CREDIBILITY THEORY AND GENERALIZED LINEAR MODELS

J A NELDER [1] and R J VERRALL [2]

ABSTRACT

This paper shows how credibility theory can be encompassed within the theory of Hierarchical Generalized Linear Models. It is shown that credibility estimates are obtained by including random effects in the model. The framework of Hierarchical Generalized Linear Models allows a more extensive range of models to be used than straightforward credibility theory. The model fitting and testing procedures can be carried out using a standard statistical package Thus, the paper contributes a further range of models which may be useful in a wide range of actuarial applications, including premium rating and claims reserving

KEYWORDS

Credibility Theory, Hierarchical Generalized Linear Models; Generalized Linear Models; Premium Rating Random-Effect Models

1 INTRODUCTION

Credibility theory began with the papers by Mowbray (1914) and Whitney (1918). In those papers, the emphasis was on deriving a premium which was a balance between the experience of an individual risk and of a class of risks Buhlmann (1967) showed how a credibility formula can be derived in a distribution-free way, using a least-squares criterion. Since then, a number of papers have shown how this approach can be extended· see particularly Buhlmann and Straub (1970), Hachemeister (1975), de Vylder (1976, 1986). The survey by Goovaerts and Hoogstad (1987) provides an excellent introduction to these papers.

[1] Department of Mathematics
Imperial College
Huxley Building
180 Queen's Gate
LONDON
SW7 2BZ

[2] Department of Actuarial Science and Statistics
City University
Northampton Square
LONDON
EC1V 0HB

ASTIN BULLETIN, Vol 27, No 1, 1997 pp 71-82

The underlying assumption of credibility theory which sets it apart from formulae based on the individual risk alone is that the risk parameter is regarded as a random variable This naturally leads to a Bayesian model, and there have been a large number of papers which adopt the Bayesian approach to credibility theory: for example Jewell (1974, 1975), Klugman (1987), Zehnwirth (1977) Klugman (1992) gives an introduction to the use of Bayesian methods, covering particularly aspects of credibility theory. A recent review of Bayesian methods in actuarial science and credibility theory is given by Makov *et al* (1996)

It can be shown that, under suitable assumptions, a credibility formula can be derived as the best linear approximation to the Bayesian estimate, using a quadratic loss function Jewell (1974) showed that for an exponential family of distributions, the credibility formula is the same as the exact formula, so long as the conjugate prior distribution and a natural parametrisation is used This result will be derived in a different way in section 3, in order to place the basic model of credibility within a wider framework. The choice of structure for the collective and the parameterisation will be discussed in more detail. Since exponential families form the basis of Generalized Linear Models (GLMs) - see McCullagh and Nelder (1989) - it is natural to seek an extension of credibility theory encompassing the full range of models which can be formulated as GLMs. This is particularly apposite as GLMs have many very natural applications in the actuarial field· see, for example Renshaw (1991), Renshaw and Verrall (1994) This will also make possible more applications of credibility theory.

The main purpose of this paper is to show how credibility theory can be incorporated into the general framework of GLMs and implemented in the statistical package Genstat Although the formulation of the credibility model is similar in many ways to the Bayesian approach, our approach is likelihood-based rather than Bayesian. The dispersion parameters will be estimated directly from the data without specifying prior distributions No prior estimates for the parameters need to be supplied. All assumptions used in the model can be checked using, for example, appropriate residual analyses Recent advances in the statistical literature on GLMs allow unobserved random effects to be estimated along with the parameter vector in the linear predictor A useful recent paper is Breslow and Clayton (1993) which covers the theory of generalized linear mixed models (GLMMs) GLMMs allow the inclusion of normally distributed random effects and have been applied to a wide variety of statistical problems We use the theory of Lee and Nelder (1996), which develops hierarchical generalized linear models (HGLMs). HGLMs also allow the inclusion of random effects, but these are not restricted to be normally distributed Pure random-effect models, in which no fixed effects are included in the linear predictor, are known in the actuarial literature as credibility models. They form one part of a much wider class of models which have many potential applications to actuarial data

Thus, the purpose of this paper is further to unify the actuarial theory; to show how modern statistical methods can be used to enable credibility theory to be applied in a standard statistical package, to allow extensions of basic credibility theory and to show how the assumptions made can be checked This last point is important, since we

regard many aspects of actuarial work as exercises is statistical modeling, rather than a dogmatic application of risk theory models

It should be noted that the theory can be applied to models that specify only the mean and variance functions, using quasi-likelihood (Wedderburn, 1974, Nelder and Pregibon, 1987) - see section 5

The paper is set out as follows. Section 2 contains a brief introduction to GLMs and derives some results which will be used elsewhere. Section 3 shows how credibility theory can be treated within the context of HGLMs Section 4 outlines more general HGLMs, and how they are likely to be used for actuarial data Section 5 outlines some extensions to the models in sections 3 and 4

2 INTRODUCTION TO GLMS

This section contains a brief introduction to GLMs, and derives some of the key results which will be used later in the paper. A complete treatment of the theory and application of GLMs can be found in McCullagh and Nelder (1989).

The basis of GLMs is the assumption that the data are sampled from a one-parameter exponential family of distributions We first describe these and some of their fundamental properties

Consider a single observation y A one-parameter exponential family of distributions has a log-likelihood of the form

$$\frac{y\theta - b(\theta)}{\varphi} + c(y, \varphi) \tag{2 1}$$

where θ is the canonical parameter
and φ is the dispersion parameter, assumed known

Haberman and Renshaw (1996) review the application of Generalized Linear Models in actuarial science, and include a section on loss distributions. In actuarial applications, many distributions belonging to one-parameter exponential families are useful However, Haberman and Renshaw (1996) show how it is also possible to fit certain heavy-tailed distributions using Generalized Linear Models

Some examples of such families are given below It is straightforward to show that

$$\mu = E(Y) = \frac{db(\theta)}{d\theta} \tag{2 2}$$

$$\text{and } Var(Y) = \frac{d^2 b(\theta)}{d\theta^2}\varphi \tag{2.3}$$

Note that $Var(Y)$ is the product of two quantities $\frac{d^2 b(\theta)}{d\theta^2}$ is called the variance function and depends on the canonical parameter (and hence on the mean) We can write this as $V(\mu)$, since equation (2 2) shows that θ is a function of μ

$$\text{Thus } V(\mu) = \frac{d^2 b(\theta)}{d\theta^2} \tag{2.4}$$

In actuarial applications, it is possible to include deterministic volume measures in the definition of *Var(Y)*. A GLM may be defined by specifying a distribution, as above, together with a link function and a linear predictor. The link function defines the relationship between the linear predictor and the mean. The linear predictor takes the form

$$\eta = X\beta \tag{2.5}$$

where β is parameter vector
and X is defined by the design.

For a single observation, X is a row vector, and for a set of observations, X is the design matrix

The linear predictor is related to the mean by $\eta = g(\mu)$ The function g is called the link function, and the special case $g(\mu) = \theta$ is called the canonical link function

By way of illustration, the log-likelihoods for some common distributions are given below

(i) Normal

The log-likelihood is $\dfrac{\mu v - \frac{1}{2}\mu^2}{\sigma^2} - \dfrac{y^2}{2\sigma^2} - \dfrac{1}{2}\log(2\pi\sigma^2)$

Thus, $\theta = \mu$ and the canonical link function is the identity function.

$b(\theta) = \dfrac{\theta^2}{2}$ and $c(y,\theta) = -\dfrac{y^2}{2\sigma^2} - \dfrac{1}{2}\log(2\pi\sigma^2)$

$V(\mu) = 1$ and $\varphi = \sigma^2$

(ii) Poisson

The log-likelihood is $y \log\mu - \mu - \log y'$

$\theta = \log \mu$ and the canonical link is the log function

$b(\theta) = e^{\theta}$ and $c(y, \varphi) = -\log y'$

$V(\mu) = \mu$ and $\varphi = 1$

(iii) Binomial

Suppose $R \sim Binomial\ (m, \mu)$. Define $Y = \dfrac{R}{m}$. Then the log-likelihood is

$$\frac{y \log \dfrac{\mu}{1-\mu} - \log(1-\mu)}{\dfrac{1}{m}} + \log\binom{m}{my}$$

Hence $\theta = \log \dfrac{\mu}{1-\mu}$, and the canonical link function is the logit function

$b(\theta) = \log(1 + e^{\theta})$ and $c(y, \varphi) = \log\binom{m}{my}$

$V(\mu) = \mu(1-\mu)$ and $\varphi = \frac{1}{m}$.

Note that this parameterisation may be unfamiliar because of the definition of Y. However, it enables us to give a coherent theory in the following section.

(iv) Gamma (with mean μ and variance $\frac{\mu^2}{\nu}$).

The log-likelihood is $\dfrac{-\frac{y}{\mu} + \log\frac{1}{\mu}}{\frac{1}{\nu}} + \nu \log y + \nu \log \nu - \log \Gamma(\nu)$

$\theta = -\frac{1}{\mu}$ and the canonical link is the reciprocal function.

$b(\theta) = -\log(-\theta)$ and $c(y, \varphi) = \nu \log y + \nu \log \nu - \log \Gamma(\nu)$.

$V(\mu) = \mu^2$ and $\varphi = \nu^{-1}$.

This section has given a brief introduction to GLMs. The following section shows how standard credibility theory can be applied in this context. Section 4 will show how more general models can be formulated.

3. THE BUHLMAN MODEL FOR EXPONENTIAL FAMILIES

In this section, we derive the credibility formulae for exponential families of distributions, under the assumptions made by Buhlmann (1967). It is possible to extend this to other models; for example the assumptions of Buhlmann and Straub (1970) can be incorporated using weight functions. This section derives just the credibility formulae. A brief discussion of the estimation of the dispersion parameters is given in section 4, where the appropriate references are cited.

Denote the data by y_{ij} for $i = 1, 2, \quad , t; j = 1, 2, \quad , n_i$. Assume for the moment, as is common in credibility applications, that $n_i = k, \ \forall \ i$, but note that this restriction is not necessary for the derivation of HGLMs.

Thus, i indexes the risks within the collective. In credibility theory, it is assumed that each risk has a risk parameter, which we denote by ξ_i for risk i.
The assumptions of the model of Buhlmann (1967) are

(i) The risks, and hence ξ_i, are independently, identically distributed.
(ii) $y_{ij} \mid \xi_i$ are independently, identically distributed.

We assume that $y_{ij} \mid \xi_i$ is distributed according to an exponential family. Define $m(\xi_i) = E\left[y_{ij} \mid \xi_i\right]$. Note that under the assumptions of the model, $E\left[y_{ij} \mid \xi_i\right]$ does not

depend on J. Hence the canonical parameter for observation y_{ij} does not depend on j, and we assume that it can be written as follows

$$\theta_i' = \theta(m(\xi_i)) = \theta(u_i) \tag{3.1}$$

where θ is the canonical link function and u_i is a random effect for group i. Thus, for the standard credibility model, $m(\xi_i) = u_i$. Define $v_i = \theta(u_i)$; then, in this case,

$$\theta_i' = v_i. \tag{3.2}$$

Again, note that there is no j dependence here. Note also that this also implies that $Var(y_{ij} | \xi_i)$ does not depend on j.

This has defined the distribution of the random variable within each risk, conditional on the risk parameter. It is also necessary to define the structure of the collective - the distribution of $\{\xi_i \quad i = 1, \quad , t\}$. This is often done by defining a Bayesian prior distribution; here we use the same form of distribution for the random effects, but do not perform a Bayesian analysis. Instead, we define a "hierarchical likelihood", h, which we maximize.

We define the conjugate hierarchical generalized linear model (HGLM) by defining the kernel of the log-likelihood for $\theta(u_i)$ as

$$a_1\theta_i' - a_2 b(\theta_i') \tag{3.3}$$

In the actuarial literature, this distribution (the distribution of the random effects) is known as the structure of the collective. Note that we define the log-likelihood of ξ_i implicitly through that of $\theta(m(\xi_i))$. We have conditioned on ξ_i through $m(\xi_i) = u_i$, since it is the latter that we wish to estimate.

From (3.3) and the distribution of $y_{ij} | \xi_i$, we may define a hierarchical log-likelihood as

$$h = \sum_{i,j} l(\theta_i', y_{ij} | v_i) + \sum_i l(v_i) \tag{3.4}$$

$$= \sum_{ij} \left(\frac{y_{ij}\theta_i' - b(\theta_i')}{\varphi} \right) + c(y_{ij}, \varphi) + a_1\theta_i' - a_2 b(\theta_i') \tag{3.5}$$

When the distribution of both the data and the random effects is normal, this is Henderson's joint log-likelihood (Henderson 1975). In other cases, it is an obvious extension of the joint log-likelihood, called the hierarchical log-likelihood. We have now defined a hierarchical generalized linear model (HGLM), in this case the conjugate HGLM. In the particular case described in this section, the linear predictor for y_{ij} consists solely of a random effects term which is modelled in the second stage of the likelihood, (3.2). It is possible to incorporate more structure into the model by including fixed effects and generalizing the form of the random effects model. However, in this section we are concerned solely with showing that the estimates obtained under the basic model described above are the usual credibility estimates. Thus, we require an estimate of $m(\xi_i) = u_i$. The mean random effects $\{u_i : i = 1, \ .., t\}$ are estimated by maximizing the hierarchical likelihood, (3.4), as follows.

Using (2.2)

$$\frac{\partial b(\theta(u_i))}{\partial v_i} = u_i.$$

Hence
$$\frac{\partial h}{\partial v_i} = \sum_{j=1}^{k}\left(\frac{y_{ij} - u_i}{\varphi}\right) + a_1 - a_2 u_i$$

Equating $\frac{\partial h}{\partial v_i}$ to 0 gives

$$y_{i+} - k\hat{u}_i + \varphi a_1 - \varphi a_2 \hat{u}_i = 0 \tag{3.6}$$

where $y_{i+} = \sum_{j=1}^{k} y_{ij}$

Hence
$$\begin{aligned}\hat{u}_i &= \frac{y_{i+} + \varphi a_1}{k + \varphi a_2} \\ &= Z\bar{y}_i + (1 - Z)m\end{aligned}$$

where $\bar{y}_i = \frac{1}{k} y_{i+}$, $Z = \frac{k}{k + \varphi a_2}$ and $m = \frac{a_1}{a_2}$.

Thus, we have shown that, with the choice of distribution for the random effects defined in (3.3), and using the canonical link function, the estimate of u_i is in the form of a credibility estimate provided $E(m(\xi_i)) = \frac{a_1}{a_2}$ This is straightforward to show, and was also proved by Jewell (1974). The density of u_i is proportional to

$$e^{a_1\theta_i' - a_2 b(\theta_i')}$$

Now
$$\begin{aligned}\frac{\partial e^{a_1\theta_i' - a_2 b(\theta_i')}}{\partial \theta_i'} &= \left(a_1 - a_2 \frac{\partial b(\theta_i')}{\partial \theta_i'}\right) e^{a_1\theta_i' - a_2 b(\theta_i')} \\ &= (a_1 - a_2 m(\xi_i)) e^{a_1\theta_i' - a_2 b(\theta_i')}\end{aligned}$$

Integrating over the natural range of θ_i', and assuming $e^{a_1\theta_i' - a_2 b(\theta_i')}$ is zero at the end points, we have

$$a_1 - a_2 E[m(\xi_i)] = 0.$$

Hence, using (2.2), $$E[m(\xi_i)] = E[u_i] = \frac{a_1}{a_2}$$

Thus, we have shown that the credibility estimate is the same as the estimate obtained using a conjugate HGLM with pure random effects. This shows that credibility theory is closely connected to the statistical theory of random-effect models Of course, it is possible to widen the scope of the models considerably. Fixed effects terms can also

be included in the model, other link functions may be considered and the form of the random-effect models can be generalized

It is possible to formulate the pure random-effect model in another way, by including a fixed effect which is constant for all the data This means that the overall mean is estimated as a fixed effect and the randomeffects model departures from this overall mean There is no effect on the credibility estimates, but the above derivation is, in some ways, closer to the actuarial theory

The results in this section are closely related to those of Jewell (1974). The present approach differs in that it is not presented as a Bayesian procedure, and the emphasis is on the modelling aspects encapsulated within Generalized Linear Models

The estimation of the dispersion parameters is discussed in section 4. This includes the estimation of φ and of a_1 and a_2. It should be noted that if a constant fixed effect is included in the model, as outlined above, there is only one parameter to estimate in the distribution of u_i For this reason we adopt this approach henceforth

By way of illustration, we consider the four exponential families outlined in section 2 Note that we can derive the density of u_i from the density of $\theta(u_i)$, defined in (3.3) The density of u_i is proportional to

$$e^{a_1\theta_i' - a_2 b(\theta_i')} \frac{\partial\theta(u_i)}{\partial u_i}$$

$$\frac{e^{a_1\theta_i' - a_2 b(\theta_i')}}{V(u_i)} \tag{3.7}$$

(ii) Normal

The random effects have log-likelihood whose kernel is

$$a_1 u_i - a_2 \frac{u_i^2}{2}$$

i e. $u_i \sim N(m, \sigma_0^2)$ $\quad a_1 = \frac{m}{\sigma_0^2}, a_2 = \frac{1}{\sigma_0^2}$ and $m = E[u_i] = \frac{a_1}{a_2}$

(ii) Poisson

u_i has a likelihood proportional to

$$\frac{e^{a_1 \log u_i - a_2 u_i}}{u_i}$$

Hence $u_i \sim$ Gamma, parameters a_1 and a_2, and $m = E[u_i] = \frac{a_1}{a_2}$

(iii) Binomial

u_i has a likelihood proportional to

$$\frac{\exp\left[a_1 \log\left(\frac{u_i}{1-u_i}\right) - a_2 \log\left(\frac{1}{1-u_i}\right)\right]}{u_i(1-u_i)}$$

i e $u_i \sim$ Beta, parameters a_1 and $a_2 - a_1$, and $m = E[u_i] = \frac{a_1}{a_2}$

(iv) Gamma

u_i has a likelihood proportional to

$$\frac{\exp\left(\dfrac{-a_1}{u_i} + a_2 \log u_i\right)}{u_i^2}$$

i e $u_i \sim$ inverse gamma and $m = E[u_i] = \dfrac{a_1}{a_2}$

Having shown that the estimates obtained using conjugate HGLMs for a simple random-effect model are the usual credibility estimates, we now define a more general framework which encompasses credibility models

4 HIERARCHICAL GENERALIZED LINEAR MODELS

Standard GLMs model differences between groups, parametric variation and other effects as fixed effects in the linear predictor Random-effect models can be combined with standard GLMs in order to formulate models with both fixed effects and the random effects of credibility models. To do this, we define an extended linear predictor for a single observation as

$$\eta' = \eta + v \tag{4 1}$$

where $\eta = X\beta$, as in (2 5)

and v is a strictly monotonic function of $u, v = v(u)$

When $v = 0$, (4.1) reduces to the standard linear predictor for GLMs. When $\eta = 0$ and $v = \theta(u)$, we have the basis credibility model described in section 3.

The hierarchical log-likelihood, (3.4), becomes

$$h = \sum_{i,j} l(\beta, y_{ij} \mid v_i) + \sum_i l(v_i)$$

where $v_i = v(u_i)$

The maximum hierarchical likelihood extimates (MHLEs) of β and u are obtained from the pair of equations

$$\frac{\partial h}{\partial \beta} = 0 \quad \text{and} \quad \frac{\partial h}{\partial v} = 0$$

which may be solved iteratively using the procedures written by the second author for the statistical package Genstat

We consider here the case when the canonical link function is used for the fixed effects and $v = \theta$(u) In this case, equation (3 1) for observation y_{ij} becomes

$$\theta'_{ij} = \theta_{ij} + \theta(u_i) \tag{4 2}$$

where $\theta_{ij} = X_{ij}\beta$

θ is the canonical link function

and X_{ij} is the row from the design matrix for the fixed effects which relates to y_{ij}

The same log-likelihood is used for $\theta(u_i)$, as in (3 3) Then the kernel of h is

$$\frac{\sum_{i,j}(y_{ij}\theta'_{ij} - b(\theta'_{ij}))}{\varphi} + \sum_i l(v_i)$$

Hence
$$\frac{\partial h}{\partial \beta_k} = \frac{\sum_{i,j}(y_{ij} - u'_{ij}))x_{kij}}{\varphi} \tag{4.3}$$

and
$$\frac{\partial h}{\partial v_i} = \frac{\sum_{i,j}(y_{ij} - u'_{ij})) + \varphi a_1}{\varphi} - a_2 u_i \tag{4 4}$$

where $u'_{ij} = E[y_{ij}|u_i] = E[y_{ij}|\xi_i]$,

β_k is the kth parameter in the fixed effects

and x_{kij} is the kth entry of the row vector X_{ij}

Note that in this case, unlike that in section 3, $E[y_{ij}|\xi_i] \neq u_i$ Instead,

$$\theta(u'_{ij}) = \eta_{ij} + \theta(u_i) \tag{4 5}$$

which implies that $\mu'_{ij} = u_i$ when $\eta_{ij} = 0$.

We include the overall mean as a fixed effect and require that the random effects then have the appropriate mean (eg 0 for the identity link function).

The dispersion parameters given the fixed and random effects are estimated by maximising the h-likelihood after a suitable adjustment. The adjustment, which results in an adjusted profile h-likelihood, is necessary because the marginal maximum likelihood estimates may be biased. Further justifications for this adjustment can be found in Cox and Reid (1987) and Lee and Nelder (1996) For the normal distribution, unbiased estimates are obtained More details on estimation theory for random-effect GLMs can be found in McGilchrist (1994) and Schall (1991).

The joint estimates of the mean effects (fixed and random) and the dispersion parameters are obtained by iterating between the two sets of estimating equations. These processes may be conveniently carried out in Genstat, for which a set of procedures is available from the second author.

For the distributions illustrated in section 4, the likelihoods of the random effects are again appropriate, but the estimate will be different because of the difference between (3 1) and (4 2)

5 DISCUSSION

It is possible to extend the class of models to which these methods may be applied by specifying just the mean and variance functions This is useful when greater flexibility is required in the modelling assumptions For example, Renshaw and Verrall (1994) show that the chain-ladder technique in claims reserving is essentially equivalent to GLM with a Poisson likelihood and an appropriate linear predictor. By specifying just the mean and variance function, this model may be applied to a much wider class of data than is implied by the Poisson assumption (which obviously requires the variance to equal the nean). This involves the use of extended quasi-likelihood (Wedderburn 1974, Nelder and Pregibon 1987). For HGLMs, the equivalent extension is the extended quasi-h-likelihood, in which the extended quasi-likelihood is used in the hierarchical likelihood This extension makes it possible, for example, to include random effects in the chain-ladder linear model to allow a connection between accident years.

HGLMs may also be of use when a particular factor is hard to model parametrically An example of this, which has been mentioned above, is claims reserving, when it is inappropriate to model the accident years as completely independent, but a parametric relationship is also inappropriate. The same comment applies in motor premium rating, when it is usual to group a factor such as the age of the policyholder. Such a grouping may be inappropriate, as it may be crude or doubtful because it has been decided before the analysis of the data (for example, according to the present rating structure). However, it is often inappropriate, because of computational and theroretical considerations, to treat the ages as completely separate or to apply a parametric model In this situation, HGLMs may be useful

Applications in life insurance include similar premium-rating situations as in general insurance, and also graduation theory The use of HGLMs for graduation would have some similarities to Whittaker graduation, which can be regarded as a GLM with a stochastic linear predictor (Verrall, 1993).

REFERENCES

BRESLOW, N E and CLAYTON D G (1993) Approximate inference in generalised linear mixed models *J Am Statist Ass*, **88**, 9-25

BUHLMANN, H (1967) Experience Rating and Credibility *ASTIN Bulletin*, **4**, 199-207

BUHLMANN, H and STRAUB, E (1970) Credibility for Loss Ratios *ARCH*, **1972.2**.

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

GOOVAERTS, M and HOOGSTAD, W (1987) *Credibility Theory* Surveys of Actuarial Studies No 4, Rotterdam Nationale-Nederlanden

HABERMAN, S and RENSHAW, A E, (1996) *Generalized Linear Models and actuarial science* The Statistician, **45**, 407-436

HACHEMEISTER, C R (1975) Credibility for Regression Models with Applications to Trend, in *Credibility Theory and Applications*, P Kalm, ed, New York Academic Press

HENDERSON, C R (1975) Best linear unbiased estimation and prediction under a selection model *Biometrics* **31**, 423-447

JEWELL, W S (1974) Credible means are exact Bayesian for exponential families *ASTIN Bulletin*, **8**, 77-90

JEWELL, W S (1975) The use of collateral data in credibility theory a hierarchical model *Giornale dell'Instituto Italiano degli Attuari*, **38**, 1-16

KLUGMAN, S (1987) Credibility for Classification Ratemaking via the Hierarchical Normal Linear Model *Proc of the Casualty Act Soc*, **74**, 272-321

KLUGMAN, S (1992) *Bayesian Statistics in Actuarial Science with Emphasis on Credibility* Kluwer Academic Publishers

LEE Y and NELDER, J A (1996) Hierarchical Generalized Linear Models *J R Statist Soc* B, 58, 619-678

MCCULLAGH, P and NELDER, J A (1989) *Generalized Linear Models* 2nd Edition London Chapman and Hall

MCGILCHRIST, C A (1994) Estimation in generalized mixed models *J R Statist Soc B*, **56**, 61-69

MAKOV, U E, SMITH, A F M and LIU, Y -H *Bayesian methods in actuarial science* The Statistician, **45**, 503-515

MOWBRAY, A H (1914) How extensive a payroll exposure is necessary to give a dependable pure premium *Proc of the Casualty Act Soc*, **1**, 24-30

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 *Biometrika*, **74**, 221-231

RENSHAW A E (1991) Actuarial Graduation Practice and Generalized Linear and Non-Linear Models *J Inst Acts*, **118**, 295-312

RENSHAW, A E and VERRALL, R J (1994) A stochastic model underlying the chain-ladder technique, *Proceedings, ASTIN Colloquium*, 1994

SCHALL, R (1991) Estimating in generalized linear models with random effects *Biometrika*, **78**, 719-727

VERRALL, R J (1993) A state space formulation of Whittaker graduation, with extensions *Insurance Mathematics and Economics*, **13**, 7-14

DE VYLDER, F (1976) Optimal semilinear credibility *Bull of the Assoc of Swiss Acts*, **78**, 27-40

DE VYLDER, F (1986) General regression in multidimensional credibility theory, in *Insurance and Risk Theory*, ed M GOVAERTS, J. HAEZENDONCK and F DE VYLDER, Reidel

WEDDERBURN, R W M (1974) Quasi-likelihood functions, generalized linear models and the Gauss-Newton method *Biometrika*, **61**, 439-447

WHITNEY, A W (1918) The theory of experience rating *Proc of the Casualty Act Soc*, **4**, 274-292

ZEHNWIRTH, B (1977) The mean credibility formula is a Bayes rule *Scandinavian Actuarial Journal*, **4**, 212-216