Theory and Applications of Generalized Linear Models in Insurance

Jun Zhou

A Thesis
in the
Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy (Mathematics) at
Concordia University
Montréal, Québec, Canada

September, 2011

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Jun Zhou**

Entitled: **Theory and Applications of Generalized Linear Models in Insurance**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Prof. E. Bloodgood	Chair
Prof. G. Venter	External Examiner
Dr. F. Nebebe	External to Program
Dr. Y.P. Chaubey	Examiner
Dr. P. Gaillardetz	Examiner
Dr. J. Garrido	Thesis Supervisor

Approved By Dr. J. Garrido, Graduate Program Director

September, 2011 Dr. B. Lewis, Dean Faculty of Arts and Science

ABSTRACT

Theory and Applications
of Generalized Linear Models in Insurance

by Jun Zhou Ph.D.
Concordia University, 2011

Generalized linear models (GLMs) are gaining popularity as a statistical analysis method for insurance data. We study the theory and applications of GLMs in insurance. The first chapter gives an introduction of the theory of GLMs and generalized linear mixed models (GLMMs) as well as the bias correction for GLM estimators. It is shown that the maximum likelihood estimators (MLEs) of the parameters in GLMs are asymptotically normal and asymptotically unbiased. However, when the sample size n or the total

Fisher information is small, the MLEs can be biased. The bias is usually ignored in practice. However, in small or moderate–size portfolios, a bias correction can be appreciable.

For segmented portfolios, as in car insurance, the question of credibility arises naturally; how many observations are needed in a risk class before the GLM estimators can be considered credible? In this thesis we study the limited fluctuations credibility of the GLM estimators as well as in the extended case of GLMMs. We show how credibility depends on the sample size, the distribution of covariates and the link function. We give a criteria for full credibility of the GLM estimators. This provides a mechanism to obtain confidence intervals for the GLM and GLMM estimators.

If the full credibility criteria cannot be satisfied, it is interesting to calculate the partial credibility matrix and the GLM estimators. Here, for a general link function the credibility matrix is not known explicitly. Un-

der certain assumptions, numerical methods are developed to compute the GLM estimators and the credibility matrix. For some specific link functions, further properties are developed. For instance, Hachemeister's credibility regression model is one such special case of our model, where the link function is linear.

Loss reserving is a major challenge for casualty actuaries due to the frequently changing business environments. Recently, some aggregate loss reserving models have been extended to or developed by research actuaries within the framework of GLMs. In this thesis we establish a structural loss reserving model which combines the exposure and loss emergence patterns and the loss development pattern, again within the framework of a GLM. Discounted loss reserves can also be obtained from this model.

Keywords: GLMs, full credibility, partial credibility, loss reserving.

Acknowledgment

First and foremost, I wish to express my utmost gratitude to Dr. José Garrido, my supervisor, for his encouragement, support and many invaluable suggestions and comments for this research. His perpetual energy and enthusiasm in research had motivated all his advisees, including me. In addition, he was always very supportive for my decisions both in my research and my career development, which make my research life very enjoyable, smooth and rewarding. Without his guidance and help, this dissertation would not have been possible.

I also want to thank Professors Chaubey, Gaillardetz, Nebebe and Venter for their work on the thesis examining committee. Finally, I would like to thank my wife Yuan Fang for her moral support throughout my Ph.D. study. I could not accomplish so much without her support.

Contents

List of Tables

List of Figures

Chapter 1

Generalized Linear Models and Mixed Models

1.1 Introduction

Nelder and Wedderburn (1972) introduced the theory of generalized linear models (GLMs). The class of generalized linear models is an extension of traditional linear models that allows the mean of a population to depend on a *linear predictor* through a nonlinear link function and allows the response probability distribution to be any member of an exponential family of distributions. Many other books and journal articles followed the cornerstone article by Nelder and Wedderburn (1972). McCullagh and Nelder (1989) (the original text was published in 1983) provided a detailed introduction to GLMs. The books by Aitkin et al. (1989) and Dobson (1990) are also excellent references with many examples of applications of GLMs. Haberman and Renshaw (1996) give a comprehensive review of the applications of GLMs to actuarial problems. Hardin and Hilbe (2007) provides a handbook of how to deal with data using GLMs and GLM extensions. Ohlsson and Johansson (2010) gives several illustrations on how to use GLMs in non–life insurance pricing, in particular for multiplicative and hierarchical models. Neuhaus et al. (1991), Neuhaus (1992), Zeger et al. (1998) and Verbeke and Molenberghs (2000) discuss the use of GLMs in longitudinal or correlated data. Anderson et al. (2005) is a very good practitioner's guide for GLMs. Hardin and Hilbe (2002) generalize the GLMs within the generalized estimating equations theory.

Lee et al. (2006) is a comprehensive reference for GLMs with random effects. Generalized linear mixed models (GLMMs) are an extension of GLMs, complicated by random effects. They have gained significant popularity in recent years for modeling binary/count, clustered and longitudinal data. McCulloch and Searle (2001) and Demindenko (2004) are useful references for details on GLMMs. Antonio and Beirlant (2007) give an application of GLMMs in actuarial statistics.

GLMs are becoming quickly the premier statistical analysis method for insurance data. We consider the question of credibility: how many observations are needed in a risk class of a segmented portfolio before the GLM estimator can be considered credible? Schmitter (2004) provides an excellent simple method to estimate the number of claims that will be needed for a tariff calculation depending on the number of risk factors and the number of levels for each factor. Another interesting question of credibility: if the observations are less than the number required for full credibility, what is the partial credibility of the GLM estimator? In this chapter we study the *limited fluctuations credibility* of GLM estimators as well as in the extended case of generalized linear mixed models (GLMMs). Here credibility depends on the sample size, the distribution of covariates and the choice of link function. This provides a mechanism to obtain confidence intervals for the estimates in GLMs and GLMMs. The results in this chapter have already appeared in Garrido and Zhou (2009). We begin by giving an introduction to GLMs and GLMMs.

1.2 Generalized linear models (GLMs)

GLMs are a natural generalization of classical linear models that allow the mean of a population to depend on a linear predictor through a (possibly nonlinear) link function. This allows the response probability distribution to be any member of the exponential family (EF) of distributions.

A GLM consists of the following components:

1. The response Y has a distribution in the EF, with density or probability function taking the form

$$f(y;\theta,\phi) = \exp\left\{ \int \frac{\left[y-\mu(\theta)\right]}{\phi V(\mu)}\, d\mu(\theta) + c(y,\phi) \right\}, \qquad (1.1)$$

where θ is called the *natural* parameter, ϕ is a *dispersion* parameter, $\mu = \mu(\theta) = \mathbb{E}(Y)$ and $\mathbb{V}(Y) = \phi\, V(\mu)$, for a given variance function V and known bivariate function c. The EF is very flexible and can model continuous, binary, or count data.

2. For a random sample $Y_1, \ldots, Y_n$, the linear component is defined as

$$\eta_i = \underline{X}_i' \underline{\beta}\,, \qquad i = 1, \ldots, n\,, \qquad (1.2)$$

for some vector of parameters $\underline{\beta} = (\beta_1, \ldots, \beta_p)'$, and covariate $\underline{X}_i = (x_{i1}, \ldots, x_{ip})'$ associated with observation Y_i.

3. A monotonic differentiable *link function* g describes how the expected response $\mu_i = \mathbb{E}(Y_i)$ is related to the linear predictor η_i

$$g(\mu_i) = \eta_i\,, \qquad i = 1, \ldots, n\,. \tag{1.3}$$

Example 1.1. *GLMs commonly used in credibility theory*

Table 1.1 below gives the different model components of the GLMs most commonly used in credibility theory for observed claim counts or claim severities.

$Y \sim$	Normal(μ, σ^2)	Gamma(α, β)	Poisson(λ)	Bin.$(m, q)/m$
$\mathbb{E}(Y) = \mu(\theta)$	$\theta = \mu$	$-\theta^{-1} = \frac{\alpha}{\beta}$	$e^\theta = \lambda$	$\frac{e^\theta}{1+e^\theta} = q$
$\mathbb{V}(Y) = V(\mu)\phi$	σ^2	$\frac{1}{\theta^2\,\alpha} = \frac{\alpha}{\beta^2}$	$e^\theta = \lambda$	$\frac{q\,(1-q)}{m}$
$V(\mu)$	1	θ^{-2}	$e^\theta = \lambda$	$q(1-q)$
ϕ	σ^2	α^{-1}	1	$1/m$
$c(y, \phi)$	$-\frac{1}{2}[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)]$	$\alpha \ln \alpha y + \ln y - \ln \Gamma(\alpha)$	$-\ln(y!)$	$\ln \binom{m}{m\,y}$
Link g	identity	reciprocal	log	logit

Table 1.1: GLMs Examples

Additional examples include inverse Gaussian and negative binomial observations, as well as multinomial proportions (for details see McCullagh and Nelder, 1989).

1.2.1 MLE of the regression parameters

For an observed independent random sample $y_1, \ldots, y_n$, consider the log–likelihood of $\underline{\beta}$:

$$l(\underline{\beta}) = \ln L(\underline{\beta}) = \sum_{i=1}^{n} \Big\{ \int \frac{\big[y_i - \mu_i(\theta)\big]}{\phi V(\mu_i)} d\mu_i(\theta) + c(y_i, \phi) \Big\} . \qquad (1.4)$$

Example 1.2. *The following log–likelihood functions are for some commonly used distributions, where w_i is a known weight for each observation (when the weight is not specified, then simply put $w_i = 1$ for each observation):*

1. Normal:

$$l(\underline{\beta}) = \sum_{i=1}^{n} -\frac{1}{2} \Big\{ \frac{w_i (y_i - \mu_i)^2}{\phi} + \ln\big(\frac{\phi}{w_i}\big) + \ln(2\pi) \Big\} .$$

2. Gamma:

$$l(\underline{\beta}) = \sum_{i=1}^{n} \Big\{ \frac{w_i}{\phi} \ln\big(\frac{w_i y_i}{\phi \mu_i}\big) - \frac{w_i y_i}{\phi \mu_i} - \ln(y_i) - \ln\Big(\Gamma\big(\frac{w_i}{\phi}\big)\Big) \Big\}.$$

3. Poisson:

$$l(\underline{\beta}) = \sum_{i=1}^{n} \Big\{ y_i \ln(\mu_i) - \mu_i \Big\}.$$

4. Inverse Gaussian:

$$l(\underline{\beta}) = \sum_{i=1}^{n} -\frac{1}{2}\Big\{ \frac{w_i (y_i - \mu_i)^2}{y_i \mu^2 \phi} + \ln\big(\frac{\phi y_i^3}{w_i}\big) + \ln(2\pi) \Big\}.$$

5. Negative Binomial:

$$l(\underline{\beta}) = \sum_{i=1}^{n} \Big\{ y_i \ln(k\mu) - (y_i + 1/k)\ln(1 + k\mu) + \ln\Big(\frac{\Gamma(y_i + 1/k)}{\Gamma(y_i + 1)\Gamma(1/k)}\Big) \Big\}.$$

6. Multinomial

$$l(\underline{\beta}) = \sum_{i=1}^{n} \Big\{ y_{ij} \ln(\mu_{ij}) \Big\}.$$

Maximizing the log–likelihood function (1.4), we solve for the MLE of the regression parameter as $\hat{\underline{\beta}}$. Take the derivative of (1.4):

$$\frac{dl(\underline{\beta})}{d\underline{\beta}} = \sum_{i=1}^{n} \frac{dl(\underline{\beta})}{d\mu_i} \frac{d\mu_i}{d\underline{\beta}} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{d\mu_i}{d\underline{X}_i'\underline{\beta}} \frac{d\underline{X}_i'\underline{\beta}}{d\underline{\beta}},$$

where

$$\frac{d\mu_i}{d\underline{X}_i'\underline{\beta}} = \frac{dg^{-1}(\underline{X}_i'\underline{\beta})}{d\underline{X}_i'\underline{\beta}} = \frac{1}{g'(\mu_i)}.$$

Hence

$$\frac{dl(\underline{\beta})}{d\underline{\beta}} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{1}{g'(\mu_i)} \underline{X}_i'. \tag{1.5}$$

Note that if Y_i has a normal distribution, then $g'(\mu_i) = 1$, and $V(\mu_i) = 1$ for all i. Setting $\frac{dl(\underline{\beta})}{d\underline{\beta}} = 0$ yields $\sum_{i=1}^{n} \underline{X}_i(y_i - \underline{X}_i'\underline{\beta}) = 0$. In other EF cases, no closed form solution is available to this system of p equations. Instead, to obtain the maximum likelihood estimator (MLE) numerically, we must resort to an iterative algorithm such as Newton–Raphson or Fisher scoring methods.

The Newton–Raphson method provides successive approximations to the root $\underline{\hat{\beta}}$ of (1.5). On the rth iteration, the algorithm updates the parameter estimate $\underline{\hat{\beta}}_r$ with

$$\underline{\hat{\beta}}_{r+1} = \underline{\hat{\beta}}_r - \mathbf{H}^{-1}\underline{s}\,, \quad r = 1, 2, \ldots,$$

where $\mathbf{H}$ is the Hessian (second derivative) matrix, and $\underline{s}$ is the gradient (first derivative) vector of the log-likelihood function. Both are evaluated at the current value of the parameter estimate and are given by

$$\underline{s} = \sum_i \frac{w_i(y_i - \mu_i)\underline{x}_i}{V(\mu_i)g'(\mu_i)\phi}\,,$$

$$\mathbf{H} = -\mathbf{X}'\mathbf{W}_o\mathbf{X}\,,$$

where $\mathbf{X}$ is the design matrix, $\underline{x}_i$ is the transpose of the ith row of $\mathbf{X}$, and V is the variance function. The matrix $\mathbf{W}_o$ is diagonal with its ith diagonal element equal to

$$w_{oi} = w_{ei} + w_i(y_i - \mu_i)\frac{V(\mu_i)g''(\mu_i) + V'(\mu_i)g'(\mu_i)}{[V(\mu_i)]^2[g'(\mu_i)]^3\phi}\,,$$

where

$$w_{ei} = \frac{w_i}{\phi V(\mu_i)[g'(\mu_i)]^2},$$

and w_i is a known weight for each observation. If the weight is not specified, then simply put $w_i = 1$ for each observation. The primes denote derivatives of g and V with respect to μ. The negative of $\mathbf{H}$ is called the observed information matrix. The expected value of $\mathbf{W}_o$ is a diagonal matrix $\mathbf{W}_e$ with diagonal values w_{ei}. If you replace $\mathbf{W}_o$ with $\mathbf{W}_e$, then the negative of $\mathbf{H}$ is called the expected information matrix. $\mathbf{W}_e$ is the weight matrix for Fisher's scoring method.

Remark 1.1. *The theory of GLM's was developed for observations in the exponential family (EF) of distributions, but the theory and numerical algorithm also extends to other non–EF distributions.*

1.2.2 Asymptotic properties of the regression MLE

The MLE $\hat{\underline{\beta}}$ for the GLM parameters has some nice asymptotic properties when n, the number of observations, tends to infinity.

Lemma 1.1. *For the MLE, $\hat{\underline{\beta}}$ that solves* (1.5), *we have:*

1. *$\hat{\underline{\beta}}$ is an asymptotically unbiased and consistent estimator of $\underline{\beta}$.*

2. *$\mathbb{V}(\hat{\underline{\beta}}) \rightarrow \mathbf{\Sigma} = -\mathbf{H}^{-1}$, as $n \rightarrow \infty$. $\mathbf{H} = -\mathbf{X}'\mathbf{W_o}\mathbf{X}$ is the Hessian matrix, while $\mathbf{W_o} = diag(w_{o1}, \ldots, w_{on})$ is a diagonal weight matrix with i-th element $w_{oi} = \frac{w_i}{\phi V(\mu_i)(g'(\mu_i))^2} + w_i(y_i - \mu_i)\frac{V(\mu_i)g''(\mu_i)+V'(\mu_i)g'(\mu_i)}{(V(\mu_i))^2(g'(\mu_i))^3\phi}$, for known weights w_i and covariate matrix $\mathbf{X} = (\underline{X}_1, \ldots, \underline{X}_n)'$.*

3. *$\hat{\underline{\beta}} \xrightarrow{d} N\big(\underline{\beta}, (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\,\phi\big)$, i.e. it converges in distribution.*

For a proof see Fahrmeir and Kaufmann (1985).

Note that for a finite sample, the MLE $\hat{\underline{\beta}}$ is usually biased. Hence its mean square error $\text{MSE}(\hat{\underline{\beta}}) = \mathbb{V}(\hat{\underline{\beta}}) + \text{bias}(\hat{\underline{\beta}})$ plays an important role. We will see in Section 2.2 that this finite–sample bias is affected by the choice of link function g (see Cordeiro and McCullagh, 1991). For a proof see McCullagh and Nelder (1989).

1.3 Generalized linear mixed models (GLMMs)

The generalized linear mixed model is an extension of the generalized linear model, complicated by random effects. It has gained significant popularity in recent years for modeling binary/count, clustered and longitudinal data. A GLMM consists of the following components:

1. For cluster data Y_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, assumed conditionally independent given the random effects $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$, consider the following EF distribution:

$$f(y_{ij}|\underline{u}_i, \theta, \phi) = \exp\left\{\frac{[y_{ij}\theta_{ij} - b(\theta_{ij})]}{\phi} + c(y_{ij}, \phi)\right\}, \qquad (1.6)$$

where $\underline{u}_i = (u_{i1}, \ldots, u_{ik})$ are variates from normally distributed k-

dimensional random vectors $\boldsymbol{U}_i \sim N(0, \mathbf{D})$, where $\mathbf{D}$ is the variance–covariance matrix and $\mu_{ij} = \mathbb{E}[Y_{ij}|\boldsymbol{U}_i = \underline{u}_i] = b'(\theta_{ij})$. The variance of the observations, conditional on the random effects, is given by $\mathbb{V}[Y_{ij}|\boldsymbol{U}_i = \underline{u}_i] = \mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2}$. The diagonal matrix $\mathbf{A}_i$ contains the variance functions of the model, which express the variance of a response Y_{ij} as a function of its mean μ_{ij}. The matrix $\mathbf{R}_i$ is the variance–covariance matrix for the random effects.

2. The linear mixed effects model is defined as:

$$\eta_{ij} = \underline{X}'_{ij}\underline{\beta} + \underline{T}'_{ij}\underline{u}_i, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \tag{1.7}$$

for the fixed effects parameter vector $\underline{\beta} = (\beta_1, \ldots, \beta_p)'$ and random effects vector $\underline{u}_i = (u_{i1}, \ldots, u_{ik})'$. Here $\underline{X}_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ and $\underline{T}_{ij} = (t_{ij1}, \ldots, t_{ijk})'$ are both covariates.

3. A link function g,

$$g(\mu_{ij}) = \eta_{ij}\,, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n_i\,, \tag{1.8}$$

completes the model.

Most estimation methods for $\underline{\beta}$ and $\underline{u}_i$ of GLMMs rest on some form of likelihood principle, and numerical methods are needed in most cases to obtain the estimates. Antonio and Beirlant (2007) give a brief review of some numerical techniques, such as a restricted pseudo–likelihood, the Gauss–Hermite quadrature and Bayesian methods. Demidenko (2004) gives four types of algorithms and methods for the GLMM: (a) maximum likelihood with numerical quadrature, (b) penalized quasi–likelihood (PQL), (c) specific methods in conjunction with a Laplace approximation or a generalized estimating equation (GEE) approach, and (d) Monte Carlo methods for integral or log–likelihood approximations.

For the GLMM defined in (1.6)–(1.8), the log–likelihood takes the form

$$l(\underline{\beta}, \mathbf{D}) = -\frac{nk}{2}\ln(2\pi) - \frac{n}{2}|\mathbf{D}| + \sum_{i=1}^{n} \ln \int_{R^k} e^{l_i(\underline{\beta}, \boldsymbol{v}) - \frac{1}{2}\boldsymbol{v}'\mathbf{D}^{-1}\boldsymbol{v}} d\boldsymbol{v}\,, \tag{1.9}$$

where

$$l_i(\underline{\beta}, \underline{u}_i) = \sum_{j=1}^{n_i} \left[(\underline{X}'_{ij}\underline{\beta} + \underline{T}'_{ij}\underline{u}_i)\, y_{ij} - b(\underline{X}'_{ij}\underline{\beta} + \underline{T}'_{ij}\underline{u}_i) \right] \tag{1.10}$$

is the i-th conditional log–likelihood (the term $c(y)$ is omitted because it does not affect the likelihood maximization).

As explained in [49, pp.119–121], there are two types of numerical algorithms to solve for (1.9). The first type is based on Taylor series and hence these algorithms are known as linearization methods. The series expansions give an approximate model based on pseudo–data, with fewer non–linear components.

This computation of the linear approximation must be repeated several times until convergence is reached, according to some criterion. Schabenberger and Gregoire (1996) give several algorithms based on Taylor series for clustered data.

These fitting techniques based on linearizations are usually doubly iterative. The GLMM is first approximated by a linear mixed model based on current values of the covariance parameter estimates. Then the resulting linear mixed model is fitted, forming an iterative process. At convergence, the new parameter estimates are used to update the linearization, generating a new linear mixed model. The process stops when parameter estimates, for successive fits of the linear mixed model, change only within a specified tolerance.

The second type of algorithm is based on integral approximations. The log–likelihood of the GLMM is first approximated before the numerical optimization. Various techniques exist to compute the approximation: Laplace and quadrature methods, Monte Carlo integration, and Markov chain Monte

Carlo methods. The advantage of these integral approximation methods is that they give an actual objective function for the optimization step. This allows for likelihood ratio tests among nested models, and the computation of likelihood–based fit statistics. The estimation requires only a single iterative process.

The disadvantage of integral approximation methods is the difficulty to study crossed random effects, multiple subject effects, and complex $\mathbf{R}_i$–side covariance structures. Also, the number of random effects must be small if the integral approximation is to be feasible.

On the other hand, linearization methods yield a simpler linearized model, for which it is sufficient to fit only the mean and variance of the linearized form. This is a great advantage for models in which the joint distribution is difficult or impossible to obtain. Models with correlated errors, a larger number or crossed random effects, and multiple types of subjects perform well under linearization methods. The main disadvantages of this approach are the absence of a true objective function for the overall optimization. Also, it can lead to potentially biased estimators of the covariance parameters, especially in the case of binary data. The objective function, after each linearization update, is dependent on the current pseudo–data. The optimization process can fail at both levels of the double iteration scheme. For details see Wolfinger and O'Connell (1993).

1.3.1 Pseudo-likelihood estimation based on linearization

From (1.7)–(1.8) and the SAS manual [49] we have that $\mathbb{E}[\boldsymbol{Y}_i|\boldsymbol{U}_i = \underline{u}_i] = g^{-1}(\underline{X}_i\underline{\beta}+\underline{T}_i\underline{u}_i) = g^{-1}(\underline{\eta}_i) = \underline{\mu}_i$ for $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$, $\underline{X}_i = (X'_{i1}, \ldots, X'_{in_i})$, $\underline{T}_i = (T'_{i1}, \ldots, T'_{in_i})$, $\underline{\eta}_i = (\eta_{i1}, \ldots, \eta_{in_i})'$ and $\underline{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})'$. The first Taylor series of $\underline{\mu}_i$ about $\hat{\underline{\beta}}$ and $\hat{\underline{u}}_i$ yields

$$g^{-1}(\underline{\eta}_i) \doteq g^{-1}(\hat{\underline{\eta}}_i) + \hat{\boldsymbol{\Delta}}_i\underline{X}_i(\underline{\beta} - \hat{\underline{\beta}}) + \hat{\boldsymbol{\Delta}}_i\underline{T}_i(\underline{u}_i - \hat{\underline{u}}_i)\,, \tag{1.11}$$

where

$$\hat{\mathbf{\Delta}}_i = \frac{\partial g^{-1}(\underline{\eta}_i)}{\partial \underline{\eta}_i}\Big|_{\underline{\hat{\beta}},\underline{\hat{u}}_i}, \tag{1.12}$$

is a diagonal matrix of derivatives of the conditional mean evaluated at the expansion locus. Rearranging terms yields the following expression

$$\hat{\mathbf{\Delta}}_i^{-1}\big(\underline{\mu}_i - g^{-1}(\underline{\hat{\eta}}_i)\big) + \underline{X}_i\underline{\hat{\beta}} + \underline{T}_i\underline{\hat{u}}_i \doteq \underline{X}_i\underline{\beta} + \underline{T}_i\underline{u}_i\,. \tag{1.13}$$

The left–hand side is the expected value, conditional on $\underline{u}_i$, of

$$\hat{\mathbf{\Delta}}_i^{-1}\big(\boldsymbol{Y}_i - g^{-1}(\underline{\hat{\eta}}_i)\big) + \underline{X}_i\underline{\hat{\beta}} + \underline{T}_i\underline{\hat{u}}_i \equiv \underline{P}_i \tag{1.14}$$

and the variance–covariance matrix

$$\mathbb{V}[\underline{P}_i|\underline{u}_i] = \hat{\mathbf{\Delta}}_i^{-1}\mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2}\hat{\mathbf{\Delta}}_i^{-1}\,. \tag{1.15}$$

One can thus consider the model

$$\underline{P}_i = \underline{X}_i\underline{\beta} + \underline{T}_i\underline{u}_i + \underline{\epsilon}_i\,, \tag{1.16}$$

which is a linear mixed model with a pseudo–response $\underline{P}_i$, fixed effects $\underline{\beta}$, random effects $\underline{u}_i$, and $\mathbb{V}[\underline{\epsilon}_i] = \mathbb{V}[\underline{P}_i|\underline{u}_i]$.

Now define

$$\boldsymbol{V}(\underline{\theta}_i) = \underline{T}_i \mathbf{D} \underline{T}_i + \hat{\boldsymbol{\Delta}}_i^{-1} \mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2} \hat{\boldsymbol{\Delta}}_i^{-1}, \tag{1.17}$$

as the marginal variance function in the linear mixed pseudo–model, where $\underline{\theta}_i$ is the $q \times 1$ parameter vector containing all unknowns in $\mathbf{D}$ and $\mathbf{R}_i$. Based on this linearized model, an objective function can be defined, assuming that the distribution of $\underline{P}_i$ is known. The maximum log pseudo–likelihood, $l(\underline{\theta}, \underline{P})$, for all $\underline{\theta}_i$ and $\underline{P}_i$ is then given by

$$l(\underline{\theta}, \underline{P}) \;=\; -\frac{1}{2}\Big[\sum_{i=1}^{n} \ln|\boldsymbol{V}(\underline{\theta}_i)| - \sum_{i=1}^{n} \underline{r}_i' \boldsymbol{V}(\underline{\theta}_i)^{-1} \underline{r}_i - f\,\ln(2\pi)\Big], \tag{1.18}$$

where $\underline{r}_i = \underline{P}_i - \underline{X}_i \big(\sum_{j=1}^{n} \underline{X}_j' \boldsymbol{V}(\underline{\theta}_j)^{-1} \underline{X}_j\big)^{-1} \big(\sum_{j=1}^{n} \underline{X}_j' \boldsymbol{V}(\underline{\theta}_j)^{-1} \underline{P}_j\big)$, while f denotes the sum of the frequencies used in the analysis. At convergence, the estimates are

$$\underline{\hat{\beta}} \;=\; \Big(\sum_{i=1}^{n} \underline{X}_i' \boldsymbol{V}(\underline{\hat{\theta}}_i)^{-1} \underline{X}_i\Big)^{-1} \Big(\sum_{i=1}^{n} \underline{X}_i' \boldsymbol{V}(\underline{\hat{\theta}}_i)^{-1} \underline{P}_i\Big), \tag{1.19}$$

$$\underline{\hat{u}}_i \;=\; \hat{\mathbf{D}} \mathbf{T}_i' \boldsymbol{V}(\underline{\hat{\theta}}_i)^{-1} (\underline{P}_i - \underline{X}_i \underline{\hat{\beta}}). \tag{1.20}$$

For more details on this pseudo-likelihood method with linearization, please see the SAS manual in [49].

1.4 Bias correction in GLMs

The MLEs of the GLM parameters are asymptotically unbiased and normal. However, when the sample size n or the total Fisher information is small the MLEs may be biased. The bias is usually ignored in practice. However, in small or moderate-size samples, a bias correction can be appreciable. Furthermore, the estimated expected responses are also biased unless the link function is linear. We give a bias correction in GLMs which makes the GLM estimators more accurate and credible.

McCullagh and Nelder (1989) and Cordeiro and McCullagh (1991) derive some general formulae for first-order biases of MLEs of the linear parameters, linear predictors, the dispersion parameter and the fitted values in GLMs.

These formulae may be used to compute bias–corrected maximum likelihood estimates. This will increase the accuracy and credibility of the GLMs estimators, the residual bias being bounded by as term of n^{-1}.

For the GLMs as defined (1.1)–(1.3), let $\underline{\hat{\beta}}$, $\underline{\hat{\eta}}$ and $\underline{\hat{\mu}} = g^{-1}(\underline{\hat{\eta}})$ be the MLEs of $\underline{\beta}$, $\underline{\eta}$ and $\underline{\mu}$. Furthermore, let $B_1(\underline{\hat{\beta}})$, $B_1(\underline{\hat{\eta}})$, $B_1(\underline{\hat{\mu}})$ be the residual bias of $\underline{\hat{\beta}}$, $\underline{\hat{\eta}}$ and $\underline{\hat{\mu}}$ respectively, after the above bias correction. Cordeiro and McCullagh (1991) give the following formulae for first-order biases as follows.

Theorem 1.1. *In matrix notation, the bias–corrected* $\underline{\hat{\beta}}$*,* $\underline{\hat{\eta}}$ *and* $\underline{\hat{\mu}}$ *have residual biases as follows*

$$B_1(\underline{\hat{\beta}}) = -(2\phi)^{-1}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Delta}_d\mathbf{F}\underline{\mathbf{1}}, \qquad (1.21)$$

$$B_1(\underline{\hat{\eta}}) = -(2\phi)^{-1}\mathbf{\Delta}\mathbf{\Delta}_d\mathbf{F}\underline{\mathbf{1}}, \qquad (1.22)$$

$$B_1(\underline{\hat{\mu}}) = -(2\phi)^{-1}(\mathbf{G}_2 - \mathbf{G}_1\mathbf{\Delta}\mathbf{F})\mathbf{\Delta}_d\underline{\mathbf{1}}, \qquad (1.23)$$

and the bias–corrected MLE values are defined as

$$\underline{\hat{\beta}}_c = \underline{\hat{\beta}} + B_1(\underline{\hat{\beta}}), \quad (1.24)$$

$$\underline{\hat{\eta}}_c = \underline{\hat{\eta}} + B_1(\underline{\hat{\eta}}), \quad (1.25)$$

$$\underline{\hat{\mu}}_c = \underline{\hat{\mu}} + B_1(\underline{\hat{\mu}}). \quad (1.26)$$

In these expression ϕ is the dispersion parameter, $\mathbf{X}$ and $\mathbf{W}$ are defined as in Lemma 1.1, $\mathbf{\Delta} = (\delta_{ij}) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'$ is, apart from the dispersion parameter ϕ, the asymptotic covariance matrix for the estimates $\hat{\eta}_1, \ldots, \hat{\eta}_n$ of the linear predictors of the model, $\mathbf{\Delta}_d = diag\{\delta_{11}, \ldots, \delta_{nn}\}$, $\underline{\mathbf{1}}$ is an $n \times 1$ vector of ones, $\mathbf{F} = diag\{\frac{1}{V}\frac{d\mu}{d\eta}\frac{d^2\mu}{d\eta^2}\}$, where V is the variance function as defined in (1.1), $\mathbf{G}_1 = diag\{\frac{d\mu}{d\eta}\}$ and $\mathbf{G}_2 = diag\{\frac{d^2\mu}{d\eta^2}\}$.

Remark 1.2. *For the binomial and Poisson distributions, expressions (1.24)–(1.26) do not involve ϕ.*

Cordeiro and McCullagh (1991) also give the following example.

Example 1.3. *Binomial proportion with a log–link function*

Suppose that $Y \sim Bin.(m, q)/m$, the binomial proportion with index m and parameter q as defined in Table 1.1, and that $\eta = \ln[q/(1-q)]$ is the logistic link of q. The MLE of η is

$$\hat{\eta} = \ln[Y/(m-Y)] \tag{1.27}$$

whereas the usual bias-corrected estimates have the form

$$\hat{\eta}_c = \ln[(Y+c)/(m-Y+c)] \tag{1.28}$$

for some constant, usually taken to be $c = \frac{1}{2}$ (Cox and Snell, 1979).

Obviously if $c > 0$, $|\hat{\eta}_c| < |\hat{\eta}|$, so that bias correction has the effect of shrinking the MLE towards the origin. The general bias-corrected estimate of the linear predictor described above is equivalent in this case to the choice $c = m/2(m-1)$. There is a finite probability that $Y = 0$ or $Y = m$, and consequently $\hat{\eta}$ is consistent for η with asymptotic bias of order $O(m^{-1})$ for large m.

Cordeiro and McCullagh (1991) give the first-order bias corrections in GLMs for small or moderate-sized samples. However, even with large enough sample size n, the estimated response $\hat{y}_i = g^{-1}(\underline{X}_i \hat{\underline{\beta}})$ is asymptotically biased due to the curvature of the link function. Cordeiro and McCullagh (1991) also give numerical examples to illustrate the computations by using the logistic model.

Chapter 2

Full Credibility Theory for GLMs and GLMMs

2.1 Full credibility criteria

Developed in the early part of the 20th century, *limited fluctuations credibility* gives formulas to assign *full* or *partial* credibility to an individual or group of policy–holders' experience. Mowbray (1914) pioneered the use of experience rating for worker's compensation premium formulas. He used a heuristic approach, based on classical statistics, to develop full credibility formulae.

Also in the context of worker's compensation, Whitney (1918) is an early attempt at a more rigorous *greatest accuracy credibility.* Bailey (1950) is also a significant contribution to this early credibility research literature.

A more statistical approach to credibility was developed in the second part of the century. Some of the important contributions to partial credibility of that period were given by Bühlmann (1967, 1969), Bühlmann and Straub (1970), Hachemeister (1975) and Jewell (1975).

More recently, Frees (2003) as well as Goulet et al. (2006) give some results in using credibility theory in loss data. Nelder and Verrall (1997) show how credibility theory can be encompassed within the theory of GLMs. In that vein Schmitter (2004) gives a simple method to estimate the number of claims needed for a GLM tariff calculation. Here we focus on full credibility with a GLM model.

The insurer may find credible the estimator $\hat{\mu}_i$ of the mean parameter μ_i it estimates if the probability of small differences $|\hat{\mu}_i - \mu_i|$ is large. If this difference is small “enough”, we say that *full credibility* is achieved. Statistically, this can be defined as

$$\pi_i = \mathbb{P}\big\{|\hat{\mu}_i - \mu_i| \le r\mu_i\big\} \ge p_i\,, \qquad i = 1, \ldots, n\,, \tag{2.1}$$

for a chosen estimation–error tolerance level $0 < r < 1$ and confidence probability p_i.

Proposition 2.1. *For any generalized linear model, as defined in (1.1)–(1.3), let g be a monotonic increasing link function. Then the probability*

$$\begin{aligned}
\pi_i &= \mathbb{P}\big\{|\hat{\mu}_i - \mu_i| \le r\mu_i\big\} = \mathbb{P}\big\{(1-r)\mu_i \le \hat{\mu}_i \le (1+r)\mu_i\big\} \\
&= \mathbb{P}\big\{g[(1-r)\mu_i] - g(\mu_i) \le g(\hat{\mu}_i) - g(\mu_i) \le g[(1+r)\mu_i] - g(\mu_i)\big\} \\
&= \mathbb{P}\big\{g[(1-r)\mu_i] - \underline{X}_i'\underline{\beta} \le \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \le g[(1+r)\mu_i] - \underline{X}_i'\underline{\beta}\big\}\,. \qquad (2.2)
\end{aligned}$$

It is reasonable to restrict g to increasing link functions. If needed, similar results would follow for decreasing link functions.

Proposition 2.1 gives some expressions equivalent to (2.1) and transfers the confidence interval from the scale of the GLM estimators $\hat{\mu}_i$, to the scale of the linear components, through the link function g.

For a general link function the lower and upper bounds to $\underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta}$ in (2.2) depend on the parameters in $\underline{\beta}$. But if the link function satisfies the condition that $g(c\,\mu_i) = g(\mu_i) + c'$ for any μ_i, where c and c' are constants with respect to μ_i, then (2.2) admits a simpler form as follows.

Proposition 2.2. *For any given error tolerance level r and any μ_i,*

$$\mathbb{P}\big\{|\hat{\mu}_i - \mu_i| \le r\mu_i\big\} = \mathbb{P}\big\{c_1 \le \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \le c_2\big\}\,, \qquad i = 1, \ldots, n\,, \quad (2.3)$$

if and only if a log–link function $g(x) = c\,\ln(x) + \tau$ is used in (2.2)*, where c and τ are scale and shift–parameters, respectively, and c_1, c_2 are given by:*

$$c_1 = c\,\ln(1-r) \quad \textit{and} \quad c_2 = c\,\ln(1+r)\,. \quad (2.4)$$

Proof: ($\Leftarrow$) If $g(x) = c\,\ln(x) + \tau$, by (2.2), it is clear that for any fixed

$i = 1, \ldots, n$,

$$g[(1-r)\mu_i] - g(\mu_i) = c\,\ln[(1-r)\mu_i] - c\,\ln(\mu_i) = c\,\ln(1-r)\,,$$

and

$$g[(1+r)\mu_i] - g(\mu_i) = c\,\ln[(1+r)\mu_i] - c\,\ln(\mu_i) = c\,\ln(1+r)\,.$$

($\Rightarrow$) Again, for any fixed $i = 1, \ldots, n$, if

$$\mathbb{P}\big\{|\hat{\mu}_i - \mu_i| \le r\mu_i\big\} = \mathbb{P}\big\{c_1 \le \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \le c_2\big\}\,,$$

then from (2.2), for any μ_i,

$$c_1 = g[(1-r)\mu_i] - g(\mu_i) \quad \text{and} \quad c_2 = g[(1+r)\mu_i] - g(\mu_i). \tag{2.5}$$

Assuming that g is differentiable, then for any μ_i

$$g'(\mu_i) = \lim_{r\to 0} \frac{g[(1-r)\mu_i] - g(\mu_i)}{-r\mu_i} = \lim_{r\to 0} \frac{c_1}{-r\mu_i} \tag{2.6}$$

but also

$$g'(\mu_i) = \lim_{r\to 0} \frac{g[(1+r)\mu_i] - g(\mu_i)}{r\mu_i} = \lim_{r\to 0} \frac{c_2}{r\mu_i}. \tag{2.7}$$

Hence $\lim_{r\to 0}\frac{c_1}{-r} = \lim_{r\to 0}\frac{c_2}{r} = c$, say. Then $g'(\mu_i) = \frac{c}{\mu_i}$, which indicates that

$g(x) = c\,\ln(x) + \tau$. □

The above proposition shows that for the log–link function, the upper and lower bounds of the full credibility rule do not depend on the estimated value $\hat{\mu}_i$. They only depend on the chosen error tolerance level r. The following example gives a concrete illustration.

Example 2.1. *Poisson distribution with a log–link function*

Let Y_i be independent Poisson distributed random variables representing the number of claims for risk $i = 1, \ldots, n$. Here $\mathbb{E}(Y_i) = \mu_i = e^{x_{i1}\beta_1 + \cdots + x_{ip}\beta_{ip}}$. With the log–link function, $g[\mathbb{E}(Y_i)] = g(\mu_i) = x_{i1}\beta_1 + \cdots + x_{ip}\beta_{ip}$. By (2.2), $|\hat{\mu}_i - \mu_i| \le r\mu_i \Leftrightarrow \ln(1-r) \le \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \le \ln(1+r)$. Since $0 < r < 1$, then

$|\ln(1+r)| < |\ln(1-r)|$ and hence

$$\begin{aligned}\mathbb{P}\{|\hat{\mu}_i - \mu_i| \le r\mu_i\} &= \mathbb{P}\{\ln(1-r) \le \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \le \ln(1+r)\} \\ &\le \mathbb{P}\{|\underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta}| \le |\ln(1-r)|\} \ . \qquad (2.8)\end{aligned}$$

Now let $s^2 = \mathbb{V}(\hat{\beta}_1 + \cdots + \hat{\beta}_p)$ and $\underline{X}_i = (1, 1, \ldots, 1)$, then (2.8) becomes

$$\begin{aligned}&\mathbb{P}\{|\underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta}| \le |\ln(1-r)|\} \\ &\quad = \mathbb{P}\{|(\hat{\beta}_1 + \cdots + \hat{\beta}_p) - (\beta_1 + \cdots + \beta_p)| \le |\ln(1-r)|\} \\ &\quad = \mathbb{P}\Big\{\Big|\frac{(\hat{\beta}_1 + \cdots + \hat{\beta}_p) - (\beta_1 + \cdots + \beta_p)}{s}\Big| \le \frac{|\ln(1-r)|}{s}\Big\}. \qquad (2.9)\end{aligned}$$

Approximating by a normal distribution, (2.9) yields $\frac{|\ln(1-r)|}{s} \ge Z_{\pi_*}$, where Z_{π_*} is the $\pi_* = 100[1 - (\frac{1-\pi}{2})]$-percentile of a standard normal distribution. Hence the following asymptotic full credibility criterion is obtained:

$$s^2 \le \Big[\frac{\ln(1-r)}{Z_{\pi_*}}\Big]^2 = s_*^2 \ ,$$

which says that the sample size n must be sufficiently large to ensure that the variance of the sum of the estimators $\hat{\beta}_1, \ldots, \hat{\beta}_p$ be at most s_*^2. For instance, if $r = 0.1$ and $\pi = 90\%$ then $s_*^2 = 0.00410$. This result is consistent with the result given by Schmitter (2004, p.258).

The following results consider the asymptotic behaviour of $\hat{\mu}_i = \underline{X}_i'\underline{\hat{\beta}}$.

Proposition 2.3. *Let* $\mathbf{\Sigma} = (\sigma_{ij})_{i,j} = (\mathbf{X}'\mathbf{W_o}\mathbf{X})^{-1}$ *and* $s_i^2 = \mathbb{V}(\hat{\mu}_i) = \mathbb{V}(\underline{X}_i'\underline{\hat{\beta}})$. *Then for every component* $i = 1, \ldots, n$,

$$s_i^2 \to \underline{X}_i'\mathbf{\Sigma}\underline{X}_i \,, \tag{2.10}$$

as $n \to \infty$, *where* $\underline{X}_i$, $\mathbf{W_o}$ *and* $\mathbf{X}$ *are given in Lemma 1.1.*

Proof: From Lemma 1.1–(2) we have that $\mathbb{V}(\underline{\hat{\beta}}) \to \mathbf{\Sigma}$, as $n \to \infty$, and the iterative $\underline{\hat{\beta}}$ converges to the true $\underline{\beta}$, then

$$\begin{aligned} s_i^2 &= \mathbb{V}(\underline{X}_i'\underline{\hat{\beta}}) = \mathbb{V}(x_{i1}\hat{\beta}_1 + \cdots + x_{ip}\hat{\beta}_p) \\ &= \sum_{j=1}^{p}\sum_{k=1}^{p} x_{ij}\, x_{ik}\, \mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_k) \to \sum_{j=1}^{p}\sum_{k=1}^{p} x_{ij}\, x_{ik}\, \sigma_{jk} = \underline{X}_i'\mathbf{\Sigma}\underline{X}_i \,. \end{aligned}$$

□

Furthermore, Lemma 1.1 states that $\underline{\hat{\beta}}$ converges to $N\big(\underline{\beta}, \mathbf{\Sigma}\big)$ in distribution. Then, the following corollary to Proposition 2.3 holds.

Corollary 2.1. $\big(\underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta}\big)/s_i$ *converges to* $N(0,1)$ *in distribution.*

We are now in a position to state the main results in this section on the asymptotic full credibility standard for $\hat{\mu}_i$.

Theorem 2.1. *For the log–link function, an asymptotic normal approximation gives*

$$\pi_i \doteq \Phi\Big(\frac{\ln(1+r)}{s_i}\Big) - \Phi\Big(\frac{\ln(1-r)}{s_i}\Big), \qquad i = 1, \ldots, n, \tag{2.11}$$

where Φ *is the cumulative distribution function (cdf) of the standard normal distribution.*

Proof: From Propositions 2.1 and 2.2,

$$\begin{aligned} \pi_i &= \mathbb{P}\big\{\ln(1-r) \le \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \le \ln(1+r)\big\} \\ &= \mathbb{P}\Big\{\frac{\ln(1-r)}{s_i} \le \frac{\underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta}}{s_i} \le \frac{\ln(1+r)}{s_i}\Big\}\,. \end{aligned}$$

Hence, by the normal approximation, $\pi_i \doteq \Phi(\frac{\ln(1+r)}{s_i}) - \Phi(\frac{\ln(1-r)}{s_i})$. $\square$

For any confidence coefficient π_i, Theorem 2.1 gives a $100(1-r)\%$ confidence interval for μ_i, the mean response from the GLM. The theorem also shows that the confidence interval varies with the value of the covariates since s_i is a function of $\underline{X}_i$. The examples in Section 2.3 illustrate the above results.

Now for a general link function g, let

$$Q_1 = g[(1-r)\mu_i] - g(\mu_i) \quad \text{and} \quad Q_2 = g[(1+r)\mu_i] - g(\mu_i)\,. \tag{2.12}$$

Theorem 2.2. *For a monotonic increasing link function g, we have the following asymptotic approximation:*

$$\pi_i \doteq \Phi\Big(\frac{Q_2}{s_i}\Big) - \Phi\Big(\frac{Q_1}{s_i}\Big)\,, \qquad i = 1, \ldots, n\,, \tag{2.13}$$

where Φ is the cdf of the standard normal distribution, Q_1 and Q_2 are given in (2.12) *and s_i in Proposition 2.3.*

Proof:

$$\begin{aligned}\pi_i &= \mathbb{P}\big\{|\hat{\mu}_i - \mu_i| \le r\mu_i\big\} = \mathbb{P}\big\{Q_1 \le \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \le Q_2\big\} \\ &= \mathbb{P}\Big\{\frac{Q_1}{s_i} \le \frac{\underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta}}{s_i} \le \frac{Q_2}{s_i}\Big\} .\end{aligned}$$

Approximating by the normal distribution gives (2.13). □

Clearly, the smaller s_i the bigger π_i (approximately), which differs for different i. If g is the log–link function, then Proposition 2.2 gives closed forms for Q_1 and Q_2. For other link functions, as the true parameter value μ_i is unknown, we can approximate Q_1, Q_2 and π_i as follows. First set

$$\hat{Q}_1 = g[(1-r)\hat{\mu}_i] - g(\hat{\mu}_i) \quad \text{and} \quad \hat{Q}_2 = g[(1+r)\hat{\mu}_i] - g(\hat{\mu}_i) \ , \tag{2.14}$$

which then implies that

$$\hat{\pi}_i \doteq \Phi\Big(\frac{\hat{Q}_2}{s_i}\Big) - \Phi\Big(\frac{\hat{Q}_1}{s_i}\Big) \ . \tag{2.15}$$

Section 2.2 discusses the effect of the choice of link function on the above approximation.

Finally, similar results hold for the confidence probability estimates in GLMMs.

Proposition 2.4. *For any generalized linear mixed model, as defined in* (1.6)–(1.8)*, let g be a monotonic increasing link function. Then*

$$\begin{aligned}
\pi_{ij} &= \mathbb{P}\big\{|\hat{\mu}_{ij} - \mu_{ij}| \leq r\mu_{ij}\big\} = \mathbb{P}\big\{(1-r)\mu_{ij} \leq \hat{\mu}_{ij} \leq (1+r)\mu_{ij}\big\} \\
&= \mathbb{P}\big\{g[(1-r)\mu_{ij}] - g(\mu_{ij}) \leq g(\hat{\mu}_{ij}) - g(\mu_{ij}) \leq g[(1+r)\mu_{ij}] - g(\mu_{ij})\big\} \\
&= \mathbb{P}\big\{g[(1-r)\mu_{ij}] - \underline{X}'_{ij}\underline{\beta} - T'_{ij}\underline{u}_i \leq \underline{X}'_{ij}\hat{\underline{\beta}} + T'_{ij}\hat{\underline{u}}_i - \underline{X}'_{ij}\underline{\beta} - T'_{ij}\underline{u}_i \\
&\qquad\qquad \leq g[(1+r)\mu_{ij}] - \underline{X}'_{ij}\underline{\beta} - T'_{ij}\underline{u}_i\big\} \, . \qquad (2.16)
\end{aligned}$$

Using the same idea as in Theorem 2.2 (see Liang and Zeger, 1986 for the asymptotic normal distribution of the GLMM estimators), we obtain the following result for GLMMs.

Theorem 2.3. *For any link function g, let $s^2_{ij} = \mathbb{V}(\underline{X}'_{ij}\underline{\beta} + T'_{ij}\underline{u}_i)$ and Q_{1j},*

Q_{2j} be defined as

$$Q_{1j} = g[(1-r)\mu_{ij}] - g(\mu_{ij}) \quad \textit{and} \quad Q_{2j} = g[(1+r)\mu_{ij}] - g(\mu_{ij})\,, \qquad (2.17)$$

then

$$\pi_{ij} \doteq \Phi\Big(\frac{Q_{2j}}{s_{ij}}\Big) - \Phi\Big(\frac{Q_{1j}}{s_{ij}}\Big)\,, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n_i\,. \qquad (2.18)$$

2.2 The choice of link function

As shown in the previous section, the main idea here is to transfer the full credibility condition (2.1) to an equivalent form that is easier to implement, as in Theorems 2.1–2.2. Expression (2.13) gives the credibility of the GLM estimator as a function of Q_1, Q_2 and s_i, which also depend on the link function g. Thus, it is natural to investigate the effect of this choice of link function.

The following lemma shows that rescaling or shifting the link function of a given GLM has no effect on the credibility of the resulting GLM estimators.

Lemma 2.1. *Rescaling or shifting a given link function g, such as in $h(x) = c\,g(x)+\tau$, does not affect the approximate confidence probabilities π_i in* (2.13).

Proof: For a link function g, (1.3) can be rewritten as $g(\mu_i) = \beta_0^{(g)} + \underline{X}_i'\underline{\beta}^{(g)}$, where $\beta_0^{(g)}$ is the intercept. Let the new link function be $h(x) = c\,g(x) + \tau$. Then $h(\mu_i) = \beta_0^{(h)} + \underline{X}_i'\underline{\beta}^{(h)} = c\,g(\mu_i) + \tau$ and hence $g(\mu_i) = \frac{\beta_0^{(h)}-\tau}{c} + \underline{X}_i'\frac{\underline{\beta}^{(h)}}{c}$. It follows that $\beta_0^{(g)} = \frac{\beta_0^{(h)}-\tau}{c}$ and $\underline{\beta}^{(g)} = \frac{\underline{\beta}^{(h)}}{c}$.

Now let $s_i^{(g)} = \sqrt{\mathbb{V}(\underline{X}_i'\hat{\underline{\beta}}^{(g)})}$, $s_i^{(h)} = \sqrt{\mathbb{V}(\underline{X}_i'\hat{\underline{\beta}}^{(h)})}$. Clearly $s_i^{(g)} = \frac{1}{c}s_i^{(h)}$, or equivalently, $s_i^{(h)} = c\,s_i^{(g)}$, while

$$Q_i^{(h)} = h[(1 \pm r)\mu_i] - h(\mu_i) = c\left\{g[(1 \pm r)\mu_i] - g(\mu_i)\right\} = c\,Q_i^{(g)}\ ,$$

for $i = 1, 2$. Refer to (2.13) and substitute $Q_i^{(h)}$ and $s_i^{(h)}$ above, to see that

$$\pi_i^{(g)} \doteq \Phi\Big(\frac{Q_2^{(g)}}{s_i^{(g)}}\Big) - \Phi\Big(\frac{Q_1^{(g)}}{s_i^{(g)}}\Big) = \Phi\Big(\frac{c\,Q_2^{(g)}}{c\,s_i^{(g)}}\Big) - \Phi\Big(\frac{c\,Q_1^{(g)}}{c\,s_i^{(g)}}\Big) \doteq \pi_i^{(h)}\ .$$

□

Example 2.4 gives a numerical illustration of Lemma 2.1. It shows how the estimated probabilities π_i, in (2.13), but where s_i is estimated with $\hat{s}_i$ given by the GLM, also remain essentially unchanged under any rescaling of the log–link function.

The choice of link function also affects the bias in GLM estimators, $\hat{\boldsymbol{\beta}}$, $\hat{\mu}_i = g^{-1}\big(\underline{X}_i'\hat{\boldsymbol{\beta}}\big)$ and in our estimated $\hat{Q}_1$, $\hat{Q}_2$ in (2.14). This is explored in the next result. We first reproduce a version of Jensen's inequality that we need. In what follows a convex function is called convex upward while a concave function is called convex downward.

Lemma 2.2. ***(Jensen's Inequality)*** *Let X be a random variable with finite mean $\mathbb{E}(X)$ and φ be a convex upward (respectively downward) function on $\mathbb{R}$. Then*

$$\mathbb{E}\big[\varphi(X)\big] \geq (\textit{resp.} \ \leq)\, \varphi\big(\mathbb{E}[X]\big)\,. \tag{2.19}$$

Now we can explore how the link function affects the estimation bias in our confidence intervals. We distinguish the cases when g is linear, convex upward and decreasing, like the inverse function $g(x) = 1/x$, or else when it is convex downward and increasing, like the log link function $g(x) = \ln(x)$.

Theorem 2.4. *$\hat{Q}_1$ and $\hat{Q}_2$ (2.14) are:*

1. *unbiased estimators if the link function g is linear,*

2. *asymptotically upward–biased if the link function g is convex upward and decreasing,*

3. *asymptotically downward–biased if the link function g is convex downward and increasing.*

Proof: Recall that $\hat{Q}_1 = g[(1-r)\hat{\mu}_i] - g(\hat{\mu}_i)$ and $Q_1 = g[(1-r)\mu_i] - g(\mu_i)$, where $g(\mu_i) = \underline{X}_i'\underline{\beta}$ and $g(\hat{\mu}_i) = \underline{X}_i'\underline{\hat{\beta}}$. Then

$$\begin{aligned} \text{bias}(\hat{Q}_1) &= \mathbb{E}(\hat{Q}_1) - Q_1 \\ &= \mathbb{E}\big\{g[(1-r)\hat{\mu}_i]\big\} - \mathbb{E}[\underline{X}_i'\underline{\hat{\beta}}] - g[(1-r)\mu_i] + \underline{X}_i'\underline{\beta} \\ &= \mathbb{E}\big\{g[(1-r)\hat{\mu}_i]\big\} - g[(1-r)\mu_i] - \underline{X}_i' \,\text{bias}(\underline{\hat{\beta}})\,. \qquad (2.20) \end{aligned}$$

Three cases need to be distinguished:

1. If g is linear then $\mathbb{E}\big\{g[(1-r)\hat{\mu}_i]\big\} - g[(1-r)\mu_i] = 0$ and $\hat{\beta}$ is unbiased, hence so is $\hat{Q}_1$.

2. If g is a convex upward decreasing function, then by Jensen's inequality in (2.19)

$$\mathbb{E}(\hat{\mu}_i) = \mathbb{E}\big[g^{-1}(\underline{X}_i'\underline{\hat{\beta}})\big] \leq g^{-1}\big[\mathbb{E}(\underline{X}_i'\underline{\hat{\beta}})\big] = g^{-1}(\underline{X}_i'\underline{\beta}) = \mu_i \ ,$$

   that is $\mathbb{E}(\hat{\mu}_i) \leq \mu_i$. Now since

$$\mathbb{E}\big\{g[(1-r)\hat{\mu}_i]\big\} \geq g\big\{\mathbb{E}[(1-r)\hat{\mu}_i]\big\} = g\big\{(1-r)\mathbb{E}[\hat{\mu}_i]\big\} \geq g[(1-r)\mu_i] \ ,$$

   and $\underline{\hat{\beta}}$ is asymptotically unbiased, then asymptotically $\mathbb{E}(\hat{Q}_1) - Q_1 \geq 0$. Hence $\hat{Q}_1$ is an asymptotically upward–biased estimator.

3. If g is a concave increasing function, the proof is similar but with the inverse inequalities. That is asymptotically $\mathbb{E}(\hat{Q}_1) - Q_1 \leq 0$ and $\hat{Q}_1$ is an asymptotically downward–biased estimator.

The proof is similar for the results on $\hat{Q}_2$. □

In practice the choice of a link function for a GLM is not a straightforward problem. Its solution heavily relies on experience and intuition. The following theorem gives a criterion for the choice of the link function.

Theorem 2.5. *For a GLM problem, $\hat{\pi}_i$ given by (2.15) can be used as a criterion to choose between two link functions g_1 and g_2. If $\hat{\pi}_i^{(g_1)} < \hat{\pi}_i^{(g_2)}$, we say that the estimator given under the link function g_1 is less credible than the estimator given under g_2, that is g_2 is better than g_1.*

2.3 Numerical examples

Example 2.2. *Car Insurance Claims Data (GLM)*

The SAS Technical Report P-243 (1993) gives the illustrative dataset in Table 2.1 of a car insurance portfolio (also reproduced in Schmitter, 2004). For earlier examples of traditional/nonlinear analysis of car insurance data

Class	Number of Risks	Number of Claims	Car Type	Age Group
$i = 1$	500	42	small	1
2	1200	37	medium	1
3	100	1	large	1
4	400	101	small	2
5	500	73	medium	2
6	300	14	large	2

Table 2.1: Car Insurance Data

see Coutts (1984), Aitkin et al. (1989) and Brockman and Wright (1992).

Now let the number of claims per risk y_i be Poisson and choose a log–link function. Furthermore, let the covariates $\underline{X}_i = (x_{i1}, \ldots, x_{i4})'$, where

$$
\begin{aligned}
x_{i1} &= 1, \\
x_{i2} &= \begin{cases} 1 & \text{if } \textit{car type} \text{ is } \textit{large} \\ 0 & \text{otherwise,} \end{cases} \\
x_{i3} &= \begin{cases} 1 & \text{if } \textit{car type} \text{ is } \textit{medium} \\ 0 & \text{otherwise,} \end{cases} \\
x_{i4} &= \begin{cases} 1 & \text{if } \textit{age group} \text{ is } \textit{1} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

In this notation $\underline{X}_4 = (1, 0, 0, 0)'$ defines the base premium $\mathbb{E}(Y_4) = e^{\beta_1}$ for a *small* car type in age group *2*. The matrix of variance–covariance $\mathbf{\Sigma}$ in Proposition 2.3 is computed with SAS for weights equal to the number of risks, i.e. $w_1 = 500, w_2 = 1200, w_3 = 100, w_4 = 400, w_5 = 500$ and $w_6 = 300$ (see Lemma 1.1–2).

$$\mathbf{\Sigma} = \begin{pmatrix} 0.008150 & -0.007772 & -0.006344 & -0.004623 \\ -0.007772 & 0.074180 & 0.006556 & 0.003113 \\ -0.006344 & 0.006556 & 0.016450 & -0.002592 \\ -0.004623 & 0.003113 & -0.002592 & 0.018470 \end{pmatrix}$$

Let the tolerance level $r = 0.1$ and $\underline{X}_3 = (1, 1, 0, 1)'$ for the third class of drivers, i.e. with a *large* car type in age group *1*. Then the asymptotic value in (2.10) for $s_3^2 = \underline{X}_3'\mathbf{\Sigma}\underline{X}_3 = 0.082236$ and from (2.11) we get $\pi_3 = \Phi(\frac{\ln(1+r)}{s_3}) - \Phi(\frac{\ln(1-r)}{s_3}) = 0.273533$. Clearly, the current experience produces GLM estimators that are not credible for this class with only one claim, as $s_3^2 = 0.273533 > 0.00410 = s_*^2$, for $r = 0.1$ and $\pi = 90\%$.

By contrast, letting $\underline{X}_1 = (1, 0, 0, 1)'$ gives $s_1^2 = 0.017374$ and $\pi_1 = 0.553138$, which indicates a higher confidence in the GLM estimator for *small* cars than for *large* cars, in age group *1*, although not sufficient for full credibility $s_1^2 = 0.017374 > 0.00410 = s_*^2$. Table 2.2 reports the asymptotic variances $s_i^2 = \mathbb{V}(\underline{X}'_{ij}\underline{\beta}) \to \underline{X}'_i \boldsymbol{\Sigma} \underline{X}_i$ and the credibility probabilities π_i for all 6 classes.

Class	$i = 1$	2	3	4	5	6
$\underline{X}'_i$	(1,0,0,1)	(1,0,1,1)	(1,1,0,1)	(1,0,0,0)	(1,0,1,0)	(1,1,0,0)
$\underline{X}'_i \boldsymbol{\Sigma} \underline{X}_i$	0.017374	0.015952	0.082236	0.008150	0.011912	0.066786
π_i	0.553138	0.572679	0.273533	0.732868	0.641557	0.302114

Table 2.2: Asymptotic Variances and Confidence Probabilities

Example 2.3. *Effect of Sample Size (GLM)*

Furthermore, if we modify Example 2.2 so that the claim experience increases proportionally, we see that so does the confidence probability π_i. For instance, in the third class we need to multiply exposures by as much as 23 times (i.e. both the risk and claim counts) to get $s_3^2 = 0.003575$ and $\pi_3 = 0.905492$ (i.e. full credibility at the 90% level). As expected, the GLM tends to full credibility as the portfolio size increases.

Example 2.4. *Effect of the Covariates Distribution (GLM)*

This example shows that credibility also depends on the distribution of the covariates. For instance, modify the above Car Insurance Data to keep the total number of claims unchanged at 268 in Table 2.1, but rearrange the claim counts in each group as in Table 2.3.

Class	Number of Risks	Number of Claims	Car Type	Age Group
$i = 1$	500	45	small	1
2	1200	108	medium	1
3	100	9	large	1
4	400	36	small	2
5	500	44	medium	2
6	300	26	large	2

Table 2.3: Modified Car Insurance Data

Then for $\underline{X}_3 = (1, 1, 0, 1)'$ we get an asymptotic $s_3^2 = 0.038200$ and $\pi_3 = 0.392182$, which differs from the value of 0.273533 obtained in Example 2.2. Clearly the credibility of GLM estimates depends on the distribution of the covariates.

Example 2.5. *Effect of the Link Function (GLM)*

Let the link function $g(x) = c\ \ln(x) + \tau$. Lemma 2.1 shows that c and τ have no effect on the calculation of Q_1, Q_2 and s_i. The same is true when these are estimated by a software implementation of the GLM, for instance the GENMOD procedure in SAS.

Choosing different rescaling parameters c, Table 2.4 shows that the estimated confidence values π_i in (2.13), for classes $i = 1$ and 3 remain essentially the same.

c	s_1	π_1	s_3	π_3
0.1	0.019139	0.552537	0.028674	0.273559
0.5	0.065901	0.553164	0.143400	0.273504
1	0.013181	0.553139	0.286768	0.273533
2	0.263610	0.553157	0.573620	0.273495
5	0.659007	0.553165	1.433855	0.273531

Table 2.4: Rescaled Car Insurance Data

Hence rescaling or shifting the link function practically does not affect the π_i values.

Example 2.6. *GLMM with Territory as Random Effect*

This example illustrates the credibility results for GLMMs. We add one more variable, called territory, to Example 2.2. It takes two values, "rural" and "urban", which will illustrate the random effect of a GLMM. In this example the territory is treated as a random–effect, because some tests at the data validation step showed that the impact of this variable depends on the risk class. This in contrast to fixed–effects that have an impact at the individual level. Here the fixed–effects parameters estimates, $\underline{\hat{\beta}}$, those for the random–effects, $\underline{\hat{u}}$, as well as their variance–covariance matrix were obtained with the GLIMMIX procedure in SAS (see [49]).

As in Example 2.2, let y_i be Poisson and g be a log–link function. Fur-

Class	Number of Risks	Number of Claims	Car Type	Age Group	Territory
$i = 1$	500	42	small	1	rural
2	1200	37	medium	1	urban
3	100	1	large	1	rural
4	400	101	small	2	urban
5	500	73	medium	2	rural
6	300	14	large	2	urban

Table 2.5: Car Insurance Data with Territory Random Effects

thermore, let the covariates $\mathbf{X}_i = (x_{i1}, \ldots, x_{i5})'$ and $\mathbf{T}_j = (t_{1j}, t_{2j})'$, be coded as:

$$
\begin{aligned}
x_{i1} &= 1,\\
x_{i2} &= \begin{cases} 1 & \text{if } \textit{car type} \text{ is } \textit{large} \\ 0 & \text{otherwise,} \end{cases}\\
x_{i3} &= \begin{cases} 1 & \text{if } \textit{car type} \text{ is } \textit{medium} \\ 0 & \text{otherwise,} \end{cases}\\
x_{i4} &= \begin{cases} 1 & \text{if } \textit{age group} \text{ is } \textit{1} \\ 0 & \text{otherwise.} \end{cases}\\
t_{1j} &= \begin{cases} 1 & \text{if } \textit{territory} \text{ is } \textit{rural} \\ 0 & \text{otherwise.} \end{cases}\\
t_{2j} &= \begin{cases} 1 & \text{if } \textit{territory} \text{ is } \textit{urban} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

The variance–covariance matrices $\mathbf{\Sigma}$ of the fixed effects and $\mathbf{D}$ in $\mathbb{V}(\underline{X}_i'\underline{\beta} + \underline{T}_i'\underline{u}_i) = \underline{X}_i'\mathbf{\Sigma}\underline{X}_i + \underline{T}_i'\mathbf{D}\underline{T}_i$ of the random effects are given by:

$$
\begin{aligned}
\mathbf{\Sigma} &= \begin{pmatrix} 0.016500 & -0.007440 & -0.007600 & -0.005230 \\ -0.007440 & 0.074590 & 0.006032 & 0.003173 \\ -0.007600 & 0.006032 & 0.017680 & -0.001150 \\ -0.005230 & 0.003173 & -0.001150 & 0.018250 \end{pmatrix}\\
\mathbf{D} &= \begin{pmatrix} 0.010262 & 0 \\ 0 & 0.010262 \end{pmatrix}
\end{aligned} \tag{2.21}
$$

Let the tolerance level $r = 0.1$ and $\mathbf{X}_3 = (1, 1, 0, 1)'$, $\mathbf{T}_1 = (1, 0)'$ for the third class of drivers with a *large* car type in age group *1* and in territory *rural*. Then the estimated variance, as given in Theorem 2.3 is $s_{13}^2 = \mathbf{X}_3'\mathbf{\Sigma}\mathbf{X}_3 + \mathbf{T}_1'\mathbf{D}\mathbf{T}_1 = 0.100608$. This gives $\pi_{13} = \Phi(\frac{\ln(1+r)}{s_{13}}) - \Phi(\frac{\ln(1-r)}{s_{13}}) = 0.248217$. Clearly, the current experience produces GLMM estimators in this class that have a low confidence.

By contrast, letting $\mathbf{X}_1 = (1,0,0,1)'$ and $\mathbf{T}_1 = (1,0)'$ gives $s_{11}^2 = 0.034552$ and $\pi_{11} = 0.410517$, which indicates a GLMM estimator with a higher confidence for *small* cars in age group *1* than for *large* cars in age group *1* and a *rural* territory.

2.4 Conclusion

This chapter studies the credibility of the estimators obtained from GLM and GLMM risk models. A closed form of the full credibility criteria is given for the log–link function, usually paired to Poisson observations (i.e. claim counts). For general link functions, we propose a credibility estimation based on an asymptotic normal approximation.

The proposed method should become useful to actuaries as it provides full credibility criteria for GLM estimators, at a time when these are becoming popular in the statistical analysis of insurance and risk data.

Chapter 3

Partial Credibility Theory for GLMs

3.1 Introduction to partial credibility theory

Clearly the applicability of the full credibility is limited. Condition (2.1) gives a mathematically description of a full credibility. To be able to compute the probability statements in (2.1), the distribution of $\hat{\mu}_j$ must be known, or an approximation be used. When the sample size n is large, the Central Limit Theorem can be used to approximate it by a normal distribution.

When the standard for full credibility is not met and so the GLMs estimators are not sufficiently accurate to be used as estimates of the expected value, we need a method for dealing with this situation. Lo et al. (2007) developed a generalized estimating equations approach to estimate structural parameters of a regression credibility model. Partial credibility theory is a natural and standard approach to solve this actuarial problem.

If we believe that the GLMs parameters are not fully credible, then it may be desirable to reflect the total portfolio experience μ in the estimation of each single contract $\hat{\mu}_j$. An intuitively appealing method for doing this is through a weighted average, that is through a linear credibility estimator as

$$\underline{\hat{\mu}}_j^{(z)} = \mathbf{Z}_j \underline{\hat{\mu}}_j + (\mathbf{I} - \mathbf{Z}_j)\underline{\mu}\,, \tag{3.1}$$

where $\mathbf{Z}_j$ is the credibility factor. The question here is how to estimate the credibility factor $\mathbf{Z}_j$?

There are many formulas for $\mathbf{Z}_j$ which have been suggested in the actuarial literatures, usually justified on intuition rather than from a rigorous mathematical basis. For instance, intuitively, it should be an increasing function of n_j, the sample size for contract j.

Of many proposals in the actuarial literature, the one that turned out to be the right choice is

$$\mathbf{Z}_j = \frac{n_j}{n_j + k_j}, \tag{3.2}$$

where k_j is a parameter to be determined. This particular choice can be proved to be theoretically justified on the basis of a statistical model. Klugman et al. (2004) gives a comprehensive review of this model as well as the proof. A variety of arguments have been used to derive the value of $\mathbf{Z}_j$. Some of these proposals maybe intuitively appealing but are theoretical difficult or not statistically sound.

Klugman et al. (2004) also points out several issues with this partial linear credibility approach. First, there is no underlying theoretical model for the distribution of $\hat{\mu}_j$, especially when the sample size is small, and thus no reason why a linear credibility weighted average estimator $\hat{\mu}_j^{(z)}$ in the form (3.1) is appropriate and preferable to μ or $\hat{\mu}_j$. Why not just apply GLMs on homogeneous data and give full credibility to the GLM estimator? This approach might be the best way to use all the available information.

While there is a practical reason for using (3.1), no model has been presented to prove that it is appropriate. Consequently, the choice of $\mathbf{Z}_j$ and $\hat{\mu}_j^{(z)}$ is quite arbitrary. Especially for GLMs, little research has been done to investigate the impact of a non-linear link function on the accuracy of the linear partial credibility estimator. Intuitively, this may introduce some bias and linear credibility estimation may not improve the accuracy of the GLM estimators. In the two chapters that follow, we carry out some theoretical and numerical investigation to answer this question.

Furthermore, this approach does not examine the difference between μ_j and μ. When (3.1) is used, we are essentially assuming that the value of (3.1) is an accurate representation of the expected value of this particular contract. However, μ is usually also only an estimate and therefore itself subject to error. Finally μ is not necessarily an unbiased estimator of μ_j.

With regards to the credibility estimator, $\hat{\mu}_j^{(z)}$ is not unbiased for μ_j. In fact, one of the characteristics that allows credibility to work empirically also forces the use of biased estimators. If $\hat{\mu}_j$ and μ are biased, then $\hat{\mu}_j^{(z)}$ is also biased, except in some very rare situations. This also means that the appropriate measure of the quality of the credibility estimator $\hat{\mu}_j^{(z)}$ is not its variance, but its mean-squared error.

In this thesis, we use the sum of squared errors to assess the adequacy of the models. However, the mean-squared error requires knowledge of the bias, and that requires knowledge of the relationship between μ_j and μ. In practice little is known about this relationship in most cases. We often need to resort to ad hoc assumptions or approximations.

In this chapter, we use the standard partial credibility approach in (3.1) to derive a theoretical solution for the credibility factor $\mathbf{Z}_j$ for GLM estimators, which minimize the mean-squared error, or equivalently, the sum of squares error (SSE). Then, we use some numerical illustrations to verify if the above issues stated in Klugman et al. (2004) apply to the linear partial credibility estimator for GLMs.

3.2 Partial credibility GLM estimators

When the GLM parameters are not fully credible, as defined in Chapter 2, then an alternate estimator to $\underline{\hat{\mu}}_j$ (or, equivalently, to $\underline{\hat{\beta}}_j$) as that given in Section 1.2, may be needed. Actuaries typically use *greatest accuracy credibility* estimators of the mean response $\underline{\mu}_j = \mathbb{E}(\underline{Y}_j)$. Here this means that we would look for a credibility estimator

$$\underline{\hat{\mu}}_j^{(z)} = \mathbf{Z}_j h(\mathbf{X}\underline{\hat{\beta}}_j) + (\mathbf{I} - \mathbf{Z}_j) h(\mathbf{X}\underline{b}) \,, \tag{3.3}$$

which is a mixture between the GLM contract estimator $h(\mathbf{X}\underline{\hat{\beta}}_j)$ and the portfolio mean $h(\mathbf{X}\underline{b})$, where $\mathbf{Z}_j$ is a credibility factor matrix. Here $\mathbf{Z}_j$ is obtained as to minimize the expected value of mean square error:

$$\begin{aligned} Q(\mathbf{Z}_j) &= \mathbb{E}\Big\{[h(\mathbf{X}\underline{\beta}_j) - \mathbf{Z}_j h(\mathbf{X}\underline{\hat{\beta}}_j) - (\mathbf{I} - \mathbf{Z}_j) h(\mathbf{X}\underline{b})]' \\ &\qquad [h(\mathbf{X}\underline{\beta}_j) - \mathbf{Z}_j h(\mathbf{X}\underline{\hat{\beta}}_j) - (\mathbf{I} - \mathbf{Z}_j) h(\mathbf{X}\underline{b})]\Big\} \,, \end{aligned} \tag{3.4}$$

where $\underline{\hat{\beta}}_j$ is the vector of estimated GLM parameters, $\underline{\beta}_j$ is the true vector of parameters, $\mathbf{Z}_j = (z_{il})_{i,l}$ is the $n \times n$ credibility matrix and $\underline{b}$ is the vector of portfolio parameters, while $h = g^{-1}$ is the inverse of the link function g.

3.2.1 Hachemeister's regression model

De Vylder (1996) gives a complete review of credibility regression models, including that of Hachemeister. Hachemeister worked on U.S. bodily injury car insurance data that showed linear inflation trends in claims. This trend differed from one state to the other and also from the average national inflation trend. Such examples suggest the generalization of weighted credibility models to accept regression variables. Hachemeister (1975) relaxes the assumptions of the model of Bühlmann and Straub (1970) and defines a credibility regression model as follows:

(1) The contract random vector $(\Theta_j, \underline{Y}_j)$, for $j = 1, \ldots, k$, are pair-wise independent and the risk parameters Θ_j identically distributed.

(2) For each $j = 1, \ldots, k$ the conditional mean claim vector $\underline{\mu}(\Theta_j) = [\mu_1(\Theta_j), \ldots, \mu_n(\Theta_j)]'$ is defined as:

$$\underline{\mu}(\Theta_j) = \mathbb{E}[\underline{Y}_j|\Theta_j] = \mathbf{X}\underline{\beta}(\Theta_j) = \mathbf{X}[\beta_0(\Theta_j), \ldots, \beta_{p-1}(\Theta_j)]' \qquad (3.5)$$

where $\mathbf{X}$ is a $n \times p$ design matrix (of rank $p < n$) and $\underline{\beta}(\Theta_j)$ is an unknown regression vector of length p.

(3) Furthermore,

$$\mathrm{Cov}[\underline{Y}_j|\Theta_j] = \sigma^2(\Theta_j)\mathbf{V}_j\,, \tag{3.6}$$

where $\mathbf{V}_j$ is a positive semi–definite $n \times n$ (inverse) matrix of weights and σ^2 a scale function of Θ_j. All matrices $\mathbf{X}$ and $\mathbf{V}_j$ are assumed known in advance.

Hachemeister (1975) gives the weighted least squares regression estimators of $\hat{\underline{\beta}}_j = \underline{\beta}(\Theta_j)$ and the related credibility results as follows.

Lemma 3.1. *The weighted least squares regression estimators of $\underline{\beta}_j$ which minimize*

$$Q_j(\underline{\beta}) = [\underline{Y}_j - \mathbf{X}\underline{\beta}(\Theta_j)]'\mathbf{V}_j^{-1}[\underline{Y}_j - \mathbf{X}\underline{\beta}(\Theta_j)] \tag{3.7}$$

for $j = 1, \ldots, k$, are given by

$$\hat{\underline{\beta}}_j = (\mathbf{X}'\mathbf{V}_j^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_j^{-1}\underline{Y}_j = \mathbf{U}_j\mathbf{X}'\mathbf{V}_j^{-1}\underline{Y}_j, \tag{3.8}$$

and

$$\mathbb{V}(\underline{\hat{\beta}}_j) = \mathbf{A} + s^2\mathbf{U}_j, \tag{3.9}$$

$$Cov(\underline{\hat{\beta}}_j, \underline{\beta}_j) = \mathbf{A}, \tag{3.10}$$

$$\mathbb{E}[\underline{\hat{\beta}}_j] = \underline{b}, \tag{3.11}$$

where $\mathbf{U}_j = (\mathbf{X}'\mathbf{V}_j^{-1}\mathbf{X})^{-1}$, $s^2 = \mathbb{E}[\sigma^2(\Theta_j)]$, $\mathbf{A} = Cov[\underline{\beta}(\Theta_j)]$ *and* $\underline{b} = \mathbb{E}[\underline{\beta}(\Theta_j)]$.

Corollary 3.1. *Weighted least squares estimator of* $\underline{\beta}$ *given by (3.8) produces the following credibility matrix*

$$\mathbf{Z}_j = \mathbf{A}(\mathbf{A} + s^2\mathbf{U}_j)^{-1}. \tag{3.12}$$

The proof of the above results can be found in De Vylder (1996). We will also show that the above results are a special case of the GLMs partial credibility model in the following sections.

3.2.2 The GLM partial credibility model

Consider a generalization of the credibility regression model of Hachemeister (1975) in the prior section, where a link function relates the linear term to the expected response. More specifically, the model for the partial credibility theory of GLMs is defined as follows:

(1) The contract random vectors $(\Theta_j, \underline{Y}_j)$, for $j = 1, 2, \ldots, k$, are pair–wise independent and the unobservable risk parameters Θ_j identically distributed.

(2) For all $j = 1, 2, \ldots, k$, let the vector $\underline{\mu}_j = \underline{\mu}(\Theta_j) = [\mu_1(\Theta_j), \ldots, \mu_n(\Theta_j)]'$ denote the conditional mean claims per contract, where

$$\underline{\mu}_j = \mathbb{E}[\underline{Y}_j | \Theta_j], \qquad j = 1, \ldots, k. \tag{3.13}$$

(3) For an observable random sample $\underline{Y}_1, \ldots, \underline{Y}_k$, the linear component is

defined as

$$\underline{\eta}_j = \underline{\eta}(\Theta_j) = \mathbf{X}\underline{\beta}(\Theta_j), \qquad j = 1, \ldots, k. \tag{3.14}$$

where $\mathbf{X}$ is an $n \times p$ design matrix (of rank $p < n$) and $\underline{\beta}(\Theta_j)$ is an unknown regression vector of length p (i.e. the vector functions $\underline{\eta}$ and $\underline{\beta}$ are the same for all contracts).

(4) A monotonic differentiable *link function* g describes how the expected response $\underline{\mu}_j = \mathbb{E}[\underline{Y}_j|\Theta_j]$ is related to the linear predictor as $\underline{\eta}_j$

$$g(\underline{\mu}_j) = \underline{\eta}_j, \qquad j = 1, \ldots, k, \tag{3.15}$$

or, equivalently,

$$\underline{\mu}_j = h(\underline{\eta}_j), \qquad j = 1, \ldots, k, \tag{3.16}$$

where $h = g^{-1}$ is the inverse function of g.

(5) Furthermore,

$$\mathrm{Cov}[\underline{Y}_j|\Theta_j] = \sigma^2(\Theta_j)\mathbf{V}_j, \qquad j = 1, \ldots, k, \tag{3.17}$$

where $\mathbf{V}_j = diag(w_{j1}^{-1}, w_{j2}^{-1}, \ldots, w_{jn}^{-1})$ is a positive semi–definite $n \times n$ (inverse) matrix of weights and $\sigma^2(\Theta_j)$ a scalar function of Θ_j, such that $\mathbb{V}(Y_j|\Theta_j) = \frac{\sigma^2(\Theta_j)}{w_j} = \frac{\phi}{w_j}\mathbb{V}[\mu(\Theta_j)]$. All matrices $\mathbf{X}$ and $\mathbf{V}_j$ are assumed known in advance.

Comparing this GLM partial credibility model with Hachemeister's regression model, the essential differences are in (3.15) and (3.16), where a non-linear function can be used to link the linear predictors and the estimators. This is an essential characteristics of GLMs.

For each $j = 1, 2, \ldots, k$, we estimate the GLM parameters $\underline{\beta}_j$, as $\hat{\underline{\beta}}_j$. Then, we calculate the credibility of each estimator $\hat{\underline{\mu}}_j = h(\mathbf{X}\hat{\underline{\beta}}_j)$ and apply (3.3) to get the linear credibility estimation. For both steps, explicit solutions only exist for very special cases. We need to resort to the numerical methods to approximate the solutions.

The likelihood method is standard to estimate $\underline{\beta}_j$, assuming that the GLM is appropriate and we have the required distributional information. However, in reality, the distribution of the response variable does not necessarily belong to the exponential family and is in fact unknown. That restricts the use of GLMs.

In this chapter, we develop an algorithm using least squares error (LSE) to estimate $\underline{\beta}_j$. This is a significant generalization of the GLMs because it relaxes the assumption on the distribution of the dependent variables $\underline{Y}_j$.

3.3 Estimation of the GLM parameters

Traditionally, GLMs parameters are estimated by numerical methods because the explicit solutions do not exist for most link functions. Newton–Raphson methods were widely used before some of more advanced algorithms were developed. One such improvement of the Newton–Raphson algorithm is the iteratively reweighted least squares (IRLS) algorithm, which is widely used now for GLM parameter estimation. Hardin and Hilbe (2007) gives a complete review and comparison of these two algorithms.

3.3.1 MLEs of the GLM parameters

Parameter estimation for standard GLMs is based on the likelihood function by assuming that the model is completely and correctly specified by the definition in Section 1.2. That is $\underline{Y}_j$ belongs to the exponential family (EF) as in (1.1), for $j = 1, 2, \ldots, k$. Then the log-likelihood function is as in (1.4) and the estimators $\underline{\hat{\beta}}_1, \ldots, \underline{\hat{\beta}}_k$ are calculated as in Section 1.2.1.

Hardin and Hilbe (2007) gives a clear explanation of the Newton–Raphson (using the observed Hessian) and IRLS (using the expected Hessian) algorithms. The characteristics of each algorithm are discussed and compared.

Fahrmeir and Gerhard (2001) gives very detailed discussion of the maximum likelihood method including the uniqueness and existence of MLEs.

The questions of whether MLEs exist, whether they lie in the interior of the parameter space and whether they are unique, are very interesting and important. Based on the concavity of the log-likelihood, the existence and uniqueness were developed by various authors.

Haberman (1974) gives the results for log-linear and binomial models. Wederburn (1976) gives the results for normal, Poisson, gamma and binomial models. Fahrmeir and Gerhard (2001) discuss the binomial and Poisson models in their book.

Furthermore, to relax the assumption that the response variable follows a member distribution of the exponential family, a quasi-likelihood model was developed. Quasi-likelihood relaxes assumption of GLMs so that only the mean and the variance function are needed. Fahrmeir and Gerhard (2001) gives a detailed description for this model as well.

However, in may applications, even the quasi-likelihood assumptions may be too restrictive. In the next section we further relax its assumptions and extend the result, without any distributional assumptions, using the least squares error estimation (LSE) approach.

3.3.2 LSEs of the GLMs parameters

Using LSE, there is no need to assume a distribution of $\underline{Y}_j$, nor a prior distribution on the Θ_j's. An empirical Bayes approach is used as in the classical credibility models of Bühlmann (1969) and Bühlmann and Straub (1970). The following results apply to this general model.

For each $j = 1, 2, \ldots, k$, we minimize the sum of squares

$$Q(\underline{\beta}_j) = [\underline{Y}_j - \underline{\mu}(\Theta_j)]' \mathbf{V}_j^{-1} [\underline{Y}_j - \underline{\mu}(\Theta_j)]. \tag{3.18}$$

Equivalently, the weighted least squares estimators for $\underline{\beta}_j$ are those that minimize

$$Q(\underline{\beta}_j) = [\underline{Y}_j - g^{-1}(\mathbf{X}\underline{\beta}_j)]' \ \mathbf{V}_j^{-1} \ [\underline{Y}_j - g^{-1}(\mathbf{X}\underline{\beta}_j)]. \tag{3.19}$$

Jan and Heinz (1988) is a comprehensive review of matrix differential calculus with application in Statistics. In particular it gives the following matrix calculus results needed in this section.

Lemma 3.2. *Let* ψ *be a scalar-valued function, and* f *be a vector-valued function of a* $p \times 1$ *vector* $\underline{x}$*. If* $\psi(\underline{x}) = (f(\underline{x}))'\mathbf{A}f(\underline{x})$*, then*

$$D\psi(\underline{x}) = (f(\underline{x}))'(\mathbf{A} + \mathbf{A}')Df(\underline{x})\,, \tag{3.20}$$

where $D = D_{\underline{x}}$ *is the differential operator with respect to* $\underline{x}$*.*

Lemma 3.3. ***(Chain rule)*** *Let* S *be a subset of* $\mathbb{R}^n$*, and assume that* $f : S \to \mathbb{R}^m$ *is differentiable at an interior point* $\underline{c}$ *of* S*. Let* T *be a subset of* $\mathbb{R}^m$ *such that* $f(\underline{x}) \in T$ *for all* $\underline{x} \in S$*, and assume that* $g : T \to \mathbb{R}^p$ *is differentiable at an interior point* $\underline{b} = f(\underline{c})$ *of* T*. Then the composite function* $h : S \to \mathbb{R}^p$ *defined by*

$$h(\underline{x}) = g(f(\underline{x}))$$

is differentiable at $\underline{c}$*, and*

$$Dh(x)|_{\underline{x}=\underline{c}} = [D_{\underline{y}}g(\underline{y})|_{\underline{y}=\underline{b}}]\ [Df(\underline{x})|_{\underline{x}=\underline{c}}]. \tag{3.21}$$

To minimize (3.19), we apply a differential operator with respect to $\underline{\beta}_j$ on $Q(\underline{\beta}_j)$ and then let the result equal to 0.

Lemma 3.4. *Take the derivative of (3.18) with respect to* $\underline{\beta}_j$*, then*

$$D_{\underline{\beta}_j} Q(\underline{\beta}_j) = 2[\underline{Y}_j - h(\mathbf{X}\underline{\beta}_j)]'\ \mathbf{V}_j^{-1}\ \mathbf{H}(\underline{\beta}_j)\mathbf{X}\,, \tag{3.22}$$

where $\mathbf{H}(\underline{\beta}_j)$ *is a diagonal matrix as*

$$\mathbf{H}(\underline{\beta}_j) = \begin{pmatrix} h^{(1)}(\underline{X}_1'\underline{\beta}_j) & 0 & \dots & 0 \\ 0 & h^{(1)}(\underline{X}_2'\underline{\beta}_j) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & h^{(1)}(\underline{X}_n'\underline{\beta}_j) \end{pmatrix}_{n\times n},$$

and $h^{(1)}(x)$ *is the first derivative of* $h(x)$*.*

Proof: Now by Lemma 3.2,

$$DQ(\underline{\beta}_j) = [\underline{Y}_j - g^{-1}(\mathbf{X}\underline{\beta}_j)]'\ [\mathbf{V}_j^{-1} + (\mathbf{V}_j^{-1})']\ D[\underline{Y}_j - g^{-1}(\mathbf{X}\underline{\beta}_j)], \tag{3.23}$$

where $\mathbf{V}_j^{-1} = (\mathbf{V}_j^{-1})'$, and

$$D[\underline{Y}_j - g^{-1}(\mathbf{X}\underline{\beta}_j)] = -Dg^{-1}(\mathbf{X}\underline{\beta}_j) = -Dg^{-1}(\mathbf{X}\underline{\beta}_j), \tag{3.24}$$

with $D = D_{\underline{\beta}_j}$ being a differential operator with respect to $\underline{\beta}_j$. Let $h = g^{-1}$, then

$$D_{\underline{\beta}_j} h(\mathbf{X}\underline{\beta}_j) = D_{\underline{\beta}_j} \begin{pmatrix} h(\underline{X}_1'\underline{\beta}_j) \\ h(\underline{X}_2'\underline{\beta}_j) \\ \vdots \\ h(\underline{X}_n'\underline{\beta}_j) \end{pmatrix}_{n\times 1}.$$

More specifically,

$$D_{\underline{\beta}_j} h(\mathbf{X}\underline{\beta}_j) = D_{\underline{\beta}_j} \begin{pmatrix} h(\beta_{j0} + \beta_{j1}x_{11} + \cdots + \beta_{j,p-1}x_{1,p-1}) \\ h(\beta_{j0} + \beta_{j1}x_{21} + \cdots + \beta_{j,p-1}x_{2,p-1}) \\ \vdots \\ h(\beta_{j0} + \beta_{j1}x_{n1} + \cdots + \beta_{j,p-1}x_{n,p-1}) \end{pmatrix}_{n\times 1},$$

that is,

$$D_{\underline{\beta}_j} h(\mathbf{X}\underline{\beta}_j) = \begin{pmatrix} h^{(1)}(\underline{X}_1'\underline{\beta}_j) & h^{(1)}(\underline{X}_1'\underline{\beta}_j)x_{11} & \dots & h^{(1)}(\underline{X}_1'\underline{\beta}_j)x_{1,p-1} \\ h^{(1)}(\underline{X}_2'\underline{\beta}_j) & h^{(1)}(\underline{X}_2'\underline{\beta}_j)x_{21} & \dots & h^{(1)}(\underline{X}_2'\underline{\beta}_j)x_{2,p-1} \\ \vdots & \vdots & \dots & \vdots \\ h^{(1)}(\underline{X}_n'\underline{\beta}_j) & h^{(1)}(\underline{X}_n'\underline{\beta}_j)x_{n1} & \dots & h^{(1)}(\underline{X}_n'\underline{\beta}_j)x_{n,p-1} \end{pmatrix}_{n\times p}$$

$$= \begin{pmatrix} h^{(1)}(\underline{X}_1'\underline{\beta}_j) & 0 & \dots & 0 \\ 0 & h^{(1)}(\underline{X}_2'\underline{\beta}_j) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & h^{(1)}(\underline{X}_n'\underline{\beta}_j) \end{pmatrix}_{n\times n} \begin{pmatrix} 1 & x_{11} & \dots & x_{1,p-1} \\ 1 & x_{21} & \dots & x_{2,p-1} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{n,p-1} \end{pmatrix}_{n\times p},$$

where $h^{(1)}(x)$ is the first derivative of $h(x)$. Now let

$$\mathbf{H}(\underline{\beta}_j) = \begin{pmatrix} h^{(1)}(\underline{X}_1'\underline{\beta}_j) & 0 & \dots & 0 \\ 0 & h^{(1)}(\underline{X}_2'\underline{\beta}_j) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & h^{(1)}(\underline{X}_n'\underline{\beta}_j) \end{pmatrix}_{n\times n},$$

then (3.23) can be written as

$$D_{\underline{\beta}_j} Q(\underline{\beta}_j) = 2[\underline{Y}_j - h(\mathbf{X}\underline{\beta}_j)]' \ \mathbf{V}_j^{-1} \ \mathbf{H}(\underline{\beta}_j)\mathbf{X}.$$

□

Generally there is no closed form solution for $D_{\underline{\beta}_j} Q(\underline{\beta}_j) = 0$. Numerical methods are used to obtain the weighted least squares (WLSE) estimator of $\underline{\beta}_j$. If the link function is the identity, then the WLSE can be derived in closed form, which is consistent with Hachemeister's classical regression model as in Section 3.2.1.

Proposition 3.1. *If the link function is the identity, then the diagonal elements of* $\mathbf{H}(\underline{\beta}_j)$ *are all 1 and the weighted least squares GLM estimators which minimize*

$$Q(\underline{\beta}_j) = [\underline{Y}_j - \mathbf{X}\underline{\beta}_j]' \mathbf{V}_j^{-1} [\underline{Y}_j - \mathbf{X}\underline{\beta}_j],$$

are given by

$$\hat{\underline{\beta}}_j = [\mathbf{X}' \mathbf{V}_j^{-1} \mathbf{X}]^{-1} \mathbf{X}' \mathbf{V}_j^{-1} \underline{Y}_j. \tag{3.25}$$

Proof: Since $\mathbf{H}(\underline{\beta}_j)$ is the identity matrix, if all its diagonal elements are equal to 1, then $D_{\underline{\beta}_j} Q(\underline{\beta}_j) = 0$ can be written as

$$[\underline{Y}_j - \mathbf{X}\underline{\beta}_j]'\mathbf{V}_j^{-1}\mathbf{X} = 0\,, \tag{3.26}$$

and solving (3.26), we have

$$\hat{\underline{\beta}}_j = [\mathbf{X}'\mathbf{V}_j^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}_j^{-1}\underline{Y}_j\,,$$

which is also consistent with the solution of Hachemeister's credibility regression model. □

3.4 Algorithms to compute LSE estimators

The least square errors equations and their derivatives are generally non-linear. For most link functions, no closed form solution of $D_{\underline{\beta}_j} Q(\underline{\beta}_j) = 0$ is available. Instead, we must resort to an iterative algorithm.

3.4.1 Newton–Raphson method

Let $U(\underline{\beta}_j) = \frac{1}{2} D_{\underline{\beta}_j} Q(\underline{\beta}_j)$, and $\hat{\underline{\beta}}_j$ be the weighted least squares GLM estimator. The Newton–Raphson method is based on a one-term Taylor expansion

$$0 = U(\hat{\underline{\beta}}_j) \doteq U(\hat{\underline{\beta}}_j^0) + D_{\underline{\beta}_j} U(\underline{\beta}_j)|_{\underline{\beta}_j = \hat{\underline{\beta}}_j^0} (\hat{\underline{\beta}}_j - \hat{\underline{\beta}}_j^0), \quad j = 1, \ldots, k. \tag{3.27}$$

Hence, denoting by $D_{\underline{\beta}_j} U(\hat{\underline{\beta}}_j^0) = D_{\underline{\beta}_j} U(\underline{\beta}_j)|_{\underline{\beta}_j = \hat{\underline{\beta}}_j^0}$, we can write

$$\hat{\underline{\beta}}_j \doteq \hat{\underline{\beta}}_j^0 - [D_{\underline{\beta}_j} U(\hat{\underline{\beta}}_j^0)]^{-1} \, U(\hat{\underline{\beta}}_j^0), \quad j = 1, \ldots, k, \tag{3.28}$$

which defines the iterative algorithm. It is started from an arbitrary but well-chosen $\hat{\underline{\beta}}_j^0$ to induce an improved approximation of $\hat{\underline{\beta}}_j$, that we denote $\hat{\underline{\beta}}_j^1$. Now using this $\hat{\underline{\beta}}_j^1$ as a starting value in (3.28) instead of $\hat{\underline{\beta}}_j^0$, produces a new improved approximation, say $\hat{\underline{\beta}}_j^2$. With this iterative use of (3.28) we can obtain successive approximations as $\hat{\underline{\beta}}_j^1$, $\hat{\underline{\beta}}_j^2$, $\hat{\underline{\beta}}_j^3$, $\ldots$, which converge to $\hat{\underline{\beta}}_j$ in the best case.

The Newton–Raphson method bears the disadvantage that the inverse of

$D_{\underline{\beta}_j} U(\hat{\underline{\beta}}_j^0)$ might not exist. Otherwise, in most cases the Newton–Raphson method is an efficient and simple algorithm to find the approximate solution of (3.22).

3.4.2 Fisher scoring method

A Newton–Raphson variant is the Fisher scoring algorithm that replaces the Hessian matrix by its expectation.

$$U(\hat{\underline{\beta}}_j) \doteq U(\hat{\underline{\beta}}_j^0) + \mathbb{E}\big[D_{\underline{\beta}_j} U(\hat{\underline{\beta}}_j^0)\big](\hat{\underline{\beta}}_j - \hat{\underline{\beta}}_j^0), \quad j = 1, \ldots, k. \tag{3.29}$$

Hence, let

$$\hat{\underline{\beta}}_j \doteq \hat{\underline{\beta}}_j^0 - \big\{\mathbb{E}[D_{\underline{\beta}_j} U(\hat{\underline{\beta}}_j^0)]\big\}^{-1} U(\hat{\underline{\beta}}_j^0) \tag{3.30}$$

define the iterative algorithm, which avoids the main disadvantage of the Newton–Raphson method.

3.4.3 Parameters in the iterative algorithms

Next we derive an iterative solution to the least squares problem in (3.18)-(3.19).

Theorem 3.1. *To solve*

$$D_{\underline{\beta}_j} Q(\underline{\beta}_j) = 2[\underline{Y}_j - h(\mathbf{X}\underline{\beta}_j)]' \ \mathbf{V}_j^{-1} \ \mathbf{H}(\underline{\beta}_j)\mathbf{X} = 0\,, \tag{3.31}$$

the Fisher scoring iteration is given by

$$\hat{\underline{\beta}}_j^{r+1} = \hat{\underline{\beta}}_j^{r} - \left[\mathbf{X}'\hat{\mathbf{W}}^r\mathbf{X}\right]^{-1} U(\hat{\underline{\beta}}_j^{r}), \quad j = 1, \ldots, k, \tag{3.32}$$

where $\hat{\mathbf{W}}^r = (\hat{w}^r_{jil})$ *and* $w_{jil} = h^{(1)}(\gamma_{ji})\, h^{(1)}(\gamma_{jl})\, v_{jil}$.

Proof: Refer to (3.23) and let it equal to 0. For notational convenience, let

$$U(\underline{\beta}_j) = [\underline{Y}_j - h(\mathbf{X}\underline{\beta}_j)]' \ \mathbf{V}_j^{-1} \ \mathbf{H}(\underline{\beta}_j)\mathbf{X} = \begin{pmatrix} U_1(\underline{\beta}_j) \\ U_2(\underline{\beta}_j) \\ \vdots \\ U_p(\underline{\beta}_j) \end{pmatrix}_{1\times p},$$

and for a fixed $j = 1, \ldots, k$,

$$\gamma_{jl} = \underline{X}_l'\underline{\beta}_j, \quad \text{for } l = 1, 2, \ldots, n. \tag{3.33}$$

Then for $l = 1, 2, \ldots, n$,

$$U_l(\underline{\beta}_j) = \sum_{k=1}^{n} \sum_{i=1}^{n} \left[Y_{ji}\, v_{jik}\, h^{(1)}(\gamma_{jk})\, x_{k,l-1} - h(\gamma_{ji})\, v_{jik}\, h^{(1)}(\gamma_{jk})\, x_{k,l-1} \right], \quad (3.34)$$

and

$$\begin{aligned}
D_{\underline{\beta}_j} U_l(\underline{\beta}_j) &= \sum_{k=1}^{n} \sum_{i=1}^{n} \big\{ Y_{ji}\, v_{jik}\, h^{(2)}(\gamma_{jk})\, x_{k,l-1}\, D_{\underline{\beta}_j} \gamma_{jk} \\
&\quad - h^{(1)}(\gamma_{ji})\, v_{jik}\, h^{(1)}(\gamma_{jk})\, x_{k,l-1}\, D_{\underline{\beta}_j} \gamma_{ji} \\
&\quad - h(\gamma_{ji})\, v_{jik}\, h^{(2)}(\gamma_{jk})\, x_{k,l-1}\, D_{\underline{\beta}_j} \gamma_{jk} \big\} \\
&= \sum_{k=1}^{n} \sum_{i=1}^{n} [Y_{ji} - h(\gamma_{ji})]\, v_{jik}\, h^{(2)}(\gamma_{jk})\, x_{k,l-1}\, D_{\underline{\beta}_j} \gamma_{jk} \\
&\quad - \sum_{k=1}^{n} \sum_{i=1}^{n} \big\{ h^{(1)}(\gamma_{ji})\, v_{jik}\, h^{(1)}(\gamma_{jk})\, x_{k,l-1}\, D_{\underline{\beta}_j} \gamma_{ji} \big\}, \quad (3.35)
\end{aligned}$$

where $h^{(1)}$ and $h^{(2)}$ are the first and second derivative, respectively, of $h = g^{-1}$.

Taking the expectation of (3.35) with respect to the observations $\underline{Y}_i$ yields

$$\begin{aligned}\mathbb{E}\Big\{D_{\underline{\beta}_j}U_l(\underline{\beta}_j)\Big\} &= \sum_{k=1}^{n}\sum_{i=1}^{n}\mathbb{E}[Y_{ji}-h(\gamma_{ji})]\,v_{jik}\,h^{(2)}(\gamma_{jk})\,x_{k,l-1}\,D_{\underline{\beta}_j}\gamma_{jk}\\ &\quad -\sum_{k=1}^{n}\sum_{i=1}^{n}\big\{h^{(1)}(\gamma_{ji})\,v_{jik}\,h^{(1)}(\gamma_{jk})\,x_{k,l-1}\,D_{\underline{\beta}_j}\gamma_{ji}\big\}\\ &= -\sum_{k=1}^{n}\sum_{i=1}^{n}\big\{h^{(1)}(\gamma_{ji})\,v_{jik}\,h^{(1)}(\gamma_{jk})\,x_{k,l-1}\underline{X}_i\big\}. \qquad (3.36)\end{aligned}$$

Arranging the derivatives in (3.35) and (3.36) in matrix form, we can write

$$D_{\underline{\beta}_j}U(\underline{\beta}_j)=\begin{pmatrix} D_{\underline{\beta}_{j0}}U_1(\underline{\beta}_j) & D_{\underline{\beta}_{j0}}U_2(\underline{\beta}_j) & \dots & D_{\underline{\beta}_{j0}}U_p(\underline{\beta}_j)\\ D_{\underline{\beta}_{j1}}U_1(\underline{\beta}_j) & D_{\underline{\beta}_{j1}}U_2(\underline{\beta}_j) & \dots & D_{\underline{\beta}_{j1}}U_p(\underline{\beta}_j)\\ \vdots & \vdots & \dots & \vdots\\ D_{\underline{\beta}_{j,p-1}}U_1(\underline{\beta}_j) & D_{\underline{\beta}_{j,p-1}}U_2(\underline{\beta}_j) & \dots & D_{\underline{\beta}_{j,p-1}}U_p(\underline{\beta}_j)\end{pmatrix}_{p\times p}$$

and

$$\begin{aligned}\mathbb{E}\big[D_{\underline{\beta}_j}U(\underline{\beta}_j)\big] &= -\begin{pmatrix}\sum_{k=1}^{n}\sum_{i=1}^{n}\big\{h^{(1)}(\gamma_{ji})\,v_{jik}\,h^{(1)}(\gamma_{jk})\,x_{k,0}\,\underline{X}_i\big\}\\ \sum_{k=1}^{n}\sum_{i=1}^{n}\big\{h^{(1)}(\gamma_{ji})\,v_{jik}\,h^{(1)}(\gamma_{jk})\,x_{k,1}\,\underline{X}_i\big\}\\ \vdots\\ \sum_{k=1}^{n}\sum_{i=1}^{n}\big\{h^{(1)}(\gamma_{ji})\,v_{jik}\,h^{(1)}(\gamma_{jk})\,x_{k,n-1}\,\underline{X}_i\big\}\end{pmatrix}_{p\times p}\\ &= \mathbf{X}'\mathbf{H}(\underline{\beta}_j)\mathbf{V}_j^{-1}\mathbf{H}(\underline{\beta}_j)\mathbf{X} \qquad (3.37)\end{aligned}$$

Now let

$$\mathbf{W}_j=(w_{jil})=\mathbf{H}(\underline{\beta}_j)\mathbf{V}_j^{-1}\mathbf{H}(\underline{\beta}_j),\quad j=1,\dots,k, \qquad (3.38)$$

where

$$w_{jil} = h^{(1)}(\gamma_{ji})\, h^{(1)}(\gamma_{jl})\, v_{jil}.$$

Similarly let

$$\hat{w}^r_{jil} = h^{(1)}(\underline{X}_i \underline{\hat{\beta}}^r_j)\, h^{(1)}(\underline{X}_l \underline{\hat{\beta}}^r_j)\, v_{jil},$$

and

$$\hat{\mathbf{W}}^r = (\hat{w}^r_{jil}),$$

then from (3.37) the Fisher scoring iteration becomes

$$\underline{\hat{\beta}}^{r+1}_j = \underline{\hat{\beta}}^r_j - \left[\mathbf{X}'\hat{\mathbf{W}}^r\mathbf{X}\right]^{-1} U(\underline{\hat{\beta}}^r_j), \quad j = 1, \ldots, k, \tag{3.39}$$

which converges to the weighted least squares GLM estimator. □

3.5 The linear credibility premium

Our objective is to find a credibility estimator of the mean $\underline{\mu}_j$, called the credibility premium, $\underline{\hat{\mu}}^{(z)}_j = \mathbf{Z}_j h(\mathbf{X}\underline{\hat{\beta}}_j) + (\mathbf{I} - \mathbf{Z}_j) h(\mathbf{X}\underline{b})$, which minimizes

$$Q(\mathbf{Z}_j) = \mathbb{E}\{[\underline{\mu}(\Theta_j) - \underline{\hat{\mu}}_j^{(z)}]'\boldsymbol{\Sigma}[\underline{\mu}(\Theta_j) - \underline{\hat{\mu}}_j^{(z)}]\}. \tag{3.40}$$

To simplify the problem, assume that the weight matrix $\boldsymbol{\Sigma} = \mathbf{I}$, then

$$\begin{aligned} Q(\mathbf{Z}_j) &= \mathbb{E}[\underline{\mu}(\Theta_j) - \underline{\hat{\mu}}_j^{(z)}]'\boldsymbol{\Sigma}[\underline{\mu}(\Theta_j) - \underline{\hat{\mu}}_j^{(z)}] \\ &= \mathbb{E}[h(\mathbf{X}\underline{\beta}_j) - \mathbf{Z}_j h(\mathbf{X}\underline{\hat{\beta}}_j) - (\mathbf{I} - \mathbf{Z}_j)h(\mathbf{X}\underline{b})]' \\ &\quad [h(\mathbf{X}\underline{\beta}_j) - \mathbf{Z}_j h(\mathbf{X}\underline{\hat{\beta}}_j) - (\mathbf{I} - \mathbf{Z}_j)h(\mathbf{X}\underline{b})]. \end{aligned} \tag{3.41}$$

3.5.1 Estimate of the credibility matrix

Theorem 3.2. *If the inverse of* $MSE(\underline{\hat{\beta}}_j)$ *exists, then the credibility matrix* $\mathbf{Z}_j$ *that minimizes* (3.41) *is given by*

$$\begin{aligned} \mathbf{Z}_j &= \left\{ Cov[h(\mathbf{X}\underline{\beta}_j), h(\mathbf{X}\underline{\hat{\beta}}_j)] + [\mathbb{E}[h(\mathbf{X}\underline{\beta}_j)] - h(\mathbf{X}\underline{b})][\mathbb{E}[h(\mathbf{X}\underline{\hat{\beta}}_j)] - h(\mathbf{X}\underline{b})] \right\} \\ &\quad \left\{ MSE(\underline{\hat{\beta}}_j) \right\}^{-1}. \end{aligned} \tag{3.42}$$

Proof: As in (3.33), now for $l = 1, 2, \ldots, n$, let

$$\gamma_l = \underline{X}_l'\underline{b}, \quad \text{and} \quad \xi_l = h(\gamma_l), \tag{3.43}$$

while for any given $j = 1, \ldots, k$,

$$\gamma_{jl} = \underline{X}_l' \underline{\beta}_j, \quad \xi_{jl} = h(\gamma_{jl}) \quad \text{and} \quad \hat{\gamma}_{jl} = \underline{X}_l' \underline{\hat{\beta}}_j, \quad \hat{\xi}_{jl} = h(\hat{\gamma}_{jl}). \tag{3.44}$$

Also let the credibility matrix

$$\mathbf{Z}_j = \begin{pmatrix} z_{j11} & z_{j12} & \cdots & z_{j1n} \\ z_{j21} & z_{j22} & \cdots & z_{j2n} \\ \vdots & \vdots & \cdots & \vdots \\ z_{jn,1} & z_{jn,2} & \cdots & z_{jn,n} \end{pmatrix}_{n \times n}. \tag{3.45}$$

Then, from (3.41) and (3.43)–(3.45) we can write $Q(\mathbf{Z}_j)$ as

$$\begin{aligned} Q(\mathbf{Z}_j) &= \sum_{k=1}^{n} [\xi_{jk} - \xi_k + \sum_{i=1}^{n} z_{jki}(\xi_i - \hat{\xi}_{ji})]^2 \\ &= \sum_{k=1}^{n} \big\{(\xi_{jk} - \xi_k)^2 + [\sum_{i=1}^{n} z_{jki}(\xi_i - \hat{\xi}_{ji})]^2 + 2(\xi_{jk} - \xi_k) \sum_{i=1}^{n} z_{jki}(\xi_i - \hat{\xi}_{ji})\big\}. \end{aligned}$$

Taking the derivative with respect to $\mathbf{Z}_j$, we can rewrite it in compact form

as

$$D_{\mathbf{Z}_j} Q(\mathbf{Z}_j) = \begin{pmatrix} D_{z_{j11}} Q(\mathbf{Z}_j) & D_{z_{j12}} Q(\mathbf{Z}_j) & \ldots & D_{z_{j1p}} Q(\mathbf{Z}_j) \\ D_{z_{j21}} Q(\mathbf{Z}_j) & D_{z_{j22}} Q(\mathbf{Z}_j) & \ldots & D_{z_{j2p}} Q(\mathbf{Z}_j) \\ \vdots & \vdots & \ldots & \vdots \\ D_{z_{jn,1}} Q(\mathbf{Z}_j) & D_{z_{jn,2}} Q(\mathbf{Z}_j) & \ldots & D_{z_{jn,p}} Q(\mathbf{Z}_j) \end{pmatrix}_{n \times n},$$

and

$$
\begin{aligned}
D_{z_{juv}}Q(\mathbf{Z}_j) &= 2\big\{\sum_{i=1}^{n} z_{jui}(\xi_i-\hat{\xi}_{ji})(\xi_v-\hat{\xi}_{jv})+(\xi_{ju}-\xi_u)(\xi_v-\hat{\xi}_{jv})\big\} \\
&= 2\big\{\sum_{i=1}^{n} z_{jui}\xi_i\xi_v-\sum_{i=1}^{n} z_{jui}\xi_i\hat{\xi}_{jv}-\sum_{i=1}^{n} z_{jui}\xi_v\hat{\xi}_{ji}+\sum_{i=1}^{n} z_{jui}\hat{\xi}_{ji}\hat{\xi}_{jv} \\
&\quad +\xi_v\xi_{ju}-\xi_{ju}\hat{\xi}_{jv}-\xi_u\xi_v+\xi_u\hat{\xi}_{jv}\big\}.
\end{aligned}
$$

Then, rearranging terms, we get

$$
\begin{aligned}
\mathbb{E}[D_{z_{juv}}Q(\mathbf{Z}_j)] &= 2\big\{\xi_v\sum_{i=1}^{n} z_{jui}[\xi_i-\mathbb{E}(\hat{\xi}_{ji})]+\sum_{i=1}^{n} z_{jui}[\mathbb{E}(\hat{\xi}_{ji}\hat{\xi}_{jv})-\xi_i\mathbb{E}(\hat{\xi}_{jv})] \\
&\quad +\xi_v[\mathbb{E}(\xi_{ju})-\xi_u]-[\mathbb{E}(\xi_{ju}\hat{\xi}_{jv})-\mathbb{E}(\xi_u\hat{\xi}_{jv})]\big\} \\
&= 2\big\{\xi_v\sum_{i=1}^{n} z_{jui}[\xi_i-\mathbb{E}(\hat{\xi}_{ji})]+\sum_{i=1}^{n} z_{jui}[\mathbb{E}(\hat{\xi}_{ji}\hat{\xi}_{jv})-\mathbb{E}(\hat{\xi}_{jv})\mathbb{E}(\hat{\xi}_{ji})] \\
&\quad +\sum_{i=1}^{n} z_{jui}[\mathbb{E}(\hat{\xi}_{jv})\mathbb{E}(\hat{\xi}_{ji})-\xi_i\mathbb{E}(\hat{\xi}_{jv})]-[\mathbb{E}(\xi_{ju}\hat{\xi}_{jv})-\mathbb{E}(\xi_{ju})\mathbb{E}(\hat{\xi}_{jv})] \\
&\quad -[\mathbb{E}(\xi_{ju})-\xi_u][\mathbb{E}(\hat{\xi}_{jv})-\xi_v]\big\}
\end{aligned}
$$

Now, for any given j, for $l=1,2,\ldots,n$, let

$$
\hat{\zeta}_{jl}=\mathbb{E}(\hat{\xi}_{jl})-\xi_l \quad \text{and} \quad \zeta_{jl}=\mathbb{E}(\xi_{jl})-\xi_l. \tag{3.46}
$$

Then

$$\mathbb{E}[D_{z_{juv}}Q(\mathbf{Z}_j)] = 2\big\{\xi_v \sum_{i=1}^{n} z_{jui}[-\hat{\zeta}_{ji}] + \sum_{i=1}^{n} z_{jui}\mathrm{Cov}(\hat{\xi}_{ji}\hat{\xi}_{jv}) + \sum_{i=1}^{n} z_{jui}\hat{\zeta}_{ji}\mathbb{E}(\hat{\xi}_{ji})$$
$$-\mathrm{Cov}(\xi_{ju}\hat{\xi}_{jv}) - \zeta_{ju}\hat{\zeta}_{jv}\big\}. \quad (3.47)$$

From (3.47), we can write $\mathbb{E}[D_{z_{juv}}Q_{Z_j}] = 0$ in matrix form as

$$\mathbf{Z}_j[-\hat{\underline{\zeta}}_j]\underline{\xi}' + \mathbf{Z}_j\mathrm{Cov}(\hat{\underline{\xi}}_j, \hat{\underline{\xi}}_j) + \mathbf{Z}_j\mathbb{E}(\hat{\underline{\xi}}_j)\hat{\underline{\zeta}}_j' - \mathrm{Cov}(\underline{\xi}_j, \hat{\underline{\xi}}_j) - \underline{\zeta}_j\hat{\underline{\zeta}}_j' = \mathbf{0}, \quad (3.48)$$

hence

$$\begin{aligned}
\mathbf{Z}_j &= [\mathrm{Cov}(\underline{\xi}_j, \hat{\underline{\xi}}_j) + \underline{\zeta}_j\hat{\underline{\zeta}}_j'][\mathrm{Cov}(\hat{\underline{\xi}}_j, \hat{\underline{\xi}}_j) + \mathbb{E}(\hat{\underline{\xi}}_j)\hat{\underline{\zeta}}_j' - \hat{\underline{\zeta}}_j\underline{\xi}']^{-1} \\
&= \big\{\mathrm{Cov}[h(\mathbf{X}\underline{\beta}_j), h(\mathbf{X}\hat{\underline{\beta}}_j)] + [\mathbb{E}[h(\mathbf{X}\underline{\beta}_j)] - h(\mathbf{X}\underline{b})][\mathbb{E}[h(\mathbf{X}\hat{\underline{\beta}}_j)] - h(\mathbf{X}\underline{b})]\big\} \\
&\quad \big\{\mathrm{Cov}[h(\mathbf{X}\hat{\underline{\beta}}_j), h(\mathbf{X}\hat{\underline{\beta}}_j)] + \mathbb{E}[h(\mathbf{X}\hat{\underline{\beta}}_j)][\mathbb{E}[h(\mathbf{X}\hat{\underline{\beta}}_j)] - h(\mathbf{X}\underline{b})]' - \\
&\quad [\mathbb{E}[h(\mathbf{X}\hat{\underline{\beta}}_j)] - h(\mathbf{X}\underline{b})]h(\mathbf{X}\underline{b})'\big\}^{-1} \\
&= \big\{\mathrm{Cov}[h(\mathbf{X}\underline{\beta}_j), h(\mathbf{X}\hat{\underline{\beta}}_j)] + [\mathbb{E}[h(\mathbf{X}\underline{\beta}_j)] - h(\mathbf{X}\underline{b})][\mathbb{E}[h(\mathbf{X}\hat{\underline{\beta}}_j)] - h(\mathbf{X}\underline{b})]\big\} \\
&\quad \big\{\mathrm{MSE}(\hat{\underline{\beta}}_j)\big\}^{-1}. \quad (3.49)
\end{aligned}$$

□

From the above results, it is clear that we need to estimate

$$\mathbb{E}[h(\mathbf{X}\hat{\underline{\beta}}_j)]\,,\ \mathbb{E}[h(\mathbf{X}\underline{\beta}_j)]\,,\ h(\mathbf{X}\underline{b})\,, \tag{3.50}$$

and

$$\mathrm{Cov}[h(\mathbf{X}\hat{\underline{\beta}}_j), h(\mathbf{X}\hat{\underline{\beta}}_j)]\,,\ \mathrm{Cov}[h(\mathbf{X}\hat{\underline{\beta}}_j), h(\mathbf{X}\underline{\beta}_j)] \tag{3.51}$$

to calculate the credibility matrix estimate from (3.42). In most situations, the explicit solution will not exist. We need to resort to the numerical methods or some approximations.

3.5.2 Credibility matrix for identity link

For the identity link function, if the credibility matrix $\mathbf{Z}_j$ applies to the GLM estimated parameter $\hat{\underline{\beta}}_j$ instead of the contract estimator $h(\mathbf{X}\hat{\underline{\beta}}_j)$, then (3.4) can be written as

$$\begin{aligned} Q(\mathbf{Z}_j) &= \mathbb{E}\Big\{[\mathbf{X}\underline{\beta}_j - \mathbf{X}\mathbf{Z}_j\hat{\underline{\beta}}_j - \mathbf{X}(\mathbf{I}-\mathbf{Z}_j)\underline{b}]' \\ &\qquad [\mathbf{X}\underline{\beta}_j - \mathbf{X}\mathbf{Z}_j\hat{\underline{\beta}}_j - \mathbf{X}(\mathbf{I}-\mathbf{Z}_j)\underline{b}]\Big\}, \end{aligned} \tag{3.52}$$

and the result (3.42) from prior section can be simplified as follows.

Theorem 3.3. *If the inverse of* $MSE(\hat{\underline{\beta}}_j)$ *exist, then the credibility matrix* $\mathbf{Z}_j$ *that minimizes* (3.52) *is given by*

$$\begin{aligned} \mathbf{Z}_j &= Cov(\underline{\beta}_j, \hat{\underline{\beta}}_j)\big[Cov(\hat{\underline{\beta}}_j) + Bias(\hat{\underline{\beta}}_j)Bias(\hat{\underline{\beta}}_j)'\big]^{-1} \\ &= Cov(\hat{\underline{\beta}}_j, \underline{\beta}_j)\big[MSE(\hat{\underline{\beta}}_j)\big]^{-1}. \end{aligned} \tag{3.53}$$

Proof: Rewrite $d(\mathbf{Z}_j)$ in (3.52) as

$$d(\mathbf{Z}_j) = \mathbb{E}\Big\{\big[(\underline{\beta}_j - \underline{b}) - \mathbf{Z}_j(\hat{\underline{\beta}}_j - \underline{b})\big]'\big[(\underline{\beta}_j - \underline{b}) - \mathbf{Z}_j(\hat{\underline{\beta}}_j - \underline{b})\big]\Big\},$$

and set the first derivative to zero:

$$\frac{\partial}{\partial \mathbf{Z}_j} d(\mathbf{Z}_j) = -2\mathbb{E}\Big\{\big[(\underline{\beta}_j - \underline{b}) - \mathbf{Z}_j(\hat{\underline{\beta}}_j - \underline{b})\big]'(\hat{\underline{\beta}}_j - \underline{b})'\Big\} = 0\,.$$

Solving gives $\mathbb{E}\big[(\underline{\beta}_j - \underline{b})(\hat{\underline{\beta}}_j - \underline{b})'\big] = \mathbf{Z}_j\,\mathbb{E}\big[(\hat{\underline{\beta}}_j - \underline{b})(\hat{\underline{\beta}}_j - \underline{b})'\big]$, or equivalently $\mathrm{Cov}(\hat{\underline{\beta}}_j, \underline{\beta}_j) = \mathbf{Z}_j\,\mathrm{MSE}(\hat{\underline{\beta}}_j)$ and the result follows. $\square$

If the link function is the identity function, then the GLM partial credibility model becomes Hachemeister's regression model. The credibility matrix in (3.53) is consistent with the result of Hachemeister (1975) as in (3.12).

Remark 3.1. *The above results provide some insight of the credibility of an estimator. Although we cannot interpret the inverse of a matrix as the reciprocal of a real number, (3.53) gives some indications that the "smaller" the MSE of the estimated value $\hat{\underline{\beta}}_j$ the "greater" the credibility matrix $\mathbf{Z}_j$. Similarly the "more" correlated the estimated value $\hat{\underline{\beta}}_j$ is with the real value $\underline{\beta}_j$ the "greater" the credibility matrix, which is consistent with our intuition.*

Furthermore, for large data sets, the properties stated in Lemma 1.1 imply that the MLE $\hat{\underline{\beta}}_j$ for the GLM parameters, as the sample size $n \to \infty$, is such that Bias$(\hat{\underline{\beta}}_j) \to 0$ and Cov$[\hat{\underline{\beta}}_j] \to$ Cov$[\hat{\underline{\beta}}_j, \underline{\beta}_j]$, hence $\mathbf{Z}_j \to \mathbf{I}$. For *smaller data sets, we need to estimate the* $Cov[\underline{\hat{\beta}}_j, \underline{\beta}_j]$, $Cov[\underline{\hat{\beta}}_j]$ *and* $Bias(\underline{\hat{\beta}}_j)$ *to calculate the credibility matrix* $\mathbf{Z}_j$.

3.6 Conclusion

In this chapter, we have extended the GLM estimation to include the use of partial credibility theory. Also, we further relax the assumption for GLMs so that the response variable does not need to follow a distribution in the exponential family. Least squares estimation (LSE) is used instead of maximum likelihood. The numerical algorithm is derived to estimate the GLM parameters by LSE. Finally, the credibility matrix is derived to minimize the sum of squares error (SSE).

Chapter 4

Analysis of fit for the GLM linear credibility estimators

4.1 Introduction

Given certain data and assumptions, in a traditional generalized linear model we need to choose an appropriate form of the linear predictor, a distribution in the exponential family for the response variable, and the link function to combine them. There are various tests and techniques to assess the goodness-of-fit of GLMs. Hardin and Hilbe (2007) gives a very informed summary of the analysis of fit. Fahrmeir and Gerhard (2001) gives complete guidelines for selecting and checking models.

In this chapter, we compare the GLM to the GLM partial credibility model estimators. Here, we assume that the link function and the dependent variables are correctly specified. We focus on the analysis of the models fit rather than on parametrization and calibration of the models. We compare the GLM estimators to the linear credibility weighted estimators. In this chapter, we will answer the question of whether the linear credibility weighted estimators are superior to the GLM estimators.

The question at hand is of theoretical interest. In Chapter 3, we summarized three issues with partial linear credibility theory as stated in Klugman et al. (2004). Most credibility factors used in empirical credibility formulas lack sound theoretical grounds. Even if the credibility factor is estimated by minimizing the sum of squared errors, there is no evidence in the GLM case that the credibility weighted estimator is appropriate and preferable to the

GLM estimators or to the alternate estimator, which is the portfolio average. Little research has been done in evaluating the effectiveness of partial credibility estimators.

We have developed a statistically sound formula in (3.42) to calculate the credibility matrix for partial credibility GLMs. Here, we want to develop an approach to evaluate whether this partial credibility estimator for GLMs can substantially improve the accuracy of the GLM estimators.

The results can also significantly impact GLM modelers in practice. A substantial amount of resources are needed to define the credibility weight and to calculate the credibility matrix. Modelers could spare time and effort if the partial linear credibility estimators for GLMs do not increase predictive power, as judged by a cost-benefit criteria. Efforts could then be directed on improving the estimation of the components of the GLM, such as variable selection, the choice of link function and of distribution for the response variable.

4.2 Model adequacy testing

Many techniques are used in evaluating GLMs, comparing and selecting models. McCullagh and Nelder (1989) gives a comprehensive survey of parametric tests, such as testing the correct form of the variance function by embedding it in a broader parametric class of variance functions. Hardin and Hilbe (2007) and Fahrmeir and Gerhard (2001) explain the analysis of fit statistics, such as residuals, Cook's distance, Akaike information criterion (AIC) and the Bayesian information criterion (BIC), R^2 measures, and other tests. Sen and Chaubey (2011) and Babu and Chaubey (1996) investigate a special case of GLMs, the inverse Gaussian (IG) regression, including the assessment of the IG regression model on gamma errors. McCullagh and Nelder (1989) gives a widely accepted process of statistical analysis of model adequacy in the form

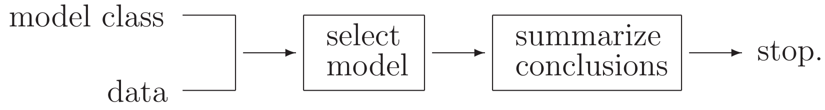

Figure 4.1: Model Checking Process

Here, model class is the GLM and the partial credibility model for GLMs, as defined in Chapter 4. This model checking process assumes that one model is superior to the other. The data is simulated and is explained in more detail in the next section. The "select model" step is a crucial part in this process. Then, based on some statistics, we can summarize conclusions and stop the process.

As explained in McCullagh and Nelder (1989, Chapter 12), model adequacy testing can be either an informal or formal process. Informal techniques rely upon tables and graphs to detect patterns. The argument is that if we can detect patterns in the residuals we can then find a better model.

Such informal methods take a successful model to be the one that has not obvious pattern of unexplained trends in the residuals. The practical problem is that all residuals can seem to exhibit some kind of pattern if we look hard enough. So, one needs to be cautious to avoid over-interpretation. Informal methods are important in model checking especially in practice.

Formal methods rely on some quantitative statistics to measure the goodness of fit. Here, we use the discrepancy between the observed and fitted value for each observation as the statistic for the fit analysis.

The differences between the GLM fitted values and observed values are called residuals. In model assessment, residual analysis is one of the most commonly used techniques. Pierce and Schafer (1986) and Cox and Snell (1968) give excellent and comprehensive surveys of many different definitions for residuals in linear regression. Here we list nine different definitions following those in Hardin and Hilbe (2007). Throughout these formulas, the quantity $\partial\eta/\partial\mu$ is evaluated at $\hat{\mu}$.

1. Response residuals:

$$r_i^R = y_i - \hat{\mu}_i \,. \tag{4.1}$$

   The response residuals are simply the difference between the observed and fitted outcomes.

2. Working residuals:

$$r_i^W = (y_i - \hat{\mu}_i)\big(\frac{\partial\eta}{\partial\mu}\big)_i \,. \tag{4.2}$$

   The working residuals are the difference between the working response and the linear predictor at convergence of the IRLS algorithm in evaluating the solution of the estimating equation.

3. Pearson residuals:

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\mathbb{V}(\hat{\mu}_i)}} \,. \tag{4.3}$$

Pearson residuals rescale the residuals on similar scales of variance. The sum of the squared Pearson residuals is equal to the Pearson chi-squared statistic.

4. Partial residuals:

$$r_{ki}^{T} = (y_i - \hat{\mu}_i)\Big(\frac{\partial\eta}{\partial\mu}\Big)_i + (x_{ik}\hat{\beta}_k)\,, \tag{4.4}$$

where k is the number of predictors. Partial residuals are used to asses the form of predictors and are thus calculated for each predictor.

5. Anscombe residuals:

$$r_i^A = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{\mathbb{V}(\hat{\mu}_i)}}\,, \tag{4.5}$$

where

$$A(\cdot) = \int \frac{d\mu}{\sqrt[3]{\mathbb{V}(\mu)}}\,. \tag{4.6}$$

Anscombe (1972) defines this residual and the choice of the function

$A(\cdot)$ was made so that the resulting residuals would be as normal as possible.

6. Deviance residuals:

$$r_i^D = sign(y_i - \hat{\mu}_i)\sqrt{\hat{d}_i^2}\,, \tag{4.7}$$

The deviance residuals are based on the χ^2 distribution and $\hat{d}_i^2$ depends on the underlying distribution of y_i.

7. Adjusted deviance residuals:

$$r_i^{D_a} = r_i^D + \frac{1}{6}\rho_3(\theta)\,, \tag{4.8}$$

where $\rho_3(\theta)$ is defined for each individual distribution. The adjusted deviance residuals improve on deviance residuals to make the convergence to normal distribution faster.

8. Likelihood residuals:

$$r_i^L = sign(y_i - \hat{\mu}_i)\big\{h_i(r_i^{P'})^2 + (1-h_i)(r_i^{D'})^2\big\}^{1/2}\,. \tag{4.9}$$

Likelihood residuals are a combination of standardized Pearson residuals and standardized deviance residuals.

9. Score residuals:

$$r_i^S = \frac{y_i - \hat{\mu}_i}{\mathbb{V}(\hat{\mu}_i)} \Big(\frac{\partial \eta}{\partial \mu}\Big)_i^{-1}. \tag{4.10}$$

The score residuals are used in calculating the sandwich estimate of variance.

Each of the residuals defined above have their advantages and disadvantages. For instance, Pearson residuals work well in detecting outliers and Anscombe residuals are preferred for the binomial family when the denominator is less than 10. Here, we want to compare the GLM estimators and the linear credibility weighted estimators. We use the response residuals, which is simple and intuitively appealing.

Mallows (1973) discusses the interpretation of C_p, which is used to assess the goodness-of-fit of a regression model estimated by the least squares error method. Mallows' C_p is a goodness-of-fit criteria including the consideration of residuals, sample size and the number of explanatory variables. Mallows' C_p address the issue of overfitting. Ye (1998) provides some other methods in counting the parameters which may lead to different results. In this paper, we follow the standard process in computing the Mallows' C_p. The C_p statistic provided a criterion for selecting among some alternative subset models. The GLM is a subset model of the linear partial credibility model. The linear partial credibility model for GLMs requires an additional estimation on top of the GLMs estimators. These additional parameters should penalize the sum of squares. Here is the formula to compute C_p:

$$C_p = \frac{SSE_p}{S^2} - N + 2p\,, \tag{4.11}$$

where $SSE_p = \sum_{i=1}^{N}(y_i - y_{pi})^2$ is the sum of squares error for the model with p explaining variables, S^2 is the sum of squares error for the full model with K regressors, N is the sample size and K is the number of all available variables.

The C_p value is often used as a stopping rule to select a model from many alternative subset models. Here the GLM is a subset of the linear partial credibility GLM. The expected value of C_p is equal to p. It is suggested that we should choose a model with a C_p value approaching p. Hence, we define the C_p value bias term as

$$C_p \text{ Value Bias Term} = C_p - p\,. \tag{4.12}$$

In the next section, the distribution of the C_p value bias term for each model is studied and compared. The one with a C_p value bias term closer to 0 is seen as a better model.

We use both a residual analysis and Mallows' C_p to assess and compare the model based on GLMs and the linear partial credibility GLM.

4.3 A numerical illustration

Now, with the above understanding of the model checking process, we compare the GLM estimators and the linear partial credibility estimators for GLMs. We use a numerical example to illustrate how the process works and summarize the conclusions of the model selection.

4.3.1 The testing process

Refer to Figure 4.1, the testing is based on simulated data. The idea is to simulate the complete modelling process, first by GLMs and then by linear credibility weighted estimation. Then, the sum of squares error (SSE) is calculated for both approaches to help select the model. We assess the two models by comparing their SSEs. We believe that if the model based on the credibility estimators is to be considered superior then it should lead to smaller SSEs. Linear credibility has been proposed as a way to reduce the variance of mean estimators for small sample risk classes. In a GLM setting, it is difficult to asses the variance of the linear credibility GLM estimator analytically. That is why we resort to simulated values and an SSE criterion.

Here are the twelve steps used in this simulation approach:

Step 1: Define the parameters. In this portfolio we simulate four contracts, that is $j = 1, \ldots, 4$, or $k = 4$. Each contract has 20 individuals, that is $n = 20$. Each individual has three independent variables, that means $p = 4$. For each individual, we simulate 25 times the response $\tilde{y}^l_{mj}$, that is $l = 1, \ldots, 25$. The link functions are selected to be the log and identity functions.

Step 2: Simulate the covariate matrix $\mathbf{X}$, which is a $n \times p$ matrix (here 20×4) with all entries in the first column equal to 1.

Step 3: Select a portfolio mean parameter vector $\underline{b}$. In this example, $\underline{b}$ is a $p = 4$ dimensional vector.

Step 4: For each $j = 1, \ldots, 4$ and $l = 1, \ldots, 25$, simulate the contract specific GLMs parameter $\underline{\beta}_j$ as $\tilde{\underline{\beta}}_j^l$, which is a four dimensional vector.

Step 5: For each given $j = 1, \ldots, 4$ and $l = 1, \ldots, 25$, simulate $n = 20$ error terms corresponding to each individual as $\tilde{\varepsilon}_{mj}^l$, for $m = 1, \ldots, 20$.

Step 6: For each given $j = 1, \ldots, 4$, $l = 1, \ldots, 25$, and $m = 1, \ldots, 20$, calculate the simulated observed response variable as

$$\tilde{y}_{mj}^l = g^{-1}(\underline{X}_{mj}' \cdot \tilde{\underline{\beta}}_j^l) + \tilde{\varepsilon}_{mj}^l \, . \tag{4.13}$$

Step 7: For each $j = 1, \ldots, 4$, and $l = 1, \ldots, 25$ we have the simulated response variable from Step 6 and the design matrix from Step 2. We can apply the LSE algorithm developed in Section 3.3.2 to estimate the GLM parameter as $\hat{\underline{\beta}}_j^l$.

Step 8: For each $j = 1, \ldots, 4$, we will have 25 pairs of parameters as $\hat{\underline{\beta}}_j^l$ and $\tilde{\underline{\beta}}_j^l$, respectively for the GLM estimated and simulated parameters. Then, we estimate

$$\mathbb{E}[h(\mathbf{X}\hat{\underline{\beta}}_j)]\,,\ \mathbb{E}[h(\mathbf{X}\underline{\beta}_j)]\,,\ h(\mathbf{X}\underline{b})\,, \tag{4.14}$$

and

$$\mathrm{Cov}[h(\mathbf{X}\hat{\underline{\beta}}_j), h(\mathbf{X}\hat{\underline{\beta}}_j)]\,,\ \mathrm{Cov}[h(\mathbf{X}\hat{\underline{\beta}}_j), h(\mathbf{X}\underline{\beta}_j)]\,, \tag{4.15}$$

which are used to calculate the credibility matrix $\mathbf{Z}_j$ as in (3.42). We estimate four credibility matrices $\mathbf{Z}_j$ for each $j = 1, \ldots, 4$.

Step 9: From Step 7, for each $j = 1, \ldots, 4$, $l = 1, \ldots, 25$, and $m = 1, \ldots, 20$, we calculate the GLM estimator of the response as

$$\hat{y}_{mj}^l = g^{-1}(\underline{X}'_{mj} \cdot \hat{\underline{\beta}}_j^l)\,. \tag{4.16}$$

Step 10: From Step 8 and Step 9, for each $j = 1, \ldots, 4$, and $l = 1, \ldots, 25$, we calculate the credibility weighted estimator of the response as

$$\underline{\hat{y}}_j^{(z)l} = \mathbf{Z}_j h(\mathbf{X}\underline{\hat{\beta}}_j^l) - (\mathbf{I} - \mathbf{Z}_j)h(\mathbf{X}\underline{b})\,. \tag{4.17}$$

Here, for notational convenience, we write in matrix form. Each element in $\underline{\hat{y}}_j^{(z)l}$ is $\hat{y}_{mj}^{(z)l}$.

Step 11: Step 6 simulates the observed value as $\tilde{y}_{mj}^l$. From Step 9 and Step 10, we have the GLM estimated responses $\hat{y}_{mj}^l$ and the credibility weighted responses $\hat{y}_{mj}^{(z)l}$ respectively. The sum of squares errors can be calculated as

$$\text{SSE (GLM)} = \sum_{m,j,l}(\hat{y}_{mj}^l - \tilde{y}_{mj}^l)^2\,, \tag{4.18}$$

and

$$\text{SSE (Credibility)} = \sum_{m,j,l}(\hat{y}_{mj}^{(z)l} - \tilde{y}_{mj}^l)^2\,. \tag{4.19}$$

Assess the model based on the difference of the SSE (GLM) and SSE

(Credibility). The SSE ratio is the SSE from GLM estimators divided by the SSE from partial linear credibility estimators. That is (4.19) divided by (4.18) as

$$\text{SSE Ratio} = \frac{\text{SSE(GLM)}}{\text{SSE(Credibility)}} = \frac{\sum_{m,j,l}(\hat{y}^l_{mj} - \tilde{y}^l_{mj})^2}{\sum_{m,j,l}(\hat{y}^{(z)l}_{mj} - \tilde{y}^l_{mj})^2}. \qquad (4.20)$$

Step 12: Repeat Step 1 to Step 11 many times and study the distribution pattern of the SSE Ratios. In this example, we repeated 200 times Steps 1 to 11. If the SSE Ratios are significantly larger than 100% most of the time, that can provide an indication that the linear credibility estimators are better than the GLM estimators. Otherwise, said, linear credibility is not a good approach to further improve GLM estimators if these SSE Ratios are too small.

4.3.2 The testing results

Intuitively, the credibility weighted estimators should give a better fit. However, to apply partial linear credibility theory to GLM estimators is a relatively new idea. We should be prudent and carry out some tests before we use this approach. As in Section 3.1, partial credibility theory itself has some issues here. Also, as discussed in Section 1.4, the estimation of the GLM parameters is not unbiased. Cordeiro and McCullagh (1991) and Neuhaus and Jewell (1993) explore this problem. Hence, we are now in a position to verify if the partial linear credibility approach necessarily improves the fit or not.

To have a complete and robust test, we simulated six different combinations of link functions and error terms. The three link functions are log, identity and inverse functions. The two error terms are simulated from normal and Poisson distributions. For each of the six combination, 25 year-to-year pattern are simulated for each of 80 individuals divided in 4 contracts 200 times. The SSE ratios are calculated for each of these 200 times simulation. The histogram and distribution of these 200 SSE ratios are studied to check the pattern. The year-to-year SSE ratio patterns are also investigated.

Table 4.1 summarizes the distribution pattern of the SSE ratios obtained from the model checking process described in the prior section for log link and normal error simulations.

SSE Ratio	Number of Simulations in the Range	Percentage of Simulations in the Range
SSE Ratio < 90%	14	7.0%
90% ≤ SSE Ratio < 98%	13	6.5%
98% ≤ SSE Ratio < 102%	53	26.5%
102% ≤ SSE Ratio < 110%	93	46.5%
SSE Ratio ≥ 110%	27	13.5%
Total	200	100%

Table 4.1: SSE Ratio Distribution

From Table 4.1, it is clear that only 13.5% SSE ratios are larger than 110%, and more than 40% are less than 102%. Which leads us to conclude that from the distribution of SSE ratios, the linear credibility estimators are not a significant improvement over the GLM estimators. That is, linear credibility theory does not necessarily reduce the variance of GLM estimators.

Figures 4.2–4.7 are histogram and distributional graphs of the SSE ratio for those six combinations, which give a visual illustration of comparison of the GLMs and the linear partial credibility model for GLMs.

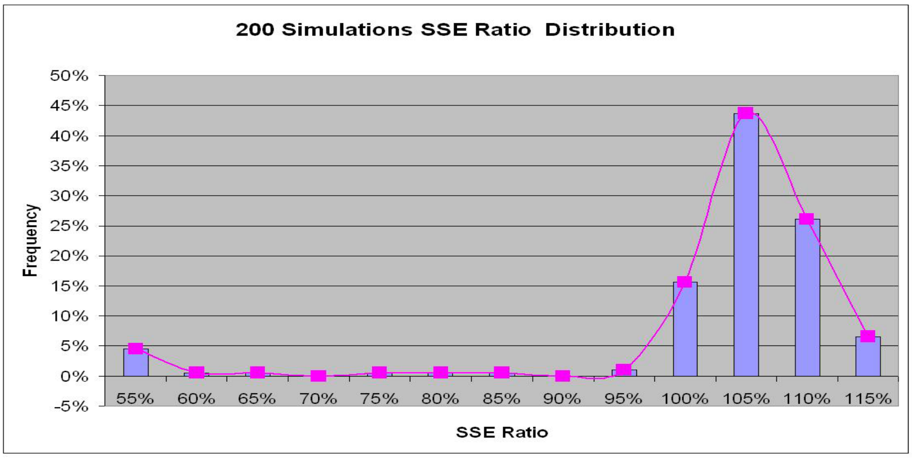

Figure 4.2: Histogram of SSE Ratios for Log Link and Normal Error

If the linear partial credibility model were significantly superior to the GLMs, then we would expect to see from Figures 4.2 to 4.7 that the SSE

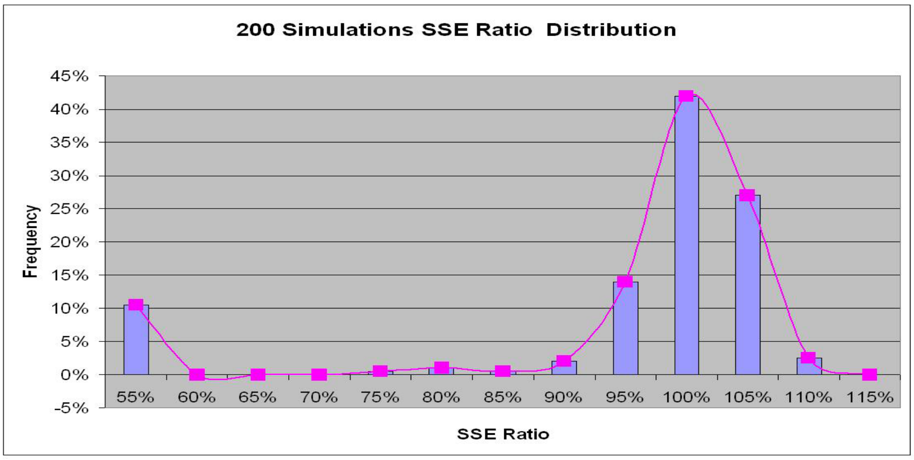

Figure 4.3: Histogram of SSE Ratios for Log Link and Poisson Error

ratios are significantly larger than 100%. However, from the six histograms and distributional graphs, the SSE ratios do not show such a pattern in any of the six combinations. Most of the SSE ratios are around 100% and there are large portions of the SSE ratios that are less than 100%.

Figure 4.8 shows the year-to-year patterns of the SSE ratios for a each of the four contracts. From Figure 4.8, we can see that there are no obvious patterns to the year-to-year SSE ratios.

We also checked the average SSE ratio instead of the ratio of the SSE.

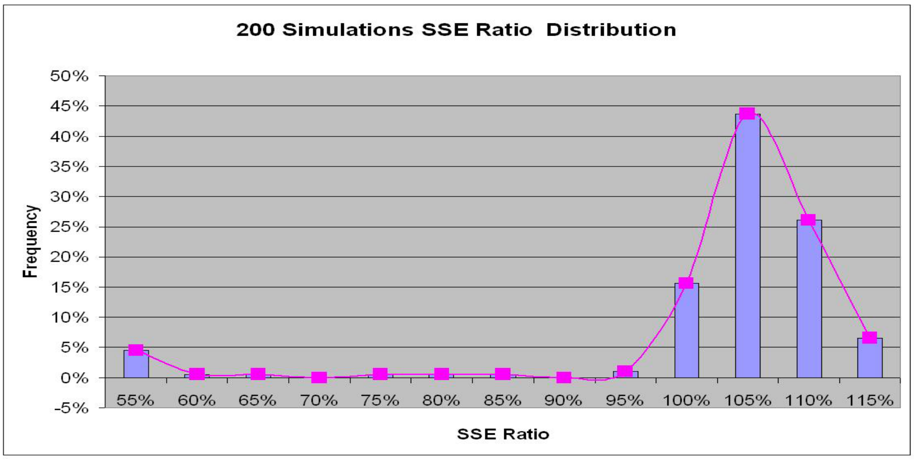

Figure 4.4: Histogram of SSE Ratios for Identity Link and Normal Error

We modify (4.18) as the ratio of the sum of the SSE ratios for each contract instead of the sum of squares errors for all contracts. There is no significant difference between these two ratios. Hence (4.18) is sufficient to compare the two models.

Mallows' C_p is also used to assess the fit of the GLMs and the linear partial credibility model for GLMs. We follow the definition in equations (4.11) and (4.12). Figures 4.9 and 4.10 display the C_p value bias term distribution as defined in (4.12). Clearly, the GLMs give a beautiful bell shape

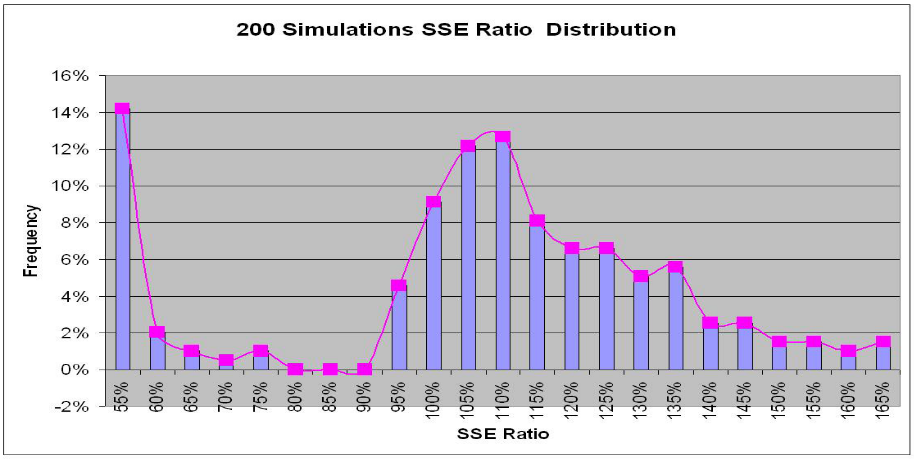

Figure 4.5: Histogram of SSE Ratios for Identity Link and Poisson Error

distribution of C_p values around 0 and the linear partial credibility model from GLMs produce significantly more biased values, which further supports the conclusion that the linear partial credibility estimators for GLMs are not necessarily better than the least squares estimators of the GLM parameters.

To further illustrate Mallows' C_p assessment, Figure 4.11 gives a residual plot of the C_p values against p. The Mallows' C_p computed as in (4.11) was plotted against p for each simulation, which forms a residual graph. The more concentrated the C_p values to 0 the better the model. Here the two

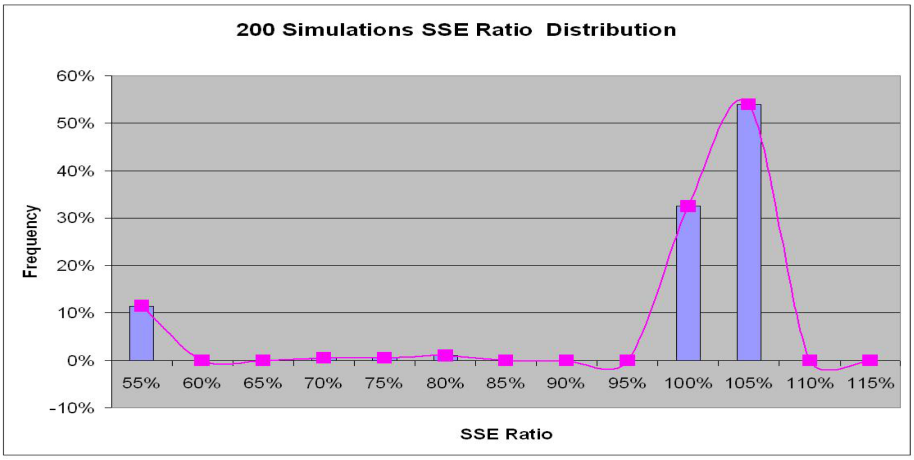

Figure 4.6: Histogram of SSE Ratios for Inverse Link and Normal Error

models we are comparing are the best GLM and the linear partial credibility GLM derived based on the best GLM and Chapter 3 results. Clearly, we can see that the C_p values are much more concentrated for the GLMs than for the credibility estimators. This further supports the conclusions from Figures 4.9 and 4.10.

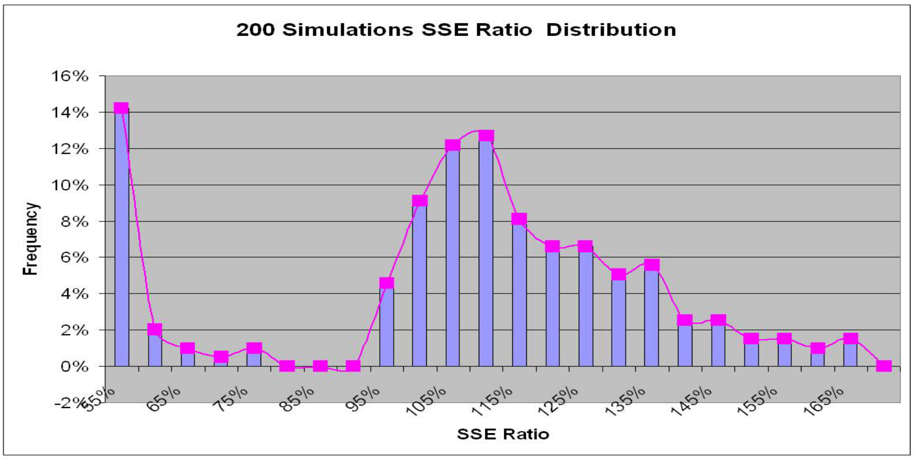

Figure 4.7: Histogram of SSE Ratios for Inverse Link and Poisson Error

4.4 Conclusion

In this chapter, we answer the question whether the linear credibility weighted estimator is superior to the GLMs estimator through a simulation study. Following a brief introduction to the analysis of goodness-of-fit, we define a simulation algorithm to compare these two approaches. A simulator coded in Visual Basic for Applications (VBA) is implemented under an Excel platform. Five thousand simulations are investigated for each of 80 individuals

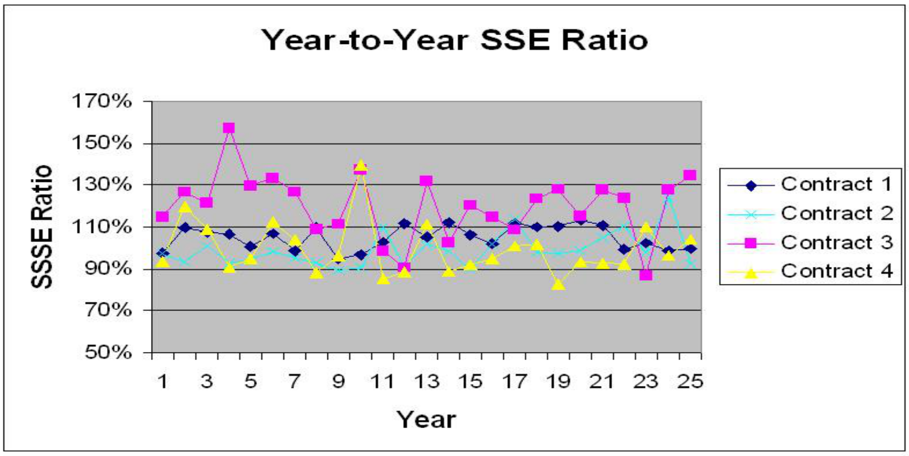

Figure 4.8: Year to Year SSE Ratios Pattern

divided in 4 contracts. The results show that credibility weighted mean estimators are not useful to further improve the accuracy (variance/SSE) of GLM estimators.

The numerical examples illustrate that in most cases for the GLM estimators, the linear partial credibility estimator can only slightly reduce the SSE, while in the 40% other cases, the linear partial credibility estimator is not even as good as the GLM estimators.

This result can be theoretically interesting, but is definitely very useful in

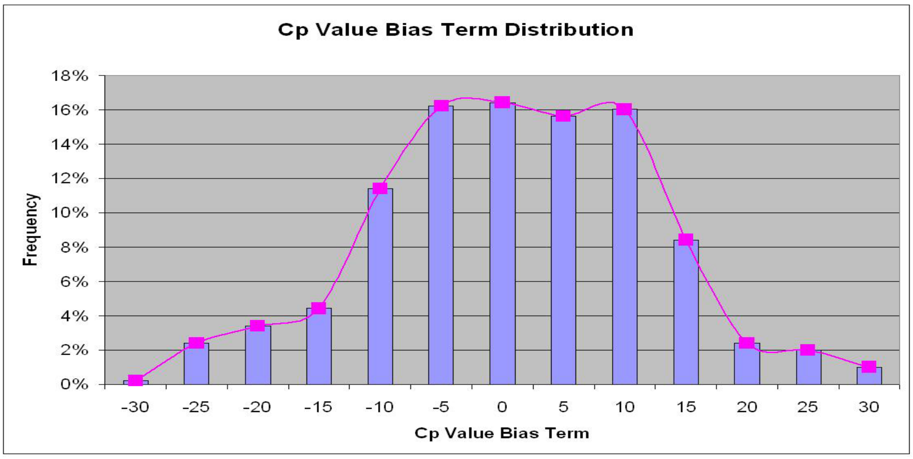

Figure 4.9: Bias of C_p Values for GLMs (Log Link and Normal Error)

practice. This gives a guidance for GLM modelers. They should focus on improving the GLM fit instead of using resources to calculate linear credibility weighted estimators that do not significantly improve the GLM results.

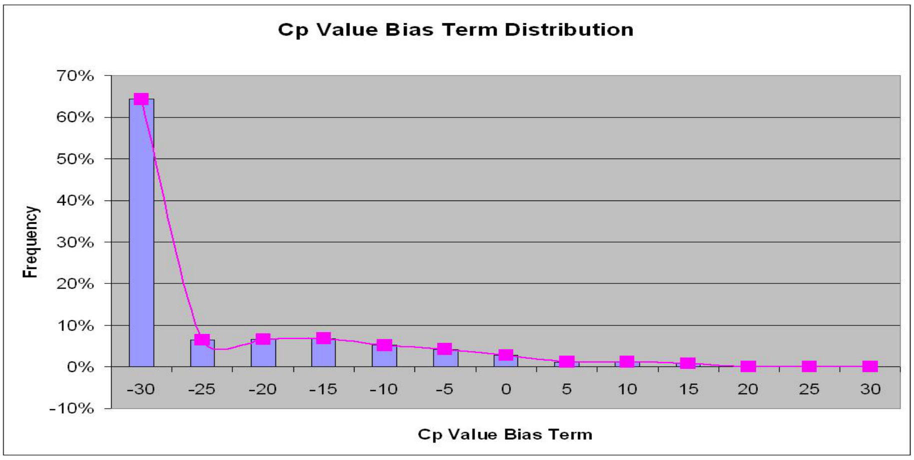

Figure 4.10: Bias of C_p Values for Credibility (Log Link and Normal Error)

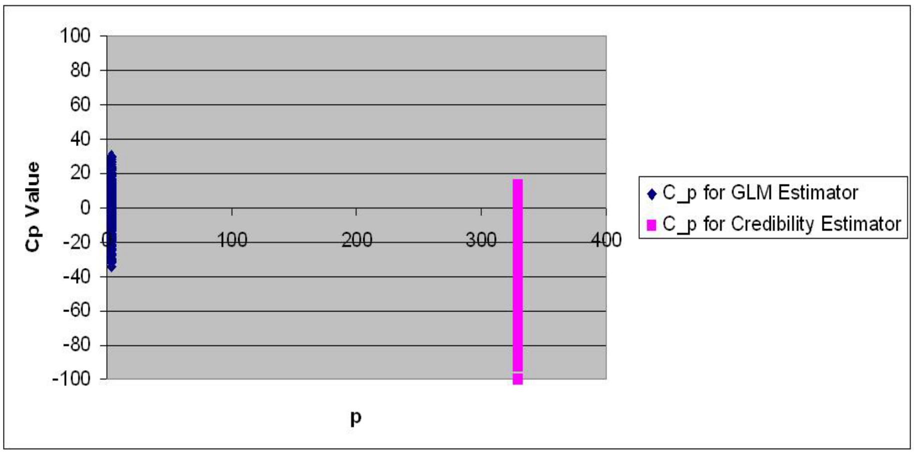

Figure 4.11: C_p vs p for GLMs and Credibility (Log Link – Normal Error)

Chapter 5

A Loss Reserving Method Based on GLMs

5.1 Introduction

Loss reserving is one of the most challenging actuarial tasks. Numerous approaches have been developed to give reasonable estimates. Wiser et al. (2001) provides a detailed introduction to loss reserving. Schmidt (2006) gives a unifying survey of some of the most important methods and models in loss reserving that are based on loss development triangles. Wüthrich and Merz (2008) give a handbook of stochastic claims reserving methods in insurance.

Generalized linear models (GLMs) have become a popular statistical analysis method also for loss reserving data. Haberman and Renshaw (1996) gives a comprehensive review of the application of GLMs to actuarial problems, including loss reserving. Taylor and McGuire (2004) reviews loss reserving models with GLMs. Hoedemakers et al. (2005) constructs bounds for the discounted loss reserves within the framework of GLMs. Verrall (2004) uses a Bayesian parametric model based on GLMs to estimate reserves. Venter (2007) extends generalized linear models beyond the exponential family and gives loss reserve applications.

Most of the above papers propose aggregate reserving methods based on loss development triangles, which do not use individual information to the actual claims processes in each risk class. Frees et al. (2009) demonstrates actuarial applications of modern statistical methods that are applied to detailed, micro-level insurance data. In this chapter we establish a more complex structural reserving method that uses more detailed information at the individual risk class level, such as the premium exposure emergence pattern, or the loss emergence and development patterns within risk classes, and embeds them in the framework of GLMs. This approach has the following advantages:

- In theory it should be more accurate than aggregate loss reserving methods based on loss development triangles, as more detailed information is used.
- It is more flexible in dealing with unusual or quickly changing situations, as variables are analyzed continuously rather than discretely.
- A discount factor can be added and the model adjusted easily.
- It provides a mechanism to analyze separately the effect of each loss

reserving factor.

- The model connects the frequency and severity estimations, both in ratemaking and loss reserving, making the work of actuaries more consistent and easier to interpret.

5.2 A Loss Reserving Model Based on GLMs

This section gives a detailed description of our GLM loss reserving method, starting with the definition of the loss, loss development and loss reserve functions, both with and without discounting.

The *loss function* $l(t)$ defines a stochastic process representing the rate at which losses occur at time t. This loss function l tells us how the in–force risk exposure, and the seasonality in the distribution of risk exposures, determine losses. A detailed treatment of exposure bases is given in Bouska (1989).

In practice we cannot observe $l(t)$ directly. However, we can approach the expected value of $l(t)$, through in–force exposures, as it directly depends on the premium emergence pattern. Then we can define the *aggregate loss* $L(t_1, t_2)$ occurred over the time period (t_1, t_2):

$$L(t_1, t_2) = \int_{t_1}^{t_2} l(t)\, dt\,. \tag{5.1}$$

In conjunction, the *loss development function* $D(t)$ forms a stochastic process which represents the percentage of losses that are paid within t years after their occurrence. It is clear that $D(t) = 0$ for $t \leq 0$, and $\lim_{t\to\infty} D(t) = 1$, almost surely. For a given time $T > t$, then $l(t)D(T - t)$ represents the aggregate paid amount at T, for losses that occurred at time t (see Figure 5.1 for a time–diagram). Assuming that the process has continuous sample paths, by integrating these aggregate paid amounts over (t_1, t_2), we can then define the aggregate paid losses from claims incurred in period (t_1, t_2), as developed to time $T \geq t_2 > t_1$:

$$L(T, t_1, t_2) = \int_{t_1}^{t_2} l(t)\, D(T - t)\, dt\,. \tag{5.2}$$

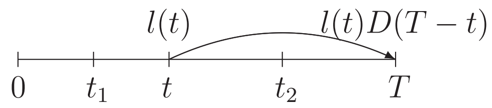

Figure 5.1: Time Scale

The integrals in (5.1) and (5.2) give the ultimate losses and the paid losses incurred in period (t_1, t_2). Their difference, $L(t_1, t_2) - L(T, t_1, t_2)$, represents the unpaid losses, or also called loss reserves.

Consequently, given a loss function l and a loss development function D, the *loss reserves* for claims incurred in period (t_1, t_2), as developed to time $T \geq t_2 > t_1$ are defined as:

$$\begin{aligned} R(T) &= L(t_1, t_2) - L(T, t_1, t_2) \\ &= \int_{t_1}^{t_2} l(t)\, [1 - D(T - t)]\, dt\,. \end{aligned} \tag{5.3}$$

In the case of discounted reserves, we need to add a discount factor in the above analysis. Let $\delta(t)$ be the stochastic force of interest at time t. Again, assuming continuous sample paths, $B(t) = \int_0^t \delta(s)\, ds$ defines the aggregate interest rate in the period of $(0, t)$, while more generally, $B(T+t) - B(t) = \int_T^{T+t} \delta(s)\, ds$ is the aggregate interest rate over $(T, T+t)$.

In the discounted case, loss reserves are no longer obtained by difference. Instead, first consider a fixed time t, where $t_1 < t < t_2$, at which we incur losses at rate $l(t)$. Then $l(t)\, d(s-t)\, ds$ of these will develop at future instant $s > t$, where we assume that $d(t) = D'(t)$, almost surely. Hence, the discounted value at an evaluation date T in (t, s) (see Figure 5.2 for a time–diagram) is given by $e^{-[B(s)-B(T)]}\, l(t)\, d(s-t)\, ds$. Finally, integrating over all future development times $s \in (T, \infty)$ yields the discounted value at time T

of the unpaid loss reserves from period (t_1, t_2), as developed to $T \geq t_2 \geq t_1$:

$$Z(T) \;=\; \int_{t_1}^{t_2} l(t) \int_{T}^{\infty} e^{-[B(s)-B(T)]}\, d(s-t)\, ds\, dt\,, \tag{5.4}$$

almost surely.

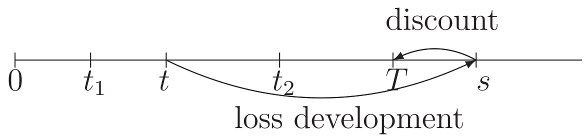

Figure 5.2: Evaluation Time T

Equations (5.3) and (5.4) give the formulas for the loss reserves and *discounted loss reserves*, respectively. In fact, note that when the aggregate interest rate $B(t) = 0$, i.e. no discounting, then (5.4) reduces to (5.3), since $\lim_{t\to\infty} D(t) = 1$, almost surely.

To conclude the definition of this loss reserving model, introduce the following assumption to calculate the expected value of the processes in (5.3) and (5.4).

(A1) The loss function l, the loss development function D and the force of interest δ are independent processes.

Assumption (A1) directly implies the following results for the expected loss reserves and discounted loss reserves: given $\mathbb{E}\big[l(t)\big]$, $\mathbb{E}\big[D(t)\big]$ and $\mathbb{E}\big[e^{-B(t)}\big]$, for fixed t, then

$$\mathbb{E}\big[R(T)\big] = \int_{t_1}^{t_2} \mathbb{E}\big[l(t)\big]\,\big\{1-\mathbb{E}\big[D(T-t)\big]\big\}\,dt\,, \qquad T \geq t_2 > t_1\,, \tag{5.5}$$

$$\mathbb{E}\big[Z(T)\big] = \int_{T}^{\infty} \mathbb{E}\big\{e^{-[B(s)-B(T)]}\big\} \int_{t_1}^{t_2} \mathbb{E}\big[l(t)\big]\,\mathbb{E}\big[d(s-t)\big]\,dt\,ds\,. \tag{5.6}$$

It is clear that the loss reserves only depend on the expected loss function and loss development function. We use GLMs to estimate these two functions.

In practice, loss and development functions can be very complex. Even when long historical data is available, the development process itself can change with time. The future force of interest is also unknown. In addition to the assumption (A1) above, the following simplifying assumptions are used henceforth:

(A2) All policy periods are one year and the amount of exposure to risk of an insurance policy spreads uniformly over the policy period.

(A3) The expected value of the loss development function D is of the exponential form $\mathbb{E}\big[D(t)\big] = 1 - a^{-t}$, where $a > 1$ is a constant.

(A4) The future force of interest is a known constant $\delta \geq 0$, that is $B(t) = \delta t$ almost surely.

The average settlement time is a key parameter for loss development. Based on Assumption (A3), above, we can estimate this parameter within the framework of GLMs.

Given Assumption (A3), the expected average loss development time is given by:

$$\mathbb{E}(\tau) = \int_0^\infty \mathbb{E}\big[1 - D(t)\big]\, dt = \int_0^\infty \frac{dt}{a^t} = \frac{1}{\ln a}\,, \qquad \text{for } a > 1\,. \tag{5.7}$$

The motivation for assumption (A1) is to estimate the expected values of the loss function, $\mathbb{E}\big[l(t)\big]$, of the loss development function, $\mathbb{E}\big[D(t)\big]$, and of the discount factor $\mathbb{E}[e^{-B(t)}\big] = e^{-\delta t}$, separately. These can then be substituted into (5.5) and (5.6) to estimate the expected loss and discounted reserves.

Assumption (A2) can be relaxed for seasonality or other distributional patterns.

Assumption (A3) states that $\mathbb{E}[D(t)]$ takes the form of the cumulative distribution function (CDF) of an exponential distribution, which is appropriate for high–frequency/low–severity business lines, such as auto insurance. For heavy tail cases such as liability claims, the CDF of a Weibull or Pareto distribution may be more adequate models for the loss development function.

Under these assumptions, we can estimate the expected value of the loss function $l(t)$ and the loss development function $D(t)$ within the framework of GLMs. A key aspect is to model the number of claims n and the claim severity as independent responses of separate GLMs.

Consider a set of observed claims under some risk classification system. Let cell $i = 1, \dots, k$ denote a generic risk class defined by this system. The GLMs for frequency and severity can be written as follows. Let

- f_i be the claim frequency, z_i the claim severity and τ_i the average settlement time, respectively, in cell $i = 1, \dots, k$,

- $w_i(t)$ be the number of exposure units (e.g. policyholders) in cell $i = 1, \dots, k$, at time t,

- η_{f_i}, η_{z_i} and η_{τ_i} be the linear predictors of claim frequency, severity and average settlement time in cell $i = 1, \dots, k$, respectively.

- g_f, g_z and g_τ be the GLM link functions for the claim frequency, severity and average settlement time, respectively.

Then the GLMs give the expected value of the claim frequency, severity and average settlement time, for each cell $i = 1, \ldots, k$, as:

$$\mathbb{E}(f_i) = g_f^{-1}(\eta_{f_i}), \quad \mathbb{E}(z_i) = g_z^{-1}(\eta_{z_i}) \quad \text{and} \quad \mathbb{E}(\tau_i) = g_\tau^{-1}(\eta_{\tau_i}). \tag{5.8}$$

Combining (5.8) with Assumption (A3) and with (5.7) gives:

$$g_\tau^{-1}(\eta_{\tau_i}) = \frac{1}{\ln a_i} \quad \Rightarrow \quad a_i = \exp\left\{\frac{1}{g_\tau^{-1}(\eta_{\tau_i})}\right\}, \qquad i = 1, \ldots, k.$$

Now with Assumption (A2), we get that the expected total loss rate $\mathbb{E}\big[l_i(t)\big]$ and loss development function $\mathbb{E}\big[D_i(t)\big]$ in cell i at time t are:

$$\mathbb{E}\big[l_i(t)\big] = w_i(t)\, g_f^{-1}(\eta_{f_i})\, g_s^{-1}(\eta_{s_i}), \qquad i = 1, \ldots, k, \tag{5.9}$$

$$\mathbb{E}\big[D_i(t)\big] = \exp\left\{\frac{1}{g_\tau^{-1}(\eta_{\tau_i})}\right\}, \qquad i = 1, \ldots, k. \tag{5.10}$$

Then (5.5) and (5.6) give the expected loss and discounted loss reserves in cell i:

$$\mathbb{E}\big[R_i(T)\big] = \int_{t_1}^{t_2} l_i(t)\left\{1 - \mathbb{E}\big[D_i(T-t)\big]\right\} dt, \qquad i = 1, \ldots, k, \tag{5.11}$$

$$\mathbb{E}\big[Z_i(T)\big] = \int_T^\infty e^{-\delta(s-T)} \int_{t_1}^{t_2} \mathbb{E}\big[l_i(t)\big]\,\mathbb{E}\big[d(s-t)\big]\, dt\, ds. \tag{5.12}$$

Hence, summing over all cells in the portfolio we have the total loss and discounted loss reserves:

$$\mathbb{E}\big[R(T)\big] = \sum_{i=1}^{k} \mathbb{E}\big[R_i(T)\big]\,, \tag{5.13}$$

$$\mathbb{E}\big[Z(T)\big] = \sum_{i=1}^{k} \mathbb{E}\big[Z_i(T)\big]\,. \tag{5.14}$$

5.3 Numerical Example

There are numerous methods used to estimate loss reserves, such as the Chain Ladder (CL) method, Bornhuetter–Ferguson (B–F), Percent of Premium, Pegged Loss, etc. Each of these methods presents advantages and disadvantages over the others, and there is no generally accepted best method. The CL and B–F methods are the most widely used in the industry. Actuaries keep on investigating the properties of these methods, for instance Wüthrich et al. (2008) recently provided upper and lower bounds on the estimation error in the chain ladder method.

For comparison purposes, we call the method proposed in the previous section the GLM–Reserving (G–R) method. The following examples illustrate three of these methods, CL, B–F and G–R, with some real industry data.

5.3.1 Accident Benefit Coverage Data from a Leading Property and Casualty Company in Canada

According to the definition of the Financial Services Commission of Ontario (FSCO), Statutory Accident Benefit Coverage (AB) “provides you with benefits if you are injured in an automobile accident, regardless of who caused the accident, including supplementary medical, rehabilitation, attendant care, caregiver, non-earner and income replacement benefits. Benefits are also provided to passengers and pedestrians who do not have their own policy under which to claim.” Here, we will study the AB coverage for Accident Years (AY’s) 2003–2005.

The empirical average settlement lag is 0.4696 year or 5.6352 months. A detailed examination of the data shows that most losses were paid within 3 years after the accident. To make a consistent comparison, we first provide the projected loss development triangles using the CL, B–F and G–R methods, and then compare them to the observed triangles with the real data, as these have now developed to year 2008. By comparing the estimated triangle given by each model with the real triangle, we evaluate the goodness–of–fit of the models through cross–validation.

For the loss development triangles, each row represents a given Accident Quarter (AQ). Each row data represents a fixed group of claims defined by the accident date (in quarters). The columns thus keep track of the losses at subsequent evaluation dates for an individual accident quarter. Each column represents the loss development age by accident quarter. Hence, in total we have twelve rows and twelve columns that represent twelve accident quarters of loss experience and twelve quarters of settlement ages.

5.3.2 G–R Method

The following steps summarize the implementation of the GLM loss reserving model described in (5.8)–(5.14). Log–link functions g_f and g_z where chosen for the frequency and severity GLMs, respectively, as this is the most usual in the auto insurance industry. Then a Poisson distribution was chosen to model the loss frequency and a gamma for the loss severity, again because these are industry standards for the portfolio at study. All parameter estimations for the GLMs components were carried out is SAS using the GENMOD procedure.

- Step 1: build a log–Gamma GLM to project the severity for each risk;
- Step 2: build a log–Poisson GLM to project the frequency for each risk;
- Step 3: combine the results from Steps 1 and 2, followed by a realloca-

tion to get the pure premium for each risk;

- Step 4: for any given risk, if there is a claim, project the expected claim settlement lags using a log–Weibull GLM;

- Step 5: for a given book of business, project the loss payment pattern.

Here Steps 1 to 3 produce standard GLMs for the pure premium. Step 4 gives an estimated settlement lag for any risk in the portfolio given there is an accident. With the expected loss and expected settlement lag for each loss, in Step 5 we can produce the expected incremental loss development triangle, as in the lower right part of Table A.1 in the Appendix.

5.3.3 C–L and B–F Methods

Applying the Chain Ladder and Bornhuetter–Ferguson methods in the usual way, we obtain the projected incremental loss development triangles illustrated in the lower right part of Tables A.2 and A.3 respectively.

5.3.4 Comparison of Methods

Table 5.1 summarizes the comparison between the empirical loss development and the projected development from the G–R, C–L and B–F methods. A more detailed comparison, per quarter, is given in Table A.5.

Empirical	C–L/Emp	C–L - Emp	B–F/Emp	B–F - Emp	G–R/Emp	G–R - Emp
Total	30.39%	763,018	-31.19%	-783,073	0.59%	14,907
St.Dev.	46.63%	217,054	32.73%	251,936	39.67%	62,828
Max	128.54%	763,018	65.91%	38,347	16.43%	179,072
Min	-64.61%	-15,104	-69.29%	-783,073	-94.01%	-55,477
Mean	32.33%	69,365	-27.93%	-71,188	3.78%	1,355

Table 5.1: Comparison of Empirical/Projected Development Triangles

The first row of Table 5.1 gives the relative differences (in percentage) and the raw differences (in Dollars) between the total projected loss reserves obtained by the C–L, B-F and G–R methods, and the total empirical losses, as ultimately developed. The G–R method gives the smallest relative difference, at only 0.59%, while the C–L method over–estimates the loss reserve by more than 30% and the B–F method under–estimates it by more than 30%. These differences are also illustrated graphically in Figures 5.3–5.4 (right–hand side vertical axis), together with the empirical exposures (left–hand side vertical axis).

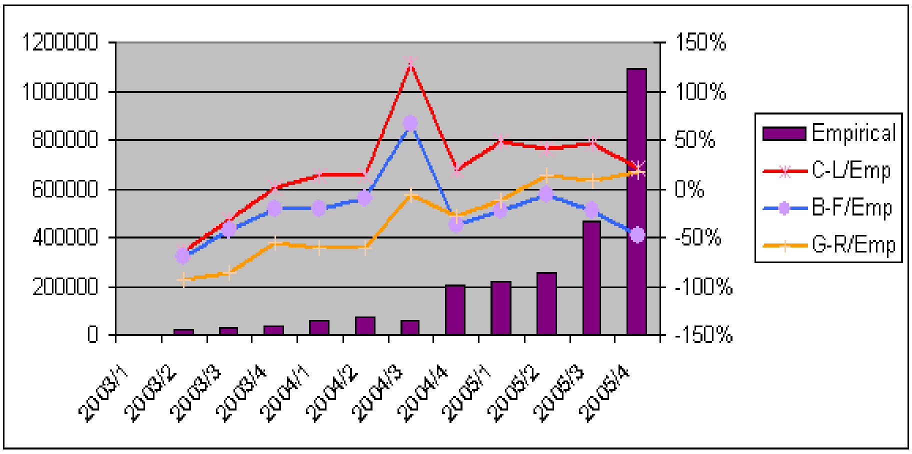

Figure 5.3: Empirical Exposures and Comparison of Percentage Differences

The projected reserves were compared also with the empirical ultimately developed losses separately for each of the twelve accident quarters. The standard deviation, maximum, minimum and mean difference for each method are reported in rows 3 to 5 of Table 5.1. For most of these criteria, the G–R

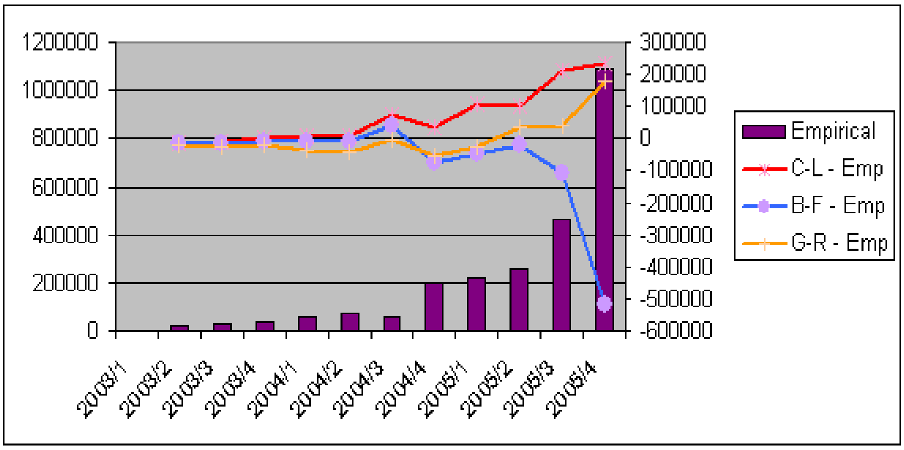

Figure 5.4: Empirical Exposures and Comparison of Dollar Differences

method performs better than the classical C–L and B–F methods.

For completeness, Tables A.1–A.3 of the Appendix give the detail of the loss reserve projections, quarter per quarter, for each of the methods, while Table A.4 gives the fully developed empirical losses used in the cross–valuation tests.

The differences between each of the methods and the empirical method are calculated in Table A.5, both in dollar and in percentage terms for each accident quarter. The standard deviation, minimum value, maximum value and mean differences are calculated to assess each of the methods. The G–R method gives the smallest mean at 3.78%, which means that the total projected future losses differs only by 3.78%, compared to the empirical value. This is much better than the C–L and B–F methods, which are 32.33% larger and -27.93% smaller respectively. The G–R method also gives the smallest maximum value of the difference. For the standard deviation, the G–R method is a little worse than the B–F method and for the minimum value of the difference, the G–R method is not as good as the C–L and B–F methods. So, over all, the G–R method gives here a superior estimation of IBNR reserves compared to the C–L and B–F methods.

5.4 Conclusion

This chapter establishes a structural loss reserving model within the framework of GLMs. It enables the estimation of the loss reserves on an individual risk class basis. Compared to traditional models based on the aggregated loss development triangles, this GLM approach uses more detailed information, such as the premium exposure emergence pattern, the loss emergence pattern and the loss development pattern. This means that we can expect that the GLM loss reserving model should give more accurate and stable estimates of the loss reserves. It is definitely true for the illustrative Bodily Injury Auto Insurance dataset analyzed in the previous section.

Appendix A

Loss Reserves Data and Projections

Accident	Development Age in Quarters											
Quarter	1	2	3	4	5	6	7	8	9	10	11	12
2003/1	955,478	613,431	348,077	164,180	82,903	72,933	66,876	63,453	27,357	27,623	26,780	13,651
2003/2	645,197	331,337	175,645	80,949	92,981	50,244	12,390	31,880	14,558	31,301	17,994	1,400
2003/3	452,590	269,145	158,831	115,536	99,189	49,882	36,852	31,902	33,287	28,187	2,929	1,230
2003/4	422,851	263,863	141,592	68,522	42,407	38,978	52,798	43,112	29,450	9,041	4,959	2,629
2004/1	446,742	264,120	169,399	87,825	66,702	48,404	37,445	19,344	7,744	6,765	4,454	2,771
2004/2	282,422	277,838	165,429	121,050	35,593	42,575	33,760	14,304	6,734	2,821	2,468	1,437
2004/3	349,989	310,533	124,391	105,095	83,739	39,971	25,890	11,679	7,150	4,564	3,165	2,366
2004/4	466,583	455,688	193,463	86,952	98,737	62,411	39,962	20,034	12,177	6,169	2,161	1,698
2005/1	445,847	416,837	218,716	133,375	87,462	49,421	27,304	14,331	8,567	4,854	2,772	636
2005/2	369,504	311,752	195,932	132,509	71,580	39,617	21,639	10,343	5,088	3,936	1,921	1,115
2005/3	463,661	430,669	241,183	123,797	64,976	33,292	18,510	9,814	6,370	2,407	1,352	1,078
2005/4	564,744	593, 559	312,606	171,918	92,460	46,651	25,371	12,686	7,265	3,656	1,813	829

Table A.1: Projected Loss Development – GLM (G–R)

Accident	Development Age in Quarters											
Quarter	1	2	3	4	5	6	7	8	9	10	11	12
2003/1	955,478	613,431	348,077	164,180	82,903	72,933	66,876	63,453	27,357	27,623	26,780	13,651
2003/2	645,197	331,337	175,645	80,949	92,981	50,244	12,390	31,880	14,558	31,301	17,994	8,274
2003/3	452,590	269,145	158,831	115,536	99,189	49,882	36,852	31,902	33,287	28,187	14,684	7,191
2003/4	422,851	263,863	141,592	68,522	42,407	38,978	52,798	43,112	29,450	18,935	12,924	6,329
2004/1	446,742	264,120	169,399	87,825	66,702	48,404	37,445	19,344	23,759	19,967	13,629	6,674
2004/2	282,422	277,838	165,429	121,050	35,593	42,575	33,760	27,135	20,546	17,267	11,785	5,771
2004/3	349,989	310,533	124,391	105,095	83,739	39,971	42,094	29,884	22,628	19,016	12,980	6,356
2004/4	466,583	455,688	193,463	86,952	98,737	58,951	56,488	40,103	30,366	25,519	17,418	8,530
2005/1	445,847	416,837	218,716	133,375	90,121	59,109	56,639	40,210	30,447	25,587	17,465	8,552
2005/2	369,504	311,752	195,932	95,725	72,178	47,340	45,362	32,205	24,385	20,493	13,987	6,850
2005/3	463,661	430,669	220,521	121,660	91,734	60,166	57,653	40,930	30,992	26,045	17,777	8,705
2005/4	564,744	511,866	265,467	146,457	110,431	72,429	69,403	49,272	37,308	31,354	21,401	10,480

Table A.2: Projected Loss Development – Chain Ladder (C–L)

Accident	Development Age in Quarters											
Quarter	1	2	3	4	5	6	7	8	9	10	11	12
2003/1	955,478	613,431	348,077	164,180	82,903	72,933	66,876	63,453	27,357	27,623	26,780	13,651
2003/2	645,197	331,337	175,645	80,949	92,981	50,244	12,390	31,880	14,558	31,301	17,994	7,180
2003/3	452,590	269,145	158,831	115,536	99,189	49,882	36,852	31,902	33,287	28,187	12,011	5,882
2003/4	422,851	263,863	141,592	68,522	42,407	38,978	52,798	43,112	29,450	15,039	10,265	5,027
2004/1	446,742	264,120	169,399	87,825	66,702	48,404	37,445	19,344	16,663	14,003	9,558	4,681
2004/2	282,422	277,838	165,429	121,050	35,593	42,575	33,760	21,666	16,405	13,787	9,410	4,608
2004/3	349,989	310,533	124,391	105,095	83,739	39,971	30,559	21,695	16,427	13,805	9,423	4,614
2004/4	466,583	455,688	193,463	86,952	98,737	31,461	30,146	21,402	16,205	13,619	9,296	4,552
2005/1	445,847	416,837	218,716	133,375	47,114	30,901	29,610	21,021	15,917	13,377	9,130	4,471
2005/2	369,504	311,752	195,932	63,363	47,777	31,336	30,027	21,317	16,141	13,565	9,259	4,534
2005/3	463,661	430,669	116,432	64,235	48,434	31,767	30,440	21,610	16,363	13,752	9,386	4,596
2005/4	564,744	221,041	114,638	63,245	47,688	31,277	29,971	21,277	16,111	13,540	9,241	4,525

Table A.3: Projected Loss Development – Bornhuetter–Ferguson (B–F)

Accident	Development Age in Quarters											
Quarter	1	2	3	4	5	6	7	8	9	10	11	12
2003/1	955,478	613,431	348,077	164,180	82,903	72,933	66,876	63,453	27,357	27,623	26,780	13,651
2003/2	645,197	331,337	175,645	80,949	92,981	50,244	12,390	31,880	14,558	31,301	17,994	23,378
2003/3	452,590	269,145	158,831	115,536	99,189	49,882	36,852	31,902	33,287	28,187	21,240	10,494
2003/4	422,851	263,863	141,592	68,522	42,407	38,978	52,798	43,112	29,450	15,210	8,745	13,971
2004/1	446,742	264,120	169,399	87,825	66,702	48,404	37,445	19,344	25,785	19,250	8,132	3,050
2004/2	282,422	277,838	165,429	121,050	35,593	42,575	33,760	36,854	16,755	10,667	8,205	0
2004/3	349,989	310,533	124,391	105,095	83,739	39,971	25,598	6,369	4,660	3,123	14,338	4,090
2004/4	466,583	455,688	193,463	86,952	98,737	40,528	40,424	29,245	36,168	21,701	4,040	27,983
2005/1	445,847	416,837	218,716	133,375	73,047	45,585	40,657	21,611	16,856	7,068	11,735	5,103
2005/2	369,504	311,752	195,932	91,563	41,952	38,389	9,056	27,861	14,955	11,689	11,852	7,736
2005/3	463,661	430,669	193,318	75,236	56,397	49,103	33,270	19,703	15,879	7,478	12,955	1,091
2005/4	564,744	454,369	180,682	100,655	101,822	76,607	32,137	46,371	45,654	18,679	24,140	8,627

Table A.4: Empirical Loss Development

Quarter	Empirical	C–L	I	II	B–F	III	IV	G–R	V	VI
2003/2	23,378	8,274	-64.61%	-15,104	7,180	-69.29%	-16,198	1,400	-94.01%	-21,978
2003/3	31,734	21,875	-31.07%	-9,859	17,892	-43.62%	-13,842	4,159	-86.89%	-27,575
2003/4	37,926	38,188	0.69%	261	30,330	-20.03%	-7,596	16,629	-56.15%	-21,297
2004/1	56,218	64,029	13.89%	7,811	44,905	-20.12%	-11,313	21,734	-61.34%	-34,484
2004/2	72,481	82,504	13.83%	10,023	65,877	-9.11%	-6,604	27,764	-61.70%	-44,718
2004/3	58,178	132,958	128.54%	74,779	96,525	65.91%	38,347	54,815	-5.78%	-3,363
2004/4	200,088	237,376	18.64%	37,287	126,680	-36.69%	-73,408	144,611	-27.73%	-55,477
2005/1	221,662	328,131	48.03%	106,469	171,541	-22.61%	-50,121	195,347	-11.87%	-26,315
2005/2	255,054	358,525	40.57%	103,470	237,318	-6.95%	-17,736	287,748	12.82%	32,694
2005/3	464,430	676,184	45.59%	211753	357,016	-23.13%	-107,414	502,777	8.26%	38,347
2005/4	1,089,742	1,325,868	21.67%	236,126	572,555	-47.46%	-517,187	1,268,814	16.43%	179,072
Total	2,510,892	3,273,910	30.39%	763,018	1,727,819	-31.19%	-783,073	2,525,800	0.59%	14,907
		St.Dev.	46.63%	217,054	St.Dev.	32.73%	251,936	St.Dev.	39.67%	62,828
		Max	128.54%	763,018	Max	65.91%	38,347	Max	16.43%	179,072
		Min	-64.61%	-15,104	Min	-69.29%	-783,073	Min	-94.01%	-55,477
		Mean	32.33%	69,365	Mean	-27.93%	-71,188	Mean	3.78%	1,355

Table A.5: Per Quarter Comparison of Empirical/Projected Loss Development

Column I: Ratio C–L minus Empirical over Empirical (in percentage).

Column II: Difference C–L minus Empirical (in Dollars).

Column III: Ratio B–F minus Empirical over Empirical (in percentage).

Column IV: Difference B–F minus Empirical (in Dollars).

Column V: Ratio G–R minus Empirical over Empirical (in percentage).

Column VI: Difference G–R minus Empirical (in Dollars).

Conclusion

The thesis focuses on the theory of GLMs and their application in insurance. The aim is to provide some further theoretical insights as well as innovative methods in practical uses of GLMs.

The first chapter gives an introduction of the theory of GLMs and GLMMs. The standard maximum likelihood estimation (MLE) method, the asymptotic properties and the bias of MLE estimators are presented to give readers the necessary background.

The second chapter gives a criteria for full credibility of the GLM and GLMM estimators. We show how credibility relates to the sample size and the components of the GLM, as well as the link function and distribution of covariates.

A closed form of the full credibility criterion is given for the log–link function, usually paired to Poisson observations (i.e. claim counts). For general link functions, we propose a credibility estimation based on an asymptotic normal approximation. This provides a method to compute confidence intervals for GLM estimators. These results can be very useful in practice as they provide full credibility criteria for the GLM estimators, at a time when GLMs are becoming popular in the statistical analysis of insurance data.

The third chapter studies the partial credibility theory for GLM estimators. Also, we further relax the assumption for GLMs so that the response variable does not need to follow a distribution in the exponential family. Least squares estimation (LSE) is used instead of maximum likelihood. The numerical algorithm is derived to estimate the GLM parameters by LSE. Finally, the credibility matrix is derived to minimize the sum of squares error (SSE).

In the fourth chapter, we answer the question whether the linear credibility weighted estimators are superior to the GLM estimators through a simulation study. Following a brief introduction to goodness-of-fit assessments, we define a simulation algorithm to compare these two approaches.

The numerical examples illustrate that in most GLM cases, the linear partial credibility estimators can only slightly reduce the SSE, and in some cases, they are not even as good as the GLM estimators. The results show that linear credibility weighted mean estimators do not further improve the accuracy of GLM estimators.

This result can be theoretically interesting, but is definitely very useful in practice, providing guidance for GLM modelers. The focus should be on improving the GLM fit instead of using resources to calculate linear credibility weighted estimators that do not significantly improve GLM results.

The last chapter establishes a structural loss reserving model within the framework of GLMs. It enables the estimation of the loss reserves on an individual risk class basis. Compared to traditional models based on the aggregated loss development triangles, this GLM approach uses more detailed information, such as the premium exposure emergence pattern, the loss emergence pattern and the loss development pattern. This means that we can expect the GLM loss reserving model to give more accurate and stable estimates of the loss reserves. A numerical example confirms this conclusion.

Overall, this thesis is theoretically innovative and has a strong potential for practical applications. It can become a useful handbook for both actuarial researchers and GLM modelers.

Bibliography

[1] Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989) *Statistical Modelling in GLIM*. Oxford Science Publications, Oxford.

[2] Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E. and Thandi, N. (2005) *A Practitioner's Guide to Generalized Linear Models*. Casualty Actuarial Society, Virginia.

[3] Anscombe, J.A. (1972) Contribution to the discussion of H.Hotelling's paper. *Journal of the Royal Statistical Society, Series B*, **15**(1), 229–230.

[4] Antonio, K. and Beirlant, J. (2007) Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, **40**(1),

58–76.

[5] Babu, G.J. and Chaubey, Y.P. (1996) Asymptotics and bootstrap for inverse Gaussian regression. *Annuals of the Institute of Statistical Mathematics*, **48**, 75–88.

[6] Bailey, A. (1950) Credibility procedures, Laplace's generalization of Bayes' rule and the combination of collateral knowledge with observed data. *Proc. of the Casualty Actuarial Society*, **37**, 7–23 (with discussions, 94–115).

[7] Bouska, A. (1989) Exposure bases revisited. *Proceedings of the Casualty Actuarial Society*, **76** (1), 1–23.

[8] Brockman, M.J. and Wright, T.S. (1992) Statistical motor rating: making effective use of your data. *Journal of the Institute of Actuaries*, **119**, 457–543.

[9] Bühlmann, H. (1967) Experience rating and credibility. *ASTIN Bulletin*, **4**(3), 199–207.

[10] Bühlmann, H. (1969) Experience rating and credibility. *ASTIN Bulletin*, **5**(2), 157–165 .

[11] Bühlmann, H. and Straub, E. (1970) Glaubwürdigkeit für Schadensätze. *Bulletin of the Swiss Association of Actuaries*, 111–133.

[12] Cordeiro, G.M. and McCullagh, P. (1991) Bias correction in generalized linear models. *Journal of the Royal Statistical Society, B*, **53**(3), 629–643.

[13] Coutts, S.M. (1984) Motor insurance rating, and actuarial approach. *Journal of the Institute of Actuaries*, **111**, 87–148.

[14] Cox, D.R. and Snell, E.J. (1968) A general definition of residuals. *Journal of the Royal Statistical Society, Series B*, **53**(1), 629–643.

[15] Cox, D.R. and Snell, E.J. (1979) On Sampling and estimation of rare errors. *Biometrika*, **66**(1), 125–132.

[16] Demindenko, E. (2004) *Mixed Models: Theory and Applications*. Hoboken, New Jersey.

[17] De Vylder, F.E. (1996) *Advanced Risk Theory*. de l'Université de Bruxelles and Swiss Association of Actuaries, Belgium.

[18] Dobson, A. (1990) *An Introduction to Generalized Linear Models*. Chapman and Hall, London.

[19] Fahrmeir, L. and Kaufmann, H. (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, **13** (1), 342–368.

[20] Fahrmeir, L. and Gerhard, T. (2001) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer, New York.

[21] Frees, E.W. (2003) Multivariate credibility for aggregate loss models. *North American Actuarial Journal*, **7**, 13–37.

[22] Frees, E.W., Shi, P. and Valdez, E.A. (2009) Actuarial applications of a hierarchical insurance claims model. *ASTIN Bulletin*, **39**(1), 165–197.

[23] Garrido, J. and Zhou, J. (2009) Full credibility with generalized linear and mixed models. *ASTIN Bulletin*, **39** (1), 61–80.

[24] Goulet, V., Forgues, A. and Lu, J. (2006) Credibility for severity revisited. *North American Actuarial Journal*, **10**, 49–62.

[25] Haberman, S.J. (1974) Loglinear models for frequency tables with ordered classifications. *Biometrics*, **30**(4), 589–600.

[26] Haberman, S. and Renshaw, A.E. (1996) Generalized linear models and actuarial science. *The Statistician*, **45**(4), 407–436.

[27] Hachemeister, C.A. (1975) Credibility for regression models with application to trend. *Credibility, Theory and Application*. Academic Press, New York, 129–163.

[28] Hardin, J.W. and Hilbe, J.M. (2007) *Generalized Linear Models and Extensions*. Chapman and Hall/CRC, Boca Raton.

[29] Hardin, J.W. and Hilbe, J.M. (2002) *Generalized Estimating Equations*. Chapman and Hall/CRC, London.

[30] Hoedemakers, T., Beirlant, J., Goovaerts, M.J. and Dhaene, J. (2005) On the distribution of discounted loss reserves using generalized linear models. *Scandinavian Actuarial Journal*, **1**, 25–45.

[31] Jan, R.M. and Heinz, N. (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, New York.

[32] Jewell, W.S. (1975) The use of collateral data in credibility theory: A hierarchical model. *Giornale dell' Instituto Italiano degli Attuari*, **38**, 1–16.

[33] Klugman, S.A., Panjer, H.H. and Willmot G.E. (2004) *Loss Models From Data to Decisions.* John Wiley & Sons, New Jersey.

[34] Lee, Y., Nelder, J.A. and Pawitan, Y. (2006) *Generalized Linear Models With Random Effects*. Chapman and Hall/CRC, Boca Raton.

[35] Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**(1), 13–22.

[36] Lo, C.H., Fung, W.K. and Zhu, Z.Y. (2007) Structural parameter estimation using generalized estimating equations for regression credibility models. *ASTIN Bulletin*, **37** (2), 323–343.

[37] Mallows, C.L. (1973) Some comments on C_p. *Technometrics*, **15**(4), 661–675.

[38] McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*. Chapman and Hall, New–York.

[39] McCulloch, C.E. and Searle, S.R. (2001) *Generalized, Linear and Mixed Models*. John Wiley & Sons, New York.

[40] Mowbray, A.H. (1914) How extensive a payroll exposure is necessary to give a dependable pure premium. *Proc. of the Casualty Actuarial Society*, **1**, 24–30.

[41] Nelder, J.A. and Verrall, R.J. (1997) Credibility theory and generalized linear models. *ASTIN Bulletin*, **27**(1), 71–82.

[42] Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of the Royal Statistical Society, A*, **135**(3), 370–384.

[43] Neuhaus, J.M. (1992) Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research*, **1**, 249–273.

[44] Neuhaus, J.M., Jewell, N.P. (1993) A geometric approach to asses bias due to omitted covariates in generalized linear models. *Biometrika*, **80**, 807–815.

[45] Neuhaus, J.M., Kalbfleisch, J.D. and Hauck, W.W. (1991) A comparison of cluster–specific and population–averaged approaches for analyzing correlated binary data. *International Statistical Review*, **59**, 25–35.

[46] Ohlsson, E. and Johansson, B. (2010) *Non-life Insurance Pricing with Generalized Linear Models*. Springer-Verlag, Berlin.

[47] Pierce, D.A. and Schafer, D.W. (1986) Residuals in generalized linear models. *Journal of the American Statistical Association*, **81**, 977–986.

[48] SAS Technical Report P–243 (1993) *SAS/STAT Software: The GENMOD Procedure, Release 6.09*, SAS Institute Inc., Cary, NC.

[49] SAS/STAT Software (2006) *GLIMMIX Procedure, Release 9.1*, June 2006, SAS Institute Inc., Cary, NC.

[50] Schabenberger, O. and Gregoire, T.G. (1996) Population–averaged and subject–specific approaches for clustered categorical data. *Journal of Statistical Computation and Simulation*, **54**, 231–253.

[51] Schmitter, H. (2004) The sample size needed for the calculation of a GLM tariff. *ASTIN Bulletin*, **34**(1), 249–262.

[52] Schmidt, K.D. (2006) Methods and models of loss reserving based on run-off triangles: a unifying survey. *Casualty Actuarial Society Forum*, 269–317.

[53] Sen, D. and Chaubey, Y.P. (2011) A class of accelerated life testing models based on the Gamma distribution. *Journal of the India Society of Agricultural Statistics*, **65**(2), 1–10.

[54] Taylor, G. and McGuire, G. (2004) Loss reserving with GLM: a case study. *Spring 2004 Meeting of Casualty Actuarial Society*, 327–392.

[55] Venter, G.G. (2007) Generalized linear models beyond the exponential family with loss reserve applications. *ASTIN Bulletin*, **37** (2), 345–364.

[56] Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. Springer, New York.

[57] Verral, R.J. (2004) A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claim reserving. *North American Actuarial Journal*, **8**(3), 67–89.

[58] Wedderburn, R.W.M. (1976) On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika*, **63**, 27–32.

[59] Whitney, A.W. (1918) The theory of experience rating. *Proc. of the Casualty Actuarial Society*, **4**, 274–292.

[60] Wiser, R.F., Cockley, J.E. and Gardner, A. (2001) Loss reserving. *Foundations of Casualty Actuarial Science*. Casualty Actuarial Society, Virginia.

[61] Wolfinger, R. and O'Connell, M. (1993) Generalized linear mixed models: a pseudo–likelihood approach. *Journal of Statistical Computation and Simulation*, **4**, 233–243.

[62] Wüthrich, M.V. and Merz, M. (2008) *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons, Chichester.

[63] Wüthrich, M.V., Merz, M. and Bühlmann,H. (2008) Bounds on the estimation error in the chain ladder method. *Scandinavian Actuarial Journal*, **4**, 25–45.

[64] Ye, J. (1998) On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, **93**(441), 120–131.

[65] Zeger, S.L., Liang K.-Y. and Albert, P.S. (1998) Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049–1060.