TECHNICAL ASPECTS OF DOMESTIC LINES PRICING

Greg Taylor

Consultant, Tillinghast-Towers Perrin,

GPO Box 3279

SYDNEY NSW 2001 AUSTRALIA

and

Professorial Associate, Centre for Actuarial Studies,

Faculty of Economics and Commerce,

University of Melbourne,

PARKVILLE VICTORIA 3052 AUSTRALIA

June 1997

Table of Contents

Summary

Recent years have seen a considerable upsurge in the use of Generalised Linear Models (GLMs) to carry out technical pricing of domestic (and other) lines. The paper by Brockman and Wright (1992) detailed some of the fundamentals of the application of GLMs.

This paper attempts to build on that earlier one. There are many practical issues which arise repeatedly in applications. An attempt is made here to identify these issues, establish a framework for dealing with them, and reduce as many as possible to a routine.

Section 3 addresses the matching of risk premiums to claims experience after a GLM has been used to estimate all premium relativities, i.e. cell-to-cell ratios. Subsequent sections deal with modelling the relativities themselves.

Section 6 covers various issues concerned with the selection of a model structure and estimation of its parameters. Section 7 discusses the validation of the model. Section 8 discusses the response of the model to No Claim Discount and deductibles, two subjects which do not fit easily into the GLM framework.

Section 9 summarises as much as possible of the discussion into a protocol, set out in flowchart form.

Keywords: Pricing, domestic lines, generalised linear models.

1 Introduction

This paper is concerned with the pricing of domestic lines of insurance, specifically with estimation of **risk premiums**. No consideration will be given to expenses, profit or any strategic adjustments made to premiums in response to competitors' prices.

Emphasis will be placed on the use of **Generalised Linear Models** (GLMs) to rate risk premiums. This paper therefore forms some sort of sequel to that of Brockman and Wright (1992), which dealt with a number of aspects of pricing using GLMs, with particular reference to the statistical package GLIM.

The purpose here will be twofold:

- to expand on some of the areas discussed by Brockman and Wright; and
- to establish, as far as possible, a **protocol** by which a risk premium rating assignment might be carried out.

Risk rating of a portfolio consists of two parts:

- rating the relativities between risks;
- ensuring that the aggregate premium pool is adequate;

as will be discussed in Section 3. This paper is concerned largely, though not solely, with the first of these two questions.

2 Notation and terminology

Brockman and Wright (1992) mention (p. 484) the variety of different claim types which can arise under a single policy. They recommend separate modelling of different claim types. This question is discussed in Section 6. Suffice to say here that where different claim types are recognised, all notation below applies to a **single claim type**.

It is supposed, conventionally, that each policy at risk has a number of **attributes** which function as **risk descriptors**, i.e. the risk premium associated with a policy may be viewed as a mapping of its attributes.

In this sense, the attributes may be regarded as **risk covariates**. Examples in motor insurance would be:

- age of insured;
- gender of insured;
- sum insured;

etc.

Consider a portfolio of policies, and denote the identified covariates by $i, j, k, \ldots$ Though there can be many covariates, it will be assumed for the sake of exposition that there are only 3. This reduces the notational load, while extension of the development below to a larger number of covariates will be obvious.

With this understanding, each policy can be labelled by its covariates $i, j, k, \ldots$ A **cell** is a collection of policies with common i, j, k, and may be labelled with these values. Thus a typical cell will be denoted by $\{i, j, k\}$.

In the following, a quantity sub-scripted by ijk will represent the summation of that basic quantity over all policies in cell $\{i, j, k\}$. Let

E_{ijk} = number of years of exposure to risk of claim;

N_{ijk} = number of claims;

C_{ijk} = amount of claims incurred.

These quantities will be measured over some defined, but arbitrary, period. The period might be a calendar year, quarter, month, etc.

In practice, some consideration will need to be given to the definition of "claims incurred". This is discussed in Section 5.1.

It is assumed that

$$E[N_{ijk}] = E_{ijk} f_{ijk}, \tag{2.1}$$

$$E[C_{ijk} | N_{ijk}] = N_{ijk} a_{ijk} \tag{2.2}$$

where $E[.]$ is the expectation operator and f_{ijk}, a_{ijk} are cell-dependent parameters:

f_{ijk} = expected **claim frequency** (per year of exposure) for the cell;

a_{ijk} = expected **claim size** for the cell.

Define the target of the investigation:

P_{ijk} = **risk premium** per year of exposure,

and note that

$$P_{ijk} = E[C_{ijk}]/E_{ijk}$$
$$= f_{ijk}\, a_{ijk}, \qquad (2.3)$$

by (2.1) and (2.2).

Occasionally, it will be convenient to have a notation for the expectation appearing in (2.3). Let

$$\gamma_{ijk} = E[C_{ijk}].$$

The **risk premium relativity** between two cells $\{i_1, j_1, k_1\}$ and $\{i_2, j_2, k_2\}$ will be defined as $P_{i_1j_1k_1}/P_{i_2j_2k_2}$. Rather than considering all pairs of cells, it will usually be convenient to express relativities in terms of a **base**, or in some sense typical, cell $\{i_0, j_0, k_0\}$.

In this case, one may speak of just the **relativity of cell** $\{i, j, k\}$, defined as:

$$r_{ijk} = \frac{P_{ijk}}{P_{i_0j_0k_0}} = \left(\frac{f_{ijk}}{f_{i_0j_0k_0}}\right)\left(\frac{a_{ijk}}{a_{i_0j_0k_0}}\right). \qquad (2.4)$$

The decomposition on the right shows that:

risk premium relativity = (claim frequency relativity) × (claim size relativity). (2.5)

As will be seen later, it is often useful to consider relativities obtained by varying only one covariate, e.g. $P_{ij_0k_0}/P_{i_0j_0k_0}$, which might reasonably be referred to as the **relativity of covariate i**.

Let

$$E = \sum_{ijk} E_{ijk} = \text{ total exposure in data;}$$
$$C = \sum_{ijk} C_{ijk} = \text{ total claim cost;}$$
$$P = \sum_{ijk} E_{ijk}P_{ijk} = \text{total risk premium.}$$

By (2.3),

$$P = E[C]. \qquad (2.6)$$

3 Premium relativities and aggregate

It is apparent from (2.4) that the whole risk premium scale is specified by:

- the complete set of relativities r_{ijk}; and
- the base premium $P_{i_0 j_0 k_0}$.

It will be seen in Section 7.4.1 that the analysis which leads to estimates of the r_{ijk} will also provide an estimate of $P_{i_0 j_0 k_0}$.

However, the remainder of the present section explains why this last estimate may be passed over in favour of an alternative, related more directly to the aggregate experience of the portfolio.

Substitution of (2.4) in (2.6) yields:

$$E[C] = P_{i_0 j_0 k_0} \sum_{ijk} E_{ijk} r_{ijk}, \qquad (3.1)$$

whence $P_{i_0 j_0 k_0}$ is estimated by

$$\hat{P}_{i_0 j_0 k_0} = \hat{C} / \sum_{ijk} E_{ijk} r_{ijk}, \qquad (3.2)$$

with $\hat{C}$ some suitable estimator of $E[C]$.

This shows that the base premium can be estimated once all relativities have been found and an estimator of the underlying (i.e. after removal of extraordinary effects) aggregate claim cost obtained.

The estimator $\hat{C}$ may be formed from historical aggregate costs. Here some care will be required to take into account any changes in portfolio composition which may have occurred within the period of the historical data.

An alternative form of (3.1) is:

$$E[C/E] = P_{i_0 j_0 k_0} \sum_{ijk} w_{ijk} r_{ijk}, \qquad (3.3)$$

where the w_{ijk} form a set of weights:

$$w_{ijk} = E_{ijk} / E. \qquad (3.4)$$

This shows that the left side of (3.3) depends on the weighting of the various cells in the experience.

Suppose that C and E are available for a number of past quarters. Denote these quantities by $C^{(t)}$, $C^{(t-1)}$, etc. in reverse time order. Suppose that this series has been corrected for inflation and seasonality. Each of the quarters $t-s$ will have an associated set of weights $\{w^{(t-s)}\}$. Equation (3.3) shows that strictly the $C^{(t-s)}$ are directly comparable if these sets of weights do not vary over time. If they do, the $C^{(t-s)}$ should be corrected accordingly.

The counterpart of (3.3) with the time index taken into account is:

$$E\left[C^{(t-s)}/E^{(t-s)}\right] = P_{i_0 j_0 k_0} \sum_{ijk} w_{ijk}^{(t-s)} r_{ijk}, \tag{3.5}$$

whence the base premium is estimated from the data of quarter $t-s$ by

$$\hat{P}_{i_0 j_0 k_0}^{(t-s)} = \frac{C^{(t-s)}}{E^{(t-s)} \sum_{ijk} w_{ijk}^{(t-s)} r_{ijk}}. \tag{3.6}$$

The summation appearing in the denominator is the **average relativity** taken over the whole portfolio.

The estimate of base premium using all the data is

$$\hat{P}_{i_0 j_0 k_0} = \frac{\sum_s C^{(t-s)}}{\sum_s \left[E^{(t-s)} \sum_{ijk} w_{ijk}^{(t-s)} r_{ijk}\right]}. \tag{3.7}$$

A numerical example will illustrate. Consider the situation set out in Table 3.1.

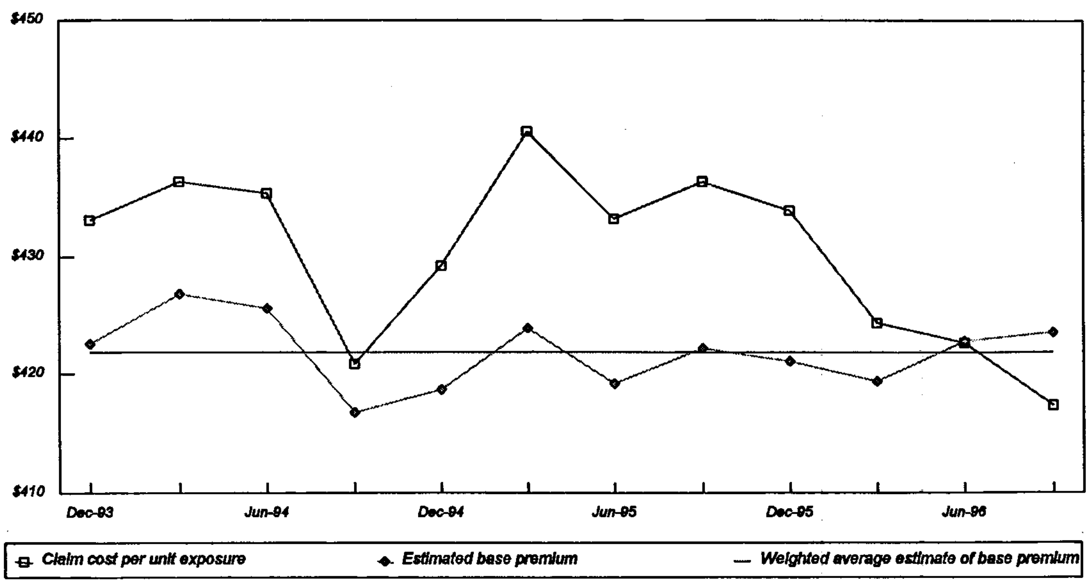
Table 3.1
Claims history adjustment

Quarter	Exposure E	Claims C \$M	Average claim cost per year of exposure C/E \$	Average relativity	Estimated base premium [a]
Dec 93	52,347	22.68	433.2	1.0251	422.6
Mar 94	52,579	22.95	436.4	1.0223	426.9
Jun 94	53,333	23.23	435.5	1.0233	425.6
Sep 94	53,012	22.31	420.8	1.0099	416.7
Dec 94	52,549	22.56	429.3	1.0253	418.7
Mar 95	52,552	23.16	440.7	1.0397	423.9
Jun 95	53,341	23.11	433.3	1.0336	419.2
Sep 95	53,994	23.57	436.4	1.0336	422.3
Dec 95	54,501	23.65	434.0	1.0305	421.1
Mar 96	54,763	23.24	424.4	1.0120	419.4
Jun 96	54,788	23.16	422.7	0.9997	422.9
Sep 96	55,002	22.95	417.3	0.9853	423.5
Weighted average[b]					421.9

Notes: [a] According to (3.6).
[b] According to (3.7).

These results are illustrated graphically in Figure 3.1.

Fig 3.1
Relativity adjustment of claims history



In practice it may be difficult to estimate the base premium with such precision. For example, while the aggregate quantities E and C may be available by quarter for the 3-year period addressed in Table 3.1, the fine detail required to calculate the series of average relativities over the whole period may not be.

Suppose that the portfolio is known to have undergone changes in the proportion of policies free of an excess, but no other systematic changes in portfolio structure have been identified. A summary of the situation is given in Table 3.2, where the "Actual" average claim cost per year of exposure is reproduced from Table 3.1.

Table 3.2
Claims history

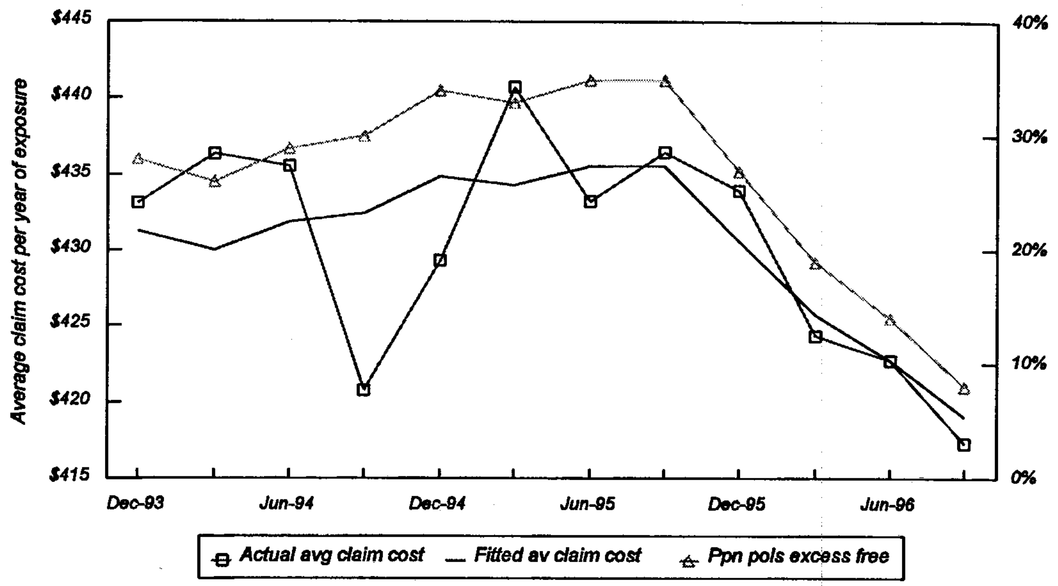
Quarter	Proportion of exposure with no excess	Average claim cost per year of exposure	
		Actual	Fitted
	%	$	$
Dec 93	28	433	431
Mar 94	26	436	430
Jun 94	29	435	432
Sep 94	30	421	432
Dec 94	34	429	435
Mar 95	33	441	434
Jun 95	35	433	436
Sep 95	35	436	436
Dec 95	27	434	431
Mar 96	19	424	426
Jun 96	14	423	423
Sep 96	8	417	419

The table reflects the marked shift away from excess free policies, beginning in the December 1995 quarter. Note that this is consistent with the decline in average relativity, over the same period, appearing in Table 3.1.

A simple regression of average claim costs against proportion of policies excess free yields the fitted average claim costs in the final column of the table. The situation is illustrated in Figure 3.2.

Fig. 3.2

Approximate adjustment of claims history



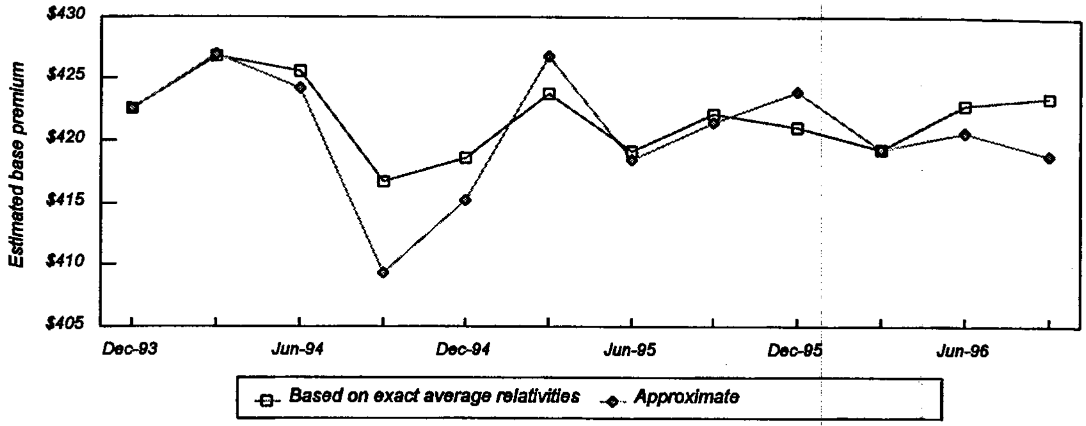
If the regression line is taken as providing an estimate of the portfolio average risk premium for each quarter, then it implies a time series average relativities. The implied average relativity for quarter $t\text{-}s$ is:

$$\text{average relativity for Dec 93 quarter} \times \frac{\text{Fitted average claim cost for quarter t-s}}{\text{Fitted average claim cost for Dec 93 quarter}}. \qquad (3.8)$$

If the average relativity of the Dec 93 quarter is taken as 1.0251, as in Table 3.1, the average relativities (3.8) yield a series of **approximate** base premium estimates. Figure 3.3 compares these with those based on exact average relativities in Table 3.1.

Fig 3.3

Comparison of claims history adjustments



Note that the approximate estimates of base premium follow the exact ones reasonably well, but display somewhat greater volatility. This is to be expected since they do not recognise any source of variation in portfolio structure other than change in the proportion of policies free of excess.

Note that the base premium of $421.9, estimated in Table 3.1, is assumed applicable to **all** past quarters, subject to inflation and seasonality adjustments. Likewise, it may be assumed applicable to all future periods, unless there is some specific reason to the contrary.

This constancy of base premium holds despite the decline in portfolio average risk premium, this decline being accounted for in the average relativity rather than the base premium.

The average risk premium may continue to decline in future, but there is no need to project this when rating future premiums. In fact, to do so would double count that effect since it is already allowed for in the relativities (the future portfolio will contain a greater proportion of policies with low relativities).

4 Outline of relativity estimation

The remainder of this paper will be concerned with relativities of the various sorts discussed in Section 2. There are several distinct stages in this procedure.

First, there will be some decisions to be made in **assembling the data** and deciding **which quantities are to be modelled**. These matters will be discussed in Section 5.

Second, there are practicalities to consider in making the **most effective use of GLMs**. These are treated in Section 6.

Third, there are some matters, specifically the rating of:

Geographic zone;
Excess;
No Claim Discount (NCD);

which **do not fit altogether comfortably into the GLM framework**. The alternative, or supplementary methodologies are discussed in Section 7.

Finally, there is a need to **validate** whatever model is produced by the methods of Sections 5 to 7. Procedures for doing so are described in Section 8.

5 Data and model set-up

5.1 Claims incurred

All of the development of amounts of claims in Sections 2 and 3 was in terms of **claims incurred**. These quantities are supposed known for one or more periods, i.e. **accident periods**.

The fact that C_{ijk} relates to an accident period raises the question of its IBNR and IBNER components. Decompose C_{ijk} as follows:

$$C_{ijk} = C_{ijk}^{r} + C_{ijk}^{u}, \tag{5.1}$$

where

C_{ijk}^{r} = the reported component of C_{ijk}, i.e. losses paid to the date of the analysis plus estimate at that date of losses unpaid in respect of reported claims;

C_{ijk}^{u} = the unreported component of C_{ijk}, i.e. the difference between C_{ijk} and its ultimate value when all claims from the accident period in question have been settled.

Let

$$\gamma_{ijk}^{r} = E[C_{ijk}^{r}], \tag{5.2}$$

$$\gamma_{ijk}^{u} = E[C_{ijk}^{u}]. \tag{5.3}$$

The risk premium relativity r_{ijk} defined by (2.4) is

$$r_{ijk} = (\gamma_{ijk}/E_{ijk}) / (\gamma_{i_0j_0k_0}/E_{i_0j_0k_0}), \tag{5.4}$$

by (2.3).

For the present sub-section only, it will be convenient to abbreviate the subscripts $i_0j_0k_0$ and ijk to just 0 and blank respectively. With this understanding, (5.4) becomes:

$$\begin{aligned} r &= (\gamma/\gamma_0) / (E/E_0) \\ &= \frac{\gamma^r + \gamma^u}{\gamma_0^r + \gamma_0^u} \Big/ \frac{E}{E_0}. \end{aligned} \tag{5.5}$$

Suppose that the unreported components in (5.5) are simply ignored, so that r is replaced by the quantity

$$r' = (\gamma^r/\gamma_0^r) / (E/E_0) \tag{5.6}$$

By (5.5) and (5.6), r' is in error by the factor

$$\frac{r}{r'} = \frac{1 + \gamma^u/\gamma^r}{1 + \gamma_0^u/\gamma_0^r} \tag{5.7}$$

which equals 1 if

$$\gamma^u/\gamma^r = \gamma_0^u/\gamma_0^r. \tag{5.8}$$

Relations (5.7) and (5.8) show that ignoring IBNR and IBNER claims creates errors in risk premium relativities only to the extent that the **rate of recognition of these components over time differs from cell to cell.**

Note that r' would not be improved by the inclusion of $\alpha\gamma^r$ (α const.) as an estimator of γ^u, since this would not change (5.6).

An example will illustrate the magnitude of the errors (5.7). Suppose the data used to estimate relativities for a motor property damage portfolio relate to two complete accident years. Typically, γ^u/γ^r, taken over the whole portfolio, might be of the order of a few percentage points, say 2½%.

Suppose in addition that, taken over all cells,

$$0.8\ \gamma_0^u/\gamma_0^r < \gamma^u/\gamma^r < 1.2\ \gamma_0^u/\gamma_0^r. \tag{5.9}$$

Then rough bounds on the error factor r/r' are:

$$0.995 < r/r' < 1.005. \qquad (5.10)$$

In this example, ignoring IBNR and IBNER claims leads to errors of less than ½%, in most cases much less, in respect of individual cell risk premium relativities.

The only way of improving on this situation would be to estimate the γ''/γ' differentially over cells. This would require considerable sophistication, with very limited scope for improvement. It would not usually be attempted.

While the above discussion has been framed in terms of risk premium relativities, parallel arguments apply to claim frequency and claim size relativities.

5.2 Claim Type

Claims in most portfolios can be separated into a number of distinct types. For example, motor accidental damage claims might be classified as a:

- collision;
- theft;
- fire;

etc.

As another example, house (building) claims would include:

- windstorm;
- flood;
- other water damage;
- fusion;

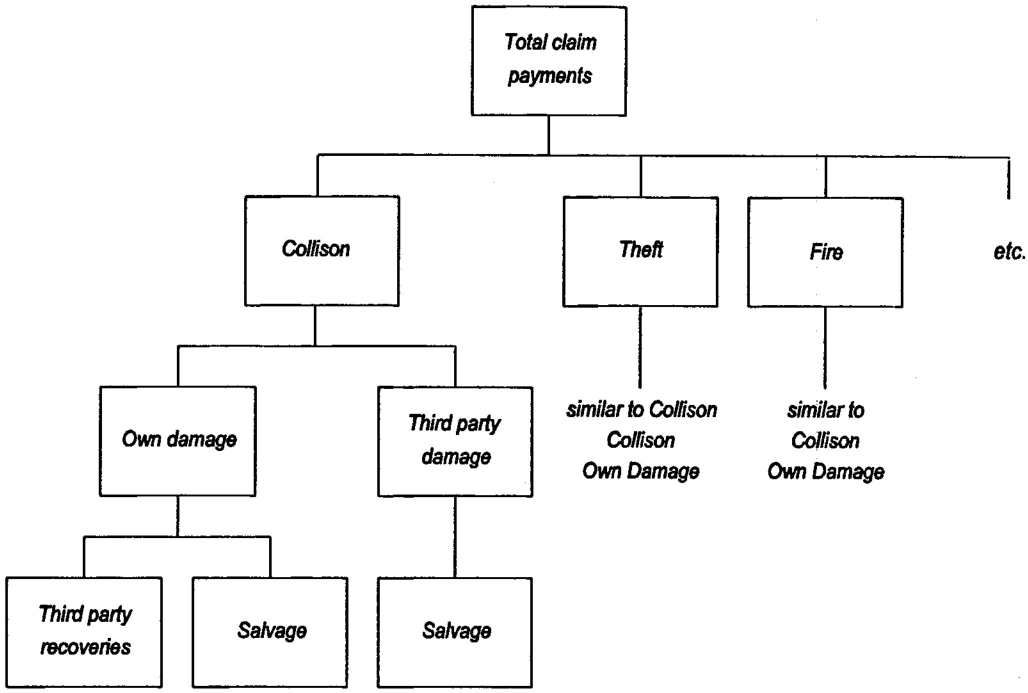
etc.

Even within these claim types, further subdivision may be possible. For example, in motor insurance, collision claim payments may be dissected into:

- insured's own damage;
- third party damage;
- third party recoveries;
- salvage.

Note that payments under some of these types would be dependent on other types, so that the tree diagram indicating causality in the modelling might look like Figure 5.1.

Figure 5.1
Motor (Property Damage) -
causality between payment types



Strictly, each claim type can be viewed as a subject to a separate model, despite the fact that certain claim types may be compulsorily bracketed together in the coverages available to policy owners. Consider the implications of this for modelling.

Let

$P_{ijk}^t =$ the value of P_{ijk} for claim type t.

Anticipating Section 6.1 somewhat, suppose that P_{ijk}^t has the **multiplicative** form:

$$P_{ijk}^t = P_{i_0 j_0 k_0}^t \, b_{1i}^t \, b_{2j}^t \, b_{3k}^t \, , \tag{5.11}$$

where b_{1i}^t, b_{2j}^t, b_{3k}^t are constant multipliers applicable to claim type t.

Consider a policy providing coverage under all claim types $t \in$ some set T. The risk premium for this policy is

$$\begin{aligned} P_{ijk} &= \sum_{t \in T} P_{ijk}^t \\ &= \sum_{t \in T} P_{i_0 j_0 k_0} \, b_{1i}^t \, b_{2j}^t \, b_{3k}^t \, , \end{aligned} \tag{5.12}$$

by (5.11).

In general, P_{ijk} **cannot** be put in multiplicative form even though its components can - in general, multiplicative models are not closed under addition.

A trivial example will illustrate. Suppose there are only 2 claim types, $T = \{1, 2\}$, and suppose there are only 2 covariates (k may be omitted from (5.11)), and each takes only 2 values. Suppose the relevant parameters are as a follows.

Claim type 1	**Claim type 2**
$P_{11}^{1} = 300$	$P_{11}^{2} = 50$
$b_{11}^{1} = 1$	$b_{11}^{2} = 1$
$b_{12}^{1} = 1.2$	$b_{12}^{2} = 1.1$
$b_{21}^{1} = 1$	$b_{21}^{2} = 1$
$b_{22}^{1} = 0.9$	$b_{22}^{2} = 1$

The resulting risk premium schedules are as a follows.

Claim type 1

		Covariate 2: 1	Covariate 2: 2
Covariate 1	1	300	270
	2	360	324

Claim type 2

		Covariate 2: 1	Covariate 2: 2
Covariate 1	1	50	50
	2	55	55

Claim types 1 and 2

		Covariate 2: 1	Covariate 2: 2
Covariate 1	1	350	320
	2	415	379

Note that this last schedule is **not** multiplicative. It would be if the figure of 379 were replaced by 379.4, or some corresponding change made elsewhere in the schedule.

Thus, the fitting of a multiplicative form to risk premiums for coverage T will introduce some distortion (on the multiplicative assumptions made in respect of the component coverages). The distortion will increase as the differences between the b^t_{mi} for different t increase.

On the other hand claim type will contribute to the non-multiplicativity of total risk premium in proportion with $P^t_{i_0 j_0 k_0}$. If the costs under claim type t are generally small, then its contribution to non-multiplicativity will also be small.

Thus, the final choice of which claim types should be modelled separately will depend very much on questions of materiality.

6 Application of GLMs

6.1 Basic concepts

The use of GLMs has been dealt with extensively by Brockman and Wright (1992) and earlier authors. Hence, only the briefest definitions are given here. Subsequent sub-sections are not intended to constitute a full discussion, but rather just to add to the commentary of the previous authors.

Let $Y_i, 1 = 1, 2, ..., n$, be observations on some random variable. Suppose that they are modelled as a follows:

$$Y_i = h^{-1}\left(\alpha_i + x_i^T \beta\right) + e_i, \tag{6.1}$$

where

h	=	some one-one function with range $(-\infty, +\infty)$, called the **link function**;
x_i	=	a p-vector of covariates associated with the i-th observation;
α_i	=	a scalar parameter, (the **offset**);
β	=	a parameter p-vector;
e_i	=	a drawing from a centred (zero mean) distribution.

The e_i are not necessarily i.i.d., but the distribution of e_i is fixed by α_i and β.

The x_i may include a component 1, applicable to all i.

Note that

$$E\left[Y_i\right] = h^{-1}\left(\alpha_i + x_i^T \beta\right) \tag{6.2}$$

Consider a log link function. In this case

$$h^{-1}(z) = \exp z. \tag{6.3}$$

Then (6.2) gives

$$\begin{aligned} E[Y_i] &= \exp(\alpha_i + x_i^T \beta) \\ &= \exp(\alpha_i) \exp(x_{i1} \beta_1) \exp(x_{i2} \beta_2) \ldots \exp(x_{ip} \beta_p) \end{aligned} \tag{6.4}$$

where the x_{ij} are components of the vector x_i and β_j are components of β. This may be rewritten as:

$$E[Y_i] = \exp(\alpha_i)\, b_{i1}\, b_{i2} \ldots b_{ip}, \tag{6.5}$$

with

$$b_{im} = \exp(x_{im} \beta_m),\ m = 1, 2, \ldots, p. \tag{6.6}$$

Note that the meaning of b_{im} is different from that used in Section 5.2.

Thus, $E[Y_i]$ is the product of $(p + 1)$ factors. Such models are called **multiplicative**. That this is the same as a the multiplicative form anticipated in (5.11) may be seen as follows. Suppose that Y is indexed by the triple *ijk*, putting (6.5) in the form:

$$E[Y_{ijk}] = \exp(\alpha_{ijk})\, b_{ijk1}\, b_{ijk2}\, b_{ijk3}, \text{ for } p = 3. \tag{6.7}$$

A special case of this is given by:

$$b_{ijk1} = b_{i1},\ b_{ijk2} = b_{j2},\ b_{ijk3} = b_{k3},$$

making (6.7) equivalent to (5.11).

Let

$x_i \otimes x_i$ = the **cross-classification** of vector x_i with itself, i.e., $x_i \otimes x_i$ is a p^2 - vector with components $x_{ir}\, x_{is}$, $r, s = 1, 2, \ldots, p$.

The GLM (6.1) can be generalised to the following:

$$Y_i = h^{-1}\left(\alpha_i + x_i^T \beta + (x_i \otimes x_i)^T \gamma\right) + e_i\,, \tag{6.8}$$

where

$\gamma \quad = \quad$ a parameter p^2-vector.

Each of the terms $\gamma_{rs}\, x_{ir}\, x_{is}$ appearing in (6.8) is called a **2-way interaction** term.

In principle, it is possible to include n-way interaction terms, involving $x_i \otimes x_i \otimes \ldots \otimes x_i$ (n factors), in the model.

Note that a log link model which includes one or more interaction terms is **not** multiplicative. If it were, the interaction term could be eliminated by absorption into the 1-way terms.

The statistical packages most commonly used to fit such models, at least in the UK and Australia, are SAS, GLIM and S-PLUS.

It is helpful to interpret the vector x_i further. It concatenates the covariate values for the i -th cell. The covariates will be of two types, conventionally called **categorical** (or **nominal**) and **continuous** (or **ordinal**). Categorical variables are referred to as a FACTOR variables in GLIM and S-PLUS, and as a CLASS variables in SAS.

The structure of the vector x_i is most easily illustrated by means of an example in which there are just 2 covariates, one continuous, one categorical. Then x_i will be conventionally represented with a double subscript ij, where i indexes the continuous variable and j the categorical variable. That is,

$$x_{ij}^T = (x_i, x_j), \tag{6.9}$$

with x_i denoting the values assumed by the two covariates in question.

A continuous variable takes values from an ordered subset of the real numbers, the ordering being physically meaningful. An example would be **age of insured.**

The direct inclusion of continuous covariates in (6.1) would cause them to contribute linearly to the bracketed term there. This may not be appropriate. They will often be subjected to some non-linear transformation (see Section 6.4) before inclusion in (6.1). In this case (6.9) is replaced by:

$$x_{ij}^T = [g(x_i), x_j], \tag{6.10}$$

where x_i is the continuous variable, and $g(.)$ is the transformation applied to it.

A categorical variable simply assigns cases to categories. For example, motor experience may be recorded according to **vehicle category**, which might take the values A, B, C, etc. Sometimes a categorical variable takes numerical values. For example, vehicle categories A, B, C, ... may be relabelled 1, 2, 3, ..., but the numerals then serve as no more than **labels**; they do not necessarily imply any particular ordinal relation between cells involving different values of the variable concerned.

Suppose the categorical variable x_j has q possible values, say $\xi_1, ..., \xi_q$, where the ξ 's are labels of any type. The convenient representation of x_j is a binary q-vector, specifically $x_j = \xi_r$ is denoted by $(0, ..., \underset{r\text{-th}}{1}, 0, ..., 0)$.

In more concise notation x_j is represented here by u_r, the r-th **natural basis vector** in q-dimensional space. The representation of x_{ij} in (6.10) is then replaced by:

$$x_{ij}^T = [g(x_i), u(x_j)], \tag{6.11}$$

where u: $x_j \rightarrow$ {natural basis q-vectors}.

Note that, in this example, although there are only 2 covariates, x_{ij} is a $(q+1)$-vector.

Now consider the meaning of the term $x_i^T\beta$ ($x_{ij}^T\beta$ in the present notation) appearing in (6.1), as it applies to the present example. One calculates:

$$x_{ij}^T\beta = \beta_1\, g(x_i) + \beta_2^T\, u(x_j), \tag{6.12}$$

where β is a $(q+1)$-vector given by

$$\beta^T = (\beta_1, \beta_2^T), \tag{6.13}$$

with β a q-vector.

In the case dealt with above, in which $x_j = \xi_r$ so that $u(x_j) = u_r$, (6.12) reduces to:

$$x_{ij}^T\beta = \beta_1\, g(x_i) + \beta_{2r}\,, \tag{6.14}$$

where β_{2r} is the r-th component of β_2.

This indicates the selector action of u. For each cell ij, u selects the component of β_2 associated with x_j.

The GLM statistical packages usually allow the stochastic term e_i in (6.1) to be chosen from the **exponential family** (see e.g. McCullagh and Nelder, 1989). The uncentred density of this family takes the general form:

$$f(y) = \exp\left\{\frac{y\theta - b(\theta)}{\phi/w} + c(y, \phi/w)\right\}, \tag{6.15}$$

for parameters θ and ϕ, functions a, b and c, and a weight w which is specific to each observation.

This family accommodates many well known distributions.

Let the mean of this distribution be denoted by μ. It is possible to shown that its variance may be expressed in the form:

$$V(\mu)\ \phi/w, \tag{6.16}$$

where $V(\cdot)$ is referred to as the **variance function.**

Let $l(y, \mu)$ be the log likelihood, expressed in terms of the observations y and model mean μ.

The **scaled deviance** of a model is defined as

$$D(y, \mu) = 2\left[l(y, y) - l(y, \mu)\right]. \tag{6.17}$$

For nested models, the difference between deviances has a limiting chi-square distribution under certain regularity conditions. Specifically, $D(y, \nu) - D(y, \mu) \sim \chi_p^2$, where μ, ν are the respective predictions of the models under comparison, with second model nested within the first, and with a difference of p degrees of freedom between the two models.

6.2 Relativities

Models such as a (6.1) and (6.8) can be interpreted in terms of the relativities introduced in Section 2.

In the case of (6.1), this is most easily illustrated by means of the example involving only 2 covariates, represented by (6.9). Assume there is no offset term in (6.1). By the definition given just after (2.5), the relativity of the **continuous covariate** i is:

$$\exp(x_{ij}^T\beta)/\exp(x_{i_0j}^T\beta)$$
$$= \exp\beta_1[g(x_i) - g(x_0)], \quad (6.18)$$

by (6.14).

Here the variable x_i ranges over all cells, i.e. over all values actually **observed**. However, the application of (6.18) may be extended to the continuum of values of x_i lying between its extremes of observation.

For example, suppose that x_i represents sum insured. It may be found that a reasonable form of g is:

$$g(x) = \log x. \quad (6.19)$$

Then (6.18) yields a relativity of:

$$(x_i/x_0)^{\beta_1}. \quad (6.20)$$

This means that, if variable x_j is held constant, the quantity with which the relativity is associated (e.g. risk premium) varies in proportion with the β_1 power of sum insured.

The relativity of the **categorical covariate** j is:

$$\exp(x_{ij}^T\beta) / \exp(x_{ij_0}^T\beta)$$
$$= \exp(\beta_{2r} - \beta_{2r_0}), \quad (6.21)$$

where $x_{j_0} = \xi_{r_0}$. It is always possible to absorb the quantity $\exp(-\beta_{2r_0})$ into the base premium, so that the base case corresponds to $\beta_{2r_0} = 0$. Then the relativity (6.21) reduces to:

$$\exp \beta_{2r}. \quad (6.22)$$

This indicates how β_2 may be regarded as a vector of **scores** associated with the corresponding levels of the categorical variable x_j. Exponentiation converts these scores to relativities.

6.3 Separate modelling of claim frequency and size

Since the ultimate objective of a pricing exercise is to calculate premiums, it is occasionally suggested, on the grounds of directness and simplicity, that modelling should focus on risk premium rather than its factors, claim frequency and average claim size.

A contrary argument would be that a greater understanding of the portfolio would be obtained by analysis of the factors. However, from a technical viewpoint, there is another, highly compelling, argument, concerned with the stochastic properties of the quantity under analysis.

If the claim counts and average claim sizes of the cells are subjected to separate analysis, it will be possible to make a reasonable assumption about the stochastic properties of each (e.g. see Sections 6.6 and 6.7). Consider, however, the situation when only the **aggregate claim amounts** for the cells are analysed.

In the notation of Section 2,

$$C_{ijk} = \sum_{n=1}^{N_{ijk}} C_{ijkn}, \tag{6.23}$$

where

$$C_{ijkn} = \text{amount of } n\text{-th claim in cell } ijk \, (>0).$$

Note that N_{ijk} in (6.23) is stochastic, and so C_{ijk} has a **compound distribution**. For example, if N_{ijk} is Poisson, C_{ijk} is compound Poisson.

Let

$$n_{ijk} = E[N_{ijk}] \tag{6.24}$$

Under the Poisson assumption,

$$\text{Prob}[N_{ijk} = 0] = \exp(-n_{ijk}), \tag{6.25}$$

and this is materially positive for n_{ijk} small.

Note that

$$\text{Prob}[C_{ijk} = 0] = \text{Prob}[N_{ijk}] = 0, \tag{6.26}$$

and so this will have positive probability (6.25).

The conditional distribution of $C_{ijk}|N_{ijk} \geq 1$ will be continuous if the individual claim size distribution is continuous.

Thus the probability density of C_{ijk} will be mixed, as illustrated in Figure 6.1.

Consider two cases of n_{ijk}.

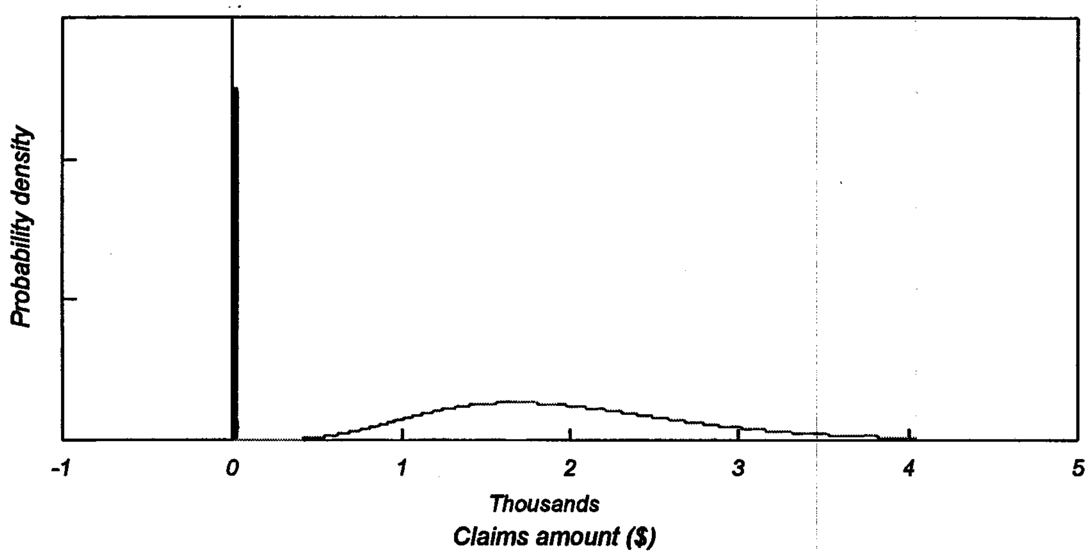
Case I: n_{ijk} small

In the case illustrated by Figure 6.1 $n_{ijk} = 0.25$, and the probability mass of 77.88% at $N_{ijk} = 0$ (shown as a **finite** density here) is visible.

The remaining mass of 21.12% is distributed continuously allowing for an individual claim size distribution which has mean \$2,000 and standard deviation \$800.

Figure 6.1

Distribution of claims amount for small exposure

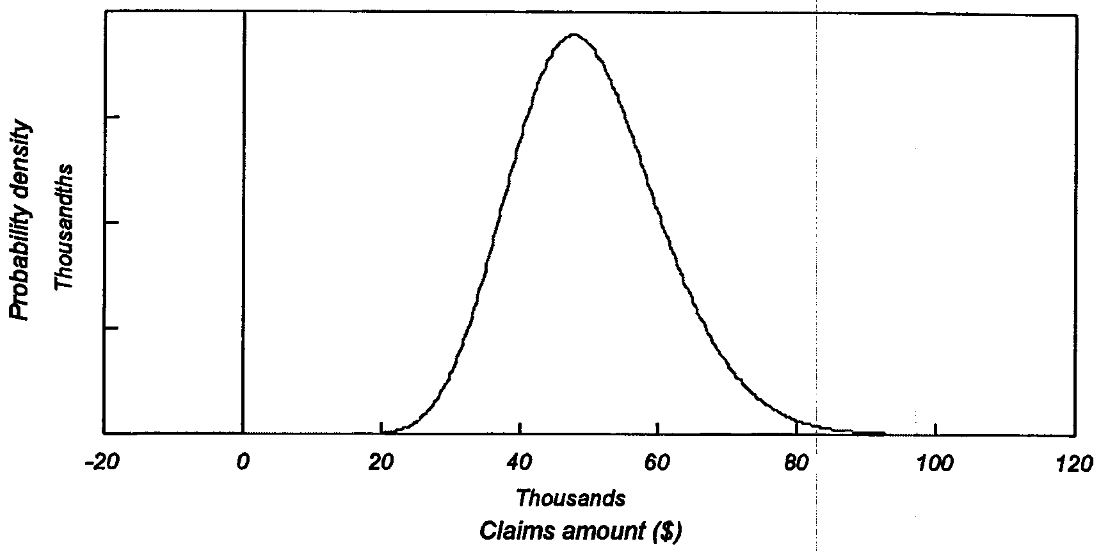


Case II: n_{ijk} large

This time, assume $n_{ijk} = 25$, so that $\text{Prob}[N_{ijk} = 0]$ is essentially zero. The probability mass at zero for C_{ijk} is eliminated, and the density is as in Figure 6.2. This illustration is based on the same individual claim size distribution as in Figure 6.1.

Figure 6.2

Distribution of claims amount for large exposure



Consider the modelling of C_{ijk} in the context of those two examples. The modeller requires a family of distributions with the following properties:

(a) mixed distributions with a probability mass at zero must be admissible;

(b) the size of this mass is not independent of the remainder of distribution, since both depend on n_{ijk}.

It is not easy to see how such a family can be obtained from the standard distributions provided with packages like SAS. For most purposes, separate modelling of claim frequency and size seems preferable.

6.4 Claim size variable

The choice of a model distribution for individual claim size will be considered in detail in Section 6.8. There are, however, some preliminary questions about this issue which should be considered at an early stage of the analysis.

Not least of these is the definition of the claim size variable. While it is always possible to work with the **amount** incurred as claim size, this will not always be the most advantageous course.

Consider, for example, the following distribution of claim sizes for house (building) insurance.

Table 6.1
House (building) claim size distribution

Claim Size $	Cumulative probability %
200	63.8
500	73.1
1,000	80.4
2,000	87.3
5,000	94.6
10,000	98.0
20,000	99.5
50,000	99.94
100,000	99.98
200,000	99.997
500,000	100
Mean	$1,210
Coefficient of variation	3.76

This distribution is very long tailed, and estimation of its mean correspondingly difficult. This situation can be eased somewhat by recognising that

$$\underset{(\$)}{\text{claim size}} = \text{sum insured} \times \underset{(\text{\% of sum insured})}{\text{claim size}} \qquad (6.27)$$

and that the distribution set out in Table 6.1 is the compound of the two in Table 6.2.

Table 6.2
Sum insured and claim size distribution

Sum insured $	Cumulative probability %	Claim size (% of sum insured) %	Cumulative probability %
20,000	2.9	2	85.9
40,000	13.3	5	94.4
60,000	28.2	10	98.3
80,000	43.9	20	99.7
100,000	58.0	50	99.95
125,000	72.0	100	100
150,000	82.1		
175,000	88.9		
200,000	93.2		
250,000	97.6		
300,000	99.2		
350,000	99.8		
400,000	99.92		
450,000	99.98		
500,000	100		
Mean	100,317	1.21	
Coefficient of variation	0.61	3.17	

The distribution of claim size is still awkwardly long tailed, but less so than in Table 6.1; a part of the variation of $ claim size has been removed as representing variation in sum insured, which is known.

Typically, a house insurance analysis would model claim size as a proportion of sum insured. This is less common in motor accidental damage, however, where third party costs tend to break the nexus between sum insured and claim size. But if a separate model of own damage is constructed, as illustrated in Figure 5.1, it might usefully relate claims to sum insured.

Similarly in that example, the amount of third party recoveries might be modelled as:

$$\text{third party recoveries} = \text{own damage} \times \text{proportion of damage recovered.} \qquad (6.28)$$

6.5 Significance testing of model

Consider a model of the form (6.8). It will be possible to find the maximum likelihood estimates of the vectors β, γ, but some assessment of the significance of their components will be required.

As in Section 6.1, it will be useful to consider continuous and categorical variables separately. Again, the example of (6.9)-(6.14) will be useful.

A statistical package such as those mentioned earlier will produce an estimate of the parameters (6.13), together with standard errors and significance statistics. For the case $q = 4$ in the example, this would typically take the form set out in Table 6.3.

Table 6.3
Parameter estimates

	Variable	Parameter Estimate	Parameter Standard error	Chi-square probability
$g(x_1)$	Age (continuous) variable	-0.0040	0.0008	0.000
x_2	Vehicle category			
	A	0.2963	0.0403	0.000
	B	0.1894	0.0554	0.002
	C	0.1115	0.2002	0.531
	D	0		

In this table, the "Chi-square probability" is the significance statistic, the probability that an estimate of the relevant β greater than that tabulated would have arisen by chance under the null hypothesis that β is zero. It is the χ_1^2 test statistic obtained by comparing the scaled deviances (6.17) with and without the variable value under test.

The treatment of the continuous variable in the table is straightforward. The coefficient -0.0040 is shown to be highly significant, and should be incorporated in the model under test.

As to the categorical variable, the form of the output requires a little discussion. Note first that the coefficient associated with vehicle category *D* is estimated as exactly zero, with no standard error or significance information.

This reflects the **redundancy** inherent in the use of categorical variables. Although the example of (6.9)-(6.14) is expressed in terms of only the 2 covariates x_1 and x_2, the model would usually contain an "intercept term" $\beta_0 x_0$ also, where x_0 is a scalar taking unit value **for all observations**. This possibility was mentioned just prior to (6.2).

With this addition, (6.12) becomes

$$x_{ij}^T \beta = \beta_0 + \beta_1 g(x_i) + \beta_2^T u(x_j). \tag{6.29}$$

Written more explicitly, with $\beta_2^T = (\beta_{21}, \beta_{22}, \beta_{23}, \beta_{24})$ this is:

$$x_{ij}^T \beta = \beta_0 + \beta_1 g(x_i) + \sum_{k=1}^{4} \beta_{2k}\, u_k\,(x_j)$$
$$= (\beta_0 + \beta_{24}) + \beta_1 g(x_i) + \sum_{k=1}^{4} (\beta_{2k} - \beta_{24})\, u_k\,(x_j), \qquad (6.30)$$

where use has been made of the fact that $u(x_j)$ is a natural basis vector, whence $\sum_{k=1}^{4} u_k\,(x_j) = 1$.

Now (6.30) shows that the fourth component of β_2 can be absorbed into the intercept β_0, with adjustment of β_{2k} to $\beta_{2k} - \beta_{24}$. In this case, the fourth component of β_2 becomes identically zero and so has no stochastic properties, as Table 6.3 indicates.

The appearance of $\beta_{2k} - \beta_{24}$ in (6.30) emphasises the fact that in the example the results for vehicle categories A, B and C measure their respective differences from D. The significance statistics indicate that A and B, but not C both differ significantly from D.

In general, whenever the model contains an intercept term, the β's associated with a categorical variable will measure the differences between the experience at different values of that variable and some base case. The most informative results will be obtained if the base case is selected judiciously. This usually means setting the base equal to a value of the relevant variable for which there is a reasonable volume of data.

An injudicious selection can produce unhelpful results. For example, suppose vehicle category C had been chosen as the base case in Table 6.3. The results might have appeared as in Table 6.4.

Table 6.4
An injudicious choice of base vehicle category

	Variable	Parameter		
		Estimate	Standard error	Chi-square probability
$g(x_1)$	Age (continuous) variable	-0.0040	0.0008	0.000
x_2	Vehicle category			
	A	0.1848	0.1924	0.688
	B	0.0779	0.1778	0.583
	C	0		
	D	-0.1115	0.2002	0.531

In this representation, no vehicle category appears significant. This is because all three tabulated standard errors are large. But recall from the comments earlier in the present sub-section that Table 6.4 in fact indicates that the vehicle categories A, B and D are

insignificantly different from C. Equivalently, vehicle category C is insignificantly different from A, B and D.

Returning to Table 6.3, recall that categories A and B were found significantly different from D. But the table provides no information on whether they differ significantly from each other.

Using the same notation as appears in (6.27), let $\hat{\beta}_{2k}$ be an estimate of β_{2k}, let σ_k^2 be its estimated standard error, and let ρ_{kl} be the estimated correlation between $\hat{\beta}_{2k}$ and $\hat{\beta}_{2l}$. Then $\beta_{2k} - \beta_{2l}$ is estimated by $\hat{\beta}_{2k} - \hat{\beta}_{2l}$ and

$$V[\hat{\beta}_{2k} - \hat{\beta}_{2l}] = \sigma_k^2 + \sigma_l^2 - 2\rho_{kl}\sigma_k\sigma_l. \tag{6.31}$$

As a very rough test of the null hypothesis $\beta_k = \beta_l$, one might examine the significance of the statistic:

$$U_{kl} = (\hat{\beta}_{2k} - \hat{\beta}_{2l}) \,/\, (\sigma_k^2 + \sigma_l^2 - 2\rho_{kl}\sigma_k\sigma_l)^{1/2}, \tag{6.32}$$

against a standard normal distribution.

Suppose the matrix of correlations ρ_{kl} associated with Table 6.3 were:

$$\begin{bmatrix} 1 & 0.61 & 0.29 & 0.12 \\ & 1 & 0.52 & 0.34 \\ & & 1 & 0.49 \\ & & & 1 \end{bmatrix}$$

The matrix of test statistics U_{kl} would be as in Table 6.5.

Table 6.5
Significance of differences between vehicle categories

	Test statistic U_{kl} for l =			
k =	A	B	C	D
A	0	2.41	0.96	7.35
B		0	0.44	3.42
C			0	0.57
D				0

The table suggests that A and B are significantly different from each other, as well as D.

Relation (6.31) also explains the large standard errors in Table 6.4. For

$$\begin{aligned} V[\hat{\beta}_{2k} - \hat{\beta}_{2C}] &= \sigma_k^2 + \sigma_C^2 - 2\rho_{kC}\sigma_k\sigma_C \\ &= \sigma_C^2[1 - 2\rho_{kC}(\sigma_k/\sigma_C) + (\sigma_k/\sigma_C)^2] \\ &= \sigma_C^2\{[1 - \rho_{kC}(\sigma_k/\sigma_C)]^2 + (1 - \rho_{kC}^2)\ (\sigma_k/\sigma_C)^2\}, \end{aligned} \qquad (6.33)$$

which is dominated by σ_C^2 when $\rho_{kC}\ (\sigma_k/\sigma_C)$ is small. Thus, in a case such as represented in Table 6.3, all values of $V[\hat{\beta}_{2k} - \hat{\beta}_{2C}]$ will be of the order 0.2, as indeed appears in Table 6.4.

6.6 Continuous covariates

By (6.8) and (6.10), a continuous covariate appears in the model in the following way:

$$Y_i = h^{-1}\left(\ldots + \beta\, g(x_i) + \ldots\right) + e_i, \qquad (6.34)$$

where β is a scalar and g is some real-valued one-one function defined over the range of the continuous covariate x_i.

Typically, there will be little, if any, prior knowledge of the form of g. This will need to be determined empirically.

The standard procedure for making this determination is to treat the covariate x_i initially as categorical. This is done by dissecting its continuous range into a number of bands, or sub-ranges. The number of these should be sufficient to provide an indication of the shape of g, but not so many that the data in each band become sparse and provide unreliable results.

As an example, consider the effect of sum insured on claim frequency in a Motor collision damage experience. Table 6.6 indicates the results obtained when sum insured is treated as a categorial covariate. The model contains a log link.

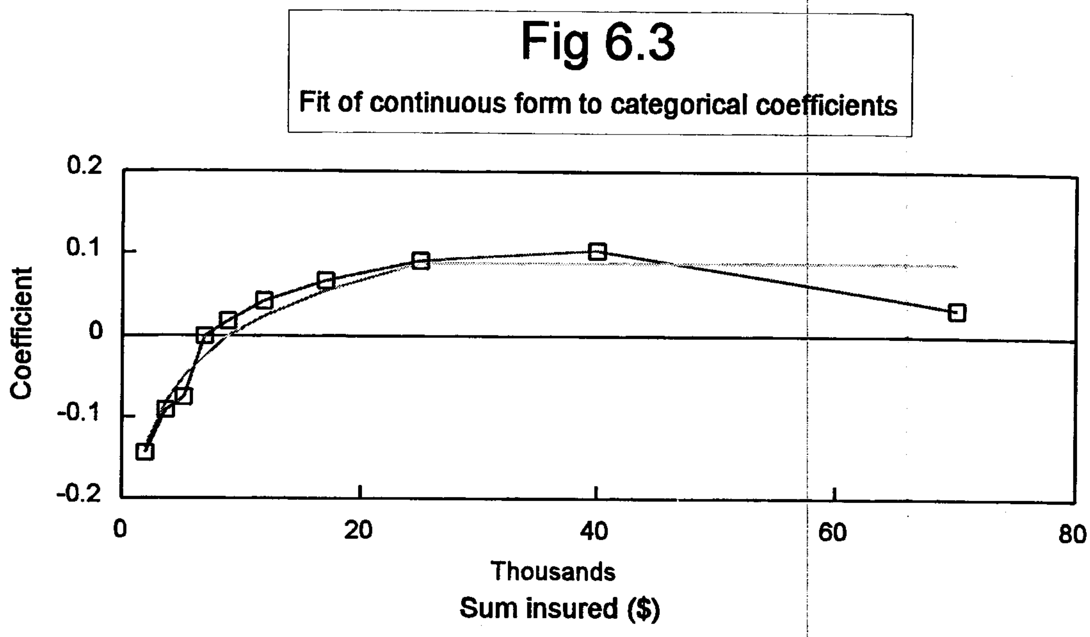
Table 6.6
Effect of sum insured on claim frequency

Range of sum insured $	Estimated β	Standard error
up to 3,000	-0.143	0.041
3,000 - 4,500	-0.090	0.032
4,500 - 6,000	-0.074	0.029
6,000 - 8,000	0	0
8,000 - 10,000	+0.018	0.027
10,000 - 14,000	+0.041	0.026
14,000 - 20,000	+0.066	0.028
20,000 - 30,000	+0.091	0.033
30,000 - 50,000	+0.103	0.041
>50,000	+0.033	0.088

Figure 6.3 plots the estimates of β parameters against the mid-values of the ranges of sums insured. The end-ranges are treated as concentrated at $2,000 and $70,000 respectively.

The plot indicates a fairly clear upward convexity. This suggests the possibility of a logarithmic representation of $g(\cdot)$. This is encouraged by Figure 6.3, which also includes a logarithmic fit to the estimated β parameters.

The fit has been obtained by simple linear regression of these estimates against the logged central values of the sum insured ranges though with sums insured in excess of $25,000 recognised as only $25,000. While this is a very crude fitting procedure, it is superseded at the next step. Once the logarithmic form of $g(\cdot)$ has been decided, it is formally included in (6.34) and β estimated properly.



Note that with $g(\cdot) = \log(\min(\cdot, 25{,}000))$ and a log link, (6.34) yields:

$$E[Y_i] = \dots [\min(x_i, 25{,}000)]^{\beta} \dots, \tag{6.35}$$

indicating that sum insured influences claim frequency according to a power law.

Table 6.7 gives a comparison of the model obtained by treating sum insured as a categorical and continuous variable respectively. It shows that adoption of the latter rather than the former increases scaled deviance by only 2.2 while increasing the d.f. by 8. The increase in deviance is

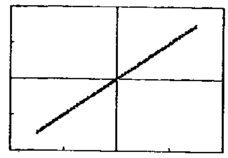
Table 6.7
Comparison of categorical and continuous forms of modelling sum insured

Representation of sum insured	Scaled deviance	Degrees of freedom used by sum insured model
Categorical	117583.8	9
Continuous	117586.0	1
Change	+2.2	-8

quite significant as a χ_8^2 variable (see the discussion following (6.17)), indicating that the continuous model is preferable.

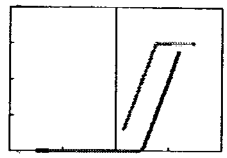
It is worthwhile carrying a small toolkit of basis functions for use in modelling continuous variables. Obvious candidates for inclusion would be as listed below.

Identity function



$$g(x) = x.$$

Broken stick functions



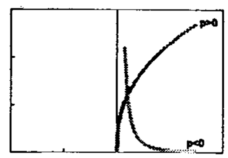
$$g(x) = \max(0, x-c)$$

and

$$g(x) = \min(x, c),$$

for some constant c .

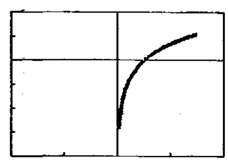
Power functions



$$g(x) = x^p,$$

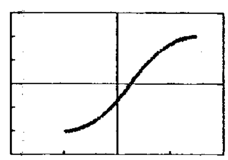
for some constant $p \neq 0$.

Logarithmic function



$$g(x) = \log x$$

Curve of squares



$$\begin{aligned} g(x) &= -1 + \left[1 + (x-c)/k\right]^2, \; c-k \le x \le c, \\ &= +1 - \left[1 - (x-c)/k\right]^2, \; c \le x \le c+k, \end{aligned}$$

for constants $k > 0$ and c .

Models for continuous covariates would be constructed as combinations of the basis functions. For example, the curve fitted in Figure 6.3 combines the logarithmic function with one of the broken stick functions.

The identity and broken stick functions may be used to generate the linear splines:

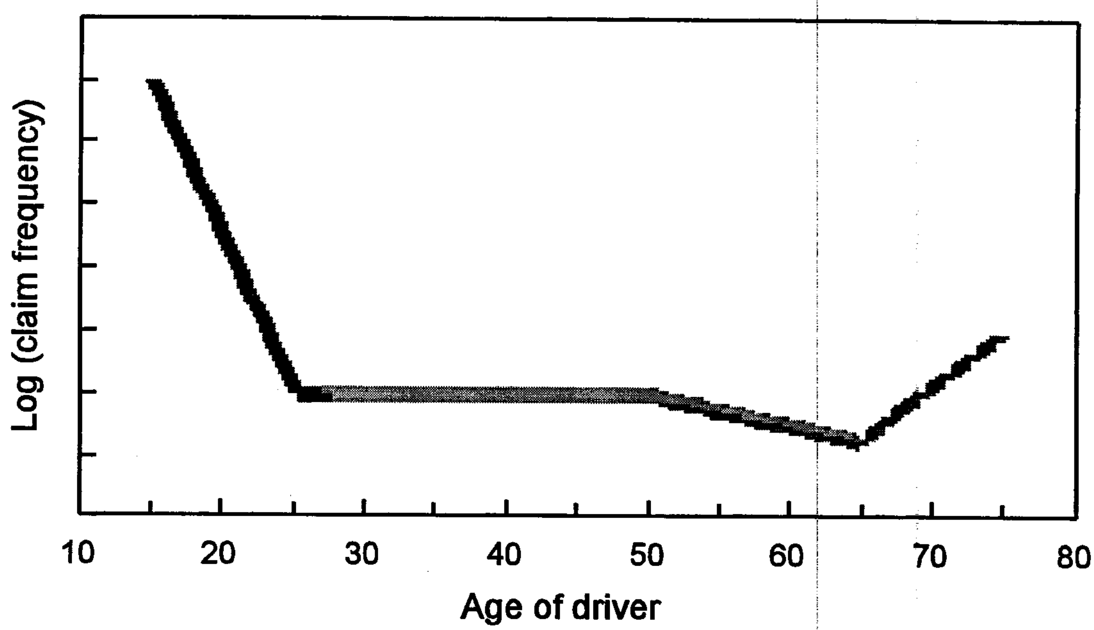
$$g(x) = \beta_0 x + \sum_{j=1}^{m} \beta_j \max(0, x - c_j), \qquad (6.36)$$

for given knots $c_1 < c_2 < \ldots < c_m$ and unknown parameters $\beta_0, \ldots, \beta_m$, which is a piecewise linear function, with gradient β_0 for $x < c_1$, gradient $\beta_0 + \beta_1$ for $c_1 < x < c_2$, etc.

Often the effect of age of insured on Motor claim frequency takes the general form indicated in Figure 6.4.

Fig 6.4

Claim frequency by age



This can be modelled in the form (6.36), although this is liable to produce a small, statistically insignificant, but non-zero value of $\beta_0 + \beta_1$ for the gradient over age range 25-50. It may be preferable to modify the representation of the linear spline to set its gradient explicitly to zero over this range:

$$g(x) = \beta_1 \max(0, 25 - x) + \beta_2 \max(0, x - 50) + \beta_3 \max(0, x - 65). \qquad (6.37)$$

The curve of squares is useful for interpolation between two levels, thus:

$$g(x) = \beta_0 + \beta_1 s(x), \quad c-k \le x \le c+k,$$
$$= \beta_0 + \beta_1, \quad x \ge c + k,$$
$$= \beta_0 - \beta_1, \quad x \le c-k, \tag{6.38}$$

where $s(x)$ is the curve of squares given earlier. Note that $g(x)$ is smooth (continuously differentiable) in (6.38).

6.7 Claim frequency

The number of claims observed in respect of a single policy over one year is typically assumed Poisson. If all policies in cell $\{i, j, k\}$ are identical, then the number of claims observed in that cell is

$$N_{ijk} \sim \text{Poiss}\left(E_{ijk} f_{ijk}\right). \tag{6.39}$$

One consequence of this is that

$$V[N_{ijk}] = E[N_{ijk}] = E_{ijk} f_{ijk}. \tag{6.40}$$

As has been observed on numerous occasions in the literature (e.g. Johnson and Hey, 1971, 216-220), (6.40) does not hold over relatively large subsets of a portfolio. In fact, a negative binomial distribution is often found to provide a superior fit (e.g. Beard, Pentikäinen and Pesonen, 1984, p.45) since it is consistent with the empirical fact that

$$V[N_{ijk}] > E[N_{ijk}], \tag{6.41}$$

instead of (6.40).

This last relation is usually attributed to within-cell heterogeneity, and there is perhaps a question as to whether such heterogeneity occurs even within small cells. It can be shown (e.g. Panjer and Willmot (1992, p.92) that N_{ijk} will be negative binomial if:

- each unit of exposure in the cell is subject to its own unique frequency; and
- these individual frequencies are gamma distributed with mean f_{ijk}.

The relation (6.4) is referred to as **over-dispersion**. There are two main methods of dealing with it.

Quasi-Poisson frequencies

First, one may replace the Poisson error assumption (6.39) by **quasi-Poisson**. This means freeing ϕ to take any positive value in (6.15) rather than fixing it at unity, as is the case for a Poisson distribution.

This leads to the following special cases of (6.15):

Poisson:

$$f(y) = e^{-f} f^y / y!, \quad y = 0, 1, 2, \text{etc.} \tag{6.42}$$

Quasi-Poisson:

$$f(y) = (e^{-f} f^y)^{1/\phi} / y!, \quad y = 0, 1, 2, \text{etc.} \tag{6.43}$$

Note that (6.43) is not a true likelihood for $\phi \neq 1$, since it does not integrate to unity. It is referred to as a **quasi-likelihood** (see e.g. McCullagh and Nelder, 1989).

The fitting of a quasi-Poisson GLM gives the same coefficients as for a Poisson GLM. However, the change in dispersion changes all standard errors and resulting significance test statistics. The major GLM packages provide a reasonably accessible quasi-Poisson option.

Negative binomial frequencies

The frequency function is:

$$f(y) = \binom{y + \alpha - 1}{y} p^y q^\alpha, y = 0, 1, 2, \text{etc.}, \tag{6.44}$$

with

$$q = 1 - p, 0 < p < 1, \alpha > 0. \tag{6.45}$$

The associated cumulant generating function is:

$$\alpha \log \left[q/(1 - pe^{-x})\right]. \tag{6.46}$$

A reasonable assumption for N_{ijk} is then

$$N_{ijk} \sim \text{Neg } Bin\ (p,\ E_{ijk}\ \alpha_{ijk}), \tag{6.47}$$

which gives

$$E[N_{ijk}] = E_{ijk}\ \alpha_{ijk}\ p/q, \tag{6.48}$$

$$V[N_{ijk}] = E_{ijk}\ \alpha_{ijk}\ p/q^2, \tag{6.49}$$

and hence an over-dispersion factor of

$$V[N_{ijk}]\ /\ E[N_{ijk}] = 1/q > 1. \tag{6.50}$$

By comparison of (6.48) with (2.1),

$$\alpha_{ijk} = f_{ijk}\ q/p, \tag{6.51}$$

and so, just as for the Poisson case, it is natural to model log α_{ijk} as a linear function of covariates.

The negative binomial distribution, by virtue of (6.50) provides a genuine distribution incorporating over-dispersion. The form of over-dispersion is simple, being constant from cell to cell. Unfortunately, however, the distribution does not fall within the exponential family (6.15). It is not, therefore, a straightforward application of the major GLM packages.

6.8 Claim size

Suppose initially that an individual claim size is modelled simply as a dollar amount. Variations of this assumption will be considered later in this sub-section.

Unlike claim frequency (Section 6.7), claim size is not subject to "natural" distributions. Virtually any distribution concentrated on the positive half-line is a candidate.

Those which fit into the exponential family and are explicit options in the major GLM packages are:

- normal;
- gamma;
- inverse Gaussian

together with their one-one transformations. Selection from this range of possibilities is discussed in Section 7.1.

Consider (6.1) applied to claim size. It would be possible, as with claim frequency, to work with cells, in which case Y_i would be the cell average claim size.

It is advisable not to do this, however, but to work with **individual** claim sizes. A number of the families of claim size distributions considered below are not closed under averaging. For example, the average of a set of log normal variables is not log normal. In this case, a log normal assumption for individual claim sizes leads to an unknown form of e_i in (6.1), as applied to cells.

Even in the case of families closed under averaging, e.g. gamma, the averaging can destroy the identity of distribution in the e_i. This would cause the standardised residuals of different cells (Section 7.3) to be differently distributed, and would cause difficulty in the construction of quantile plots.

The allowance of claim size transformations generalises (6.1) to the following:

$$g(Y_i) = h^{-1}(\alpha_i + x_i^T\beta) + e_i, \tag{6.52}$$

with one-one g: $(0, \infty) \to (-\infty, +\infty)$.

These transformations considerably widen the family of claim size distributions available. For example, if $g(\cdot) = \log(\cdot)$, then e_i normal (gamma) gives $Y_i \sim$ log normal (log gamma). Note that log gamma includes Pareto as a special case (when the gamma distribution degenerates to negative exponential).

Considerable care is needed in the implementation of non-linear transformations g in (6.52), since

$$\begin{aligned} E[Y_i] &= E\{g^{-1}[h^{-1}(\alpha_i + x_i^T\beta) + e_i]\} \\ &\neq g^{-1}[h^{-1}(\alpha_i + x_i^T\beta)], \end{aligned} \tag{6.53}$$

even though $E[e_i] = 0$.

Write instead

$$E[Y_i] = g^{-1}[h^{-1}(\alpha_i + x_i^T\beta)] \times b_i \tag{6.54}$$

where b_i is the **bias correction factor**:

$$b_i = \frac{E\{g^{-1}[h^{-1}(\alpha_i + x_i^T\beta) + e_i]\}}{g^{-1}[h^{-1}(\alpha_i + x_i^T\beta)]}. \tag{6.55}$$

Not all transformations g yield multiplicative models in conjunction with a given link h. Pairs (g, h) which do so are as follows.

link	g
identity	log
log	power

The power transformation appearing here takes the form

$$g(y) = y^p, p \neq 0. \tag{6.56}$$

Note that the identity transformation is included as a special case ($p = 1$) of the power transformation. In this case, (6.55) reduces immediately to $b_i = 1$, and no bias correction is required.

This renders the inverse Gaussian distribution a favourable choice for long tailed distributions, since it accommodates such distributions without a data transformation, and eliminates the potential difficulties arising from bias correction (see e.g. Section 7.4.2).

The identity-log case is dealt with in (6.1)-(6.6). For the log-power case, note that the first member on the right side of (6.54) is

$$\begin{aligned} g^{-1}[h^{-1}(\alpha + x_i^T\beta)] &= [\exp\ (\alpha + x_i^T\beta)]^{1/p} \\ &= \exp\ [(\alpha/p) + x_i^T(\beta/p)]. \end{aligned} \tag{6.57}$$

This demonstrates the model to be multiplicative, but also shows that the β coefficients need to be divided through by p before they can be regarded as scores in the sense described at the end of Section 6.2.

Appendix A calculates the bias correction factor (6.55) for three useful cases, viz.

(1) log transformation - identity link

(a) log normal distribution
(b) log gamma distribution

(2) power transformation - log link

power gamma distribution

The results of Appendix A may be summarised as follows.

Log normal Assume that

$$\log Y_i \sim N(\mu_i, \sigma^2), \tag{6.58}$$

where μ_i is modelled linearly, as in (6.52):

$$\mu_i = \alpha_i + x_i^T \beta, \tag{6.59}$$

and σ^2 is independent of i .

This last assumption is equivalent to assuming that all claim sizes have the same coefficient of variation. It then follows (see (A.3)) that all claim sizes have the same bias correction factor.

Power gamma Assume that Y_i^p, $p \neq 0$, is gamma distributed with mean (6.59) and coefficient of variation independent of i . It follows once again (see (A.13)) that all claim sizes have the same bias correction factor.

Log gamma Assume that $\log Y_i$ is gamma distributed with mean (6.59) and coefficient of variation v, independent of i . In this case (see (A.10)), the bias correction factor is **not** the same for all claim sizes. It is, in fact,

$$b_i = (1 - \mu_i v^2)^{-1/v^2} \exp(-\mu_i), \tag{6.60}$$

which is shown in Appendix A.1.2 to increase with μ_i.

It is possible to arrange that the log gamma bias correction factor is constant for all claim sizes by allowing v to vary with μ. In this case, the constant v is replaced by $v(\mu_i)$ and (6.60) by:

$$b_i = \left[1 - \mu_i v^2(\mu_i)\right]^{-1/v^2(\mu_i)} \exp(-\mu_i). \tag{6.61}$$

Then $v(\cdot)$ is chosen to render b_i independent of μ_i. This is awkward, however, as the required $v(\cdot)$ cannot be expressed in closed form.

In any event, the application of this functional relationship between v and μ_i would be valid only if justified by the data. Indeed, the same remark may be made about the assumptions of constant coefficient of variation which occur above in connection with log normal and power gamma claim sizes. The testing of such assumptions forms the subject of Section 7.2.

Bias corrections that are constant for all claim sizes have the great attraction that they factor out of claim size relativities. In this case, all claim size relativities may be calculated without any consideration of bias correction.

If, however, one wishes to compare actual and model claim sizes (see Section 7.3), the correction factors will be required.

6.9 Interactions

Consider the interaction terms appearing in (6.8). A **saturated** second order model, one including all possible 2-way interactions, will contain a great many parameters. It is virtually certain to be over-parameterised, in which case even the estimates of the main effects (the β's in (6.8)) will be unstable.

Care is thus needed in the inclusion of interactions. For the most part, it seems advisable to defer their consideration until after completion of modelling the main effects.

Exceptions to this dictum may be made in respect of specific preconceptions. For example, one may commence a Motor study with a strong view than an age-gender interaction will be present. Figure 6.5 might exemplify the expectation.

Fig 6.5

Claim frequency age-gender interaction

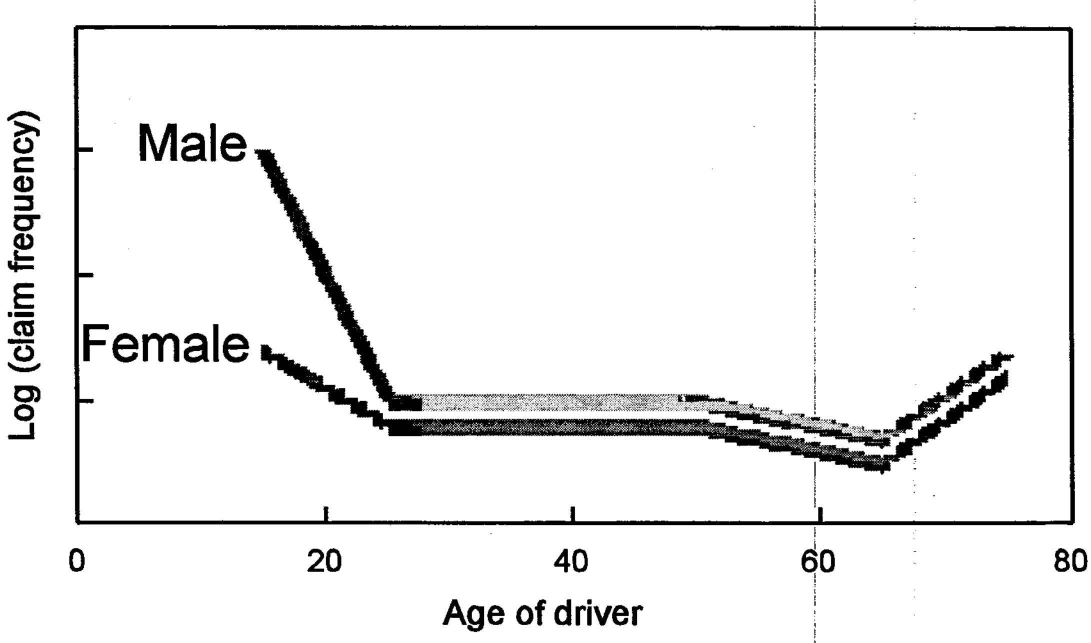


Figure 6.5 reproduces Figure 6.4 as the male age profile of log (claim frequency), and adds a female age profile to it. The two profiles are represented by the following variation of (6.37).

$$\begin{aligned} g(x,y) &= \beta_1 \max(0, 25-x) + \beta_2 \max(0, x-50) + \beta_3 \max(0, x-65) \text{ [age effect]} \\ &+ \beta_4 I(y) \qquad \text{[gender effect]} \\ &+ \gamma I(y) \max(0, 25-x) \qquad \text{[age-gender interaction]} \end{aligned} \tag{6.62}$$

where $I(\cdot)$ is the binary gender variate:

$$\begin{aligned} I(y) &= 0 \text{ if } y = \text{male}; \\ &= 1 \text{ if } y = \text{female}. \end{aligned} \tag{6.63}$$

Note that the existence of an interaction is indicated by non-parallelism of the two profiles in Figure 6.5. Parallelism ($\gamma = 0$) would have indicated a constant female/male relativity (of $\exp \beta_4$). The interaction allows this relativity to vary at ages below age 25, where Figure 6.5 indicates that male frequency increases more rapidly than female with decreasing age.

Even if interactions other than exceptions such as above are deferred until the main effects have been modelled, there remain many of them to be considered. The most practical procedure appears to consist of testing 2-way interactions one by one for significance, and then testing a model which incorporates simultaneously just those interactions which are significant in isolation.

Usually, 3-way (and higher order) interactions will be too numerous for exhaustive testing, and would only be tested on the basis of some preconception; e.g. vehicle age-policy age-sum insured, where there is likely to be a high degree of association between the three nominated variables.

Moreover, the feasibility of reliable estimation of interactions will be heavily dependent on the volume of experience under analysis.

In summary, the suggested procedure is as follows.

1. Model all main effects.
2. Test 2-way interactions one by one for significance.
3. Test a model which includes main effects and all 2-way interactions which pass a significance test in isolation.
4. Possibly eliminate some of the terms in model 3.
5. Consider whether any 3-way interactions should be included.

7 Model validation

7.1 General

The procedures of model formulation (Section 6) and model validation are presented sequentially here. In practice, however, there is likely to be a good deal of cycling between the two.

For example, the selection of a family of claim size distributions with which to work in Section 6.8 can be made properly only with application of the claim size validation technique discussed in Section 7.2.

Similarly, application of the further validation techniques set out in Sections 7.3 and 7.4 might lead to modification of the model under consideration.

7.2 Quantile plots

These are used to check whether observed claim sizes are consistent with the family of distributions assumed for them in the error term of (6.52). Details of the plot's construction depend on the family tested, but the general procedure is as follows.

Let $s(Y_i)$ be a **standardised observation** associated with the i-th claim size. This will not necessarily be of the same form as the standardised residuals conventionally defined in connection with General Linear (Gaussian) Models, but will be constructed in such a way that the $s(Y_i)$ are ***iid*** when (6.52) holds. Let D denote the common distribution.

The $s(Y_i)$ are then ordered to produce an empirical distribution. Specifically, let

Q_p = value of standardised observation which has 100p% of these values below it;

p^D = the probability of falling below Q_p in the distribution D.

The quantile plot consists of a plot of p^D against p, taken over all residuals.

Now

$$p^D = D(Q_p), \tag{7.1}$$

if $D(\cdot)$ is used to denote the ***d.f.*** of distribution D. Further, if (6.52) holds, then the definition of Q_p yields:

$$D(Q_p) = p + \text{sampling error}, \tag{7.2}$$

so that

$$p^D = p + \text{sampling error.} \tag{7.3}$$

It is evident from (7.3) that linearity of the quantile plot (subject to sampling error) supports the family of distributions hypothesised for e_i in (6.52); and non-linearity contradicts this family.

A different form of $s(\cdot)$ needs to be defined for each choice of e_i in (6.52).

Log normal error term

Let e_i be log normal, i.e.

$$\log g(Y_i) = (\alpha_i + x_i^T\beta) + \epsilon_i, \tag{7.4}$$

with

$$\epsilon_i \sim N(0, \sigma^2), \tag{7.5}$$

for some constant $\sigma^2 > 0$.

Usually $g(\cdot)$ would be the identity transformation.

By (7.4) and (7.5),

$$\left[\log g(Y_i) - (\alpha_i + x_i^T\beta)\right] \sim N(0, \sigma^2). \tag{7.6}$$

It is also true that the $s(Y_i)$ are equi-distributed when defined as:

$$s(Y_i) = \log g(Y_i) - \hat{\mu}_i, \tag{7.7}$$

with $\hat{\mu}_i$ the GLM estimate of $\alpha_i + x_i^T\beta$. These $s(Y_i)$ will then serve for the construction of the quantile plot.

In fact, it is more usual to define

$$s(\mathrm{Y_i}) = \left[\log g(\mathrm{Y_i}) - \hat{\mu}_i\right] / \hat{\sigma}, \tag{7.8}$$

where $\hat{\sigma}$ is the GLM estimate of σ. The standardised observation then becomes equal to the standardised residual (Section 7.3).

Gamma error term

Let e_i be gamma in (6.52). This covers the case of $\mathrm{Y_i}$ power gamma distributed. Specifically, suppose that

$$\mathrm{Y_i^p} \sim \Gamma(\mu_i, \gamma), \tag{7.9}$$

meaning that $\mathrm{Y_i^p}$ is gamma distributed with a mean of μ_i and a coefficient of variation $1/\gamma^{½}$, i.e. $\mathrm{Y_i^p}$ (= Z_i, say) has the *pdf*

$$\left[\Gamma(\gamma)\right]^{-1} c_i^{\gamma} z^{\gamma-1} \exp - c_i z, \tag{7.10}$$

with

$$c_i = \gamma / \mu_i. \tag{7.11}$$

By (7.9),

$$\mathrm{Y_i^p} / \mu_i \sim \Gamma(1, \gamma), \tag{7.12}$$

which is independent of i. Thus, define equi-distributed

$$s(\mathrm{Y_i}) = \mathrm{Y_i^p} / \hat{\mu}_i, \tag{7.13}$$

where $\hat{\mu}_i$ is again the GLM estimate of μ_i.

Log gamma error term

Let e_i be log gamma in (6.52), i.e.

$$\log g(Y_i) \sim \Gamma(\mu_i, \gamma). \tag{7.14}$$

As for log normal, $g(\cdot)$ would usually be the identity transformation.

By the same reasoning as for the power gamma error, define

$$s(Y_i) = (\log Y_i)/\hat{\mu}_i. \tag{7.15}$$

Inverse Gaussian error term

Let e_i be inverse Gaussian in (6.52):

$$Y_i \sim IG(\mu_i, \gamma) \tag{7.16}$$

meaning that Y_i has the *pdf*

$$\mu_i(2\pi\gamma x^3)^{-1/2} \exp[-(y-\mu_i)^2/2\gamma y]. \tag{7.17}$$

It may be shown (Panjer and Willmot, 1992, p. 116) that

$$(Y_i - \mu_i)^2/Y_i \sim \Gamma(\gamma, 1/2). \tag{7.18}$$

By the same reasoning as given in the gamma case, define

$$s(Y_i) = (Y_i - \hat{\mu}_i)^2/\hat{\gamma} Y_i. \tag{7.19}$$

Figures 7.1 to 7.3 illustrate the use of quantile plots. They are drawn from a Motor Collision experience. Figure 7.3 deals with the power gamma case with $p \neq ½$. Each diagram also includes the targeted linear plot.

Figure 7.1
Gamma Quantile Plot

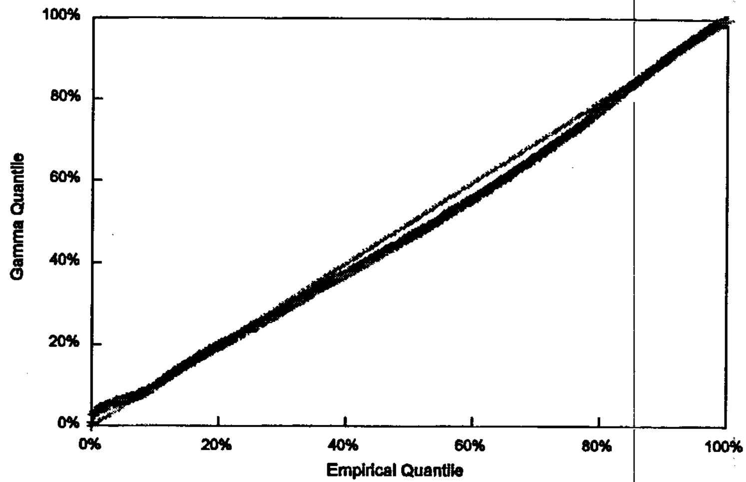


Figure 7.2
Log Normal Quantile Plot

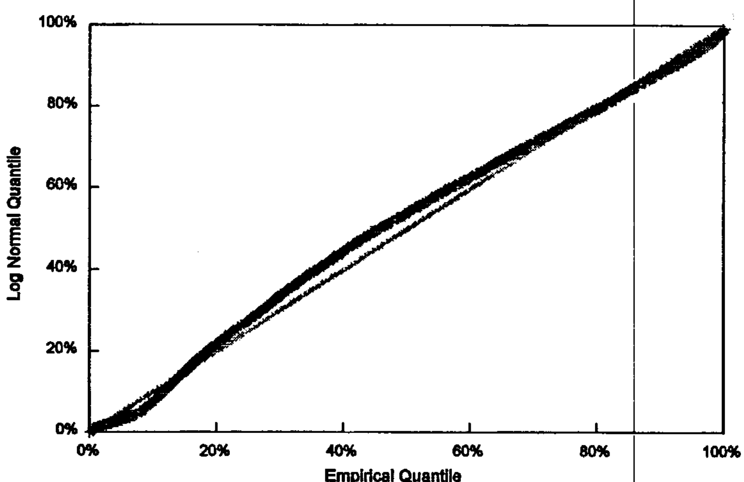
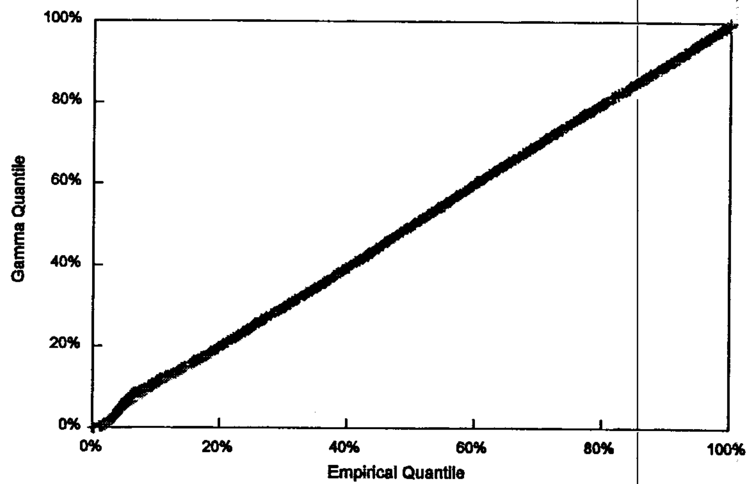


Figure 7.3
Root Gamma Quantile Plot



Note the opposite curvatures in Figures 7.1 and 7.2. When the quantile plot lies above the straight line, it indicates that the observed number of larger observations is greater than expected on the basis of the hypothesised distribution.

Thus, Figures 7.1 to 7.3 indicate that:

- the gamma distribution has too short a tail;
- the log normal distribution has too long a tail;
- the root gamma distribution fits reasonably well.

A protocol that is often useful in testing for distribution consists of the following steps:

1. Test a short tail (gamma) and a long tail (log normal) distribution.

2. If either appears close, adopt it or perhaps search for similar but superior variants, e.g. inverse Gaussian instead of log normal.

3. If Step 1 indicates a distribution intermediate between gamma and log normal, experiment with power gamma distributions with $0 < p < 1$.

4. If Step 1 indicates a shorter tailed distribution than gamma (rare), experiment with power gamma distributions with $p > 1$.

At Step 3 it is advisable to test only moderate values of p, say no lower than ¼. Choosing extreme values of p, e.g. $1/10$, can create the illusion through the quantile plots of approximating the claim size distribution, but (6.57) indicates that in this case the bias correction factor is large. This in itself introduces instability into the results.

7.3 Residual plots

Using the same notation as in Section 7.2, define the **residual** associated with the i-th observation as

$$R_i = h(g(Y_i)) - \hat{\mu}_i. \tag{7.20}$$

Define the **standardised residual**

$$r_i = R_i / \{V[R_i]\}^{1/2}, \tag{7.21}$$

which has mean zero and variance 1. The standardised residual is closely related to the standardised observation of Section 7.2 (though that did not necessarily have zero mean), which may be used to produce specific forms of (7.21).

Log normal error terms

In this case the standardised residual is the same as the standardised observation, given by (7.8).

Power gamma and log gamma error terms

By (7.9) and (7.12) for power gamma, and by (7.14) and (7.12) for log gamma,

$$r_i = \hat{\gamma}^{1/2} \left[h(g(Y_i)) - \hat{\mu}_i\right] / \hat{\mu}_i, \tag{7.22}$$

for $h(g(Y_i)) \sim \Gamma(\mu_i, \gamma)$, and with $\hat{\gamma}$ an estimator of γ.

Inverse Gaussian

When (7.16) holds,

$$E[Y_i] = \mu_i, \; V[Y_i] = \gamma \, \mu_i. \tag{7.23}$$

Then (7.21) yields

$$r_i = (Y_i - \hat{\mu}_i) / [\hat{\gamma} \, \hat{\mu}_i]^{1/2}. \tag{7.24}$$

A plot of the standardised residuals r_i against any particular variable is called a **residual plot**. The purpose of residual plots is mentioned by Brockman and Wright (1992, p. 470), and their discussion is expanded slightly here.

The purpose is essentially three-fold, checking for:

- bias
- heteroscedasticity
- outliers.

A residual plot can be constructed for

- each covariate; as well as
- fitted average claim size.

Bias

Since $E[r_i] = 0$, a residual plot should be generally centred at zero, and display no trend from left to right.

These plots are particularly useful in checking the validity of the functional forms chosen for continuous covariates, as discussed in Section 6.6.

Consider, for example, Figure 7.4, a residual plot against age, when the latter has been modelled as piecewise linear, with constant gradient over the age ranges less than 30, 30-70, and more than 70 respectively. Figure 7.5 gives the corresponding residual plot for $g(x) = x$.

Figure 7.4
Age modelled piecewise linearly

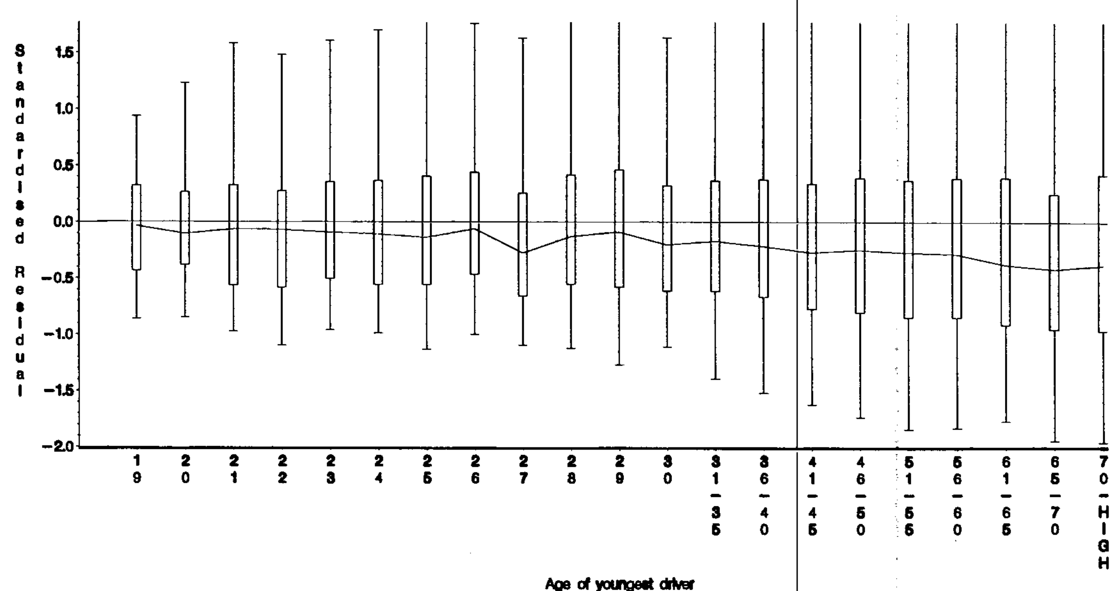
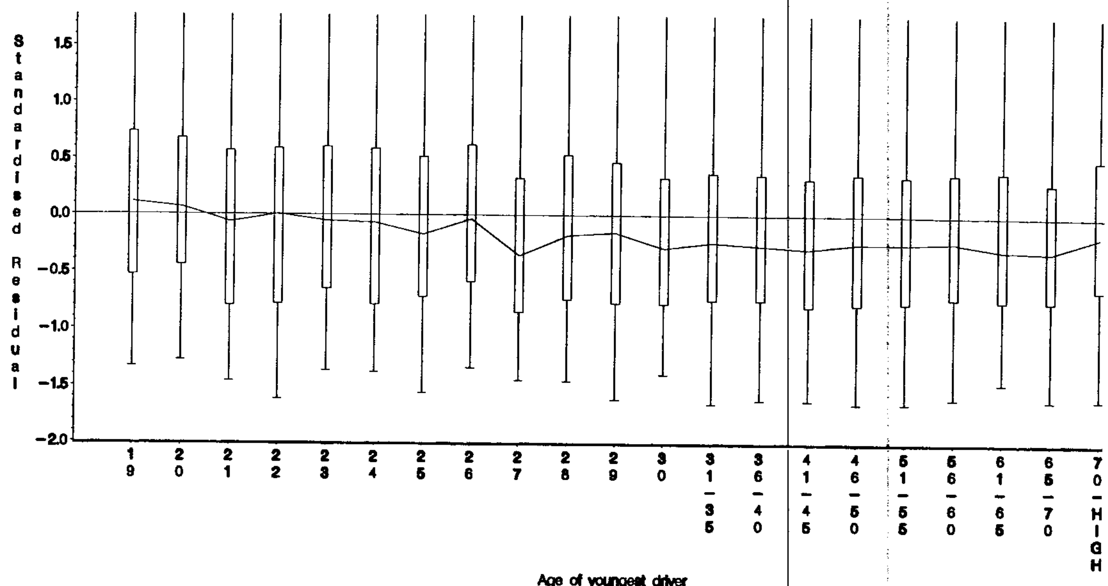


Figure 7.5
Age modelled linearly



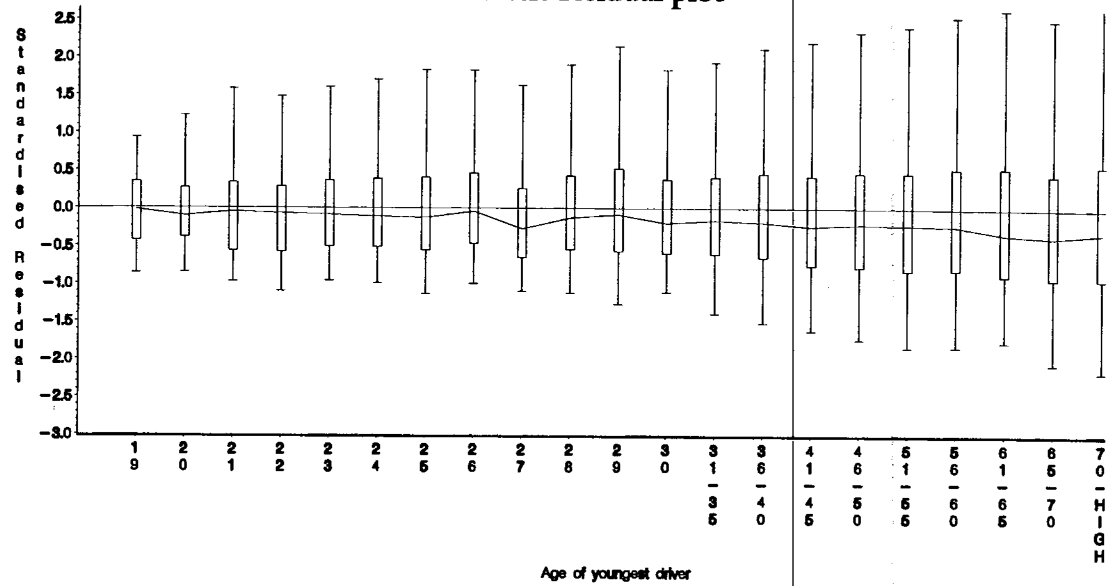
It can be seen that Figure 7.4 satisfies the no-trend requirement, while Figure 7.5 does not.

Heteroscedasticity

Since $V[r_i] = 1$, a residual plot should display no trend from left to right in dispersion. Again Figure 7.4 satisfies the requirement.

Suppose, however, that this plot had appeared as in Figure 7.6.

Figure 7.6
Heteroscedastic residual plot



Here there is fairly clear evidence of heteroscedasticity, i.e. change in $V[r_i]$ with change in the covariate. By (7.20) and (7.21), this indicates violation of the assumption that $V[g(Y_i)]$ varies with i in the systematic manner assumed by the model.

Consider the general situation in which the dispersion of r_i varies with covariate x_{ij}. Suppose that this variation takes the functional form:

$$V[r_i] = \phi(x_{ij}). \tag{7.25}$$

Define a new standardised residual:

$$r_i' = r_i / [\phi(x_{ij})]^{1/2}, \tag{7.26}$$

so that

$$V[r_i'] = 1 \tag{7.27}$$

By (7.21) and (7.26),

$$r_i' = R_i / \{V\phi(x_{ij})\}^{1/2}, \tag{7.28}$$

with V denoting $V[r_i]$, assumed constant in (7.21).

Thus, by (7.20), r_i' is consistent with an assumption that

$$V[h(g(Y_i))] \text{ proportional to } \phi(x_{ij}). \qquad (7.29)$$

By (6.16), variance structure of this sort may be recognised through the weight function w:

$$w_i \text{ proportional to } 1/\phi(x_{ij}). \qquad (7.30)$$

Thus, the steps in correcting heteroscedasticity are:

1. Construct residual plots and search for heteroscedasticity.
2. If found in any plot, express it approximately in functional form, like (7.25).
3. Introduce a weight function (7.30).
4. Re-fit the model using this weight function.
5. Return to Step 1.

In practice, if heteroscedasticity occurs, it may occur in conjunction with several covariates, so that in place of (7.25):

$$V[r_i] = \phi_j(x_{ij}) \text{ for various } j. \qquad (7.31)$$

One might then experiment with

$$w_i \text{ proportional to } 1/\phi_{j_i}(x_{ij_1})\ \phi_{j_2}(x_{ij_2}) \ \dots\ . \qquad (7.32)$$

Some trial and error with weight functions will often be required. It is simple enough to determine w_i from (7.32) for a particular model, but Step 4 in the above procedure then changes that model.

For this reason, one should not try to be too precise in determining the functions ϕ_j. Moreover, excessive precision here would be likely to amount to over-fitting (i.e. excessive parameterisation) of the model.

Outliers

Residual plots will reveal any observations that are seriously inconsistent with the model. For example, in the log normal case, with r_i given by (7.8), observations producing $|r_i| > 3$ need to be treated with caution.

Observations whose r_i are sufficiently large that they are thought to exert a distorting influence on the regression may need to be excluded. This can be done by assigning them zero weight, $w_i = 0$, (equivalently $V[h(g(Y_i))] = \infty$).

Outliers are, by their nature, isolated occurrences. The appearance in the residual plots of more than a handful of points of unduly large numerical value is likely to be symptomatic of an inappropriately chosen distribution. The quantile plots should be checked.

Note, however, that the quantile plot will not usually be sensitive to outliers since these are likely to represent just a couple of points at one or other end of the plot.

7.4 Comparison of experience with model

7.4.1 Ratios of experience to model

By (2.3),

$$E_{ijk}\, P_{ijk} = E[C_{ijk}]. \tag{7.33}$$

Let $\hat{P}_{ijk}$ be the estimate of P_{ijk}, and write

$$\hat{C}_{ijk} = E_{ijk}\, \hat{P}_{ijk}, \tag{7.34}$$

an estimate of $E[C_{ijk}]$.

Let a dot suffix indicate the summing out of that suffix, e.g.

$$C_{ij\cdot} = \sum_k C_{ijk}. \tag{7.35}$$

Then $\hat{C}_{i..}$ is an estimate of $E[C_{i..}]$, and so one wishes that

$$C_{i..} / \hat{C}_{i..} \sim 1. \tag{7.36}$$

One therefore examines the progression of actual/model (A/M) ratios (7.36) for the different values of covariate i , hoping to see them close to 100%. Further, one pays particular attention to whether any trend appears in this progression. For example, if the covariate concerned is age of insured, and the A/M ratios showed an upward trend, this would indicate that the model under-charges older insureds at the expense of the young.

Indeed, one will often be more concerned with the **trend** of the A/M ratios than with their **level**. Consider, for example, the results set out in Table 7.1.

Table 7.1
A/M ratios by sum insured

Sum insured $	Claim cost Actual $M	Claim cost Model $M	A/M %
0 - 40,000	18.9	15.1	125
40,000 - 60,000	49.2	43.2	114
60,000 - 80,000	36.5	35.8	102
80,000 - 100,000	22.0	16.7	132
100,000 - 125,000	11.8	9.0	131
125,000 - 150,000	11.3	9.4	120
150,000 - 200,000	7.6	6.3	121
200,000 - 250,000	1.9	1.6	119
over 250,000	0.4	0.3	131
Total	159.6	137.4	116

Although the model appears to under-estimate claim costs consistently, there is no obvious trend in the degree of under-estimation with changing sum insured.

Such a situation can arise in connection with a long tailed claim size distribution (e.g. house insurance (building)) for which an off-the-shelf distribution (e.g. log normal) fits reasonably well, but not precisely. A failure to fit, especially in the tail, can generate inaccurate bias correction factors (6.55).

Note, however, that this type of consistent under-estimation does not affect the sum insured **relativities**. The base premium (3.2) will also remain accurate provided that $\hat{C}$ appearing there is a reasonable estimate of $E[C]$, i.e. is more like \$159.6M than \$137.4M.

While the above discussion has applied A/M ratios to total claim cost, similar ratios can be calculated for claim numbers and claim sizes. This assists in identifying the source of any irregularities in the claim cost A/M ratios.

Decompose (7.34) by means of (2.3):

$$\hat{C}_{ijk} = E_{ijk}\,\hat{f}_{ijk}\,\hat{a}_{ijk}, \tag{7.37}$$

where $\hat{f}_{ijk}$, $\hat{a}_{ijk}$ are estimators of f_{ijk}, a_{ijk} respectively. Then the A/M ratio (7.36) may be written as:

$$C_{i..}/\hat{C}_{i..} = \sum_{jk} E_{ijk}\,F_{ijk}\,A_{ijk} \Big/ \sum_{jk} E_{ijk}\,\hat{f}_{ijk}\,\hat{a}_{ijk}, \tag{7.38}$$

where F_{ijk}, A_{ijk} are the observed values of f_{ijk}, a_{ijk}. This may be expressed in the form:

$$C_{i..}/\hat{C}_{i..} = \sum_{jk} N_{ijk}\,A_{ijk} \Big/ \sum_{jk} \hat{N}_{ijk}\,\hat{a}_{ijk}, \tag{7.39}$$

where

$$\hat{N}_{ijk} = E_{ijk}\,\hat{f}_{ijk} = \text{model fitted number of claims.} \tag{7.40}$$

Now the A/M ratio may be decomposed further:

$$C_{i..}/\hat{C}_{i..} = \frac{\sum_{jk} E_{ijk}\,F_{ijk} \big/ \sum_{jk} E_{ijk}}{\sum_{jk} E_{ijk}\,\hat{f}_{ijk} \big/ \sum_{jk} E_{ijk}} \times \frac{\sum_{jk} N_{ijk}\,A_{ijk} \big/ \sum_{jk} N_{ijk}}{\sum_{jk} N_{ijk}\,\hat{a}_{ijk} \big/ \sum_{jk} N_{ijk}} \times \frac{\sum_{jk} N_{ijk}\,\hat{a}_{ijk} \big/ \sum_{jk} N_{ijk}}{\sum_{jk} \hat{N}_{ijk}\,\hat{a}_{ijk} \big/ \sum_{jk} \hat{N}_{ijk}}, \tag{7.41a}$$

or alternatively

$$C_{i..}/\hat{C}_{i..} = \frac{\sum_{jk} E_{ijk} F_{ijk} / \sum_{jk} E_{ijk}}{\sum_{jk} E_{ijk} \hat{f}_{ijk} / \sum_{jk} E_{ijk}} \times \frac{\sum_{jk} \hat{N}_{ijk} A_{ijk} / \sum_{jk} \hat{N}_{ijk}}{\sum_{jk} \hat{N}_{ijk} \hat{a}_{ijk} / \sum_{jk} \hat{N}_{ijk}}$$

$$\times \frac{\sum_{jk} N_{ijk} A_{ijk} / \sum_{jk} N_{ijk}}{\sum_{jk} \hat{N}_{ijk} A_{ijk} / \sum_{jk} \hat{N}_{ijk}}, \qquad (7.41b)$$

Both forms of (7.41) may be read as follows:

$$C_{i..}/\hat{C}_{i..} = (F_{i..}/\hat{f}_{i..})\ (A_{i..}/\hat{a}_{i..}) \times \text{correction factor.} \qquad (7.42)$$

The first factor is common to (7.41a) and (7.41b) and is the A/M ratio for category i claim frequency.

The second factor is an A/M ratio for category i average claim size, but note that it takes different forms in (7.41a) and (7.41b). In both cases it is the ratio of a weighted average A to the same weighted average $\hat{a}$. However, the weights differ as between (7.41a) and (7.41b); in the former case they are actual claim numbers, in the latter, fitted claim numbers.

The third factor quantifies the difference between these two weighted averages of the $\hat{a}$.

It is useful to tabulate the claim frequency and claim size A/M ratios appearing in (7.42). This is particularly the case when anomalies in the claim cost A/M ratios average, because it enables them to be tracked to their source in frequency or size. It is apparent from (7.42), however, that the claim cost A/M ratios factor in three directions, not just the two represented by frequency and size.

Probably (7.41b) is the preferable form of (7.41) for this factorisation. The use of fitted claim numbers as weights in the average claim sizes helps to stabilise the claim size A/M ratio.

It was foreshadowed at the beginning of Section 3 that an alternative to (7.37) would be favoured for evaluation of the base premium. The alternative developed these was such as to ensure equality between the resulting pool of risk premiums and aggregate claims cost. Any other approach, including (7.37), would not guarantee this equality.

It is instructive to examine the reasons why (7.37) might fail systematically (i.e. not just as a result of sampling error) in this respect.

An adaptation of (7.42) yields:

$$C_{...}/\hat{C}_{...} = (F_{...}/\hat{f}_{...})\ (A_{...}/\hat{a}_{...}) \times \text{correction factor}, \tag{7.43}$$

which quantifies the difference between the contending estimates of base premium. A value markedly different from unity indicates a wide difference.

Consider the factor $A_{...}/\hat{a}_{...}$. This is particularly vulnerable to significant difference from unity whenever a data transformation (6.52) is used, since its inversion requires the bias correction (6.55). Any error in the choice of model claim size distribution might introduce error into this bias correction, and hence into $\hat{a}_{...}$.

Even in cases where claim sizes have not been transformed, and even when each of the three factors in the above formula has a unit expectation, it should be noted that any correlation between F_{ijk} and A_{ijk} will cause the expectation of $C_{...}/\hat{C}_{...}$ to assume a value other than unity.

There has been an assumption throughout this paper that F_{ijk} and A_{ijk} are stochastically independent (see Section 2), but if it should be wrong then (7.37) will be correspondingly wrong and $C_{...}/\hat{C}_{...}$ will not have unit expectation.

Errors of this nature may be difficult to control within relativities. They may be ignored there, but it is desirable that they not be permitted to distort the premium pool in such a way that it fails to match aggregate claim cost.

7.4.2 Data transformations

Data transformations for claim sizes were introduced in (6.52). Their inversion is dealt with in (6.53)-(6.55).

The claim size A/M ratios introduced in Section 7.4.1 relate to **untransformed** claim sizes. In some cases these will be drawn from a long tailed distribution, with large sampling error involved in the sample mean. The A/M ratios can be statistically unstable in these circumstances.

This is, of course, the reason why data transformations are taken. They function as mean stabilisers among other things. Where a claim size transformation $g(\cdot)$ has been taken, it will sometimes be useful to construct A/M ratios in respect of transformed claim sizes since these are likely to be more stable than their untransformed counterparts.

The mean transformed claim size may have no physical meaning, and so construction of the A/M ratios is necessarily somewhat *ad hoc*. Nonetheless, analogy with (7.41b) yields a ratio:

$$\sum_{jk} \hat{N}_{ijk}\, g(A_{ijk}) \Big/ \sum_{jk} \hat{N}_{ijk}\, \hat{\mu}_{ijk}, \tag{7.44}$$

where $\hat{\mu}_{ijk}$ is the value fitted to $g(A_{ijk})$ by the model.

An example appears in Table 7.2, which gives A/M ratios for both transformed and untransformed data for fire, flood and earthquake claims from a House (buildings) insurance portfolio. A log transform is involved.

Table 7.2
Claim size A/M ratios

State of insured risk	A/M ratio for claim size	
	Transformed %	Untransformed %
New South Wales	100	128
Victoria	100	123
Queensland	102	175
South Australia	100	136
Western Australia	99	95
Tasmania	97	84
Australian Capital Territory	106	309
Northern Territory	99	56

The states are listed in descending order of volume of experience.

The first comment to be made on Table 7.2, in the spirit of Section 7.4.1, is that the untransformed A/M ratios display no trend, and so are satisfactory in that sense.

They do display a high degree of variation, however. The transformed ratios, on the other hand, are clustered tightly around 100%, indicating very satisfactory fitting of transformed claim sizes.

It follows that the variability of the transformed ratios has been introduced at the stage of the inverse transformation g^{-1}, which denotes exponentiation . One is entitled to conclude that the fit is perfectly satisfactory (in relation to state of insured risk, at least), and that the variable A/M ratios are due to the inherent variability of individual claim sizes. Indeed, the coefficient of variation of individual claim size was estimated at 1595% in the above example.

This same variability provides a healthy warning of the dangers and difficulties involved in **not** transforming claim size before analysis of cases such as in Table 7.2.

7.4.3 Multi-way analysis

The whole of the above discussion of A/M ratios has been in terms of **1-way** analysis, in the sense that the ratios are tabulated as a function of a single covariate. There is an obvious extension to multi-way analyses.

For example, **2-way** A/M ratios are obtained when (7.38) is replaced by the following:

$$C_{ij.}/\hat{C}_{ij.} = \sum_k E_{ijk}\, C_{ijk} \Big/ \sum_k E_{ijk}\, \hat{C}_{ijk}. \qquad (7.45)$$

Multi-way A/M ratios can be useful in checking that subtle effects do not compound to the undue advantage or disadvantage of some of the more extreme segments of the portfolio. For example, one might be interested to check the A/M ratio in a Motor Accidental Damage portfolio for the segment consisting of:

- young male drivers;
- high powered vehicles;
- low NCD.

Each of these characteristics attracts a penalty. The concern would be that the compounding of these penalties might be excessive.

8 Variables which fall outside the GLM framework

8.1 Deductible

Suppose that a policy is subject to a **deductible** d, so that, when damage of amount X occurs, the size of claim payable is $\max(0, X-d)$.

The application of a deductible reduces the risk premium. If $c(d)$ denotes the risk premium in the presence of a deductible d and $f(.)$ the pdf of claims amount, then

$$c(d) = \int_d^\infty (x-d)\, f(x)\, dx. \qquad (8.1)$$

The corresponding quantity in the absence of a deductible is

$$c(0) = \int_0^\infty x\, f(x)\, dx, \qquad (8.2)$$

and $c(d)/c(0)$ is the factor by which to adjust a deductible-free risk premium to recognise a deductible of d.

The representation (8.1) involves an implicit assumptions that $f(x)$ does not vary with d. In practice, this is probably incorrect. For example, the likelihood of lodgement of a claim for $5 is very small, irrespective of whether the relevant policy carries a deductible of nil or $500.

In other words, the imposition of a deductible modifies claimant behaviour. This is particularly so in the presence of NCD.

It would be possible to treat deductible as simply one more coordinate in GLMs describing claim frequency and claim size. This empirical treatment would have the desirable effect of incorporating the behavioural aspects mentioned above in the model.

On the other hand, however, there are distinct difficulties with this approach. A typical distribution of Motor deductibles might be as in Table 8.1.

Table 8.1
Motor deductibles

Deductible $	Percentage of Exposure %
0	21
300	19
350	4
400	47
500	6
750	2
1,000	0.7
1,500	0.3
Total	100

The eight values of deductible listed are likely to fall into four groups:

- The nil deductible.
- The group $300-$400, comprising the current standard of $400, together with the obsolete standards $300 and $350.
- The $500 deductible, the standard for policies carrying a young driver risk.
- The $750 and higher group, reflecting penalty deductibles, applied to policy owners with:
    - poor claims experience;
    - traffic offence convictions;
    - high risk vehicles;
    - etc.

These four groups are likely to be distinct, behaviourally as well as in their underlying claims experience. Within each group, claimants are likely to be behaviourally similar, though experience could still be expected to vary with variation in deductible.

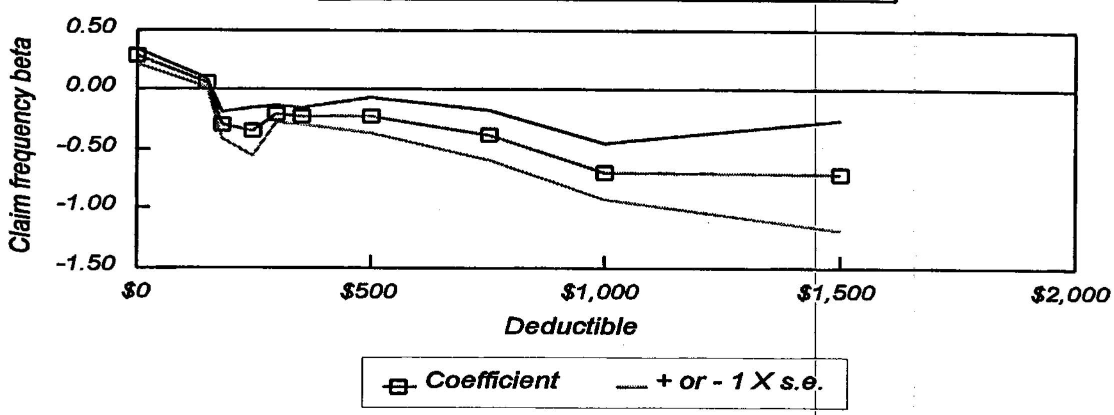
The gradation of claim frequency and claim size as deductible varies from \$300 to \$400 represents some subtle effects, and it may be too much to hope that these will be accurately identified by a GLM. These doubts are increased by the low exposure at deductible \$350.

The fact that some deductibles are age related, while others are related to vehicle type raises the possibility of multi-collinearity in the GLM.

For all these reasons, the inclusion of deductible as a covariate in a GLM can be expected to yield only moderate success. Figure 8.1 provides an example based on real data.

Figure 8.1

Regression of claim frequency on deductible



The figure plots the claim frequency β parameters (in the sense of (6.1)) for ranges of deductibles against the mid-values of those ranges. It also plots one standard error of each β up and down from its estimate. The β parameters are all relative to $\beta = 0$ (with zero standard error) at a nil deductible.

Figure 8.1 indicates the general downward trend in claim frequency with increasing deductible, though the functional form of this trend is unclear. The anomalous shape around the region \$200-300 is apparent.

The theoretical representation (8.1) might be helpful in attempting to determine this functional form. Appendix B calculates the ratio $c(d)/c(0)$ for each of the following distributions of claim size:

- log normal;
- power gamma;
- inverse Gaussian.

In each case, it is found that

$$c(d)/c(0) = D(d/a;\ \mathrm{v}), \tag{8.3}$$

for some function $D(.;\ .)$, average claim size (before application of the deductible) $\boldsymbol{a}$, and with ν a parameter measuring some form of dispersion of a claim size distribution.

This is a convenient form when ν is constant over all cells, as will often be the case. For then (8.3) indicates that the effect of the deductible on risk premium depends only on the proportion of average claim size represented by that deductible.

It is also of interest to note the implications of this for the manner in which the premium reductions vary from cell to cell. To the extent that an increase in premium from one cell to another reflects an increase in claim frequency (with no change in average claim size), a constant deductible will generate a constant percentage premium reduction.

On the other hand, to the extent that the increase in premium reflects an increase in average claim size, the premium reduction for the deductible will:

- increase in absolute terms;
- decrease as a percentage of premium;

as premium increases.

Since cell-to-cell increases in premium will typically reflect both increased frequency and increased average claim size, the usual application of (8.3) will yield premium reductions for a deductible with the properties described in the preceding paragraph.

These remarks must be read, of course, in conjunction with the qualification of (8.1) given early in the present sub-section.

Subject to this, however, they raise doubts about the common market practice of awarding a flat dollar premium reduction to all policies carrying a fixed dollar excess.

Specific versions of the function D are as follows.

Log normal

$$D(x;\ \sigma) = \left[1 - \Phi\left((\log x)/\sigma - \tfrac{1}{2}\sigma\right)\right] - x\left[1 - \Phi\left((\log x)/\sigma + \tfrac{1}{2}\sigma\right)\right], \qquad (8.4)$$

where $\exp(\sigma^2) - 1$ is the squared coefficient of variation of claim size, and $\Phi(\cdot)$ is the unit normal d.f. Expansion (8.4) may be recognised as a Black-Scholes formula.

Power gamma

Suppose that Y^p is gamma distributed, where Y denotes claim size. Then

$$D(x;\ \gamma) = \left[1 - \Gamma\left(\gamma (bx)^p;\ \gamma + 1/p\right)\right] - x\left[1 - \Gamma\left(\gamma (bx)^p;\ \gamma\right)\right], \qquad (8.5)$$

where $1/\gamma$ is the squared coefficient of variation of claim size and b is the bias correction factor (A.13), and $\Gamma(.;\ .)$ the incomplete gamma function

$$\Gamma(x;\ \gamma) = \left[\Gamma(\gamma)\right]^{-1} \int_0^x z^{\gamma-1} e^{-z}\, dz.$$

Inverse Gaussian

The precise form of $D(x;\ \nu)$ is given by (B.24) with ν equal to the variance of claim size. It appears that $D(.;\ .)$ needs to be evaluated numerically.

Numerical example

Consider the power gamma case in which

$$\gamma = 2.5,$$
$$p = 0.5,$$

giving

$$b = 1.4$$

Then application of (8.4) yields the following results.

Table 8.2
Effect of deductible on claim costs

Ratio of deductible to average claim size %	Estimated reduction in claim cost %
0	0
10	9
20	17
30	24
40	31
50	36
60	41
70	46

Suppose the average claim size, taken over a whole Motor experience, is \$2,000. The corresponding average claim cost per vehicle year is \$250. Table 8.2 suggests that for a "typical" policy, a standard deductible of \$300 (about 15% of average claim size) would justify a premium reduction of about 14%, or \$35.

8.2 No claim discount

An NCD system comprises:

(1) a set of rules which assigns a policy owner to an NCD **category** (or **rating**) on the basis of prior **claim occurrence** experience (i.e. without reference to the sizes of any claims; and

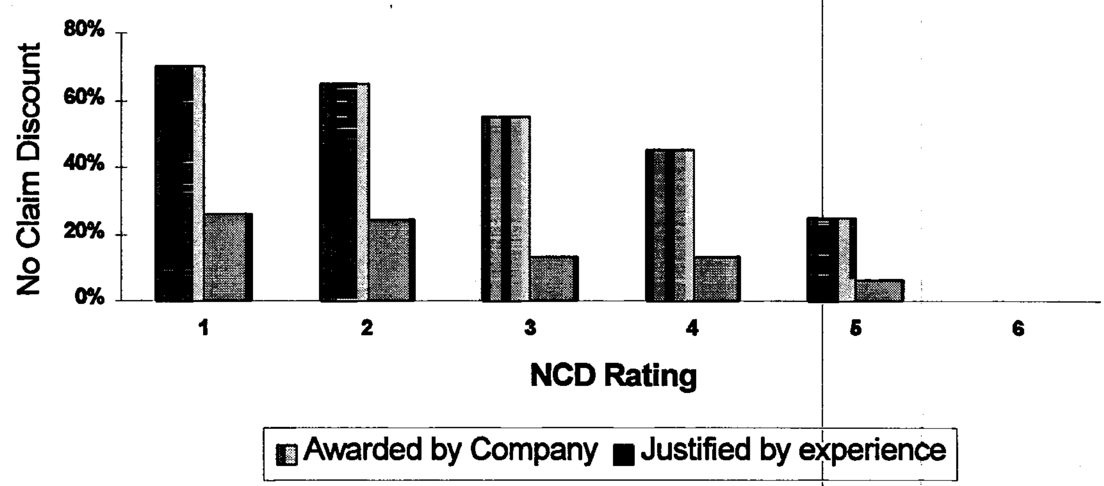
(2) a prescribed level of discount for each NCD category.

In view of (1), a policy owner's probability of occupancy of any particular NCD category at any particular future time is a function of claim frequency (plus claiming behavioural qualities). It follows that the appropriate level of discount is also a function of claim frequency.

This problem has been addressed many times in the literature. For a summary, see Lemaire (1985, 1995).

In practice, it is usual to find the NCD scale (2) set more by reference to market convention than to the above theoretical considerations, and also commonly in conflict with these considerations. For example, anticipating the latter part of the present subsection slightly, Figure 8.2 compares, for a particular Motor (accidental damage) portfolio, the NCD scale in use with that justified by experience in the various NCD ratings.

Figure 8.2
Comparison of NCD scale with experience



It is evident that the awarded discounts are generally considerably greater than can be justified on the basis of the experience. This is a typical result.

Thus, NCD is not usually just another covariate, to be treated in a manner parallel to other covariates. For one will not usually be required to estimate NCD relativities as is required in connection with other covariates; these relativities will be given.

There are then two main responses to any disparity between the NCD scale in force and that justified by claims experience:

- ignore it; or
- recognise it, by optimising the fit of premium to other covariates, taking into account the NCD scale.

The formal description of these two possibilities is as follows. Consider (6.1) in its usual application to claim frequency, with NCD recognised explicitly:

$$Y_i = \exp(\alpha_i + x_{NCD,i}^T \beta_{NCD} + x_{other,i}^T \beta_{other}) + e_i, \tag{8.6}$$

with Y_i the number of claims in the i -th cell and e_i a centred Poisson variate.

Ignoring the NCD effect consists of either:

- dropping the NCD terms from (8.6) and estimating just β_{other}; or
- estimating both β_{NCD} and β_{other}, but then ignoring β_{NCD}.

The first approach effectively estimates all effects other than NCD on the assumption that there is **no NCD effect**, i.e. claim frequency does not vary with NCD category. This will distort the estimate of β_{other}.

The second approach explicitly estimates the NCD effect, thereby eliminating distortion in the estimated β_{other}, but then incorporates a different NCD effect (according to the scale) in premiums. The β_{NCD} estimated in this manner are those reflected in the discounts "justified by experience" in Figure 8.2.

To recognise the disparity between the NCD scale and experience, recognise that β_{NCD} in (8.6) is fixed. Specifically, the component of β_{NCD} corresponding to a discount of d% will be $\log(1-d/100)$. Then (8.6) may be put in the form:

$$Y_i = \exp\left[\left(\log E_i + x_{NCD,i}^T \beta_{NCD}\right) + \left(\alpha_i' + x_{other,i}^T \beta_{other}\right)\right] + e_i, \tag{8.7}$$

where α_i has been separated into its exposure term and the rest.

The purpose of representation (8.7) is to recognise **all known** effects in the first round bracket. Everything here can be calculated; it does not require estimation, i.e.

$$Y_i = \exp\left[K_i\left(\alpha_i' + x_{other,i} \beta_{other}\right)\right] + e_i, \tag{8.8}$$

where the K_i constitute a set of known constants and α_i' (usually independent of i) and β_{other} are to be estimated. Note that

$$K_i = \log\left[E_i(1-d_i/100)\right] = \log\left[\text{discounted exposure}\right], \tag{8.9}$$

where d_i is the NCD applying to the i -th cell.

In a system such as (8.8), the K_i are known as an **offset**, to use GLM terminology. Thus, when NCD scale is given, optimal fit of claim frequency to experience with respect to other covariates can be carried out by modifying exposure to discounted exposure in the offset term.

Occasionally, a technical pricing analysis will require investigation of an appropriate NCD scale. It will be supposed here that the set of rules (1) mentioned at the start of the present sub-section is given, and only the scale (2) requires determination.

One can simply estimate the β_{NCD} as described above, or in some smoothed form. However, care would be needed in the implementation of the resulting scale. While Figure 8.2 shows that Rating 1 policy owners experience a claim frequency some 26% lower than Rating 6, this occurs in the presence of a 70% discount.

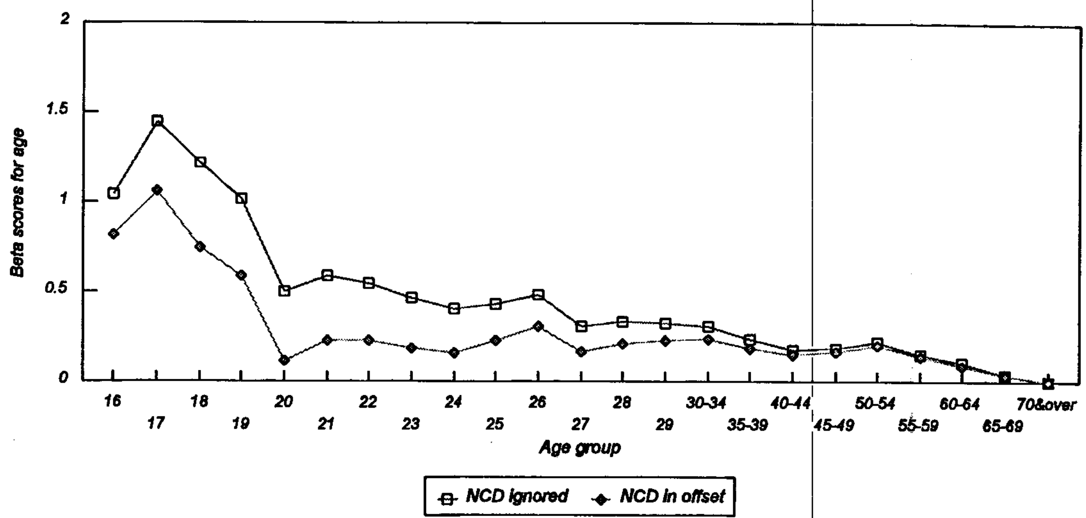
This large discount serves as a strong disincentive to claim lodgement. If it were reduced to 26%, a change in claiming behaviour might be induced, whereby the justified discount reduced to less than 26%. If this further reduced discount were implemented, a further change in claiming behaviour might be induced, ... and so on.

Rather than relying entirely on the empirical figure of 26%, it is preferable to attempt to build an understanding of why that figure arises from the NCD system in force (see e.g. Taylor, 1997). Armed with this understanding, one will be in a stronger position to predict the outcome of changes to the system.

It should be noted that some relativities can be significantly influenced by the inclusion or exclusion of β_{NCD} in (8.8). These will normally be relativities associated with covariates correlated with NCD. The most obvious of these is age of insured.

For a particular real example, Figure 8.3 graphs the categorical age effect estimated with and without β_{NCD} in (8.8).

Fig 8.3
NCD effect on age rating



It is seen that much of the effect of young ages on frequency is eliminated by the recognition of NCD. This reflects the fact that the short driving record of most young insured prevents them from having accumulated NCD. Although their claim frequencies are high, this is largely compensated by the lack of NCD.

9 Summary of procedures

No pretence is made here that GLM pricing can be reduced to cookbook style. Nevertheless, many of the issues canvassed in preceding sections recur from one assignment to the next.

Figure 9.1 is an attempt to place as much as possible of the foregoing discussion into a statement of routine. Numbers appearing at the nodes of the diagram refer to the sections dealing with the subjects named there.

Figure 9.1
Summary of pricing procedures

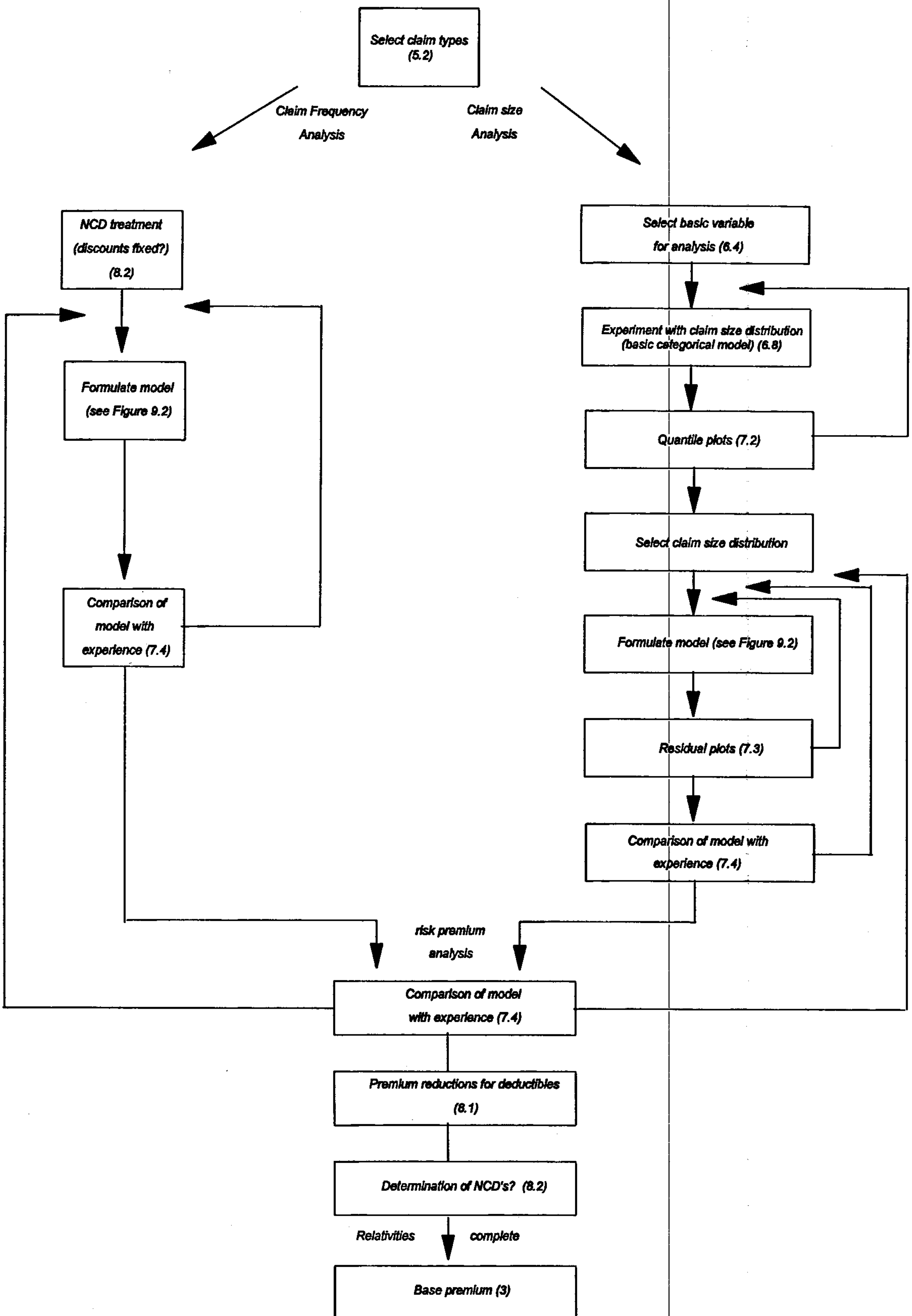
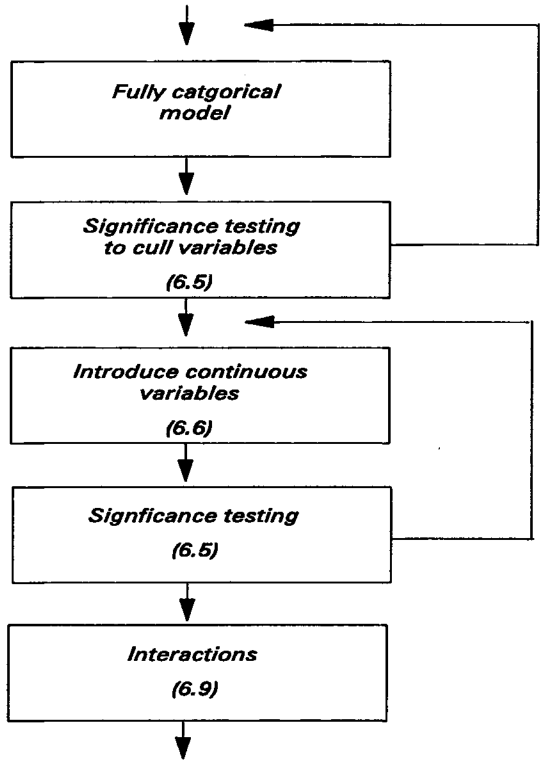


Figure 9.2 represents the model formulation stage of Figure 9.1, and applies to both claim frequency and size.

Figure 9.2
Model formulation



10 Acknowledgements

This paper chronicles a number of lessons learnt in real life applications of GLMs to technical pricing. The various assignments concerned with this have involved contributions from others.

Inevitably, therefore, a number of people have contributed to the paper, sometimes to the point where the boundaries between their ideas and the author's are not altogether clear. Special mention in this respect is due to Steven Lim, who has been directly involved in much of the development and its conversion to computer code, and also to David Rasmussen, Kevin Gomes and Geoff Trahair.

Appendix A
Bias correction for average claim size

A.1 Log transformation - identity link

A.1.1 Normal error term

Write the bias correction factor (6.55) in the simplified form:

$$b_i = E\left[\exp(\mu_i + e_i)\right]/\exp \mu_i, \tag{A.1}$$

with

$$e_i \sim N(0, \sigma^2). \tag{A.2}$$

Note the assumption that $\sigma^2 = V\left[\log Y_i\right]$ is independent of i. Note that the squared coefficient of variation of $\log Y_i$ is $\exp(\sigma^2) - 1$.

By (A.1) and (A.2),

$$b_i = \exp \tfrac{1}{2}\sigma^2, \tag{A.3}$$

independent of i.

A.1.2 Gamma error term

Consider the case in which (A.1) holds with $\mu_i + e_i$ gamma distributed, specifically with *pdf*

$$\left[\Gamma(\gamma)\right]^{-1} c_i^{\gamma} y^{\gamma-1} \exp - c_i y, \tag{A.4}$$

with $\gamma > 0$ independent of i.

The mean of (A.4) is γ/c_i. Therefore,

$$\mu_i = \gamma/c_i, \tag{A.5}$$

since $Ee_i = 0$.

Now

$$E \exp(\mu_i + e_i) = \int_0^\infty [\Gamma(\gamma)]^{-1} c_i^\gamma y^{\gamma-1} \exp - (c_i-1) y \, dy$$

$$= [c_i/(c_i-1)]^\gamma, \text{ provided } c_i > 1. \tag{A.6}$$

By (A.1), (A.5) and (A.6),

$$b_i = [c_i/(c_i-1)]^\gamma \exp(-\gamma/c_i). \tag{A.7}$$

An alternative form of (A.7) is obtained by noting that

$$v^2 = 1/\gamma, \tag{A.8}$$

with v denoting the coefficient of variation of $\mu_i + e_i$.

Then

$$b_i = (1 - \mu_i v^2)^{-1/v^2} \exp(-\mu_i), \text{ provided } \mu_i v^2 < 1, \tag{A.9}$$

which depends on μ_i. In fact,

$$d(\log b_i)/d\mu_i = (1 - \mu_i v^2)^{-1} - 1 > 0,$$

showing that b_i increases with μ_i.

A.2 Power transformation - log link

Write the bias correction factor in the simplified form:

$$b_i = E[\exp(\mu_i) + e_i]^{1/p} / [\exp \mu_i]^{1/p}. \tag{A.10}$$

Consider the case in which $\exp(\mu_i) + e_i$ is gamma distributed with *pdf* (A.4). Then its mean is

$$\exp \mu_i = \gamma / c_i, \tag{A.11}$$

and

$$E\left[\exp(\mu_i) + e_i\right]^{1/p} = \int_0^\infty \left[\Gamma(\gamma)\right]^{-1} c_i^{\gamma}\, y^{\gamma+1/p-1} \exp - c_i y$$

$$= \Gamma(\gamma + 1/p) / c_i^{1/p}\, \Gamma(\gamma). \tag{A.12}$$

By (A.10) - (A.12),

$$b_i = \Gamma(\gamma + 1/p) / \gamma^{1/p}\, \Gamma(\gamma), \tag{A.13}$$

independent of i .

Appendix B
Effect of deductible on risk premium

B.1 Log normal claim sizes

Let

$c(d)$	=	risk premium per unit exposure when the deductible is set at d;
Y	=	claim size, **before** application of the deductible;
λ	=	claim frequency, before application of the deductible.

Assume

$$\log Y \sim N(\mu, \sigma^2). \tag{B.1}$$

Then

$$c(d) = \lambda \int_{\log d}^{\infty} (e^{x} - d)\ \phi(x; \mu, \sigma^2)\ dx, \tag{B.2}$$

where $\phi(.; \mu, \sigma^2)$ is the unit normal p.d.f. with parameters μ, σ^2.

Then

$$c(d) = \lambda \int_{k}^{\infty} \left[\exp\ (\mu + \sigma z) - d\right] \phi(z; 0, 1)\ dz \tag{B.3}$$

with

$$k = (\log d - \mu)/\sigma. \tag{B.4}$$

It follows that

$$c(d)/\lambda = \exp(\mu + \tfrac{1}{2}\sigma^2)\left[1 - \Phi(k - \sigma)\right] - d\left[1 - \Phi(k)\right], \tag{B.5}$$

where $\Phi(.)$ is the unit normal d.f., and so

$$c(d)/c(0) = \left[1 - \Phi(k-\sigma)\right] - d\left[1 - \Phi(k)\right]/\exp(\mu + \tfrac{1}{2}\sigma^2)$$
$$= \left[1 - \Phi(k_1)\right] - (d/a)\left[1 - \Phi(k_2)\right], \qquad (B.6)$$

where

$$a = \exp\,(\mu + \tfrac{1}{2}\sigma^2) = \text{average claim size}, \qquad (B.7)$$

$$k_1 = k - \sigma = \left[\log\,(d/a)\right]/\sigma - \tfrac{1}{2}\sigma, \qquad (B.8)$$

$$k_2 = k_1 + \sigma. \qquad (B.9)$$

B.2 Power gamma claim sizes

Replace (B.1) by an assumption that Y^p is gamma distributed as in (A.4). Then

$$c(d) = \lambda\, c^{\gamma}\left[\Gamma(\gamma)\right]^{-1} \int_{d^p}^{\infty} (x^{1/p} - d)\, x^{\gamma-1}\, e^{-cx}\, dx$$
$$= \lambda\left\{\left[1 - \Gamma(cd^p;\ \gamma+1/p)\right]\Gamma(\gamma+1/p)/c^{1/p}\,\Gamma(\gamma) - d\left[1 - \Gamma(cd^p;\ \gamma)\right]\right\}, \qquad (B.10)$$

where $\Gamma(.;\ .)$ is the incomplete gamma function

$$\Gamma(x;\ \gamma) = \left[\Gamma(\gamma)\right]^{-1} \int_0^x z^{\gamma-1}\, e^{-z}\, dz. \qquad (B.11)$$

The subscript i has been dropped from c for convenience, and will be consistently omitted through the remainder of this appendix.

Thus

$$c(d)/c(0) = \left[1 - \Gamma(cd^p;\ \gamma+1/p)\right] - \left[1 - \Gamma(cd^p;\ \gamma)\right] d c^{1/p}\, \Gamma(\gamma)/\Gamma(\gamma+1/p). \qquad (B.12)$$

As in Appendix B.1, it is possible to express this in terms of the ratio d/a. In the present case

$$a = (\gamma/c)^{1/p}\, b, \tag{B.13}$$

where b is the bias correction factor given by (A.13):

$$b = \Gamma(\gamma+1/p)/\gamma^{1/p}\,\Gamma(\gamma). \tag{B.14}$$

By (B.13),

$$cd^{p} = \gamma\left[b\,(d/a)\right]^{p}, \tag{B.15}$$

and, by (B.13) and (B.14) together,

$$dc^{1/p}\,\Gamma(\gamma)\Big/\Gamma(\gamma+1/p) = d/a. \tag{B.16}$$

Substitution of (B.15) and (B.16) into (B.12) yields

$$c(d)/c(0) = \left[1 - \Gamma(k;\ \gamma+1/p)\right] - (d/a)\left[1-\Gamma(k;\ \gamma)\right], \tag{B.17}$$

with

$$k = \gamma\left[b\,(d/a)\right]^{p}. \tag{B.18}$$

Special case: gamma distribution ($p = 1$)

In this case, (B.14) gives

$$b = 1, \tag{B.19}$$

whence (B.18) gives

$$k = \gamma\,(d/a). \tag{B.20}$$

B.3 Inverse Gaussian claim sizes

Replace (B.1) by the assumption that Y is inverse Gaussian distributed with p.d.f.

$$\mu(2\pi\beta y^3)^{-1/2} \exp\left[-(y-\mu)^2/2\beta y\right], \; y>0. \tag{B.21}$$

Note that

$$E[Y] = \mu, \; V[Y] = \mu\beta. \tag{B.22}$$

Then

$$c(d)/c(0) = \int_d^\infty (2\pi\beta y)^{-1/2} \exp\left[-(y-\mu)^2/2\beta y\right] dy - (d/\mu)\int_d^\infty \mu(2\pi\beta y^3)^{-1/2} \exp\left[-(y-\mu)^2/2\beta y\right] dy. \tag{B.23}$$

Apply the transformation $z = y/\mu$ in (B.23) to obtain

$$c(d)/c(0) = \int_{d/\mu}^\infty (2\pi v z)^{-1/2} \exp\left[-(z-1)^2/2vz\right] dz - (d/\mu)\int_{d/\mu}^\infty (2\pi v z^3)^{-1/2} \exp\left[-(z-1)^2/2vy\right] dz, \tag{B.24}$$

with

$$v = \beta/\mu = \mu\beta/\mu^2 = [\text{coefficient of variation}]^2. \tag{B.25}$$

Thus $c(d)/c(0)$ takes the general form:

$$c(d)/c(0) = D(d/\mu; \; v). \tag{B.26}$$

The second integral in (B.24) may be evaluated (see Panjer and Willmot, 1992, p.114) as:

$$\Phi\left[(d/\mu-1)(\nu d/\mu)^{-1/2}\right] + e^{2/\nu}\,\Phi\left[-(d/\mu+1)(\nu d/\mu)^{-1/2}\right]. \qquad \text{(B.27)}$$

The first integral in (B.24) is more difficult to evaluate. It is the integral of a reciprocal inverse Gaussian ***pdf*** (Panjer and Willmot, 1992, p.117), which is in turn the integrated ***pdf*** of the sum of independent Gaussian and gamma variates.

References

Beard, R.E., Pentikäinen, T. & Pesonen, E. (1984). **Risk theory, the stochastic basis of insurance (3rd edition)**. Chapman and Hall.

Brockman, M.J. and Wright, T.S. (1992). **Statistical motor rating: making effective use of your data**. Journal of the Institute of Actuaries, 119, 457-526.

Johnson, P.D. & Hey, G.B. (1971). **Statistical Studies in Motor Insurance**. Journal of the Institute of Actuaries, 97, 199.

Lemaire, J. (1985) **Automobile insurance: actuarial models**. Kluwer-Nijhoff Publishing, Boston.

Lemaire, J. (1995). **Bonus-malus systems in automobile insurance**. Kluwer Academic Publishers, Boston.

McCullagh, P. & Nelder, J.A. (1989). **Generalised Linear Models (3rd edition)**. Chapman and Hall.

Panjer, H.H. and Willmot, G.E. (1992). **Insurance risk models**. Society of Actuaries.

Taylor, G.C. (1997). Setting a bonus-malus scale in the presence of other rating factors. **Astin Bulletin** (to appear).

RESEARCH PAPER SERIES

No.	Date	Subject	Author
1	MAR 93	AUSTRALIAN SUPERANNUATION : THE FACTS, THE FICTION, THE FUTURE	David M Knox
2	APR 93	AN EXPONENTIAL BOUND FOR RUIN PROBABILITIES	David C M Dickson
3	APR 93	SOME COMMENTS ON THE COMPOUND BINOMIAL MODEL	David C M Dickson
4	AUG 93	RUIN PROBLEMS AND DUAL EVENTS	David CM Dickson Alfredo D Egidio dos Reis
5	SEP 93	CONTEMPORARY ISSUES IN AUSTRALIAN SUPERANNUATION - A CONFERENCE SUMMARY	David M Knox John Piggott
6	SEP 93	AN ANALYSIS OF THE EQUITY INVESTMENTS OF AUSTRALIAN SUPERANNUATION FUNDS	David M Knox
7	OCT 93	A CRITIQUE OF DEFINED CONTRIBUTION USING A SIMULATION APPROACH	David M Knox
8	JAN 94	REINSURANCE AND RUIN	David C M Dickson Howard R Waters
9	MAR 94	LIFETIME INCOME, TAXATION, EXPENDITURE AND SUPERANNUATION (LITES): A LIFE-CYCLE SIMULATION MODEL	Margaret E Atkinson John Creedy David M Knox
10	FEB 94	SUPERANNUATION FUNDS AND THE PROVISION OF DEVELOPMENT/VENTURE CAPITAL: THE PERFECT MATCH? YES OR NO	David M Knox
11	JUNE 94	RUIN PROBLEMS: SIMULATION OR CALCULATION?	David C M Dickson Howard R Waters
12	JUNE 94	THE RELATIONSHIP BETWEEN THE AGE PENSION AND SUPERANNUATION BENEFITS, PARTICULARLY FOR WOMEN	David M Knox
13	JUNE 94	THE COST AND EQUITY IMPLICATIONS OF THE INSTITUTE OF ACTUARIES OF AUSTRALIA PROPOSED RETIREMENT INCOMES STRATEGY	Margaret E Atkinson John Creedy David M Knox Chris Haberecht
14	SEPT 94	PROBLEMS AND PROSPECTS FOR THE LIFE INSURANCE AND PENSIONS SECTOR IN INDONESIA	Catherine Prime David M Knox

15	OCT 94	PRESENT PROBLEMS AND PROSPECTIVE PRESSURES IN AUSTRALIA'S SUPERANNUATION SYSTEM	David M Knox
16	DEC 94	PLANNING RETIREMENT INCOME IN AUSTRALIA: ROUTES THROUGH THE MAZE	Margaret E Atkinson John Creedy David M Knox
17	JAN 95	ON THE DISTRIBUTION OF THE DURATION OF NEGATIVE SURPLUS	David C M Dickson Alfredo D Egidio dos Reis
18	FEB 95	OUTSTANDING CLAIM LIABILITIES: ARE THEY PREDICTABLE?	Ben Zehnwirth
19	MAY 95	SOME STABLE ALGORITHMS IN RUIN THEORY AND THEIR APPLICATIONS	David C M Dickson Alfredo D Egidio dos Reis Howard R Waters
20	JUN 95	SOME FINANCIAL CONSEQUENCES OF THE SIZE OF AUSTRALIA'S SUPERANNUATION INDUSTRY IN THE NEXT THREE DECADES	David M Knox
21	JUN 95	MODELLING OPTIMAL RETIREMENT IN DECISIONS IN AUSTRALIA	Margaret E Atkinson John Creedy
22	JUN 95	AN EQUITY ANALYSIS OF SOME RADICAL SUGGESTIONS FOR AUSTRALIA'S RETIREMENT INCOME SYSTEM	Margaret E Atkinson John Creedy David M Knox
23	SEP 95	EARLY RETIREMENT AND THE OPTIMAL RETIREMENT AGE	Angela Ryan
24	OCT 95	APPROXIMATE CALCULATION OF MOMENTS OF RUIN RELATED DISTRIBUTIONS	David C M Dickson
25	DEC 95	CONTEMPORARY ISSUES IN THE ONGOING REFORM OF THE AUSTRALIAN RETIREMENT INCOME SYSTEM	David M Knox
26	FEB 96	THE CHOICE OF EARLY RETIREMENT AGE AND THE AUSTRALIAN SUPERANNUATION SYSTEM	Margaret E Atkinson John Creedy
27	FEB 96	PREDICTIVE AGGREGATE CLAIMS DISTRIBUTIONS	David C M Dickson Ben Zehnwirth
28	FEB 96	THE AUSTRALIAN GOVERNMENT SUPERANNUATION CO-CONTRIBUTIONS: ANALYSIS AND COMPARISON	Margaret E Atkinson
29	MAR 96	A SURVEY OF VALUATION ASSUMPTIONS AND FUNDING METHODS USED BY AUSTRALIAN ACTUARIES IN DEFINED BENEFIT SUPERANNUATION FUND VALUATIONS	Des Welch Shauna Ferris
30	MAR 96	THE EFFECT OF INTEREST ON NEGATIVE SURPLUS	David C M Dickson Alfred D Egídio dos Reis

31	MAR 96	RESERVING CONSECUTIVE LAYERS OF INWARDS EXCESS-OF-LOSS REINSURANCE	Greg Taylor
32	AUG 96	EFFECTIVE AND ETHICAL INSTITUTIONAL INVESTMENT	Anthony Asher
33	AUG 96	STOCHASTIC INVESTMENT MODELS: UNIT ROOTS, COINTEGRATION, STATE SPACE AND GARCH MODELS FOR AUSTRALIA	Michael Sherris Leanna Tedesco Ben Zehnwirth
34	AUG 96	THREE POWERFUL DIAGNOSTIC MODELS FOR LOSS RESERVING	Ben Zehnwirth
35	SEPT 96	KALMAN FILTERS WITH APPLICATIONS TO LOSS RESERVING	Ben Zehnwirth
36	OCT 96	RELATIVE REINSURANCE RETENTION LEVELS	David C M Dickson Howard R Waters
37	OCT 96	SMOOTHNESS CRITERIA FOR MULTI-DIMENSIONAL WHITTAKER GRADUATION	Greg Taylor
38	OCT 96	GEOGRAPHIC PREMIUM RATING BY WHITTAKER SPATIAL SMOOTHING	Greg Taylor
39	OCT 96	RISK, CAPITAL AND PROFIT IN INSURANCE	Greg Taylor
40	OCT 96	SETTING A BONUS-MALUS SCALE IN THE PRESENCE OF OTHER RATING FACTORS	Greg Taylor
41	NOV 96	CALCULATIONS AND DIAGNOSTICS FOR LINK RATION TECHNIQUES	Ben Zehnwirth Glen Barnett
42	DEC 96	VIDEO CONFERENCING IN ACTUARIAL STUDIES - A THREE YEAR CASE STUDY	David M Knox
43	DEC 96	ALTERNATIVE RETIREMENT INCOME ARRANGEMENTS AND LIFETIME INCOME INEQUALITY: LESSONS FROM AUSTRALIA	Margaret E Atkinson John Creedy David M Knox
44	JAN 97	AN ANALYSIS OF PENSIONER MORTALITY BY PRE-RETIREMENT INCOME	David M Knox Andrew Tomlin
45	Jul 97	TECHNICAL ASPECTS OF DOMESTIC LINES PRICING	Greg Taylor