FITTING TWEEDIE'S COMPOUND POISSON MODEL TO INSURANCE CLAIMS DATA: DISPERSION MODELLING

BY

GORDON K. SMYTH[1] AND BENT JØRGENSEN

ABSTRACT

We reconsider the problem of producing fair and accurate tariffs based on aggregated insurance data giving numbers of claims and total costs for the claims. Jørgensen and de Souza (*Scand. Actuarial J.*, 1994) assumed Poisson arrival of claims and gamma distributed costs for individual claims. Jørgensen and de Souza (1994) directly modelled the risk or expected cost of claims per insured unit, μ say. They observed that the dependence of the likelihood function on μ is as for a linear exponential family, so that modelling similar to that of generalized linear models is possible. In this paper we observe that, when modelling the cost of insurance claims, it is generally necessary to model the dispersion of the costs as well as their mean. In order to model the dispersion we use the framework of double generalized linear models. Modelling the dispersion increases the precision of the estimated tariffs. The use of double generalized linear models also allows us to handle the case where only the total cost of claims and not the number of claims has been recorded.

KEYWORDS

Car insurance, Claims data, Compound Poisson model, Exposure, Generalized linear models, Dispersion modelling, Double generalized linear models, Power variance function, REML, Risk theory, Tarification

1. INTRODUCTION

We reconsider the problem considered by Jørgensen and de Souza (1994), namely that of producing fair and accurate tariffs based on aggregated insurance data giving numbers of claims and total costs for the claims. Jørgensen and de Souza (1994) assumed Poisson arrival of claims and gamma distributed costs for individual claims. These assumptions imply that the total cost of

[1] Address for correspondence: Dr G.K. Smyth, Bioinformatics, Walter and Eliza Hall Institute of Medical Research, Post Office, Royal Melbourne Hospital, Parkville, VIC 3050, Australia.

claims in each category over a given time period follows a Tweedie compound Poisson distribution. Jørgensen and de Souza (1994) directly modelled their parameter of interest, namely the risk or expected cost of claims per insured unit, μ say. They observed that the dependence of the likelihood function on μ is as for a linear exponential family, so that modelling similar to that of generalized linear models is possible.

In this paper we observe that, when modelling the cost of insurance claims, it is generally necessary to model the dispersion of the costs as well as their mean. In order to model the dispersion we use the framework of double generalized linear models developed by Nelder and Pregibon (1987), Smyth (1989) and Smyth and Verbyla (1999). Modelling the dispersion increases the precision of the estimated tariffs. The use of double generalized linear models also allows us to handle the case where only the total cost of claims and not the number of claims has been recorded.

The method used by Jørgensen and de Souza (1994) implicitly assumes that explanatory variables affect the expected cost of claims μ by simultaneously increasing or decreasing both the frequency of claims and the average claim size. In practice however, some explanatory factors will have a greater impact on the frequency of claims than on their size, while other variables may impact more on the size of claims. It is also possible for certain factors, such as a no-claims bonus, to affect the frequency of claims and the claim size in opposite directions. This does not invalidate the method of Jørgensen and de Souza (1994), which continues to provide consistent estimators of the risk. It does mean though that insurance claims data are likely to display non-constant dispersion, so that it is necessary to model the dispersion as well as the mean in order to obtain efficient estimation of μ. We add that refinement to the method in this paper.

Double generalized linear models allow the simultaneous modelling of both the mean and the dispersion in generalized linear models. Estimation of the dispersion is affected by a second generalized linear model, the dispersion submodel, in which the responses are the unit deviances from the original model. The unit deviances are approximately $\varphi_i \chi_1^2$, where φ_i is the dispersion parameter, so that the dispersion submodel is a gamma generalized linear model with its own dispersion parameter, which is 2. When modelling insurance data with counts of claims as well as total costs, we use the same double generalized linear model framework, but modify the definition of the response and the weights in the dispersion submodel. When only the total claim costs are observed and not the claim counts, the definitions of the response and weights in the dispersion submodel revert to their customary values.

Excellent recent reviews of generalized linear models and their actuarial applications are given by Renshaw (1994), Haberman and Renshaw (1998), Millenhall (1999) and Murphy, Brockman and Lee (2000). Of these, Millenhall (1999) gives most information on the compound Poisson models used in this application. Mathematical details on the compound Poisson distributions themselves are given by Jørgensen (1997) and by Rolski, Schmidli, Schmidt and Teugels (1999). McCullagh and Nelder (1989) and Dobson (2001) give thorough general introductions to generalized linear models and the first of these books includes in Sections 8.4.1 and 12.8.3 the earliest example of non-normal generalized linear modelling of insurance claims.

In normal regression and multivariate modelling it is well known that there are advantages to using residual maximum likelihood (REML) for estimating the variances rather than maximum likelihood estimation. The idea of REML is to adjust the variance estimators to take account of the fact that the means were estimated and are therefore closer to the data than the true means can be expected to be. REML produces more nearly unbiased estimators for the variances, and can produce consistent estimators of the variances when the number of parameters affecting the mean grows with the sample size, a situation in which maximum likelihood estimation fails. Lee and Nelder (1998), Smyth and Verbyla (1999) and Smyth, Huele and Verbyla (2001) study in some detail the problem of approximate REML for double generalized linear models, where the interest is to modify estimation of the dispersion submodel for estimation of the means. In this paper we extend the REML method of Smyth, Huele and Verbyla (2001) to the insurance claims context.

In the next section we review the Tweedie compound Poisson model. Section 3 reviews double generalized linear models and describes the case when the claim counts are not observed. Section 4 considers the joint likelihood using the counts and the costs. In Section 5 we estimate tariffs from the Swedish third party automobile portfolio of 1977.

2. The Compound-Poisson Model

Let N_i be the number of claims observed in the ith classification category and Z_i be the total claim size for that category. Suppose that the number of units at risk (typically measured in policy years) is w_i, and write $Y_i = Z_i/w_i$ for the observed claim per unit at risk. We suppose that N_i is Poisson distributed with mean $\lambda_i w_i$, and that the size of each claim is gamma distributed with mean τ_i and shape parameter α. It follows that N_i and Y_i are zero with probability $e^{-\lambda_i w_i}$ and that Y_i is otherwise continuous and positive. Individual claims are assumed to arrive independently so that the conditional distribution of Y_i given N_i is also gamma distributed with mean $N_i\tau_i / w_i$ whenever N_i is positive. We suppose that independent observations (n_i, y_i) are available for categories $i = 1, \ldots, m$.

Jørgensen and de Souza (1994) observed that the parameter of interest from the point of view of setting tariffs is $\mu_i = E(Y_i) = \lambda_i\tau_i$. From Jørgensen (1987, 1997) it is known that the distribution of Y_i forms a linear exponential family as μ_i varies, and that $\text{var}(Y_i) = \varphi_i \mu_i^p / w_i$ where $p = (\alpha + 2) / (\alpha + 1)$ and φ_i is the so-called dispersion parameter. The positivity of α implies that $1 < p < 2$. The joint density of N_i and Y_i can usefully be parametrized in terms of μ_i, φ_i and p, which describe the mean and variance of the claim per unit risk. The variance parameters φ_i and p are statistically orthogonal to μ_i, meaning that the off-diagonal elements of the Fisher information matrix are zero. This parametrization has the advantage, over the alternative parametrization in terms of λ_i, τ_i and α, that it focuses attention of the parameter of interest and two other parameters which are orthogonal to it.

The variance of Y_i can be obtained directly as $E_{N_i}\operatorname{var}(Y_i \mid N_i) + \operatorname{var}_{N_i} E(Y_i \mid N_i) = \lambda_i \tau_i^2/(\alpha w_i) + \lambda_i \tau_i^2/w_i = (1/\alpha + 1)\lambda_i \tau_i^2/w_i$. Equating this to the alternative expression $\varphi_i \mu_i^p/w_i$ for the variance gives the dispersion parameter in terms of λ and τ as

$$\varphi_i = w_i \operatorname{var}(Y_i) / \mu_i^p = \lambda_i^{1-p}/(2-p)$$

The exponent $1-p$ for λ_i here is negative, so it can be seen that any factor which increases the frequency of claims λ_i without affecting their average size will decrease the dispersion φ_i while increasing the mean μ_i. On the other hand the exponent $2-p$ for τ_i is positive, so any factor which increases the average claim size τ_i without increasing their frequency will increase both the mean and the dispersion. Any factor which affects the mean but not the dispersion must affect λ_i and τ_i in such a way that $\lambda_i^{1-p}\tau_i^{2-p}$ remains constant.

We therefore assume a model which allows both μ_i and φ_i to very depending on the values of covariates. As in generalized linear models, we assume a link-linear model for the mean cost

$$g(\mu_i) = \mathrm{x}_i^T \beta. \tag{1}$$

Here g is a known monotonic link function, x_i is a vector of covariates, and β is a vector of regression coefficients. As in double generalized linear models, we simultaneously assume another link-linear model

$$g_d(\varphi_i) = \mathrm{z}_i^T \gamma \tag{2}$$

for the dispersion, where z_i is a vector of covariates thought to affect the dispersion and γ is another vector of regression parameters to be estimated.

In many cases it will be convenient to take both g and g_d to be logarithmic, in which case (1) and (2) imply log-linear models also for the expected claim frequency λ_i and for the expected claim size τ_i. The model we describe is then equivalent to separate log-linear modelling of the claim frequency and the claim size, with the added-value that complete information is used for all inferences and the results are automatically collated for the cost per unit risk which is of direct interest.

In all of the following we assume that α, and hence also p, does not vary between cases.

3. Claim Cost Only is Observed

3.1. Maximum Likelihood

Consider now the case in which only the total cost of claims in each category and not the actual number of claims has been recorded, i.e., we observe w_i and $Y_i = y_i$, $i = 1, \ldots, m$, but not N_i. The amount of information available is rather lower than when N_i is observed as well but, as Jørgensen and de Souza

(1994) observed, the information in the frequencies is directed mainly at the φ_i and p, and is therefore of second order regarding the estimation of μ_i and the tariffs. In this case we have a double generalized linear model (Smyth, 1989; Smyth and Verbyla, 1999) in which the response, Y, follows a Tweedie compound Poisson distribution. Approximate maximum likelihood estimates of the mean coefficients β and the dispersion coefficients γ can be obtained by alternating between two generalized linear models. With γ and p fixed, β can be estimated from a generalized linear model with response y_i, mean μ_i, variance function $V(\mu_i) = \mu_i^p$, link function g, linear predictor $\mathrm{x}_i^T \beta$, weights w_i / φ_i and dispersion parameter 1. Let d_i be the unit deviances from this generalized linear model. With β and p fixed, γ can be estimated from a generalized linear model with the d_i as responses.

The saddlepoint approximation ensures that the d_i are approximately distributed as $\varphi_i \chi_1^2$ for φ_i reasonably small (Nelder and Pregibon, 1987; Jørgensen, 1997; Smyth and Verbyla, 1999). The d_i therefore follow approximately a gamma generalized linear model, with mean φ_i, variance function $V_d(\varphi_i) = \varphi_i^2$, link function g_d, linear predictor $\mathrm{z}_i^T \gamma$ and dispersion parameter 2.

The Fisher scoring equations for β and γ are as follows. The Fisher scoring update equation for β is

$$\beta^{k+1} = (X^T W X)^{-1} X^T W \mathrm{z} \tag{3}$$

where W is the diagonal matrix of working weights

$$W = \mathrm{diag}\left\{\left[\frac{\partial g(\mu_i)}{\partial \mu}\right]^{-2} \frac{w_i}{\varphi_i V(\mu_i)}\right\}$$

with variance function $V(\mu) = \mu^p$, z is the working vector with components

$$z_i = \frac{\partial g(\mu_i)}{\partial \mu}(y_i - \mu_i) + g(\mu_i)$$

and all terms on the right-hand-side of (3) are evaluated at the previous iterate β^k (McCullagh and Nelder, 1989, Section 2.5). Standard errors for $\hat{\beta}$ are obtained from the inverse of the Fisher information matrix

$$\Im_\beta = X^T W X.$$

The unit deviances for the generalized linear model can be defined as

$$d_i = 2\varphi_i \{\log f_Y(y_i; y_i, \varphi_i / w_i, p) - \log f_Y(y_i; \mu_i, \varphi_i / w_i, p)\},$$

where $f_Y(y; \mu, \varphi, p)$ is the marginal density function of the Y_i, which in our case gives

$$d_i = 2w_i\left(y_i \frac{y_i^{1-p} - \mu_i^{1-p}}{1-p} - \frac{y_i^{2-p} - \mu_i^{2-p}}{2-p}\right).$$

Note that d_i does not depend on φ_i. The approximate Fisher scoring iteration for γ is

$$\gamma^{k+1} = (Z^T W_d Z)^{-1} Z^T W_d \mathrm{z}_d \tag{4}$$

where W_d is the diagonal matrix of working weights

$$W_d = \mathrm{diag}\left\{\left[\frac{\partial g_d(\varphi_i)}{\partial \varphi}\right]^{-2} \frac{1}{2V_d(\varphi_i)}\right\}$$

with variance function $V_d(\varphi) = \varphi^2$, z_d is the working vector with components

$$z_{di} = \frac{\partial g_d(\varphi_i)}{\partial \varphi}(d_i - \varphi_i) + g_d(\varphi_i)$$

and all terms on the right-hand-side of (4) are evaluated at the previous iterate γ^k (Smyth, 1989). Standard errors for $\hat{\gamma}$ are obtained from the inverse of the Fisher information matrix

$$\Im_\gamma = Z^T W_d Z.$$

Since β and γ are orthogonal, alternating between (3) and (4) results in an efficient algorithm with typically rapid convergence (Smyth, 1996). The iteration can be initiated at $\mu_i = y_i$ and $\varphi_i = 1$. Score tests and estimated standard errors from each generalized linear model are correct for the combined model (Smyth, 1989). Finally, estimation of p can be obtained by maximizing the saddlepoint profile likelihood for p (Nelder and Pregibon, 1987; Smyth and Verbyla, 1999). We have not adjusted the standard errors for γ for estimation of p, although this could be done as in Jørgensen and de Souza (1994). The standard errors for β, which are of most interest, do not require such adjustment as β is orthogonal to p.

The accuracy of the saddlepoint approximation for the density $f_Y(y; \mu, \varphi, p)$ has been discussed by Smyth and Verbyla (1999) and by Dunn (2001). In the context of the likelihood calculations in this Section, the saddlepoint approximation is most accurate when the number of claims per risk category is large or when the estimated variability $\hat{\varphi}_i / w_i$ is small. In particular, the approximation is likely to be satisfactory when the proportion of categories with zero claims is small. When there are many categories with zero claims, the φ_i will tend to be overestimated. However this will have only a secondary effect on the estimated values for μ_i and corresponding risk factors.

Use of the saddle-point approximation for estimation of γ and p is essentially equivalent to the extended quasi-likelihood (EQL) of Nelder and Pregibon (1987) and Nelder and Lee (1992). The EQL approach emphasises the fact that the estimators depend only on second moment assumptions about the distribution of the Y_i. The properties of the estimators therefore are not highly dependent in the compound Poisson distribution assumptions about the Y_i, as long as the mean and dispersion are correctly specified.

3.2. Approximate REML

It is well known in linear regression that the maximum likelihood variance estimators are biased downwards when the number of parameters used to estimate the fitted values is large compared with the sample size. The same principle applies to double generalized linear models. The maximum likelihood *estimators $\hat{\varphi}_i$ are biased downwards and the estimated variances $\hat{\varphi}_i\hat{\mu}_i^{\hat{p}}/w_i$ are* too small by an average factor of about k/m where k is the dimension of β and m is the sample size. In normal linear models, restricted or residual maximum likelihood (REML) is usually used to estimate the variances, and this produces estimators which are approximately and sometimes exactly unbiased.

Let the h_i be the diagonal elements of the hat matrix

$$W^{1/2}X(X^TWX)^{-1}X^TW^{1/2},$$

often called the leverages for the generalized linear model for the y_i. Approximately unbiased estimators of the φ_i may be obtained by modifying the scoring update for γ as follows. The leverage adjusted scoring update is

$$\gamma^{k+1} = (Z^TW_d^*Z)^{-1}Z^TW_d^*z_d^* \tag{5}$$

where W_d^* is the diagonal matrix

$$W_d^* = \text{diag}\left\{\left[\frac{\partial g_d(\varphi_i)}{\partial\varphi}\right]^{-2}\frac{1-h_i}{2V_d(\varphi_i)}\right\}$$

and

$$z_{di}^* = \frac{\partial g_d(\varphi_i)}{\partial\varphi}\left(\frac{d_i}{1-h_i}-\varphi_i\right)+g_d(\varphi_i).$$

See Lee and Nelder (1998) and Smyth, Huele and Verbyla (2001) for a discussion of this leverage adjustment. The appearance of the factor $1-h_i$ in the information in a reflection of the fact that an observation with leverage $h_i = 1$ provides no information about φ_i. The scoring iteration (5) approximately maximizes with respect to γ the penalized profile log-likelihood

$$\ell_p^*(y;\gamma,p) = \ell(y;\hat{\beta}_\gamma,\gamma,p)+\frac{1}{2}\log\left|X^TWX\right| \tag{6}$$

where $\ell(y;\beta,\gamma,p)$ is the ordinary log-likelihood function, $\hat{\beta}_\gamma$ is the maximum likelihood estimator of β for given values of γ and p, and W is evaluated at $\beta = \hat{\beta}_\gamma$. This penalized log-likelihood reduces to the REML likelihood in the normal linear case and can be more generally justified as an approximate conditional log-likelihood (Cox and Reid, 1987). Approximately unbiased estimation of p can be obtained by maximizing (6) with respect to both γ and p.

4. Claim Cost and Frequency are Both Observed

4.1. The Joint Likelihood Function

Consider N_i and Y_i for a particular classification category, and for ease of notation drop the subscript i for most of the remainder of this section. The joint probability density function of N and Y is given by Jørgensen and de Souza (1994, equation 11). It can be written as

$$f(n,y;\mu,\varphi/w,p) = a(n,y,\varphi/w,p)\exp\left\{\frac{w}{\varphi}t(y,\mu,p)\right\}$$

with

$$a(n,y;\mu,\varphi/w,p) = \left\{\frac{(w/\varphi)^{\alpha+1}y^{\alpha}}{(p-1)^{\alpha}(2-p)}\right\}^{n}\frac{1}{n!\Gamma(n\alpha)y}$$

and

$$t(y,\mu,p) = y\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}.$$

The log-likelihood function for the unknown parameters β, γ and p is

$$\ell(\mathrm{n},\mathrm{y};\beta,\gamma,p) = \sum_{i=1}^{m}\log f(n_i,y_i;\mu_i,\varphi_i/w_i,p).$$

It can be seen that y is sufficient for μ, and that the density follows a linear exponential family as μ varies. We have

$$\frac{\partial\log f(n,y;\mu,\varphi/w,p)}{\partial\mu} = \frac{w}{\varphi}\frac{\partial t(y,\mu,p)}{\partial\mu} = \frac{w}{\varphi}\frac{y-\mu}{\mu^{p}}. \tag{7}$$

At $n = y = 0$ the distribution has probability mass given by

$$\log f(0,0;\mu,\varphi/w,p) = -w\lambda = -\frac{w}{\varphi}\frac{\mu^{2-p}}{2-p} = \frac{w}{\varphi}t(0,\mu,p)$$

so (7) holds over the whole range of the distribution. It follows, by differentiating (7) again with respect to φ or p, that the cross derivatives with respect to μ and either φ or p have expectation zero. In other words, μ is orthogonal to both φ and p.

Now consider the estimation of φ. Although the joint density is not a linear exponential family, we can fit the likelihood equations into a generalized linear model structure by creating suitable pseudo working responses and working weights. This will allow us to make use of the double generalized linear model framework in computations and in data analysis. We have

$$\frac{\partial\log f(n,y;\mu,\varphi/w,p)}{\partial\varphi} = -\frac{n}{(p-1)\varphi} - \frac{w}{\varphi^{2}}t(y,\mu,p)$$

and

$$\frac{\partial^2 \log f(n,y;\mu,\varphi/w,p)}{\partial\varphi^2} = -\frac{n}{(p-1)\varphi^2} + \frac{2w}{\varphi^3}t(y,\mu,p).$$

Now

$$E\{t(Y,\mu,p)\} = \frac{\mu^{2-p}}{(1-p)(2-p)}$$

and

$$E(N) = \frac{w}{\varphi}\frac{\mu^{2-p}}{2-p}$$

so the Fisher information for φ from a single (n_i, y_i) pair is

$$E\left(-\frac{\partial^2 \log f}{\partial\varphi^2}\right) = \frac{w\mu^{2-p}}{(2-p)(p-1)\varphi^3}.$$

Define dispersion-prior weights to be

$$w_d = \frac{2w\mu^{2-p}}{(2-p)(p-1)\varphi}.$$

Then

$$E\left(-\frac{\partial^2 \log f}{\partial\varphi^2}\right) = \frac{w_d}{2V_d(\varphi)}$$

with $V_d(\varphi) = \varphi^2$. The choice of $2\varphi^2$ in the denominator is in order to match the dispersion model in Section 2. In insurance applications we will almost always have $w_d > 1$, in which case we interpret $(w_d - 1)/\{2V_d(\varphi)\}$ as the extra information about φ arising from observation of the number of claims n_i. If $w_d < 1$, then the saddlepoint approximation which underlies the computations in Section 3.1 is poor, and the true information about φ_i arising from y_i is less than that indicated in Section 3.1. Define dispersion-responses to be

$$d = \frac{2\varphi^2}{w_d}\frac{\partial \log f}{\partial\varphi} + \varphi = -\frac{2}{w_d}\left(\frac{n\varphi}{p-1} + wt\right) + \varphi.$$

We can now write the first derivative of the log-density in the form

$$\frac{\partial \log f(n,y;\mu,\varphi/w,p)}{\partial\varphi} = \frac{w_d(d-\varphi)}{2V_d(\varphi)}.$$

The above definitions for w_d and d are somewhat artificial, but have the effect of putting the likelihood calculations into the form of a double generalized linear model. The components of the likelihood score vector $\partial\ell/\partial\varphi$ can now be written as

$$\frac{\partial\ell}{\partial\varphi_i} = \sum_{i=1}^{m}\frac{w_{di}(d_i-\varphi_i)}{2V_d(\varphi_i)}$$

https://doi.org/10.2143/AST.32.1.1020 Published online by Cambridge University Press

and the Fisher information matrix for the φ_i is

$$\Im_{\varphi} = \operatorname{diag}\left\{\frac{w_{di}}{2V_d(\varphi_i)}\right\}.$$

4.2. Maximum Likelihood

Since μ is orthogonal to φ and p, it follows that β is orthogonal to γ and p. It is sensible therefore to consider estimation of the parameters separately. For γ and p fixed, the y_i are sufficient for β and estimation of β can proceed exactly as in Section 2. The estimating equations and information matrix for β are exactly as when the n_i are not observed.

Now consider the estimation of γ for fixed β and p. Since

$$\frac{\partial\varphi}{\partial\gamma} = \operatorname{diag}\left\{\left[\frac{\partial g_d(\varphi_i)}{\partial\varphi}\right]^{-1}\right\} Z$$

where Z is the design matrix with rows z_i^T, the information matrix for γ is

$$\Im_{\gamma} = \frac{\partial\varphi}{\partial\gamma}^T \Im_{\Phi} \frac{\partial\varphi}{\partial\gamma} = Z^T W_d Z$$

where

$$W_d = \operatorname{diag}\left\{\left[\frac{\partial g_d(\varphi_i)}{\partial\varphi}\right]^{-2} \frac{w_{di}}{2V_d(\varphi_i)}\right\}.$$

This is the same weight matrix that we would obtain from a generalized linear model with link function g_d, variance function $V_d(\varphi_i) = \varphi_i^2$ and prior weights w_{di}. The score vector for γ is

$$\frac{\partial\ell}{\partial\gamma} = \frac{\partial\varphi}{\partial\gamma}^T \frac{\partial\ell}{\partial\varphi} = Z^T W_d \mathrm{r}_d$$

with $r_{di} = \{\partial g_d(\varphi) / \partial\varphi\}(d_i - \varphi_i)$. The scoring iteration for γ can be written in the standard generalized linear model form

$$\gamma^{k+1} = (Z^T W_d Z)^{-1} Z^T W_d \mathrm{z}_d$$

where $z_{di} = \{\partial g_d(\varphi_i) / \partial\varphi\}(d_i - \varphi_i) + g_d(\varphi_i)$ is the dispersion-working vector. In this equation, all terms on the right hand size are evaluated at the current working estimate γ^k and γ^{k+1} is the updated estimate. Note that w_{di} and d_i are as defined in Section 4.1.

4.3. Approximate REML

When the n_i are observed there is more information available for the estimation of γ and p, and correspondingly less need to adjust the estimation of γ for estimation of β. The adjustment may still be useful however and is relatively straightforward. Since the information matrix for β is unchanged by observation of n_i, we may use the same adjustment to the profile likelihood as in Section 3.2. Therefore we need to adjust the score vector for γ by the same quantity as in Section 3.2. We adjust the working weight matrix to

$$W_d^* = \text{diag}\left\{\left[\frac{\partial g_d(\varphi_i)}{\partial \varphi}\right]^{-2} \frac{|w_{di} - h_i|_+}{2V_d(\varphi_i)}\right\}$$

where $|w_{di} - h_i|_+$ is the maximum of $w_{di} - h_i$ and zero, and replace d_i with

$$d_i^* = \frac{w_{di}}{w_{di} - h_i} d_i.$$

The adjusted scoring iteration for g is then

$$\gamma^{k+1} = (Z^T W_d^* Z)^{-1} Z^T W_d^* z_d^*$$

with $z_{di}^* = \{\partial g_d(\varphi_i) / \partial\varphi\}(d_i^* - \varphi_i) + g_d(\varphi_i)$.

5. Swedish Third Party Motor Insurance

We consider the Third Party Motor Insurance data for Sweden for 1977 described by Andrews and Herzberg (1985) and previously analysed by Hallin and Ingenbleek (1983). The data can be obtained from the URL www.statsci.org/data/general/motorins.html. We consider only the data for Zone 1, which consists of the three largest cities, Stockholm, Göteborg and Malmö with surroundings, and exclude Make class 9 which is a miscellaneous category of all makes other than the first eight. This leaves 5406 claims over the period in 280 categories. Of the 280 categories, 20 had no claims in 1977. The explanatory factors are the Make of the car (8 classes), the number of kilometres travelled per year (in 5 ordered categories) and the no claims bonus class. Bonus represents the number of years since last claim, from 1 up to 7.

Exploratory analyses of claim frequency and claim size show that Bonus affects frequency and size in different directions, while the other two factors affect claim frequency more than claim size. We therefore expect to find strong factor effects on the dispersion as well as on the mean. We fit log-linear models (with g and g_d both equal to the logarithmic function) to both the mean and the dispersion. When main effects only are fitted all three factors for both the mean and the dispersion, the maximum likelihood estimator of p is found to be 1.725. We find that all three factors have highly significant main effects on both the mean and the dispersion (Table 1). The dispersion effects are rather more significant than those for the mean, emphasizing the importance of including the dispersion model.

There is also definite evidence of interactions in the Swedish claims data. The likelihood ratio statistics to add interactions are given in Table 2. Although

TABLE 1

JOINT MODELLING OF FREQUENCY AND SIZE OF CLAIMS:
DIFFERENCES IN TWICE THE LOG-LIKELIHOOD FOR REMOVING FACTORS.

Factor	*df*	*Deviance to Remove Factor*	
		from Mean	*from Dispersion*
Bonus	6	363.9	1190.9
Make	7	78.1	179.8
Kilometres	4	24.2	45.1

TABLE 2

JOINT MODELLING OF FREQUENCY AND SIZE OF CLAIMS:
DIFFERENCES IN TWICE THE LOG-LIKELIHOOD FOR ADDING INTERACTIONS.

Factor	*df*	*Deviance to Add Interaction*	
		to Mean	*to Dispersion*
Bonus: Make	42	63.3	78.3
Bonus: Kilometres	24	37.1	36.0
Make: Kilometres	28	35.6	52.1

the interactions are statistically significant, they are far less so than the main effects. Since the interactions produce a model which is too complex for practical use in setting insurance tariffs, we will treat the main effects model as the final model. Some exploration of the data failed to find any way to explain the interactions with a small number of degrees of freedom.

The effects for Bonus and Kilometres are monotonic, as would be expected from their meaning, except that Bonus level 6 and Kilometres level 3 are out of sequence in the mean model. To achieve monotonic effects for the Bonus and Kilometres, Bonus levels 5 and 6 and Kilometres levels 2 and 3 were combined. This increases minus twice the likelihood by only 1.0 on 4 df. The resulting model for the mean is given in the first column of Table 3. The base risk is estimated to be 694.5 Swedish kroner per car-year. This corresponds to drivers without a no claim bonus, driving Make 1, who drive fewer than 1000 km per year. For the other categories the base risk should be multiplied by the factors given in the table. Increasing the no-claims bonus decreases the mean cost per unit risk but increases the dispersion. Increasing kilometres travelled increases the mean cost but decreases the dispersion. Make 8 is the most expensive while Make 4 (the Volkswagen bug) is the cheapest. Make 4 also has the smallest dispersion.

To investigate the stability of the results, the mean-dispersion main-effects model was fitted using four methods: maximum likelihood and approximate REML using the full data and approximate maximum likelihood and REML

TABLE 3

BASE RATE RISK AND MULTIPLIERS FOR VARIOUS RISK CATEGORIES.
THE RISKS ARE ESTIMATED USING FIVE DIFFERENT METHODS. DISTANCE IS IN 1000S OF KILOMETRES PER YEAR. THE BASE RISK SHOULD BE MULTIPLIED BY THE FACTORS GIVEN FOR EACH RISK CATEGORY.

Category	*Cost & Frequency*		*Cost Only*		*Constant Dispersion*
	ML	*REML*	*ML*	*REML*	
Base rate	694.527	694.536	697.346	692.904	691.105
Bonus 1	1.000	1.000	1.000	1.000	1.000
Bonus 2	0.734	0.734	0.725	0.725	0.730
Bonus 3	0.685	0.685	0.676	0.676	0.686
Bonus 4	0.500	0.500	0.516	0.517	0.518
Bonus 5-6	0.418	0.418	0.401	0.406	0.430
Bonus 7	0.268	0.268	0.267	0.267	0.272
Make 1	1.000	1.000	1.000	1.000	1.000
Make 2	1.260	1.260	1.284	1.294	1.181
Make 3	0.960	0.960	1.015	1.031	0.902
Make 4	0.536	0.536	0.540	0.543	0.547
Make 5	1.005	1.005	1.035	1.044	1.034
Make 6	0.685	0.685	0.692	0.690	0.712
Make 7	0.768	0.768	0.808	0.819	0.765
Make 8	1.557	1.557	1.661	1.701	1.391
Kilometres ≤ 1	1.000	1.000	1.000	1.000	1.000
Kilometres 1-20	1.282	1.282	1.274	1.272	1.268
Kilometres 20-25	1.399	1.400	1.367	1.356	1.401
Kilometres ≥ 25	1.663	1.663	1.598	1.603	1.734

using the total cost of claims only. The value of the variance power p was estimated by the four methods to be 1.725, 1.735, 1.775 and 1.775 respectively. In addition, the main-effects model for the mean was also fitted with a constant dispersion model using approximate REML on the full data. This is the fifth estimation method and produces the last column of Table 3. It can be seen from the table that there is very little difference between the maximum likelihood and REML methods using the full data. The difference between the full data results and those using the total claim costs only is more noticeable but still not large. The difference between modelling the dispersion and assuming constant dispersion is of a similar magnitude to that between using the full data and using the claim costs only.

The method has been implemented as an S-Plus function tariff available from the URL www.statsci.org/s/tariff.html. In this function the number of claims in an optional argument. When it is not given, the method defaults to the double generalized linear model method outlined in Section 3. Software to fit double generalized linear models and ordinary generalized linear models with power variance functions was previously described by Smyth and Verbyla (1999).

6. Concluding Comments

The approach based on Tweedie's compound Poisson distribution provides a highly efficient method of analysing insurance claims data. The distributional assumptions can be assessed using standard data analysis techniques and in any case the relationship with extended quasi-likelihood suggests that the method will not be very sensitive to moderate deviations from the assumed Poisson and gamma distributions for the counts and claim sizes.

One side-effect of the efficiency is that more terms are likely to be found to be significant in the fitted model compared with approximate methods or methods based on the univariate likelihoods. In particular, it may be that significant interactions will be found which are too complicated for practical insurance applications. In most cases the main effects will be dominant, so that the interactions might be neglected as of lesser importance, as for the Swedish motor insurance data.

References

Andrews, D.F., and Herzberg, A.M. (1985) *Data. A collection of problems from many fields for the student and research worker*. Springer-Verlag, New York, pp. 413-421.

Cox, D.R. and Reid, N. (1987) Parameter orthogonality and approximate conditional inference. *J.R. Statist. Soc* B **49**, 1-39.

Dobson, A.J. (2001) *An Introduction to Generalized Linear Models*, Second Edition. Chapman and Hall/CRC, London.

Dunn, P.K. (2001) *Likelihood-Based Inference for Tweedie Generalized Linear Models*. PhD Thesis, Department of Mathematics, University of Queensland, Brisbane, Australia.

Haberman, S., and Renshaw, A.E. (1998) Actuarial applications of generalized linear models. In Statistics in Finance, D.J. Hand and S.D. Jacka (eds), Arnold, London.

Hallin, M., and Ingenbleek, J.-F. (1983) The Swedish automobile portfolio in 1977. A statistical study. *Scandinavian Actuarial Journal*, 49-64.

Jørgensen, B. (1987) Exponential dispersion models. *J. R. Statist. Soc.* B **49**, 127-162.

Jørgensen, B. (1997) *Theory of Dispersion Models*. Chapman & Hall, London.

Jørgensen, B. and de Souza, M.C.P. (1994) Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journal*, 69-93.

Lee, Y. and Nelder, J.A. (1998) Generalized linear models for the analysis of quality-improvement experiments. *Canadian Journal of Statistics* **26**, 95-105.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd Edition. Chapman & Hall, London.

Millenhall, S.J. (1999) A systematic relationship between minimum bias and generalized linear models. 1999 Proceedings of the Casualty Actuarial Society **86**, 393-487.

Murphy, K.P., Brockman, M.J. and Lee, P.K.W. (2000) Using generalized linear models to build dynamic pricing systems. Casualty Actuarial Forum, Winter 2000.

Nelder, J.A., and Lee, Y. (1992) Likelihood, quasi-likelihood and pseudo-likelihood: some comparisons. *J. R. Statist. Soc.* B **54**, 273-284.

Nelder, J.A. and Pregibon, D. (1987) An extended quasi-likelihood function. *Biometrik* **74**, 221-231.

Renshaw, A.E. (1994) Modelling the claims process in the presence of covariates. *ASTIN Bulletin* **24**, 265-286.

Rolski, T., Schmidli, H., Schmidt, V., and Teugels, J. (1999) Stochastic Processes for Insurance and Finance. John Wiley & Sons, Chichester.

Smyth, G.K. (1989) Generalized linear models with varying dispersion. *J. Roy. Statist. Soc.* B **51**, 47-60.

SMYTH, G.K. (1996) Partitioned algorithms for maximum likelihood and other nonlinear estimation. *Statistics and Computing* **6**, 201-216.
SMYTH, G.K., and VERBYLA, A.P. (1999) Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environments* **10**, 696-709.
SMYTH, G.K., HUELE, F., and VERBYLA, A.P. (2001) Exact and approximate REML for heteroscedastic regression. *Statistical Modelling* **1**, 161-175.

GORDON K. SMYTH
Walter and Eliza Hall Institute of Medical Research
Melbourne, Australia

BENT JØRGENSEN
Department of Statistics and Demography
University of Southern Denmark