Credibility Theory for Generalized Linear and Mixed Models

José Garrido[*] and Jun Zhou
Concordia University, Montreal, Canada
E-mail: garrido@mathstat.concordia.ca,
junzhou@mathstat.concordia.ca

August 2006

Abstract

Generalized linear models (GLMs) are gaining popularity as a statistical analysis method for insurance data. 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 paper we study the limited fluctuations credibility of the GLM estimators as well as in the extended case of generalized linear mixed model (GLMMs). We show how credibility depends on the sample size, the distribution of covariates and the link function. This provides a mechanism to obtain confidence intervals for the GLM and GLMM estimators.

Keywords: GLMs, GLMMs, limited fluctuations credibility, confidence intervals

1 Introduction

Generalized linear models (GLMs) are becoming the premier statistical analysis method for insurance data. We consider the question of credibility: how

[*]This research is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

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 number of levels for each factor. In this paper 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 and the distribution of covariates. This provides a mechanism to obtain confidence intervals for the estimates in GLMs and GLMMs.

The paper is organized as follows. Section 2 briefly recalls the basic concepts of GLMs and GLMMs. Section 3 gives the *limited fluctuations credibility* results for GLMs and GLMMs. Section 4 is devoted to the choice of the link function and its effect on credibility. Section 5 illustrates with some numerical examples the main results of the paper. Details on calculations and applications (in SAS) are provided.

2 GLMs and GLMMs

This section provides a short summary of the main characteristics of GLMs and GLMMs. McCullagh and Nelder (1989) provide 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. McCulloch and Searle (2001) and Demindenko (2004) are useful references for details on GLMMs. Antonio and Beirlant (2006) give an application of GLMMs in actuarial statistics.

2.1 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, taking the form

$$f(y;\theta,\phi) = \exp\Big\{ \int \frac{\big[y-\mu(\theta)\big]}{\phi V(\mu)}\, d\mu(\theta) + c(y,\phi)\Big\}, \tag{2.1}$$

where θ is called the *natural* parameter, ϕ is a known *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

$$\eta_i = \underline{X}_i' \underline{\beta}, \qquad i = 1, \ldots, n, \tag{2.2}$$

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

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

$$g(\mu_i) = \eta_i, \qquad i = 1, \ldots, n. \tag{2.3}$$

Example 2.1 *GLMs commonly used in credibility*

The table below gives the different model components of the GLMs most commonly used in credibility 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: GLM Examples

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

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

$$l(\underline{\beta}) = \ln L(\underline{\beta}) = \sum_{i=1}^{n} \Big\{ \int \frac{\big[y_i - \mu_i(\theta)\big]}{\phi\, V(\mu_i)} d\mu_i(\theta) + c(y_i, \phi) \Big\}, \tag{2.4}$$

and its derivative:

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

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

$$\frac{dl(\underline{\beta})}{d\underline{\beta}} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\phi\, V(\mu_i)} \frac{1}{g'(\mu_i)} \underline{X}_i'. \tag{2.5}$$

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), we must resort to an iterative algorithm, such as Newton–Raphson or Fisher scoring methods to obtain the MLE numerically.

The MLE $\hat{\underline{\beta}}$ for the GLM parameters has some nice properties.

Lemma 2.1 *For the MLE, $\hat{\underline{\beta}}$, solution of* (2.5)*, we have:*

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

2. *$\mathbb{V}(\hat{\underline{\beta}}) \to (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\phi$ consistently, as the iteratively estimated $\hat{\underline{\beta}}$ converges to the true $\underline{\beta}$, where $\mathbf{W} = diag(w_i, \ldots, w_n)$ with weights $w_i = \big[\phi\, g'(\mu_i)\, V(\mu_i)\big]^{-1}$, and 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 with the iterative algorithm.*

For a proof see McCullagh and Nelder (1989). The bias of $\hat{\underline{\beta}}$ is affected by the choice of link function g (see Cordeiro and McCullagh, 1991). This problem is further discussed in Section 4.

2.2 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 $\underline{U}_1, \ldots, \underline{U}_n$, consider the following EF distribution:

$$f(y_{ij}|u_i, \theta, \phi) = \exp\left\{\frac{[y_{ij}\theta_{ij} - b(\theta_{ij})]}{\phi} + c(y_{ij}, \phi)\right\}, \tag{2.6}$$

   where $\underline{u}_i = (u_{i1}, \ldots, u_{ik})$ are variates from normally distributed k-dimensional random vectors $\underline{U}_i \sim N(0, \mathbf{D})$, where $\mathbf{D}$ is the variance–covariance matrix and $\mu_{ij} = \mathbb{E}[Y_{ij}|\underline{U}_i]$.

2. The linear mixed effects model is defined as:

$$\eta_{ij} = \underline{X}'_{ij}\underline{\beta} + \underline{T}'_{ij}\underline{u}_i, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \tag{2.7}$$

   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,

$$g(\mu_{ij}) = \eta_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \tag{2.8}$$

   completes the model.

The derivation of the likelihood function is also straightforward for GLMMs. However, numerical methods are needed in most cases to obtain the MLEs. Antonio and Beirlant (2006) give a brief review of some numerical techniques, such as a restricted pseudo–likelihood, the Gauss–Hermite quadrature, and Bayesian methods. Demidenko (2004) gives a detailed Monte Carlo method. Most of these techniques are available in SAS.

3 Credibility Theory for GLMs

Developed in the early part of the 20th century, *limited fluctuations credibility* gives formulas to assign full or partial credibility to a policyholder's, or group of policyholders' experience. Bühlmann (1967, 1969), Bühlmann and Straub (1970), Hachemeister (1975), Jewell (1975) and Frees (2003) give several credibility formulas. Goulet et al. (2006) gives a review of four different formulas. Nelder and Verrall (1997) shows how credibility theory can be encompassed within the theory of GLMs.

If the probability of a small difference between the estimator $\hat{\mu}_i$ and the parameter it estimates, say m_i, is large, then the insurer may find $\hat{\mu}_i$ credible. If this difference is small "enough", we say that "full credibility" is achieved. Statistically, this can be defined as

$$\mathbb{P}\big\{|\hat{\mu}_i - m_i| \leq r m_i\big\} \geq \pi_i, \tag{3.1}$$

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

Proposition 3.1 *For any generalized linear model, as defined in (2.1)–(2.3), let g be a monotonic increasing link function. Then*

$$\begin{aligned}
\pi_i &= \mathbb{P}\big\{|\hat{\mu}_i - m_i| \leq r m_i\big\} = \mathbb{P}\big\{(1-r)m_i \leq \hat{\mu}_i \leq (1+r)m_i\big\} \\
&= \mathbb{P}\big\{g[(1-r)m_i] - g(m_i) \leq g(\hat{\mu}_i) - g(m_i) \leq g[(1+r)m_i] - g(m_i)\big\} \\
&= \mathbb{P}\big\{g[(1-r)m_i] - \underline{X}_i'\underline{\beta} \leq \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \leq g[(1+r)m_i] - \underline{X}_i'\underline{\beta}\big\}. \qquad (3.2)
\end{aligned}$$

It is reasonable to restrict g to increasing link functions. Similar results follow for decreasing link functions.

Proposition 3.1 gives some expressions equivalent to (3.1) and transfers the confidence interval from the space of the GLM estimators $\hat{\mu}_i$, to the space of the linear components, through the link function g. If the latter satisfies the condition $g(c\, m_i) = g(m_i) + c'$ for any m_i, where c and c' are constants with respect to m_i, then (3.2) admits a simpler form as follows.

Proposition 3.2 *For any given error tolerance level r and any m_i,*

$$\mathbb{P}\big\{|\hat{\mu}_i - m_i| \leq r m_i\big\} = \mathbb{P}\big\{c_1 \leq \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \leq c_2\big\}, \tag{3.3}$$

where c_1 and c_2 are constants given by (3.6), *if and only if a log–link function $g(x) = c\ln(x) + \tau$ in used in* (3.2), *where c is a scale- and τ is a shift–parameter.*

Proof: $(\Rightarrow)$

If $g(x) = c\ \ln(x) + \tau$, by (3.2), it is clear to see that

$$g[(1-r)m_i] - g(m_i) = c\ \ln[(1-r)m_i] - c\ \ln(m_i) = c\ \ln(1-r), \tag{3.4}$$

and

$$g[(1+r)m_i] - g(m_i) = c\ \ln[(1+r)m_i] - c\ \ln(m_i) = c\ \ln(1+r). \tag{3.5}$$

$(\Leftarrow)$

If $\mathbb{P}\{|\hat{\mu}_i - m_i| \leq rm_i\} = \mathbb{P}\{c_1 \leq \underline{X}_i'\hat{\underline{\beta}} - \underline{X}_i'\underline{\beta} \leq c_2\}$, then from (3.2), for any m_i,

$$c_1 = g[(1-r)m_i] - g(m_i) \quad \text{and} \quad c_2 = g[(1+r)m_i] - g(m_i). \tag{3.6}$$

Assuming that g is differentiable, then for any m_i

$$g'(m_i) = \lim_{r\to 0} \frac{g[(1-r)m_i] - g(m_i)}{-rm_i} = \lim_{r\to 0} \frac{c_1}{-rm_i}, \tag{3.7}$$

but also,

$$g'(m_i) = \lim_{r\to 0} \frac{g[(1+r)m_i] - g(m_i)}{rm_i} = \lim_{r\to 0} \frac{c_2}{rm_i}. \tag{3.8}$$

Hence $\lim_{r\to 0} \frac{c_1}{-r} = \lim_{r\to 0} \frac{c_2}{r} = c$, say. Then $g'(m_i) = \frac{c}{m_i}$, 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 m_i. These only depend on the chosen error tolerance level r. The following example gives a concrete illustration.

Example 3.1 *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 $\mathbb{E}(Y_i) = m_i = e^{x_{i1}\beta_1 + \cdots + x_{ip}\beta_{ip}}$. With the log–link function, $g[\mathbb{E}(Y_i)] = g(m_i) = x_{i1}\beta_1 + \cdots + x_{ip}\beta_{ip}$. By (3.2), $|\hat{\mu}_i - m_i| \leq rm_i \Leftrightarrow \ln(1-r) \leq \underline{X}_i'\hat{\underline{\beta}} - \underline{X}_i'\underline{\beta} \leq \ln(1+r)$. Since $0 < r < 1$, then $|\ln(1+r)| < |\ln(1-r)|$ and hence

$$\begin{aligned} \mathbb{P}\{|\hat{\mu}_i - m_i| \leq rm_i\} &= \mathbb{P}\{\ln(1-r) \leq \underline{X}_i'\hat{\underline{\beta}} - \underline{X}_i'\underline{\beta} \leq \ln(1+r)\} \\ &\leq \mathbb{P}\{|\underline{X}_i'\hat{\underline{\beta}} - \underline{X}_i'\underline{\beta}| \leq |\ln(1-r)|\}\ . \end{aligned} \tag{3.9}$$

Let $s^2 = \mathbb{V}(\hat{\beta}_1 + \cdots + \hat{\beta}_p)$ and $\underline{X}_i = (1, 1, \ldots, 1)$, then (3.9) becomes

$$\begin{aligned}
&\mathbb{P}\big\{|\underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta}| \leq |\ln(1-r)|\big\} \\
&\quad = \mathbb{P}\big\{|(\hat{\beta}_1 + \cdots + \hat{\beta}_r) - (\beta_1 + \cdots + \beta_r)| \leq |\ln(1-r)|\big\} \\
&\quad = \mathbb{P}\Big\{\Big|\frac{(\hat{\beta}_1 + \cdots + \hat{\beta}_r) - (\beta_1 + \cdots + \beta_r)}{s}\Big| \leq \frac{|\ln(1-r)|}{s}\Big\}. \qquad (3.10)
\end{aligned}$$

Approximating by a normal distribution, (3.10) yields $\frac{|\ln(1-r)|}{s} \geq Z_{\frac{\pi}{2}}$, where $Z_{\frac{\pi}{2}}$ is the $100\frac{\pi}{2}$-percentile of a standard normal distribution. Hence the following full–credibility criteria is obtained:

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

which says that the sample size n must be sufficiently large to ensure that the standard deviation of the (sum of the) estimators $\hat{\beta}_1, \ldots, \hat{\beta}_p$ be at most s^*. This result is consistent with the result given by Schmitter (2004).

Proposition 3.3 *Let* $\mathbf{\Sigma} = (\sigma_{ij})_{i,j} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\,\phi$ *and* $s_i^2 = \mathbb{V}(\underline{X}_i'\underline{\hat{\beta}})$, *then*

$$s_i^2 \to \underline{X}_i'\mathbf{\Sigma}\underline{X}_i, \qquad (3.11)$$

consistently, for $\underline{X}_i$, $\mathbf{W}$ *and* $\mathbf{X}$ *as in Lemma 2.1.*

Proof: Since $\mathbb{V}(\underline{\hat{\beta}}) \to (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\phi$ consistently, as the iterative $\underline{\hat{\beta}}$ converges to the true $\underline{\beta}$, then

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

consistently. □

As stated in Lemma 2.1, $\underline{\hat{\beta}}$ converges to $N\big(\underline{\beta}, (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\,\phi\big)$ in distribution. Then, the following corollary to Proposition 3.3 holds.

Corollary 3.1 $\big(\underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta}\big)/s_i$ *converges to* $N(0,1)$ *in distribution.*

Theorem 3.1 *For the log–link function, an approximation with the normal distribution gives*

$$\pi_i \doteq \Phi\Big(\frac{\ln(1+r)}{s_i}\Big) - \Phi\Big(\frac{\ln(1-r)}{s_i}\Big)\ , \tag{3.12}$$

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

Proof: From Propositions 3.1 and 3.2,

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

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

For any confidence coefficient π_i, Theorem 3.1 gives a $100(1-r)\%$ confidence interval for $\hat{\mu}_i$, the regression estimate 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 5 illustrate the above results.

Now for a general link function g, let

$$Q_1 = g[(1-r)m_i] - g(m_i) \quad \text{and} \quad Q_2 = g[(1+r)m_i] - g(m_i)\ . \tag{3.13}$$

Theorem 3.2 *For any link function g,*

$$\pi_i \doteq \Phi\Big(\frac{Q_2}{s_i}\Big) - \Phi\Big(\frac{Q_1}{s_i}\Big)\ , \tag{3.14}$$

where Φ is the cdf of the standard normal distribution, Q_1 and Q_2 are given in (3.13) *and s_i in Proposition 3.3.*

Proof:

$$\begin{aligned}\pi_i &= \mathbb{P}\Big\{|\hat{\mu}_i - m_i| \leq rm_i\Big\} = \mathbb{P}\Big\{Q_1 \leq \underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta} \leq Q_2\Big\} \\ &= \mathbb{P}\Big\{\frac{Q_1}{s_i} \leq \frac{\underline{X}_i'\underline{\hat{\beta}} - \underline{X}_i'\underline{\beta}}{s_i} \leq \frac{Q_2}{s_i}\Big\}\ .\end{aligned}$$

Approximating by the normal distribution gives (3.14). □

Clearly, the smaller s_i the bigger π_i (approximately), which differs for different i. If g is the log–link function, then Proposition 3.2 gives closed forms for Q_1 and Q_2. For other link functions, as the true parameter value m_i is unknown, we can approximate Q_1, Q_2 and π_i as follows:

$$\hat{Q}_1 = g[(1-r)\hat{\mu}_i] - g(\hat{\mu}_i) \quad \text{and} \quad \hat{Q}_2 = g[(1+r)\hat{\mu}_i] - g(\hat{\mu}_i)\ , \tag{3.15}$$

which implies that

$$\hat{\pi}_i \doteq \Phi\left(\frac{\hat{Q}_2}{s_i}\right) - \Phi\left(\frac{\hat{Q}_1}{s_i}\right)\ . \tag{3.16}$$

Section 4 discusses further the effect of the choice of link function on the above approximation.

Finally, similar results hold for the estimates credibility in GLMMs.

Proposition 3.4 *For any generalized linear mixed model, as defined in* (2.6)–(2.8)*, let g be a monotonic increasing link function. Then*

$$\begin{aligned}
\pi_i &= \mathbb{P}\big\{|\hat{\mu}_i - m_i| \le r m_i\big\} = \mathbb{P}\big\{(1-r)m_i \le \hat{\mu}_i \le (1+r)m_i\big\} \\
&= \mathbb{P}\big\{g[(1-r)m_i] - g(m_i) \le g(\hat{\mu}_i) - g(m_i) \le g[(1+r)m_i] - g(m_i)\big\} \\
&= \mathbb{P}\big\{g[(1-r)m_i] - \underline{X}'_{ij}\underline{\beta} - T'_{ij}\underline{u}_i \le \underline{X}'_{ij}\hat{\underline{\beta}} + T'_{ij}\hat{\underline{u}}_i - \underline{X}'_{ij}\underline{\beta} - T'_{ij}\underline{u}_i \\
&\qquad\qquad \le g[(1+r)m_i] - \underline{X}'_{ij}\underline{\beta} - T'_{ij}\underline{u}_i\big\}\ .
\end{aligned} \tag{3.17}$$

Using the same idea as in Theorem 3.2 we obtain the following result for GLMMs.

Theorem 3.3 *For any link function g, let $s_i^2 = \mathbb{V}(\underline{X}'_{ij}\underline{\beta} + T'_{ij}\underline{u}_i)$ and Q_1, Q_2 be defined as in* (3.13)*, then*

$$\pi_i \doteq \Phi\left(\frac{Q_2}{s_i}\right) - \Phi\left(\frac{Q_1}{s_i}\right)\ , \tag{3.18}$$

where Φ is the cdf of the standard normal distribution.

4 The Choice of Link Function

As shown in the above sections, the main idea here is to transfer the "full credibility" condition (3.1) to an equivalent form easier to implement, as in Proposition 3.1. Expression (3.14) 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 the 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 4.1 *Rescaling or shifting a given link function g, such as in $h(x) = c\,g(x) + \tau$, does not affect the approximate π_i in* (3.14).

Proof: For a link function g, (2.3) 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 $h(\mu_i) = \beta_0^{(h)} + \underline{X}_i'\underline{\beta}^{(h)}$, that is $c\,g(x) + \tau = \beta_0^{(h)} + \underline{X}_i'\underline{\beta}^{(h)}$ and hence $g(x) = \frac{\beta_0^{(h)} - \tau}{c} + \underline{X}_i'\frac{\underline{\beta}^{(h)}}{c}$. It follows that $\beta_0^{(g)} = \frac{\beta_0^{(h)} - \tau}{c}$ and $\underline{\beta}^{(g)} = \frac{\underline{\beta}^{(h)}}{c}$.

Now let $(s_i^{(g)})^2 = \mathbb{V}(\underline{X}_i'\hat{\underline{\beta}}^{(g)})$, $(s_i^{(h)})^2 = \mathbb{V}(\underline{X}_i'\hat{\underline{\beta}}^{(h)})$. Clearly $(s_i^{(g)})^2 = \frac{1}{c^2}(s_i^{(h)})^2$, or equivalently, $s_i^{(h)} = c\,s_i^{(g)}$, while

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

for $i = 1, 2$. Refer to (3.14), to see that

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

from the definitions of $Q_i^{(h)}$ and $s_i^{(h)}$. $\square$

Example 5.3 gives a numerical illustration of Lemma 4.1. It shows how the estimated probabilities π_i, from (3.14) but with the estimated 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, $\hat{\underline{\beta}}$, $\hat{\mu}_i = g^{-1}\big(\underline{X}_i'\hat{\underline{\beta}}\big)$ and in our estimated $\hat{Q}_1$, $\hat{Q}_2$ in (3.15). This is explored in the next result, but we first reproduce a version of Jensen's inequality needed in what follows.

Lemma 4.2 ***(Jensen Inequality)*** *Let φ be a convex upward (respectively concave) function on $(-\infty, \infty)$ and f an integrable function on $[0, 1]$. Then*

$$\int \varphi\big(f(t)\big)\,dt \geq (resp.\ \leq)\, \varphi\Big[\int f(t)\,dt\Big] \ . \tag{4.1}$$

The usual corollary of Jensen's inequality is to let f be the density function of a random variable X. Then

$$\mathbb{E}\big[\varphi(X)\big] \geq (\text{resp. } \leq)\, \varphi\big(\mathbb{E}[X]\big)\ . \tag{4.2}$$

Now we can explore how the link function affects the estimation bias in our confidence intervals.

Theorem 4.1 *$\hat{Q}_1$ and $\hat{Q}_2$ (3.15) are:*

1. *unbiased estimators if the link function g is linear,*
2. *asymptotically upward–biased if the link function g is convex and decreasing,*
3. *asymptotically downward–biased if the link function g is concave and increasing.*

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

$$\begin{aligned}
\text{bias}(\hat{Q}_1) &= \mathbb{E}(\hat{Q}_1) - Q_1 \\
&= \mathbb{E}\big\{g[(1-r)\hat{\mu}_i] - g(\hat{\mu}_i)\big\} - g[(1-r)m_i] + g(m_i) \\
&= \mathbb{E}\big\{g[(1-r)\hat{\mu}_i]\big\} - g[(1-r)m_i] + \underline{X}_i'\underline{\beta} - \mathbb{E}[\underline{X}_i'\hat{\underline{\beta}}] \\
&= \mathbb{E}\big\{g[(1-r)\hat{\mu}_i]\big\} - g[(1-r)m_i] - \underline{X}_i'\,\text{bias}(\hat{\beta})
\end{aligned} \tag{4.3}$$

Three cases need to be distinguished:

1. If g is linear then $\mathbb{E}\big\{g[(1-r)\hat{\mu}_i]\big\} - g[(1-r)m_i] = 0$ and $\hat{\beta}$ is unbiased, hence so is $\hat{Q}_1$.
2. If g is a convex decreasing function, then by Jensen's inequality in (4.2)

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

   that is $\mathbb{E}[\hat{\mu}_i] \leq m_i$. Now since

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

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

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

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

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

Theorem 4.2 *For a GLM problem, $\hat{\pi}_i$ given by (3.16) can be used as a criteria 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.*

5 Some Numerical Examples

Example 5.1 *Car Insurance Claims Data*

The SAS Technical Report P-243 (1993) gives the following illustrative dataset of a car insurance portfolio (also reproduced in Schmitter, 2004). For earlier examples of nonlinear analysis of car insurance data see Aitkin et al. (1989).

risk	claims	car type	age group
500	42	small	1
1200	37	medium	1
100	1	large	1
400	101	small	2
500	73	medium	2
300	14	large	2

Table 2: Car Insurance Data

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

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

The matrix of variance–covariance $\mathbf{\Sigma}$ is given by SAS as:

$$\mathbf{\Sigma} = \begin{pmatrix} 0.008150 & -0.007772 & -0.006344 & -0.004623 \\ -0.007772 & 0.07418 & 0.006556 & 0.003113 \\ -0.006344 & 0.006556 & 0.01645 & -0.002592 \\ -0.004623 & 0.003113 & -0.002592 & 0.01847 \end{pmatrix}$$

Let the tolerance level $r = 0.1$ and $\underline{X}_1 = (1, 1, 0, 1)'$. Then $s_1^2 = \underline{X}_1' \mathbf{\Sigma} \underline{X}_1 = 0.082236$ and $\pi_1 = \Phi(\frac{\ln(1+r)}{s_1}) - \Phi(\frac{\ln(1-r)}{s_1}) = 0.273533$. Clearly, the current experience produces GLM estimators that are not credible.

By contrast, letting $\underline{X}_2 = (1, 0, 1, 0)'$ gives $s_2^2 = 0.011912$ and $\pi_2 = 0.641557$, which indicates a more credible GLM estimator for *medium cars* in group *2* than for *small cars* in group *1*.

Furthermore, if the claim experience increases proportionally 23 times, i.e. the risk and claim counts for each car and age group increase 23 times, then $s_1^2 = 0.003575$ and $\pi_1 = 0.905492$. This shows that as the portfolio size increases the GLM tends to full credibility, as expected.

Example 5.2 *Modified Car Insurance Data*

This example shows that credibility also depends on the distribution of the covariates. For instance, keep the total number of claims unchanged in Table 2 at 268, but rearrange the claim counts in each group as in Table 3. Then for $\underline{X}_1 = (1, 1, 0, 1)'$ we get $s_1^2 = 0.038200$ and $\pi_1 = 0.392182$, which differs from the value of 0.273533 obtained in Example 5.1. Clearly the credibility of GLM estimates depends on the distribution of the covariates.

Example 5.3 *Rescaled Car Insurance Data*

risk	claims	car type	age group
500	45	small	1
1200	108	medium	1
100	9	large	1
400	36	small	2
500	44	medium	2
300	26	large	2

Table 3: Modified Car Insurance Data

Let the link function $g(x) = c\ \ln(x) + \tau$. Lemma 4.1 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, like SAS.

Choosing different rescaling parameters c, Table 4 shows that the estimated credibility values π_i in (3.14) remain essentially the same. Hence

c	s_1	π_1	s_2	π_2
0.1	0.028674	0.273559	0.012676	0.570980
0.5	0.143400	0.273504	0.063143	0.572737
1	0.286768	0.273533	0.126301	0.572679
2	0.573620	0.273495	0.252725	0.572455
5	1.433855	0.273531	0.631411	0.572749

Table 4: Rescaled Car Insurance Data

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

6 Conclusion

This paper 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 a 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.

References

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

[2] Antonio, K. and J. Beirlant, 2006. Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, in press.

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

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

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

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

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

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

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

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

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

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

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

[14] McCullagh, P. and J.A. Nelder, 1989. *Generalized Linear Models*. Chapman and Hall, New York.

[15] McCulloch, C.E. and S.R. Searle, 2001. *Generalized, Linear and Mixed Models*. Wiley, New York.

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

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

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