MODELLING THE CLAIMS PROCESS IN THE PRESENCE OF COVARIATES

By Arthur E. Renshaw

Department of Actuarial Science & Statistics
The City University, London

ABSTRACT

An overview of the potential of Generalized Linear Models as a means of modelling the salient features of the claims process in the presence of rating factors is presented. Specific attention is focused on the rich variety of modelling distributions which can be implemented in this context.

KEYWORDS

Claims Process; Rating Factors; Generalized Linear Models; Quasi-Likelihood; Extended Quasi-Likelihood.

1. INTRODUCTION

The claims process in non-life insurance comprises two components, claim frequency and claim serverity, in which the product of the underlying expected claim rate and expected claim severity defines the pure or risk premium. Specifically, considerable attention is given to the probabalistic modelling of various aspects of a single batch of claims, often focusing on the aggregate claims accruing in a time period of fixed duration, typically one year, under a variety of assumptions imposed on the claim frequency and claim severity mechanisms.

In this paper, attention is refocused on the considerable potential of generalized linear models (GLMs) as a comprehensive modelling tool for the study of the claims process in the presence of covariates. Section 2 contains a brief summary of the main features of GLMs which are of potential interest in modelling various aspects of the claims process. Particular attention is drawn to the rich variety of modelling distributions which are available and to the parameter estimation and model fitting techniques based on the concepts of quasi-likelihood and extended quasi-likelihood. Sections 3 and 4 focus respectively on the modelling of the claim frequency and claim severity components of the process in the presence of covariates. An overview of the potential of GLMs as a means of modelling these two aspects of the claims process is discussed. Relevant published applications are referenced, although an exhaustive search of the literature has not been conducted. A number of the suggested modelling techniques are illustrated in Section 5.

2. GLMs. QUASI-LIKELIHOOD. EXTENDED QUASI-LIKELIHOOD

Focus intially on independent response variables $\{Y_i: i = 1, 2, \ldots, n\}$ with either density or point mass function, as the case may be, of the type

$$(2.1) \qquad f(y_i \mid \theta_i, \phi_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i)\right\}$$

for specified functions $a(.)$, $b(.)$ and $c(.)$, where θ_i is the canonical parameter and ϕ_i the dispersion parameter. The cumulant function $b(.)$ plays a central role in characterising many of the properties of the distribution. It gives rise to the cumulant generating function, K, of the random variable Y_i, assuming it exits, according to the equation

$$(2.2) \qquad K_{Y_i}(t) = \frac{b\{a(\phi_i)t + \theta_i\} - b\{\theta_i\}}{a(\phi_i)}.$$

Our immediate concern therefore is with distributions with at most two parameters.

Let $\mu_i = E(Y_i)$ throughout. Comparison of the density or point mass function of a standard distribution with expression (2.1) establishes membership or otherwise of this class of distributions. It also determines the specific nature of the canonical parameter θ_i and function $a(.)$ up to a constant, as well as the nature of the dispersion parameter ϕ_i and the other two functions $b(.)$ and $c(.)$. To uniquely determine θ_i and $a(.)$ it is also necessary to compare the variance of the standard distributions with the general expression (2.6) or, more specifically, expression (2.8) for the variance of Y_i.

For inference, the log-likelhood is

$$(2.3) \qquad l = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} \left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i)\right\}.$$

The identity

$$(2.4) \qquad E\left\{\frac{\partial l_i}{\partial \theta_i}\right\} = 0 \implies E(Y_i) = \mu_i = b'(\theta_i)$$

where dash denotes differentiation. Thus, provided the function $b'(.)$ has an inverse, which is defined to be the case, the canonical parameter $\theta_i = b'^{-1}(\mu_i)$, a known function of μ_i.

The identity

$$E\left\{\frac{\partial^2 l_i}{\partial \theta_i^2}\right\} + E\left\{\left(\frac{\partial l_i}{\partial \theta_i}\right)^2\right\} = 0 \implies \text{Var}(Y_i) = b''(\theta_i)\, a(\phi_i)$$

the product of two functions. Noting that $b''(.)$ is a function of the canonical parameter θ_i and hence of μ_i, the identity

$$(2.5) \qquad b''(\theta_i) = V(\mu_i)$$

is established and hence the so-called variance function $V(.)$ defined. Hence the variance or second cumulant is

(2.6) $$\mathrm{Var}\,(Y_i) = \kappa_2^{(i)} = V(\mu_i)\,a(\phi_i).$$

The other function $a(.)$ is commonly of the type

(2.7) $$a(\phi_i) = \frac{\phi}{\omega_i}$$

with constant scale parameter ϕ and prior weights ω_i so that

(2.8) $$\mathrm{Var}\,(Y_i) = \frac{\phi V(\mu_i)}{\omega_i}.$$

This is assumed to be the case throughout. We remark that by setting $\phi = 1$, $1/\omega_i = \phi_i$, the reciprocals of the weights may also be re-interpreted as non-constant scale parameters ϕ_i.

We shall also have occasion to examine the degree of skewness in the $Y_i s$. Here the identity

$$E\left\{\frac{\partial^3 l_i}{\partial \theta_i^3}\right\} + 3E\left\{\frac{\partial^2 l_i}{\partial \theta_i^2}\frac{\partial l_i}{\partial \theta_i}\right\} + E\left\{\left(\frac{\partial l_i}{\partial \theta_i}\right)^3\right\} = 0 \Rightarrow E\{(Y_i - \mu_i)^3\} = b'''(\theta_i)\,a^2(\phi_i)$$

so that, in terms of the variance function $V(.)$, on using equation (2.5), the third cumulant of Y_i is

$$\kappa_3^{(i)} = V\frac{dV}{d\mu_i}\{a(\phi_i)\}^2.$$

Hence the coefficient of skewness

(2.9) $$\frac{\kappa_3^{(i)}}{\{\kappa_2^{(i)}\}^{3/2}} = V^{-1/2}\frac{dV}{d\mu_i}\{a(\phi_i)\}^{1/2}.$$

The expressions for the second and third cumulants can also be derived from the cumulant generating function (2.2).

Covariates may be either explanatory variables, or explanatory factors, or a mixture of both. In all three cases, covariates enter through a linear predictor

$$\eta_i = \sum_j x_{ij}\beta_j$$

with known covariate structure (x_{ij}) and unknown regression parameters β_j and are linked to be mean, μ_i, of the modelling distribution through a monotonic, differentiable (link) function g with inverse g^{-1}, such that

$$g(\mu_i) = \eta_i \quad \text{or} \quad \mu_i = g^{-1}(\eta_i).$$

To fit such a model structure, maximum likelihood estimates for the $\beta_j s$ are normally sought. These are obtained through the numerical solution of the equations

$$(2.10) \qquad \sum_{i=1}^{n} \omega_i \frac{y_i - \mu_i}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0 \qquad \forall j$$

derived by setting the partial derivatives

$$\frac{\partial l}{\partial \beta_j} = \sum_i \frac{\partial l_i}{\partial \beta_j} = \sum_i \frac{\partial l_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} = \sum_i \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}$$

of the log-likelihood with respect to the unknown parameters β_j to zero. Equations (2.3), (2.4), (2.5) and (2.7) are needed in the evaluation of the first two partial derivative terms on the right hand side. These estimates are sufficient in the case of the canonical link function, defined by $g = b'^{-1}$.

To broaden the genesis of equations (2.10) by relaxing the constraints imposed by the full log-likelhood assumption (2.3) and its associated distribution assumption (2.1), define

$$(2.11) \qquad q = q(\underline{y}; \underline{\mu}) = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} \omega_i \int_{y_i}^{\mu_i} \frac{y_i - s}{\phi V(s)} \, ds$$

to be the quasi-likelihood (strictly quasi-log-likelihood) function. Then by setting the partial derivatives of q (rather than l) with respect to β_j to zero, equations (2.10) are again reproduced. Equations (2.10) are called the Wedderburn quasi-likelihood estimating equations. The resulting quasi-likelihood parameter estimates have similar asymptotic properties to maximum likelihood parameters estimates and are identical to maximum likelihood parameter estimates for the class of distributions defined by equation (2.1). This latter class of distributions includes the binomial, Poisson, gamma and inverse Gaussian distributions, all of which are of potential interest in a claims context. The individual details are summarised in Table 2.1. The overriding feature of both the quasi-likelihood expression (2.11) and the Wedderburn quasi-likelihood estimating equations (2.10) is that a knowledge of only the first and second moments is required of the modelling distribution of the $Y_i s$. Hence, by this means, it is possible to relax the full log-likelihood assumption (2.3) and extend the range of distributions which can be readily linked to covariates in practice with an attendant shift in emphasis from maximum likelihood estmation to maximum quasi-likelihood estimation. This has important implications for the claims process which are discussed in context later.

The goodness-of-fit of different hierarchical model predictor structures is monitored, in the first instance, by comparing the differences in model deviances. To do this, compare the current model structure, denoted by c, and whose fitted values are denoted by $\hat{\mu}_i$; with the full or saturated model structure, denoted by f, and which is characterised by the fitted values $\tilde{\mu}_i = y_i$, the perfect fit. Let $\hat{\theta}_i$ and $\tilde{\theta}_i$ denote the corresponding values of the canonical parameter, defined by $\theta_i = b'^{-1}(\mu_i)$, the inverse of b'. Since we are concerned here exclusively with changes to the structure

TABLE 2.1

DETAILS OF SPECIFIC GLM DISTRIBUTIONS

	Poisson	binomial	gamma	inverse Gaussian
d.f./p.m.f.	$\dfrac{e^{-\mu}\mu^y}{y!}$	$\binom{m}{y} p^y (1-p)^{m-y}$	$\dfrac{1}{\Gamma(\nu)}\left(\dfrac{\nu}{\mu}\right)^\nu y^{\nu-1} e^{-\nu y/\mu}$	$\dfrac{1}{\sqrt{2\pi\tau y^3}}\exp\left\{\dfrac{-1}{2\tau\mu^2 y}(y-\mu)^2\right\}$
parameters	$\mu > 0$	$p \in (0, 1)$	$\mu, \nu > 0$	$\mu, \tau > 0$
range	$y = 0, 1, 2, \ldots$	$y = 0, 1, 2, \ldots, m$	$y > 0$	$y > 0$
canonic par.	$\theta = \log(\mu)$	$\theta = \log\left(\dfrac{p}{1-p}\right)$	$\theta = -\mu^{-1}$	$\theta = -\dfrac{1}{2}\mu^{-2}$
scale par.	$\phi = 1$	$\phi = 1$	$\phi = \nu^{-1}$	$\phi = \tau$
weights	$\omega = 1$	$\omega = 1$	$\omega = 1$	$\omega = 1$
$b(\theta)$	$\exp(\theta)$	$m\log(1+e^\theta)$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$c(y,\theta)$	$-\log(y!)$	$\log\left\{\binom{m}{y}\right\}$	$\nu\log(\nu) + (\nu-1)\log(y)$ $-\log\{\Gamma(\nu)\}$	$-\dfrac{1}{2}\left\{\log(2\pi y^3) + \dfrac{1}{\tau y}\right\}$
$\mu(\theta)$	$\exp(\theta)$	$\dfrac{me^\theta}{1+e^\theta}$	$-\dfrac{1}{\theta}$	$(-2\theta)^{-1/2}$
$V(\theta)$	μ	$\mu\left(1-\dfrac{\mu}{m}\right)$	μ^2	μ^3
κ_3	μ	$\mu\left(1-\dfrac{\mu}{m}\right)\left(1-\dfrac{2\mu}{m}\right)$	$2\mu^3\phi^2$	$3\mu^5\phi^2$

of the predictor, the scale parameter ϕ remains the same throughout. Then define

$$(2.12) \quad d^*(\underline{y};\hat{\underline{\mu}}) = -2(l^{(c)} - l^{(f)}) = -2\sum_{i=1}^{n} \frac{\omega_i}{\phi}\{(y_i\hat{\theta}_i - b(\hat{\theta}_i)) - (y_i\tilde{\theta}_i - b(\tilde{\theta}_i))\},$$

minus twice the log-likelihood ratio, of c relative to f, based on equations (2.3) and (2.5), to be the scaled deviance and

$$(2.13) \qquad\qquad d(\underline{y};\hat{\underline{\mu}}) = \phi d^*(\underline{y};\hat{\underline{\mu}})$$

to be the (unscaled) deviance of the current model c. Using the identity

$$\int_{y_i}^{\mu_i} \frac{y_i - s}{b''(b'^{-1}(s))}\, ds = \int_{\tilde{\theta}_i}^{\theta_i} (y_i - b'(t))\, dt$$

it follows from equations (2.12), (2.13) and (2.5) that the expression for the

deviance can be written as

$$(2.14) \qquad d(\underline{y};\underline{\mu}) = \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} 2\omega_i \int_{\mu_i}^{y_i} \frac{y_i - s}{V(s)} \, ds = -2\phi q(\underline{y};\underline{\mu})$$

where $q(\underline{y};\underline{\mu})$ is the quasi-likelihood function. Hence in common with the construction of the quasi-likelihood and quasi-likelihood estimating equations, a knowledge of only the first and second moments is required of the modelling distribution of the $Y_i s$ to construct the model deviance.

A trivial re-arrangement of equation (2.14) implies the the quasi-likelihood, q, satisfies

$$-2q = \sum_{i=1}^{n} \frac{d_i}{\phi}.$$

To accommodate inference on any parameters, such as ϕ, which might be present in the variance of the response variables Y_i, define the extended quasi-likelihood (strictly the extended-quasi-log-likelihood) q^+ where

$$(2.15) \qquad -2q^+ = \sum_{i=1}^{n} \frac{d_i}{\phi} + \sum_{i=1}^{n} \log \{\phi V(y_i)\} \qquad \left(+ \sum_{i=1}^{n} \log (2\pi/\omega_i) \right).$$

Note that this expresssion is minus twice the log-likelihood for independent normally distributed responses $Y_i \sim N(\mu_i, \sigma^2)$, for which $\sigma^2 = \phi$, $V(\mu_i) = 1$; but is not an exact log-likelihood expression for any other case. The final term in the brackets is constant for a given data set, and may be omitted.

Diagnostic checks, based on a thorough graphical analysis of residuals, are conducted before the final adoption of a specific model structure. Deviance residuals

$$r_i = \text{sign} (y_i - \hat{\mu}_i) \cdot \sqrt{d_i}$$

where d_i is ith component of the deviance defined in equation (2.14), are advocated. A suitable estimate for the constant scale parameter ϕ, if required, is provided by the moment estimator based on generalized Pearson residuals

$$(2.16) \qquad \hat{\phi} = \frac{1}{\nu} \sum_{i=1}^{n} \omega_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

where ν denotes the number of degrees of freedom associated with the fit.

Implementation is possible using the GLIM software package, BAKER & NELDER (1985) which is expressly designed to fit models of this type, while the reader is referred to the text by MCCULLAGH & NELDER (1989) for further detail.

3. CLAIM FREQUENCY

Claim frequency data are denoted throughout by (u, n_u, e_u), comprising the observed number of claims, n_u, accruing from exposures, e_u, defined for a set of

units $\{u\}$. Typically the units are of the type $u \equiv (i_1, i_2, i_3, \ldots)$, a cross-classified grid of cells defined for preselected levels of appropriate covariates, often rating factors. A number of different possible modelling scenarios can be implemented.

Focus first on targetting the underlying or expected claim rates, denoted by λ_u, based on the Poisson modelling assumption $N_u \mid \lambda_u \sim \text{Poi}(e_u \lambda_u)$, with independence over all cells or units u, and where n_u denotes the realisation of the random variable N_u. Here it is assumed that the claim rates, λ_u, are constant within cells. In the notation of Section 2, the responses $Y_i \equiv N_u$ with

mean $\mu_u = E(N_u) = e_u \lambda_u$, variance function $V(\mu_u) = \mu_u$, scale parameter $\phi = 1$

and log-likelihood

$$(3.1) \qquad l = \sum_{u=1}^{n} \{ -\mu_u + n_u \log (\mu_u) \} + \text{constant}.$$

Two link functions are of particular interest in this context, namely the log-link and the parameterised power-link.

To implement the canonical log-link, for which

$$\eta_u = \log (\mu_u) = \log (e_u) + \log (\lambda_u) = \log (e_u) + \sum_j x_{uj} \beta_j$$

the vector of $\log (e_u)$ terms is declared as an offset. Such terms from part of the linear predictor and are characterised by a known regression coefficient with value one. Thus the target, λ_u, is linked to covariates through the relationship

$$\log (\lambda_u) = \sum_j x_{uj} \beta_j \quad \Leftrightarrow \quad \lambda_u = \exp \left(\sum_j x_{uj} \beta_j \right),$$

giving rise, possibly, to a multiplicative model structure for rating factors.

A number of applications appear in the literature. Thus McCULLAGH & NELDER (1983 & 1989), using data provided by the Lloyd's Register of Shipping concerning damage incidents caused to the forward section of cargo-carrying vessels, model the reported number of damage incidents classified by the three factors- ship type, year of construction, period of operation. To allow for possible inter-ship variability in accident proneness, over-dispersion is introduced into the model through the retention of the scale parameter which is then estimated as described in Section 2, rather than setting its value to one. This modelling refinement has an impact on the standard errors of the parameter estimates but not on the parameter estimates themselves (the solutions to equations (2.10)). ANDRADE E SILVO (1989), BROCK-MAN & WRIGHT (1992) and BOSKOV (1992) have each applied this same model to motor claims data using a variety of potential rating factors in the predictor. CENTENO & ANDRADE E SILVO (1991) discuss the case when there are certain fixed linear relationships between covariates in the predictor. STROINSKI & CURRI (1989) discuss the selection of rating factors in automobile claim frequency modelling. RENSHAW & HABERMAN (1992) have modelled both sickness inception and sickness recovery rates as well as death rates from sickness with the predictor reflecting both

age at sickness inception and, where applicable, sickness duration. A feature of some of this work involves the use of break-point predictor terms in which the positions of the knots or hinges are determined by deviance profiles, constructred by scanning the positional choices of the knots. RENSHAW (1991) has also demonstrated the potential for this model in the graduation of the force of mortality in the construction of life tables.

To implement the parameterised power-link function in this context, the alternative form of the log-likelihood expression:

$$l = \sum_{u=1}^{n} e_u \left\{ -\lambda_u + \frac{n_u}{e_u} \log(\lambda_u) \right\} + \text{constant};$$

obtained by substituting $\mu_u = e_u \lambda_u$ into expression (3.1), is exploited. This implies the declaration of $y_u = n_u/e_u$ as Poisson responses with prior weights e_u, while the predictor link is denoted

$$\lambda_u^\gamma = \sum_j x_{uj} \beta_j \Leftrightarrow \lambda_u = \left(\sum_j x_{uj} \beta_j \right)^{1/\gamma}$$

with link parameter γ. The case $\gamma = 1$ corresponds to the identity-link, while the case $\gamma = 0$ corresponds to the log-link. The optimum value of γ for a specific predictor structure is determined by constructing the deviance profile over the viable range of values of γ. Examples of this are to be found in RENSHAW & HABERMAN (1992) and in RENSHAW (1990).

The Poisson model (with $\phi = 1$) assumes that the claim rate, λ_u, is constant within cells. Heterogeneity across risks as opposed to time heterogeneity discussed by BERG & HABERMAN (1992) is historically introduced into the claim frequency process by modelling λ_u as a random variable. Focus on the weighted Poisson responses $Y_u (= N_u/e_u)$ with $Y_u \sim \text{Poi}(\lambda_u)$ so that

(3.2) $$E(Y_u) = E\{E(Y_u | \lambda_u)\} = E(\lambda_u),$$

$$\text{Var}(Y_u) = E\{\text{Var}(Y_u | \lambda_u)\} + \text{Var}\{E(Y_u | \lambda_u)\}$$

and hence

(3.3) $$\text{Var}(Y_u) = E(\lambda_u) + \text{Var}(\lambda_u).$$

Note that when λ_u is constant, $E(\lambda_u) = \lambda_u$, $\text{Var}(\lambda_u) = 0$ and the within cell homogeneous Poisson model is reproduced. For the heterogeneous case, $\text{Var}(Y_u) > E(Y_u)$, that is, the model is over dispersed. There are a number of feasible practical possibilities available:

1) Allow for heterogeneity through the introduction of a constant scale parameter ϕ as described in some of the applications identified above.
2) Allow for heterogeneity through the introduction of non-constant scale parameters ϕ_u and generate their values through the introduction of a second stage

GLM chosen to model identifiable patterns of heterogeneity across cells; a technique known as joint modelling. An example, applied to life insurance, is to be found in RENSHAW (1992).

3) Allow for heterogeneity by nominating a specific distribution for the claim rate λ_u. Thus commonly in the claims context, λ_u is given a gamma distribution with mean $E(\lambda_u)$ and variance

(3.4) $$\text{Var}(\lambda_u) = \frac{1}{\nu} \{E(\lambda_u)\}^2 = \frac{1}{\nu} \{E(Y_u)\}^2 \qquad \nu > 0,$$

on using equation (3.2). Then, it is well known that Y_u has the negative binomial distribution, for which the

mean $\mu_u = E(Y_u)$, variance function $V(\mu_u) = \mu_u + \frac{1}{\nu}\mu_u^2$, scale parameter $\phi = 1$

on substituting expression (3.4) into equation (3.3). Note that as $\nu \to \infty$, then for finite μ_u, the distribution reverts to the Poisson distribution. Another possibility, BESSON & PARTRAT (1992), TREMBLAY (1992), is to assign an inverse Gaussian distribution with mean $E(\lambda_u)$ and scale parameter τ. It then follows from the relevant column of Table 2.1 that

$$\text{Var}(\lambda_u) = \tau \{E(\lambda_u)\}^3 = \tau \{E(Y_u)\}^3 \qquad \tau > 0$$

so that Y_u now has the Poisson-inverse Gaussian distribution with

mean $\mu_u = E(Y_u)$, variance function $V(\mu_u) = \mu_u + \tau\mu_u^3$, scale parameter $\phi = 1$.

This reverts to the Poisson distribution as $\tau \to 0$. Neither of these cases are members of the class of distributions defined by expression (2.1) so that their implementation lead to quasi-likelihood estimators for the $\beta_j s$ in the predictor. If these models are to be implemented, explicity expressions are needed for the deviance components defined in equation (2.14). These are

$$2\omega_u \left\{ y_u \log\left(\frac{y_u}{\mu_u}\right) + (y_u + \nu) \log\left(\frac{\mu_u + \nu}{y_u + \nu}\right) \right\}$$

for the negative binomial distribution, and

$$2\omega_u \left\{ y_u \log\left(\frac{y_u}{\mu_u}\right) + \frac{y_u}{2} \log\left(\frac{1 + \tau\mu_u^2}{1 + \tau y_u^2}\right) + \frac{1}{\sqrt{\tau}} \sin^{-1}\sqrt{\frac{\tau(\mu_u - y_u)^2}{1 + (\tau y_u \mu_u)^2 + \tau(y_u^2 + \mu_u^2)}} \right\}$$

for the Poisson-inverse Gaussian distribution. Implementation also requires a knowledge of the variance function parameters ν and τ. This is discussed in Section 5.

Focus secondly on targetting the probability of a claim (or at least one claim), denoted by p_u, based on the binomial modelling assumption $N_u | p_u \sim \text{bin}(e_u, p_u)$, with independence over all cells or units u, where again n_u denotes a realisation of

the random variable N_u. In the notation of Section 2, the responses $Y_i \equiv N_u$ with

mean $\mu_u = E(N_u) = e_u p_u$, variance function $V(\mu_u) = \mu_u \left(1 - \dfrac{\mu_u}{e_u} \right)$, scale parameter

$\phi = 1$ and log-likelihood

$$l = \sum_{u=1}^{n} \left\{ n_u \log \left(\frac{\mu_u}{e_u} \right) + (e_u - n_u) \log \left(\frac{e_u - \mu_u}{e_u} \right) \right\} + \text{constant}.$$

The canonical log-odds or logit link

$$\eta_u = \log \left(\frac{\mu_u}{e_u - \mu_u} \right) = \log \left(\frac{p_u}{1 - p_u} \right) \Leftrightarrow p_u = \frac{e^{\eta_u}}{1 + e^{\eta_u}}$$

with linear predictor

$$\eta_u = \sum_j x_{uj} \beta_j$$

is likely to be of interest in a non-life claim frequency context, while its application in this context would appear to be somewhat limited. An application of its use in targetting the probability of at least one claim in the context of (Belgium) car insurance claims is given by BEIRLANT et al. (1991). A number of researchers, including COUTTS (1984), have used this predictor-link structure to target claim proportions, over a network of cells but with estimation by weighted least squares. The binomial modelling distribution assumption, used in conjunction with the logit and other link functions, has wide application in the construction of life tables, RENSHAW (1991).

4. CLAIM SEVERITY

Claim severity or loss distributions, defined on the positive real line, are invariably positively skewed. There is an extensive literature, see for example, HOGG and KLUGMAN (1984), documenting the modelling of homogeneous batches of empirical claim amounts by specific parameterised distributions. These include the gamma, Pareto, log-normal, log-gamma and Weibull distributions only the first of which is of the type defined by expression (2.1). HABERMAN & RENSHAW (1989) have indicated how certain loss distributions, not of the type defined by expression (2.1), may be fitted in the absence of covariates by the adaption of the associated likelihood function in a special way. Here we address the question: which loss distributions are capable of sustaining covariates?

    Mean claim amounts are denoted throughout by x_u, categorised over a set of units $\{u\}$. Thus data summaries take the form (u, n_u, x_u) where x_u denotes the claim average in cell u based on n_u claims. The independence of the number n_u and the claim average x_u within each unit u is assumed. Again typically the units $u \equiv (i_1, i_2, i_3, \ldots)$, a cross-classified grid of cells defined for preselected levels of appropriate covariates, often rating factors. Denoting the underlying expected claim

severity in cell u by μ_u and assuming the independence of individual claim amounts, the cells means X_u are modelled as the responses of a GLM with $E(X_u) = \mu_u$ and $\text{Var}(X_u) = \phi V(\mu_u)/n_u$. Covariates defined on $\{u\}$ enter through a linear predictor, linked to the mean μ_u.

Focus first on the gamma distribution. Precedence for its use in this context is to be found in McCULLAGH and NELDER (1983 & 1989) in which a re-analysis of the celebrated car insurance data of BAXTER, COUTTS and ROSS (1979) is presented. The data comprise (u, n_u, x_u) the number n_u and average cost of claims x_u, cross-classified by policy holder's age, car group and vehicle age. Modelling is based on independent gamma distributed individual claim amounts, for which the mean responses, X_u, satisfy

$$\text{mean } \mu_u = E(X_u), \text{ variance function } V(\mu_u) = \mu_u^2, \text{ scale parameter } \phi = \nu^{-1} > 0$$

with weights n_u so that $\text{Var}(X_u) = \phi\mu_u^2/n_u$. In the analysis, a linear predictor η_u, composed of the additive main effects of the three factors only, is linked to μ_u through the canonical reciprocal link function. Factor interaction terms are found not to be significant. Use of the reciprocal-link function, a member of the parameterised family of power link function

$$(4.1) \qquad\qquad\qquad \mu_u^\gamma = \eta_u$$

with $\gamma = -1$, is justified on the basis of the deviance profile constructed by allowing for incremental changes in γ. The model proposed by MACK (1991) for rating automobile insurance makes identical distributional assumptions to these (formulated in terms of cell sums rather than in terms of cell averages) but restricts the modelling to the log-link, the limiting form of the power link as γ tends to zero in order to focus on a multiplicative structure. The detail is presented in terms of two rating factors so that $u \equiv (i, j)$ with model structure

$$\log(\mu_{ij}) = \alpha_i + \beta_j;$$

while the maximum likelihood parameter estimating equations discussed by MACK (1991) are special cases of equations (2.10). Implementation is readily achieved using the GLIM software package. MACK (1991) also goes on to apply the same model structure in the claims reserving context. BROCKMAN & WRIGHT (1992) use the identical model structure to MACK (1991) in their analysis of the severity of motor claims data.

Focus next on the Pareto distribution with density

$$f(x \mid \beta, \lambda) = \frac{\beta\lambda^\beta}{(\lambda + x)^{\beta+1}}, \qquad x > 0$$

and parameters $\beta, \lambda > 0$. It follows that

$$E(X) = \frac{\lambda}{\beta - 1}, \qquad \text{Var}(X) = \frac{\beta\lambda^2}{(\beta - 1)^2 (\beta - 2)}$$

provided $\beta > 2$. Introducing the reparameterisation

$$\mu = \frac{\lambda}{\beta - 1}, \qquad \phi = \frac{\beta}{\beta - 2};$$

a $1:1$ mapping $(\beta, \lambda) \mapsto (\mu, \phi)$ with domain $R_{>2} \times R_{>0}$ and image set $R_{>0} \times R_{>1}$, implies that we can construct a GLM based on independent Pareto distributed claim amounts for which the mean responses, X_u, satisfy

mean $\mu_u = E(X_u)$, variance function $V(\mu_u) = \mu_u^2$, scale parameter $\phi > 1$

and weights n_u so that again $\text{Var}(X_u) = \phi \mu_u^2 / n_u$. Apart from the mild extra constraint on the scale parameter, these details are identical to those of the GLM based on independent gamma responses, and the two different modelling assumptions lead to essentially identical GLMs. They differ only in respect of the nature of the parameter estimation criterion, maximum likelihood in the case of the gamma response model and maximum quasi-likelihood in the case of the Pareto response model.

Focus next on the generalization of these distributional assumptions through the introduction of the parameterised power variance function

(4.2) $$V(\mu) = V(\mu, \zeta) = \mu^\zeta.$$

The gamma and Pareto distributions are essentially identical special cases with $\zeta = 2$. The characteristics of this family of distributions are discussed in detail by, for example, JORGENSEN (1987). Simplifying the notation slightly by writing equation (2.4) as

$$\mu = \mu(\theta) = b'(\theta) \qquad \text{with inverse} \qquad \theta = \mu^{-1}(\mu),$$

it follows that the cumulant function $b(.)$, corresponding to a specific variance function $V(.)$, is determined by solving the equations

$$\frac{db}{d\theta} = \mu(\theta), \qquad \frac{d}{d\mu} \mu^{-1}(\mu) = \frac{1}{V(\mu)}.$$

First, the solution of the second of these equations determines $\mu^{-1}(.)$. This is then inverted to provide the expression for the right hand side of the first equation, which is then solved in turn for $b(.)$. For the power variance function defined by equation (4.2), the special solution of these equation obtained by setting the arbitrary constant of integration to zero is given by

(4.3) $$b(\theta) = \begin{cases} \exp(\theta) & \zeta = 1 \\[2ex] \dfrac{\alpha - 1}{\alpha} \left(\dfrac{\theta}{\alpha - 1} \right)^\alpha & \zeta \neq 1, 2 \\[2ex] -\log(-\theta) & \zeta = 2 \end{cases}$$

where

(4.4) $$\alpha = \frac{\zeta - 2}{\zeta - 1}.$$
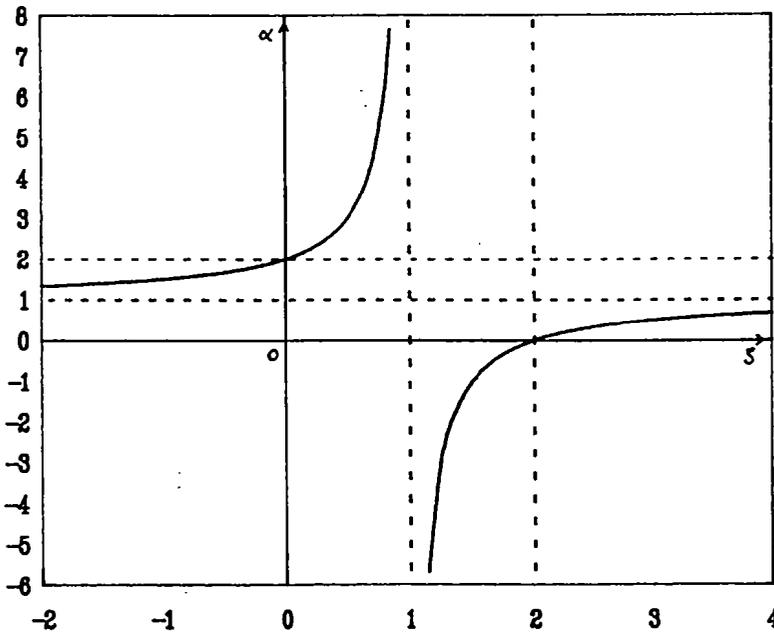
FIGURE 4.1. Graph of α against ζ.

Equation (4.3) characterises the properties of the distribution in question, while equation (4.4) is reproduced graphically in Figure 4.1.

For $\zeta > 2$ $(0 < \alpha < 1)$, $b(\theta)$ is the cumulant function of an extreme stable distribution with index α, see EATON, MORRIS and RUBIN (1971). The cumulant generating function and hence moment generating function is obtained by substituting expression (4.3) into equation (2.2). It generates GLMs with parameterised power variance function, equation (4.2) with $\zeta > 2$; has positive support, and is positive skewed. Equations (4.2), (2.7) and (2.9) determine the coefficient of skewness

$$\zeta \mu_u^{\frac{\zeta}{2} - 1} \left(\frac{\phi}{\omega_u} \right)^{1/2} \qquad \zeta > 2.$$

The family of distributions has therefore all the major characteristics of a loss distribution. It includes the inverse Gaussian distribution ($\zeta = 3$) and has the gamma distribution ($\zeta = 2$) as a limiting case. It represents a generalisation of these two cases. COUTTS (1984) suggests the potential of the two specific cases in the claim severity modelling context. For a given predictor-link structure, the optimum value for $\zeta \geq 2$ is determined by scanning (minus twice) the extended-quasi-likelihood profile, expression (2.15) namely

$$-2q^+ = \sum_{u=1}^{n} \frac{d_u}{\phi} + \sum_{u=1}^{n} \log \{\phi V(y_u)\}$$

for incremental changes in ζ. To compute these values, ϕ is estimated by expression (2.16) and (4.2), while evaluation of the integral in expression (2.14) for the power variance function given by equation (4.2), yields the deviance components

$$d_u = 2\omega_u \left\{ -\frac{x_u}{1-\zeta}(\mu_u^{1-\zeta} - x_u^{1-\zeta}) + \frac{1}{2-\zeta}(\mu_u^{2-\zeta} - x_u^{2-\zeta}) \right\} \qquad \zeta \neq 1, 2.$$

Implementation is possible using the GLIM software package. McCULLAGH and NELDER (1989) illustrate the extended-quasi-likelihood profile for the BAXTER et al. (1980) car insurance data set, which is optimal in the vicinity of $\zeta = 2.4$. They also demonstrate for these data how contour plots of the extended-quasi-likelihood determine the joint optimum position for the parameters (γ, ζ) when the parameterised power link function (4.1) is used in combination with the parameterised power variance function (4.2).

So far we have dealt with the cases $\zeta \geq 2$. The case $2 > \zeta > 1$ $(\alpha < 0)$ is also of considerable interest but in the slightly different context of aggregate claims. It is discuss by RENSHAW (1993). Of the remaining cases, $\zeta = 1$ reproduces the Poisson modelling distribution, $1 > \zeta > 0$ $(\alpha > 2)$ does not generale GLMs with distributions of the type defined by equation (2.1), $\zeta = 0$ generates the normal model, while finally $0 > \zeta$ $(1 < \alpha < 2)$ generates extreme stable distributions with support on the whole of the real line which, for this and other reasons, are of no practical consequence here.

Other claim severity modelling distributions capable of supporting covariates are the log-normal and the log-gamma distributions.

5. AN APPLICATION

The motor insurance claims experience for a recent calendar year, made available by a leading U.K. insurance company, is available for analysis. By way of illustration, the data have been edited to read as follows:

$$(t, u, e_u^{(t)}, n_u^{(t)}, x_u^{(t)})$$

where

t- claim type (1- insured vehicle, 2- third party property, 3- third party injury, 4- others)

$u \equiv (i, j, k, l, m)$ – units or cells based on 5 cross-classified rating factors

$pa : i = 1, 2, 3, 4, 5$ – polyholders age (levels arranged in increasing age bands)

$vt : j = 1, 2, 3, 4, 5$ – vehicle type (levels arranged in perceived order of increasing risk)

$va : k = 1, 2, 3, 4,$ – vehicle age (levels arranged in increasing age bands)

$rd : l = 1, 2, 3, 4, 5$ – rating district (levels arranged in perceived order of increasing risk)

$cd : m = 1, 2, 3, 4,$ – no claims discount (4 levels, arranged in order of increasing discount)

$$e_u^{(t)} - \text{exposures}$$

$$n_u^{(t)} - \text{number of claims}$$

$$x_u^{(t)} - \text{mean claim severity}.$$

The independence of the number and the claim average within each cell for each claim type is assumed. The banding of the rating factors is deliberately left ill defined, and only selective features of the ensuing analysis presented by way of illustration. Other groupings of the rating factors are possible.

Residual histogram.

```
+------------+------:---------:---------:--------:--------:--------:-+
|(-5.00,-4.50)| 1    |         |         |        |        |        | |
|(-4.50,-4.00)| 1    |         |         |        |        |        | |
|(-4.00,-3.50)| 0    |         |         |        |        |        | |
|(-3.50,-3.00)| 3    |         |         |        |        |        | |
|(-3.00,-2.50)| 20   |+++      |         |        |        |        | |
|(-2.50,-2.00)| 34   |+++++    |         |        |        |        | |
|(-2.00,-1.50)| 110  |++++++++++++++++  |        |        |        | |
|(-1.50,-1.00)| 196  |+++++++++++++++++++++++++++++ |        |        | |
|(-1.00,-0.50)| 282  |+++++++++++++++++++++++++++++++++++++++++   | |
|(-0.50, 0.00)| 347  |++++++++++++++++++++++++++++++++++++++++++++++++++++++++ |
|( 0.00, 0.50)| 360  |++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++|
|( 0.50, 1.00)| 254  |++++++++++++++++++++++++++++++++++++++++   | |
|( 1.00, 1.50)| 179  |+++++++++++++++++++++++++++  |        |        | |
|( 1.50, 2.00)| 110  |++++++++++++++++  |        |        |        | |
|( 2.00, 2.50)| 51   |+++++++  |         |        |        |        | |
|( 2.50, 3.00)| 19   |+++      |         |        |        |        | |
|( 3.00, 3.50)| 9    |+        |         |        |        |        | |
|( 3.50, 4.00)| 3    |         |         |        |        |        | |
|( 4.00, 4.50)| 0    |         |         |        |        |        | |
+------------+------:---------:---------:--------:--------:--------:-+
             0      70        140       210      280      350
```

Residual histogram.

```
+------------+------:---------:---------:--------:--------:--------:+
|(-5.40,-4.80)| 2    |         |         |        |        |       |
|(-4.80,-4.20)| 7    |+        |         |        |        |       |
|(-4.20,-3.60)| 16   |++       |         |        |        |       |
|(-3.60,-3.00)| 49   |+++++++  |         |        |        |       |
|(-3.00,-2.40)| 64   |++++++++++|        |        |        |       |
|(-2.40,-1.80)| 121  |++++++++++++++++++|        |        |       |
|(-1.80,-1.20)| 211  |+++++++++++++++++++++++++++++++  |        |       |
|(-1.20,-0.60)| 285  |+++++++++++++++++++++++++++++++++++++++++++  |
|(-0.60, 0.00)| 302  |++++++++++++++++++++++++++++++++++++++++++++  |
|( 0.00, 0.60)| 334  |+++++++++++++++++++++++++++++++++++++++++++++++++|
|( 0.60, 1.20)| 212  |++++++++++++++++++++++++++++++  |        |       |
|( 1.20, 1.80)| 173  |+++++++++++++++++++++++++  |        |        |       |
|( 1.80, 2.40)| 104  |+++++++++++++++  |        |        |        |       |
|( 2.40, 3.00)| 60   |+++++++++|        |        |        |        |       |
|( 3.00, 3.60)| 23   |+++      |         |        |        |       |
|( 3.60, 4.20)| 4    |+        |         |        |        |       |
|( 4.20, 4.80)| 7    |+        |         |        |        |       |
|( 4.80, 5.40)| 1    |         |         |        |        |       |
|( 5.40, 6.00)| 0    |         |         |        |        |       |
+------------+------:---------:---------:--------:--------:--------:+
             0      70        140       210      280
```

FIGURE 5.1. Deviance residual histograms. Poisson claim frequency (top), gamma claim severity models.
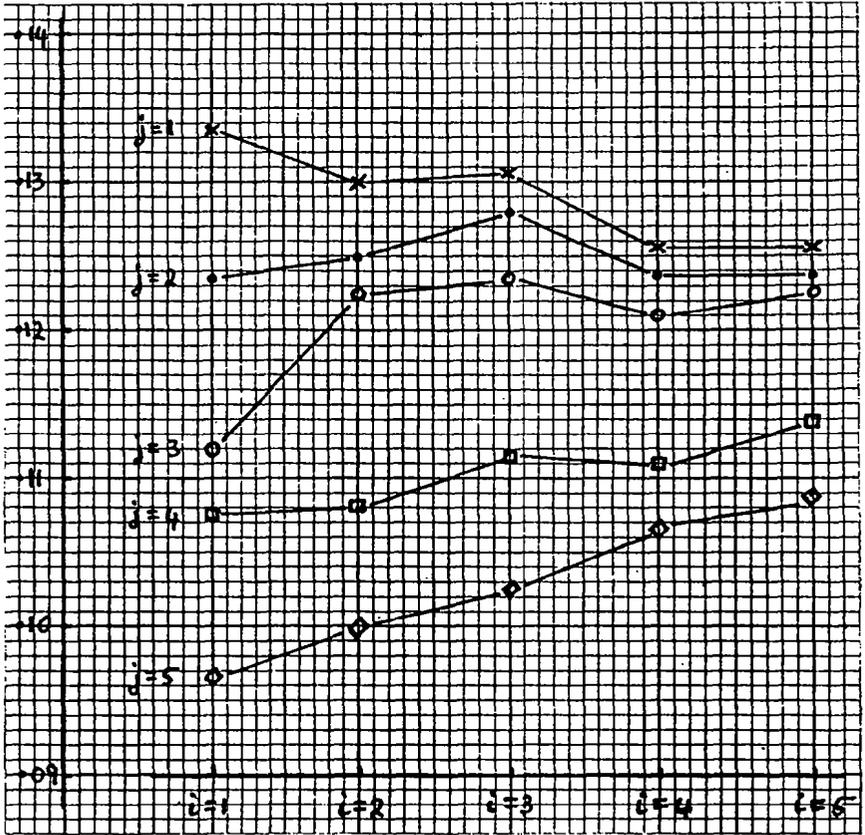
FIGURE 5.2. Contribution of the interactive factors $pa * ut$ to the linear predictor. Claim severity model.
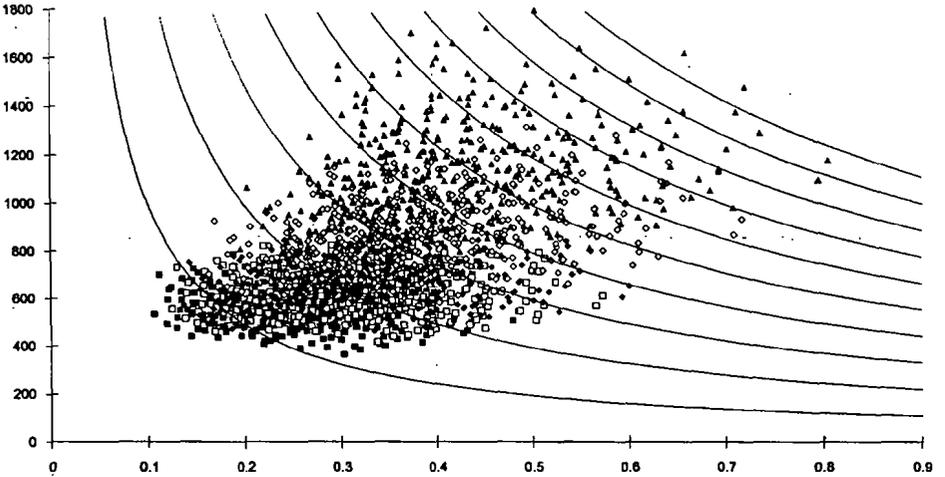
FIGURE 5.3. Expected claim severity $\hat{\mu}_u^{(1)}$ plotted against expected claim rate $\hat{\lambda}_u^{(1)}$.
Classification by specific rating factor.

The modelling of the claim frequency and claim severity patterns across units u, for different claim types, provides estimates $\hat{\lambda}_u^{(t)}$ and $\hat{\mu}_u^{(t)}$ of the expected claim rates and expected claim severities respectively. The contribution, $rp_u^{(t)}$, to the risk premium for claim type t is then the product $\hat{\lambda}_u^{(t)}\hat{\mu}_u^{(t)}$ and the risk premium, rp_u, determined by summation over all risk types t. Thus we have the sequences of mappings:

$$(t, u, e_u^{(t)}, n_u^{(t)}, x_u^{(t)}) \mapsto (t, u, \hat{\lambda}_u^{(t)}, \hat{\mu}_u^{(t)}) \mapsto (t, u, rp_u^{(t)}) \mapsto (u, rp_u)$$

where

$$rp_u^{(t)} = \hat{\lambda}_u^{(t)}\hat{\mu}_u^{(t)} \qquad \text{and} \qquad rp_u = \sum_t rp_u^{(t)}.$$

We focus attention on the first mapping which represents the modelling stage and illustrate the application of various of the suggested modelling assumptions for damage to insured vehicle claim types ($t = 1$).

Consider first the Poisson claim frequency and gamma claim severity models, each with log-link functions and predictor structures composed of main effects and paired interaction terms. The improvements in the quality of the fits, monitored by the changes in deviance, as first main effects and then interaction effects are added to the predictor structures are examined. One such sequence of fits is reproduced in Table 5.1. An examination of such tables coupled with an examination of the resulting parameter estimates and their standard errors for each fitted model prompted the adoption of the predictor structure (expressed in GLIM notation)

$$pa * (vt + rd + cd) + va$$

comprising all five main effects and three second order interactions, all involving policyholders age pa, for the claim frequency model structure; and

$$pa * vt + va + rd + cd$$

comprising all five main effects and just one second order interaction term for the claim severity model structure. The various deviance residual plots are also highly supportive of the two fits. By way of illustration, only the deviance residual histograms are reproduced in Figure 5.1. The impact of the single interaction term on the claim severity model structure with parametric representation

$$pa * vt + va + rd + cd : \eta_{ijklm} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \delta_l + \varepsilon_m$$

is illustrated in Figure 5.2 in which the estimated values of $\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ are plotted against each level i of pa for each level j of vt. Without the interaction terms $(\alpha\beta)_{ij}$ a series of parallel 'lines' would result as the structure is then additive in these factors on the log scale. For this model the expected claim severities are determined by $\mu_{ijklm} = \exp(\eta_{ijklm})$. It can be informative to plot the resulting estimated claim frequencies $\hat{\lambda}_u^{(1)}$ against their corresponding estimated claim severities $\hat{\mu}_u^{(1)}$. One such scatter plot is illustrated in Figure 5.3. In addition the contours displayed represent those of constant risk premium levels, $\hat{\lambda}_u^{(1)}\hat{\mu}_u^{(1)} = $ constant. Here also the impact of the different levels of a rating factor are highlighted by the use of a different symbol to represent each level of this factor. The clustering of the different symbols is informative.

TABLE 5.1

EXAMINATION OF MAIN EFFECTS AND TWO FACTOR INTERACTIONS ON THE DEVIANCES

	claim frequency		claim severity		
	dev.	diff.	dev.	diff.	d.f.
1	15371		12111		
		3295		5268	4
+ vt	12076		6842.5		
		4359		253.2	4
+ pa	7716.8		6589.3		
		327.3		773.2	3
+ va	7389.5		5816.1		
		2039		192.8	4
+ rd	5350.3		5623.3		
		2172		572.3	3
+ cd	3178.1		5050.9		
		95.57		355.7	16
+ vt · pa	3082.5		4695.2		
		26.67		19.3	12
+ vt · va	3055.8		4675.9		
		17.85		35.8	16
+ vt · rd	3038.0		4640.1		
		70.88		41.3	12
+ vt · cd	2967.1		4598.8		
		91.83		31.5	12
+ pa · va	2875.3		4567.3		
		69.45		70.15	16
+ pa · rd	2805.8		4497.1		
		457.4		35.1	12
+ pa · cd	2348.4		4462.0		
		24.79		18.6	12
+ va · rd	2323.6		4443.4		
		28.16		22.6	9
+ va · cd	2295.4		4420.8		
		42.63		29.9	12
+ rd · cd	2252.8		4390.9		

Reverting next to the power-link in combination with the same predictor structures as above, the resulting deviance profiles, constructed over a range of values of the power index, are reproduced in Figure 5.4. For the Poisson claim frequency model, the optimum power is at $\gamma = -0.17$, which is sufficiently close to zero to lend support to the log-link. Indeed if the one remaining paired interaction term involving the rating factor pa (and va) is added to the predictor structure, then the optimum value of the power is essentially at zero. For the gamma claim severity model the optimum value of the power in the link is at $\gamma = -0.31$, somewhat intermediate between the canonical reciprocal-link and the log-link. If the one interaction term is omitted from the model structure, the optimum power value shifts much closer to be reciprocal-link, a similar conclusion to that reported in McCULLAGH and NELDER (1989) in their reanalysis of the BAXTER et al. (1979) data set involving a main effects structure.
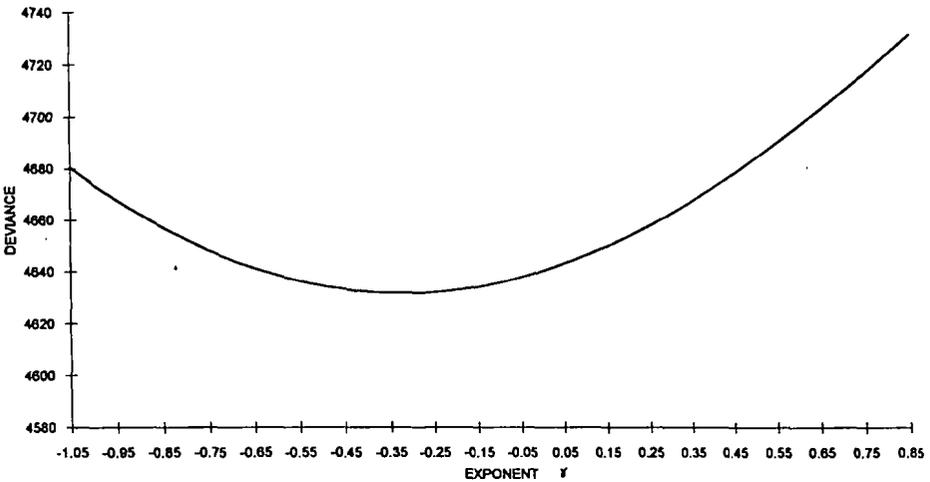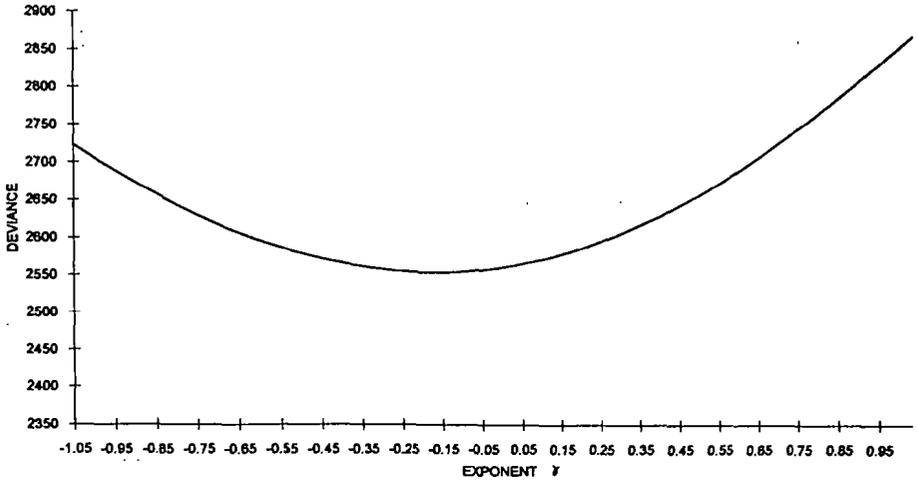
FIGURE 5.4. Deviance profiles, power links. Poisson claim frequency (top),
gamma claim severity models.

For the claim severity model with power variance function and exponent $\zeta \geq 2$, in combination with the log-link and the above predictor structure, the deviance profile over $\zeta \geq 2$ has a similar U-shape to Figure 5.4 with a minimum at $\zeta = 2.63$. Thus the optimum positions of both the exponents in the power link and in the power variance function have so-far been chosen separately by allowing for variation in just one of the exponents while keeping the other exponent fixed. The joint optimum positions of the two exponents (γ, ζ) when the power link function (4.1) is used in combination with the power variance function (4.2) is determined by
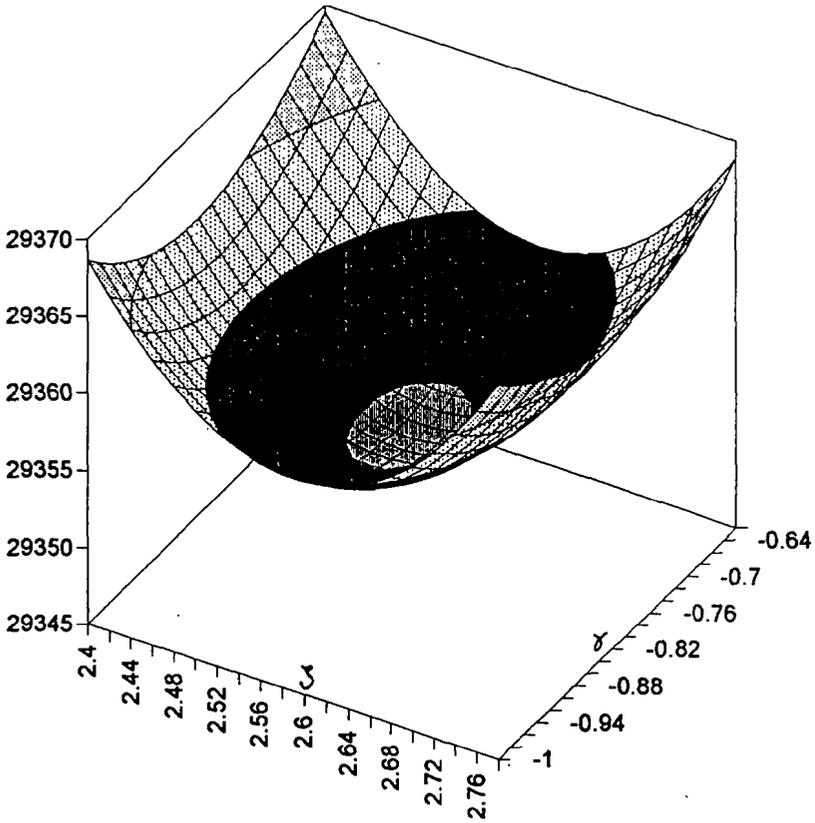
FIGURE 5.5. Extended quasi deviance profile, power link and variance function.
Link exponent γ, variance function exponent ζ.

scanning the extended quasi-deviance profile defined by equation (2.15), part of which is reproduced in Figure 5.5 showing an optimum at $(-0.75, 2.54)$.

We revert finally to the introduction of heterogeneity into the claim frequency model through the use of either the Poisson-gamma or Poisson-inverse Gaussian distributions as described in Section 3. Each of the choices involves an unknown parameter, denoted respectively by v or τ, in the corresponding variance function. One possible strategy for setting the value of the unknown parameter might be to optimise the extended quasi-likelihood but further work is needed in this respect.

6. SUMMARY

Discussion has focused on providing an overview of the variety of response variables available for modelling both the claim frequency and claim severity components of the claims process in general insurance in the presence of rating factors. Working within the rich class of GLMs it is necessary to specify only the first two moments of the associated response variables rather than the full likelihood

in order to effect implementation. By this means, it is indicated how suitably selected parameterised variance functions can be used to model heterogeniety in the claim frequency process and to provide a parameterised family of claim response variables which include the gamma response variable as a limiting case. Additional modelling flexibility is achieved through the introduction of the parameterised power link function which includes the log-link as a special case. The salient characteristics in the implementation of these features are illustrated.

REFERENCES

ANDRADE E SILVA, J.M. (1989) An application of Generalized Linear Models to Portuguese Motor Insurance. *Proceedings XXI ASTIN Colloquium, New York*, 633.

BAKER, R.J. and NELDER, J.A. (1985) The GLIM System, Release 3.77, Generalized Linear Interactive Modelling Manual. *Oxford: Numerical Algorithms Group.*

BAXTER, L.A., COUTTS, S.M. and ROSS, G.A.F. (1979) Applications of Linear Models in Motor Insurance. *21st International Congres of Actuaries* **2**, 11.

BEIRLANT, J., DERVEAUX, V., DE MEYER, A.M., GOOVAERTS, M.J., LABIE, E. and MAENHOUDT, B. (1991) Statistical Risk Evaluation Applied to (Belgian) Car Insurance. *Insurance: Mathematics and Economics* **10**, 289.

BERG, M.P. and HABERMAN, S. (1992) Trend Analysis and Prediction Procedure for Time Nonhomogeneous Claim Process. *Submitted.*

BESSON, J.-L. and PARTRAT, Ch. (1992) Trend et Systèmes de Bonus-Malus. *ASTIN Bulletin* **22**, 11.

BOSKOV, M. (1992) Private communications.

BROCKMAN, M.J. and WRIGHT, T.S. (1992) Statistical Motor Rating: Making Effective use of Your Data. *J.I.A.* **119**, 457.

CENTENO, L. and ANDRADE E SILVA, J.M. (1991) Generalized Linear Models under Constraints. *Proceedings XXIII ASTIN Colloquium, Stockholm.*

COUTTS, S.M. (1984) *Motor Insurance Rating, An Actuarial Approach. J.I.A.* **111**, 87.

EATION, M.L., MORRIS, C. and RUBIN, H. (1971) On Extreme Stable Laws and Some Applications. *J. Appl. Prob.* **8**, 794.

HABERMAN, S. and RENSHAW, A.E. (1989) Fitting Loss Distributions using Generalized Linear Models. *Proceedings XXI ASTIN Colloquium, New York*, 149.

HOGG, R.V. and KLUGMAN, S.A. (1984) Loss Distributions. *John Wiley & Sons.*

JORGENSEN, B. (1987) Exponential Dispersion Models (with discussion). *J. R. Statist. Soc. B* **49**, 127.

MACK, T. (1991) A Simple Parametric model for Rating Automobile Insurance or Estimating IBNR Claims Reserves. *ASTIN Bulletin* **22**, 93.

MCCULLAGH, P. and NELDER, J.A. (1983, 1989) Generalized Linear Models. *Chapman & Hall.*

RENSHAW, A.E. (1990) Graduation by Generalized Linear Modelling Techniques. *Proceedings XXII ASTIN Colloquium, Montreux.*

RENSHAW, A.E. (1991) Actuarial Graduation Practice and Generalized Linear & Non-Linear Models. *J.I.A.* **118**, 295.

RENSHAW, A.E. (1992) Joint Modelling for Actuarial Graduation and Duplicate Policies. *J.I.A.* **119**, 69.

RENSHAW, A.E. (1993) An Application of Exponential Dispersion Models to Premium Rating. *ASTIN Bulletin* **23**, 145.

RENSHAW, A.E. and HABERMAN, S. (1992) Graduations Associated with a Multiple State Model for Permanent Health Insurance. *Under review.*

STROINSKI, K.J. and CURRIE, I.D. (1989) Selection of Variables for Automobile Insurance Rating. *Insurance: Mathematics and Economics* **8**, 35.

TREMBLAY, L. (1992) Using the Poisson Inverse Gauassian in Bonus-Malus Systems. *ASTIN Bulletin* **22**, 97.

ARTHUR E. RENSHAW
Department of Actuarial Science & Statistics,
The City University, London.