there are to be rate changes, their timing, and the likely reaction of the market;
(iv) any special marketing campaign proposed, e.g. introducing new factors, offering discounts for older drivers, or introducing protected no-claim discount.

As a final thought, the market place for selling motor insurance is changing, owing to the advent of direct telecommunication on television. The time will come (when the next generation of televisions are made) when quotations will be available to the public via the television in their own homes. Insurance companies and brokers will have to consider the implications of this new dimension. This just highlights the dynamic world of marketing.

The issues raised by marketing considerations are very important, and in fact dominate any decision process. However, the subject is outside the scope of this paper.

SECTION 14

Analysis of surplus

14.1 *Theory*

Premium rates are simply an attempt to forecast claim and expense experience. For proper control an analysis of surplus should be regularly performed to measure where the forecasts are failing and the effects on the profitability of business. Lee (1973) gives an example which is the basis of this work.

The following analysis of surplus for motor is based on a note by Grant (1981). Additional papers by Brennan (1968) and Taylor (1974) are also relevant. These latter papers discuss the problem of ordering of the analysis. This is similar to the problem encountered in the analysis of variance for the unbalanced design. Also the same problem is found in Pensions fund analysis of surplus when salary and withdrawal are ordered Lee (1973).

In 1977, a projection based on data then available was made for 1978; the main categories of data are listed below. The results actually experienced differ from those projected because the assumptions made in the projection do not wholly agree with the actual experience. What is needed is a method of analysing these differences into their component contributions which can later be added back to yield the differences between actual and projected results. This problem is similar to an analysis of surplus and the information gained will be useful in fine-tuning the assumptions in the model, as the monetary effect of the differences between the assumptions made in the model and the actual experience is discovered. It would also provide a useful format for analysing these differences for management.

Companies use models of which the simplest involve only premiums and claim ratios, while the more complex ones involve assumptions about exposure, average premium rate, average claim cost, claim frequency and expenses. In the model considered, assumptions on the following factors are made for each period and risk group:

SB = Standing Business (number of vehicles at the end of the year)
AWP = Average Written Premium
ACF = Average Claim Frequency
ACC = Average Claims Cost
VER = Variable Expense Rates (this would include the commission rate)
FE = Fixed Expenses

The Earned Premium (EP) and the Exposure (EX) can be easily calculated using the $\frac{1}{8}$ method for quarterly projections or the $\frac{1}{24}$ method for monthly projections (Benjamin, 1977). The Average Earned Premium (AEP) can be calculated by dividing the EP by the EX. (At present, tax and investment income are ignored.)

The result from the year's Underwriting Surplus (US) (i.e. excluding any adjustment to outstanding claims for prior years) can be expressed in the form:

$$\text{US} = \text{Earned Premium} - \text{Claims} - \text{Expenses (inclusive of commission)}$$

The Underwriting Surplus projected by the model is

$$\text{US} = \text{EX} \times \text{AEP} - \text{EX} \times \text{ACF} \times \text{ACC} - \text{SB} \times \text{VER} \times \text{AWP} - \text{FE}$$

Denote by $'$ the figures derived from the actual results. It is possible to derive the AEP$'$ from the Earned Premium EP$'$ and the Exposure EX$'$. The claims can easily be expressed in the form EX$'$ × ACF$'$ × ACC$'$ as the Exposure EX$'$ and the Claim Frequency ACF$'$ are known. The Underwriting Surplus can, therefore, be expressed as

$$\text{US}' = \text{EX}' \times \text{AEP}' - \text{EX}' \times \text{ACF}' \times \text{ACC}' - \text{SB}' \times \text{VER}' \times \text{AWP}' - \text{FE}'$$

The differences between the actual experience and that projected are the result of differences in the exposure, claim frequency, level of expenses, inflation rate, etc. The formulae for a possible analysis of assessing the numerical contribution from each of the factors are given below.

Effects of

Exposure
$(\text{EX}' - \text{EX}) \times (\text{AEP} - \text{ACF} \times \text{ACC}) - (\text{SB}' - \text{SB}) \times \text{VER} \times \text{AWP}$

Average premium
$\text{EX}' \times (\text{AEP}' - \text{AEP}) - \text{SB}' \times \text{VER} \times (\text{AWP}' - \text{AWP})$

Claims frequency
$-(\text{EX}' \times (\text{ACF}' - \text{ACF}) \times \text{ACC})$

Claims inflation
$-(\text{EX}' \times \text{ACF}' \times (\text{ACC}^* - \text{ACC}))$

Claims cost
$-(\text{EX}' \times \text{ACF}' \times (\text{ACC}' - \text{ACC}^*))$

Expenses
$-(\text{FE}' - \text{FE} + \text{SB}' \times \text{VER}' \times \text{AWP}' - \text{SB}' \times \text{VER} \times \text{AWP}')$

(where ACC* is the forecast of the average claims cost using either the known or the latest estimates of claims cost inflation rates).

The formulae above can easily be checked from the expression for Underwriting Surplus.

The effect of claim frequency and claims cost and claims inflation can be expressed in several different ways. Two of these are given below:

	Original	Alternative
Claim frequency	$-EX' \times (ACF' - ACF) \times ACC$	$-EX' \times (ACF' - ACF) \times ACC'$
Claims cost	$-EX' \times ACF' \times (ACC' - ACC^*)$	$-EX' \times ACF \times (ACC' - ACC^*)$
Claims inflation	$-EX' \times ACF' \times (ACC^* - ACC)$	$-EX' \times ACF \times (ACC^* - ACC)$

The Original is preferred and is used in the analysis because the effect of the claim frequency is independent of the actual claims cost. The actual claims cost will be partly based on the outstanding claims estimated and will change over time until the ultimate settlement. Any adjustment to these estimates in the analysis will affect only the result attributed to claims cost.

Various other alternative breakdowns of the analysis exist and the choice largely depends on the projection procedure, the data and the factors to be highlighted. For example, Grant (1981) considers the effect of dividing average claims cost into its component parts, namely: AD, TPPD, TPBI, M and inflation.

14.2. *Analysis of surplus—example*

The following example is taken from the forecasts of part of the private car portfolio of an insurance company. The forecasts are made on a year of accident basis, this example being 1978. Four sets of forecasts for 1978 are given, corresponding to a projection at the end of each quarter in 1977.

The actual results for claims cost include the latest case estimates and therefore this analysis would be subject to some adjustments as the claim payments run off.

(Figures in '000s)	1st Forecast	2nd Forecast	3rd Forecast	4th Forecast	Results
Standing business	101·4	88·4	87·4	85·4	78·3
Exposure	97·5	87·8	87·8	85·8	80·8
Written premium	6244	5702	6114	6339	5794
Earned premium	6061	5684	5902	6035	5600
No. of claims	13·1	12·6	12·4	12·2	11·7
Claims costs	3993	3938	3821	3712	3348
Expenses†	2077	1987	2184	2295	2342
Profit (Loss)*	(9)	(241)	(103)	28	(90)

* Profit = Earned Premium − Claims cost − Expenses

† including variable expenses (15% of written premium)

To calculate the effects for the analysis of surplus we need, in addition, to recalculate the forecasts of claims cost using the latest estimates of claims inflation. These averages are given on the next page:

	1st Forecast	2nd Forecast	3rd Forecast	4th Forecast	Results
Forecast average	304·81	312·54	308·15	304·26	286·15
Adjusted forecast	289·85	300·13	304·13	302·84	

The effect of each factor is then:

(Figures in '000s)	1st Forecast	2nd Forecast	3rd Forecast	4th Forecast
Exposure	−141	−41	−71	−56
Premiums	431	258	121	−81
Claim frequency	−257	−33	−89	−64
Claims inflation	175	145	47	17
Claims costs	43	164	210	195
Expenses	−333	−341	−206	−129

The initial underestimation of the average premium is partly due to the premium increases during 1977 and 1978 which were not allowed for in the earlier forecasts.

Initially, claim frequency is very important. However, as the forecast approaches 1978, the estimate of frequency improves, however a 4% over prediction in frequency produces −64 contribution to surplus.

The effect of claims cost unexpectedly increases between the first and last forecasts, highlighting either a need to examine the methods used to project claims cost, or possibly an unexpected change in the claims experience.

The relatively large effect of expenses is due to two problems. Firstly, there is a slight difference between forecasts and results in the basis for allocating fixed expenses between classes; also, the inflation rates used to project expenses during 1978 could be out of line with those actually experienced. An obvious extension of these methods would be to separate the effect of inflation from effect of expenses in a similar manner to that employed for claims costs.

Section 15

Conclusion

15.1. *Summary*

This paper has put forward a framework for the analysis of motor insurance premium rating, which combines both practical and technical judgements. It has argued for a detailed breakdown of the data. This allows for an in-depth analysis of claim proportions and claims cost. Since the data-set analysed were relatively small, sophisticated statistical modelling techniques were employed to get an understanding of the underlying structure and smooth out variations in the observed data. It is suggested that the results from this part of the paper encourage similar analyses by insurance companies, and they need not rely on industry statistics to obtain input for a premium basis.

Then by combining claims experience data together with economic views concerning inflation and expenses an office premium was calculated. The advantages and disadvantages of converting the office premium into a 'points table' were briefly discussed. A mathematical method to arrive at the 'points table' was developed and put into a simple algorithm. One of the advantages of a 'points table' is the ease of comparing different sets of assumptions in the premium basis. As an example several sets of assumptions were analysed and discussed.

Finally as one of the themes of this paper is to encourage explicit account of the various assumptions in the premium basis, an analysis of surplus was developed. This type of analysis is important as a learning process since the assumptions can be monitored by comparing actual against expected; future premium analysis will benefit.

15.2. *Acknowledgements*

I would like to acknowledge my appreciation of the insurance company concerned for supplying me with this data and allowing me to reproduce it here.

I also wish to thank Sydney Benjamin, who encouraged me to submit a paper, based on my thesis, for publication in *J.I.A.*, and Bernard Benjamin and David Hart who examined the original thesis.

Finally I am indebted to numerous colleagues, particularly the late Bobbie Beard, who over the years have helped to lay the foundations on which this work is based.

APPENDIX

Orthogonal weighted least squares analysis

The general problem is to minimize the following expression:

$$\sum_{ijk} \frac{1}{V_{ijk}} [Z_{ijk} - (\hat{S} + \hat{R}_i + \hat{U}_j + \hat{V}_k + (\hat{\mathrm{I}}\mathrm{n}))]^2 \qquad \text{(A.1)}$$

where Z_{ijkl} is the dependent variable for cell $ijkl$

V_{ijkl} is the weights for cell $ijkl$

S is the overall mean

R_i is the effect of R at level i

U_j is the effect of U at level j

V_k is the effect of V at level k

(In) is the effect of the interaction terms

$\hat{}$ is the least squared estimate

e.g. Johnson and Hey (1971) used $1/n$ as the weight; and Seal (1968) used $P(1-P)/(n)$ as the weight.

Please (1974), suggested that the minimization could be simplified by assuming that the weight $1/(V_{ijk})$ factorizes i.e. if V_{ijk} satisfies

$$\frac{1}{V_{ijk}} \propto r_i u_j v_k \text{ (for all } ijk) \qquad \text{(A.2)}$$

and various concomitant constraints. The design becomes orthogonal, enabling the following estimators to be derived:

$$\hat{S} = \sum_i \sum_j \sum_k \left(\frac{r_i}{r.} \frac{u_j}{u.} \frac{v_k}{v.} \right) (Z_{ijk})$$

$$R_i = \sum_j \sum_k \frac{u_j}{u.} \frac{v_k}{v.} (Z_{ijk}) - \hat{S}$$

$$(RU)_{ij} = \sum_k \frac{v_k}{v.} (Z_{ijk}) - \hat{S} - \hat{R}_i - \hat{U}_j$$

and similar equations for $(RV)_{ik}$ and $(UV)_{jk}$

where $$r. = \sum_i r_i,\ u. = \sum_j u_j \text{ and } v. = \sum_k v_k$$

and the constraints are

$$\sum_i r_i R_i = \sum_j u_j U_j = \sum_k v_k V_k = 0$$

$$\sum_i r_i (RU)_{ij} = \sum u_j (RU)_{ij} = 0 \text{ etc.}$$

One method of justifying the factorization, is by an analysis of $-\log V_{ijk}$. Under the assumption of perfect factorization

$$-\log V_{ijk} = s' + r'_i + u'_j + v'_k$$

where s' is an arbitrary constant and r'_i, u'_j and v'_k correspond to r_i, u_j and v_k in equation (A.2). Thus, if the interactions in an analysis of variance of $-\log V_{ijk}$ are negligible, there is evidence supporting the factorization and hence the main effects can be used to calculate the weights.

In practice the weights do not always factorize perfectly, but experience shows that the model is robust with respect to the weights chosen.

REFERENCE AND BIBLIOGRAPHY

AITKIN, M. (1978) The analysis of unbalanced cross-classifications. *J.R.S.S.* A, **141,** 195.

AJNE, B. (1974) A note on the multiplicative rate making model. *ASTIN Bulletin*, **8,** 144.

AJNE, B. (1982) Home-owners comprehensive, as an example of multiplicative ratemaking. Paper presented at a GISG Conference, Stratford-upon-Avon, November 1982.

ALBRECHT, P. (1982) Parametric multiple regression risk models theory and statistical analysis. A paper presented at 16th ASTIN Colloquium, Loen, September 1982.

ALMER, B. (1957) Risk analysis in theory and practical statistics. *15th International Congress of Actuaries*, **2,** 314.

ASCOMBE, F.S. & TUKEY, J. W. (1963) The examination and analysis of residuals. *Technometrics*, **5,** 141.

ASTIN NETHERLANDS GROUP (1982) (Chairman WIT, G. W.) *New Motor Rating Structure in the Netherlands*. Publication of the ASTIN-GROEP Nederland.

BAILEY, R. A. & SIMON, L. (1960) Two studies in automobile insurance ratemaking. *ASTIN Bulletin* **1,** 192.

BAXTER, L. A. & COUTTS, S. M. (1977) An analysis of motor insurance claims data. A paper presented at the 13th ASTIN Colloquium, Washington, June 1977.

BAXTER, L. A., COUTTS, S. M. & ROSS, S. A. F. (1979) Applications of linear models in motor insurance. *21st International Congress of Actuaries*, **2,** 11.

BEARD, R. E. (1964) Some statistical problems arising from the transaction of motor insurance business. *J.S.S.* **17,** 279.

BENJAMIN, B. (1977) *General Insurance*. London: Heinemann.

BENKTANDER, G. (1982) Personal communication.

BENNETT, M. C. (1978) Models in motor insurance. *J.S.S.* **22,** 134.

BOHMAN, H. (1979) Insurance economics. *Scand. Actuarial J.* 57.

BRENNAN, A. R. (1968) A note on the theory of analysis of surplus. *Transactions of Institute of Actuaries, Australia and New Zealand*, 134.

CHANG L. & FAIRLEY, W. B. (1980) Pricing automobile insurance under multivariate classification of risks. Additive versus multiplicative. *J. of Risk and Insurance*, **7,** 75.

CHRISTENSEN, J. E. (1982) Analysis of bonus systems in motor insurance. A paper presented at a GISG Conference, Stratford-upon-Avon, November 1982.

COUTTS, S. M. (1975) Some methods of predicting the number of motor car accidents. A M.Sc. dissertation: University College London, Department of Statistics.

COUTTS, S. M. (1982) Points system in motor insurance. A paper presented at a GISG Conference, Stratford-upon-Avon, November 1982.

COUTTS, S. M. (1983) An actuarial approach to motor insurance rating: Ph.D. thesis The City University, Department of Mathematics.

COX, D. R. (1970) *Analysis of Binary Data*. Methuen Press.

DAWID, A. P. (1980) Statistical models for smoothing and predictions. Talk given at a post qualification seminar on the application of linear models to the analysis of motor insurance data. The City University.

DAWID, A. P. (1981) Personal communication.

FERRARA, G. (1971) Review of mathematical models for the construction and control of tariffs. *Istituo Italiano degli Attuari 1971*: 1.

FISHER, H. F. & YOUNG, J. (1965) *Actuarial Practice of Life Assurance*. Cambridge University Press.

FISHER, R. A. (1946) *Statistical Methods for Research Workers*. Oliver & Boyd.

FRANCIS, I. (1973) Comparison of several analysis of variance programs. *J. American Statist. Assoc.* **68,** 860.

GLIM (Generalized Linear Interactive Modelling) (1975) *General Linear Interactive Modelling*: Numerical Algorithms Group, Oxford.

GRADY, D. (1982) Premium rating for personal line coverages in the United States. A paper presented at a GISG Conference, Stratford-upon-Avon, November 1982.

GRANT, A. (1981) Analysis of surplus in motor insurance. *GIRO Bulletin* December 1981.

GRIMES, T. (1971) Claims frequency analysis in motor insurance. *J.S.S.* **19,** 147.

HALLIN, M. & INGENBLEEK, J.-F. (1981) Etude statistique des facteurs influençant le risque automobile. Université Libre de Bruxelles, Institute Statistique—Série actuarielle no. 10.

HARVILLE, D. A. (1978) Alternative formulations and procedures for the two way mixed model. *Biometrics* **34,** 441.

Insurance Companies (Accounts and Statements) Regulations (1981). H.M.S.O.

JOHNSON, P. D. (1980) Personal Communication.

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.

KAHANE, Y. & LEVY, H. (1975) Regulation in the insurance industry: determination of premiums in automobile insurance. *Journal of Risk and Insurance* **62,** 117.

LEE, E. M. (1973) *An Introduction to Pension Funds.* Published by the Institute of Actuaries.

LITTLE, R. J. A. (1978) Generalized Linear Models for Cross-classified data. World Fertility Study. Technical Bulletin Series No. 5.

MEHRING, J. (1964) A mathematical method of statistical analysis and premium calculation in motor insurance. *Blatter der Deutschen Gesellschaft fur Versicherungsmathematik*, **7,** 111.

NELDER, J. A. & WEDDERBURN, R. W. M. (1972) Generalized linear models. *J.R.S.S.*, A, **135,** 370.

NELDER, J. A. (1980) Personal Communication.

PITKANEN, P. (1974) Tariff Theory. *The ASTIN Bulletin*, **8,** 204.

PLEASE, N. W. (1974) Personal Communication.

PRICE COMMISSION (1977) Government White Paper. H.M.S.O.

RUSHTON, I. L. (1977) Expenses in general insurance. A paper presented to the Institute of Actuaries Students Society, London.

SCURFIELD, H. H. (1968) Motor insurance statistics. *J.S.S.* **18,** 207.

SEAL, H. L. (1968) The use of multiple regression in risk classification based on proportionate losses. *18th International Congress of Actuaries*, **2,** 659.

SHREWSBURY, B. (1983) Personal communication.

TAYLOR, G. C. (1974) Symmetry between components of analysis of surplus *Transactions of the Institute of Actuaries, Australia and New Zealand. 374.*

TRAYHORN, M. (1980) Report of working party on fluctuation reserves. Presented at GISG conference at Cambridge, October 1980.

WESTENBERG, R. (1983) Some statistical investigations in general insurance. Ph.D. thesis: The City University Department of Mathematics.

ZIAI, Y. (1979) Statistical models of claim amount distributions in general insurance: Ph.D thesis: The City University, Department of Mathematics.