FULL CREDIBILITY WITH GENERALIZED LINEAR AND MIXED MODELS

BY

JOSÉ GARRIDO* AND JUN ZHOU

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 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. 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, 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.

* This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant 36860–06.

The paper is organized as follows: Section 2 briefly reviews the basic concepts of GLMs and GLMMs. Section 3.1 gives *limited fluctuations credibility* results for GLMs and GLMMs. Section 3.2 studies the choice of the link function and its effect on credibility. Section 3.3 illustrates with some numerical examples the main results of the paper. Detailed 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.

Hardin and Hilbe (2007) provide a handbook for data analysis with GLMs and GLM extensions. Lee et al. (2007) is a comprehensive reference for GLMs with random effects. GLMMs are an extension of GLMs, complicated by random effects. 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.

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, with density function taking the form

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

   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

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

for some vector of parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and covariate $\boldsymbol{X}_i = (x_{i1}, \ldots, x_{ip})'$ associated to the 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

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

Example 2.1. *GLMs commonly used in credibility theory*

Table 1 below gives the different model components of the GLMs most commonly used in credibility theory for observed claim counts or claim severities.

TABLE 1

GLM EXAMPLES

$Y\sim$	**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	$1/\mathrm{m}$
$c(y, \phi)$	$-\frac{1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]$	$\alpha \ln \alpha y + \ln y - \ln\Gamma(\alpha)$	$-\ln(y!)$	$\ln\binom{m}{my}$
Link g	identity	reciprocal	log	logit

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

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

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

and its derivative:

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

where

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

Hence

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

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

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

Lemma 2.1. *For the MLE, $\widehat{\boldsymbol{\beta}}$ that solves* (2.5)*, we have:*

1. *$\widehat{\boldsymbol{\beta}}$ is an asymptotically unbiased and consistent estimator of $\boldsymbol{\beta}$.*
2. *$\mathbb{V}(\widehat{\boldsymbol{\beta}}) \to \Sigma = -\mathbf{H}^{-1}$, as $n \to \infty$. $\mathbf{H} = -\mathbf{X}'\mathbf{W_o}\mathbf{X}$ 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 $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$.*
3. *$\widehat{\boldsymbol{\beta}} \xrightarrow{d} N(\boldsymbol{\beta}, \Sigma)$, hence there is convergence in distribution.*

For a proof see Fahrmeir and Kaufmann (1985).

Note that for a finite sample, the MLE $\widehat{\boldsymbol{\beta}}$ is usually biased. Hence its mean square error $\mathrm{MSE}(\widehat{\boldsymbol{\beta}}) = \mathbb{V}(\widehat{\boldsymbol{\beta}}) + \mathrm{bias}(\widehat{\boldsymbol{\beta}})^2$ plays an important role. We will see in Section 3.2 that this finite-sample bias is affected by the choice of link function g (see Cordeiro and McCullagh, 1991).

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

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

where $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ik})$ 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 $\mu_{ij} = \mathbb{E}[Y_{ij} \mid \boldsymbol{U}_i = \boldsymbol{u}_i] = b'(\theta_{ij})$. The variance of the observations, conditional on the random effects, is given by $\mathbb{V}[Y_{ij} \mid \boldsymbol{U}_i = \boldsymbol{u}_i] = \mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2}$. 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:

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

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

3. A link function g,

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

completes the model.

Most estimation methods for $\boldsymbol{\beta}$ and $\boldsymbol{u}_i$ 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 (2.6)-(2.8), the log-likelihood takes the form

$$l(\boldsymbol{\beta}, \mathbf{D}) = -\frac{nk}{2}\ln(2\pi) - \frac{n}{2}|\mathbf{D}| + \sum_{i=1}^{n} \ln \int_{R^k} e^{l_i(\boldsymbol{\beta}, \boldsymbol{\upsilon}) - \frac{1}{2}\boldsymbol{\upsilon}'\mathbf{D}^{-1}\boldsymbol{\upsilon}} d\boldsymbol{\upsilon}, \tag{2.9}$$

where

$$l_i(\boldsymbol{\beta}, \boldsymbol{u}_i) = \sum_{j=1}^{n_i} \left[\left(\boldsymbol{X}_{ij}'\boldsymbol{\beta} + T_{ij}'\boldsymbol{u}_i\right) y_{ij} - b\left(\boldsymbol{X}_{ij}'\boldsymbol{\beta} + T_{ij}'\boldsymbol{u}_i\right)\right] \tag{2.10}$$

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

As explained in SAS/STAT (2006, pp. 119-121), there are two types of numerical algorithms to solve for (2.9). 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 covergence 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).

2.2.1. Pseudo-likelihood estimation based on linearization

From (2.7)-(2.8) and SAS/STAT (2006) we have that $\mathbb{E}[\boldsymbol{Y}_i | \boldsymbol{U}_i = \boldsymbol{u}_i] = g^{-1}(\boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{T}_i \boldsymbol{u}_i) = g^{-1}(\boldsymbol{\eta}_i) = \boldsymbol{\mu}_i$ for $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})'$, $\boldsymbol{X}_i = (X'_{i1}, \dots, X'_{in_i})$, $\boldsymbol{T}_i = (T'_{i1}, \dots, T'_{in_i})$, $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{in_i})'$ and $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{in_i})'$. The first Taylor series of $\boldsymbol{\mu}_i$ about $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{u}}_i$ yields

$$g^{-1}(\boldsymbol{\eta}_i) \doteq g^{-1}(\widehat{\boldsymbol{\eta}}_i) + \widehat{\boldsymbol{\Delta}}_i \boldsymbol{X}_i (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{\Delta}}_i \boldsymbol{T}_i (\boldsymbol{u}_i - \widehat{\boldsymbol{u}}_i), \tag{2.11}$$

where

$$\widehat{\boldsymbol{\Delta}}_i = \left.\frac{\partial g^{-1}(\boldsymbol{\eta}_i)}{\partial \boldsymbol{\eta}_i}\right|_{\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{u}}_i}, \tag{2.12}$$

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

$$\widehat{\boldsymbol{\Delta}}_i^{-1}\left(\boldsymbol{\mu}_i - g^{-1}(\widehat{\boldsymbol{\eta}}_i)\right) + \boldsymbol{X}_i \widehat{\boldsymbol{\beta}} + \boldsymbol{T}_i \widehat{\boldsymbol{u}}_i \doteq \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{T}_i \boldsymbol{u}_i. \tag{2.13}$$

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

$$\widehat{\boldsymbol{\Delta}}_i^{-1}\left(\boldsymbol{Y}_i - g^{-1}(\widehat{\boldsymbol{\eta}}_i)\right) + \boldsymbol{X}_i \widehat{\boldsymbol{\beta}} + \boldsymbol{T}_i \widehat{\boldsymbol{u}}_i \equiv \boldsymbol{P}_i \tag{2.14}$$

and the variance-covariance matrix

$$\mathbb{V}[\boldsymbol{P}_i | \boldsymbol{u}_i] = \widehat{\boldsymbol{\Delta}}_i^{-1} \mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2} \widehat{\boldsymbol{\Delta}}_i^{-1}. \tag{2.15}$$

One can thus consider the model

$$\boldsymbol{P}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{T}_i \boldsymbol{u}_i + \varepsilon_i, \tag{2.16}$$

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

Now define

$$\boldsymbol{V}(\boldsymbol{\theta}_i) = \boldsymbol{T}_i \mathbf{D} \boldsymbol{T}_i + \widehat{\boldsymbol{\Delta}}_i^{-1} \mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2} \widehat{\boldsymbol{\Delta}}_i^{-1}, \tag{2.17}$$

as the marginal variance function in the linear mixed pseudo-model, where $\boldsymbol{\theta}_i$ 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 $\boldsymbol{P}_i$ is known. The maximum log pseudo-likelihood, $l(\boldsymbol{\theta}, \boldsymbol{P})$, for all $\boldsymbol{\theta}_i$ and $\boldsymbol{P}_i$ is then given by

$$l(\boldsymbol{\theta}, \boldsymbol{P}) = -\frac{1}{2}\left[\sum_{i=1}^{n} \ln\left|\boldsymbol{V}(\boldsymbol{\theta}_i)\right| - \sum_{i=1}^{n} \boldsymbol{r}_i' \boldsymbol{V}(\boldsymbol{\theta}_i)^{-1} \boldsymbol{r}_i - f \ln(2\pi)\right], \tag{2.18}$$

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

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{V}(\widehat{\boldsymbol{\theta}}_i)^{-1} \boldsymbol{X}_i\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{V}(\widehat{\boldsymbol{\theta}}_i)^{-1} \boldsymbol{P}_i\right), \tag{2.19}$$

$$\widehat{\boldsymbol{u}}_i = \widehat{\mathbf{D}} \mathbf{T}_i' \boldsymbol{V}(\widehat{\boldsymbol{\theta}}_i)^{-1} (\boldsymbol{P}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}). \tag{2.20}$$

3. Full Credibility Theory for GLMs and GLMMs

3.1. 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, Nelder and Verrall (1997) showed how credibility theory can be encompassed within the theory of GLMs. In that vein Schmitter (2004) gave 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

$$\mathbb{P}\{|\hat{\mu}_i - \mu_i| \leq r\mu_i\} \geq p_i, \quad i = 1, \ldots, n, \tag{3.1}$$

for a chosen estimation-error tolerance level $0 < r < 1$ and confidence probability p_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 the probability*

$$\begin{aligned}\pi_i &= \mathbb{P}\{|\hat{\mu}_i - \mu_i| \leq r\mu_i\} = \mathbb{P}\{(1-r)\,\mu_i \leq \hat{\mu}_i \leq (1+r)\,\mu_i\} \\ &= \mathbb{P}\{g[(1-r)\,\mu_i] - g(\mu_i) \leq g(\hat{\mu}_i) - g(\mu_i) \leq g[(1+r)\,\mu_i] - g(\mu_i)\} \\ &= \mathbb{P}\{g[(1-r)\,\mu_i] - \boldsymbol{X}_i'\boldsymbol{\beta} \leq \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta} \leq g[(1+r)\,\mu_i] - \boldsymbol{X}_i'\boldsymbol{\beta}\}. \end{aligned} \tag{3.2}$$

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

Proposition 3.1 gives some expressions equivalent to (3.1) 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 $\boldsymbol{X}_i'\widehat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta}$ in (3.2) depend on the parameters in $\boldsymbol{\beta}$. 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 (3.2) admits a simpler form as follows.

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

$$\mathbb{P}\{|\hat{\mu}_i - \mu_i| \le r\mu_i\} = \mathbb{P}\{c_1 \le \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta} \le c_2\}, \quad i = 1, \ldots, n, \tag{3.3}$$

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

$$c_1 = c\ln(1-r) \quad \textit{and} \quad c_2 = c\ln(1+r). \tag{3.4}$$

Proof: ($\Leftarrow$) If $g(x) = c\ln(x) + \tau$, by (3.2), 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

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

then from (3.2), for any μ_i,

$$c_1 = g[(1-r)\mu_i] - g(\mu_i) \quad \text{and} \quad c_2 = g[(1+r)\mu_i] - g(\mu_i). \tag{3.5}$$

Assuming that g is differentiable, then for any μ_i

$$g'(\mu_i) = \lim_{r \to 0} \frac{g[(1-r)\mu_i] - g(\mu_i)}{-r\mu_i} = \lim_{r \to 0} \frac{c_1}{-r\mu_i} \tag{3.6}$$

but also

$$g'(\mu_i) = \lim_{r \to 0} \frac{g[(1+r)\mu_i] - g(\mu_i)}{r\mu_i} = \lim_{r \to 0} \frac{c_2}{r\mu_i}. \tag{3.7}$$

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 μ_i. They 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 $\mathbb{E}(Y_i) = \mu_i = e^{x_{i1}\beta_1 + \cdots + x_{ip}\beta_{ip}}$. With the log-link function, $g[\mathbb{E}(Y_i)] = g(\mu_i) = x_{i1}\beta_1 + \cdots + x_{ip}\beta_{ip}$. By (3.2), $|\hat{\mu}_i - \mu_i| \le r\mu_i \Leftrightarrow \ln(1-r) \le \boldsymbol{X}_i'\hat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta} \le \ln(1+r)$. Since $0 < r < 1$, then $|\ln(1+r)| < |\ln(1-r)|$ and hence

$$\mathbb{P}\{|\hat{\mu}_i - \mu_i| \le r\mu_i\} = \mathbb{P}\{\ln(1-r) \le \boldsymbol{X}_i'\hat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta} \le \ln(1+r)\}$$
$$\le \mathbb{P}\{|\boldsymbol{X}_i'\hat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta}| \le |\ln(1-r)|\}. \tag{3.8}$$

Now let $s^2 = \mathbb{V}(\hat{\beta}_1 + \cdots + \hat{\beta}_p)$ and $\boldsymbol{X}_i = (1, 1, \ldots, 1)$, then (3.8) becomes

$$\mathbb{P}\{|\boldsymbol{X}_i'\hat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta}| \le |\ln(1-r)|\}$$
$$= \mathbb{P}\{|(\hat{\beta}_1 + \cdots + \hat{\beta}_p) - (\beta_1 + \cdots + \beta_p)| \le |\ln(1-r)|\}$$
$$= \mathbb{P}\left\{\left|\frac{\left(\hat{\beta}_1 + \cdots + \hat{\beta}_p\right) - \left(\beta_1 + \cdots + \beta_p\right)}{s}\right| \le \frac{|\ln(1-r)|}{s}\right\}. \tag{3.9}$$

Approximating by a normal distribution, (3.9) yields $\frac{|\ln(1-r)|}{s} \ge Z_{\pi*}$, where $Z_{\pi*}$ is the $\pi_* = 100[1 - \left(\frac{1-\pi}{2}\right)]$-percentile of a standard normal distribution. Hence the following asymptotic full credibility criterium is obtained:

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

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 $\hat{\mu}_i = \boldsymbol{X}_i'\hat{\boldsymbol{\beta}}$.

Proposition 3.3. *Let* $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j} = (\mathbf{X}'\mathbf{W_0}\mathbf{X})^{-1}$ *and* $s_i^2 = \mathbb{V}(\hat{\mu}_i) = \mathbb{V}(\boldsymbol{X}_i'\hat{\boldsymbol{\beta}})$. *Then for every component* $i = 1, \ldots, n$,

$$s_i^2 \to \boldsymbol{X}_i'\Sigma\,\boldsymbol{X}_i, \tag{3.10}$$

as $n \to \infty$, *where* $\boldsymbol{X}_i$, $\mathbf{W_0}$ *and* $\mathbf{X}$ *are given in Lemma 2.1.*

Proof: From Lemma 2.1-(2) we have that $\mathbb{V}(\hat{\boldsymbol{\beta}}) \to \boldsymbol{\Sigma}$, as $n \to \infty$, and the iterative $\hat{\boldsymbol{\beta}}$ converges to the true $\boldsymbol{\beta}$, then

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

□

Furthermore, Lemma 2.1 states that $\widehat{\boldsymbol{\beta}}$ converges to $N(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ in distribution. Then, the following corollary to Proposition 3.3 holds.

Corollary 3.1. $(\boldsymbol{X}_i'\widehat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta})/s_i$ *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 3.1. *For the log-link function, an asymptotic normal approximation gives*

$$
\pi_i \doteq \Phi\left(\frac{\ln(1+r)}{s_i}\right) - \Phi\left(\frac{\ln(1-r)}{s_i}\right), \quad i = 1, \ldots, n, \tag{3.11}
$$

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

Proof: From Propositions 3.1 and 3.2,

$$
\begin{aligned}
\pi_i &= \mathbb{P}\left\{\ln(1-r) \le \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta} \le \ln(1+r)\right\} \\
&= \mathbb{P}\left\{\frac{\ln(1-r)}{s_i} \le \frac{\boldsymbol{X}_i'\widehat{\boldsymbol{\beta}} - \boldsymbol{X}_i'\boldsymbol{\beta}}{s_i} \le \frac{\ln(1+r)}{s_i}\right\}.
\end{aligned}
$$

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

For any confidence coefficient π_i, Theorem 3.1 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 $\boldsymbol{X}_i$. The examples in Section 3.3 illustrate the above results.

Now for a general link function g, let

$$
Q_1 = g[(1-r)\mu_i] - g(\mu_i) \quad \text{and} \quad Q_2 = g[(1+r)\mu_i] - g(\mu_i) \tag{3.12}
$$

Theorem 3.2. *For a monotonic increasing link function g, we have the following asymptotic approximation:*

$$
\pi_i \doteq \Phi\left(\frac{Q_2}{s_i}\right) - \Phi\left(\frac{Q_1}{s_i}\right), \quad i = 1, \ldots, n, \tag{3.13}
$$

where Φ is the cdf of the standard normal distribution, Q_1 and Q_2 are given in (3.12) *and s_i in Proposition 3.3.*

Proof:

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

Approximating by the normal distribution gives (3.13). □

Clearly, the smaller s_i the bigger π_i (approximately), which differs for different i. If g is the log-link function, then Proposition 3.2 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

$$\widehat{Q}_1 = g[(1-r)\hat{\mu}_i] - g(\hat{\mu}_i) \quad \text{and} \quad \widehat{Q}_2 = g[(1+r)\hat{\mu}_i] - g(\hat{\mu}_i), \tag{3.14}$$

which then implies that

$$\hat{\pi}_i \doteq \Phi\left(\frac{\widehat{Q}_2}{s_i}\right) - \Phi\left(\frac{\widehat{Q}_1}{s_i}\right). \tag{3.15}$$

Section 3.2 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 3.4. *For any generalized linear mixed model, as defined in* (2.6)-(2.8)*, let g be a monotonic increasing link function. Then*

$$\begin{aligned}\pi_{ij} &= \mathbb{P}\{|\hat{\mu}_{ij} - \mu_{ij}| \le r\mu_{ij}\} = \mathbb{P}\{(1-r)\mu_{ij} \le \hat{\mu}_{ij} \le (1+r)\mu_{ij}\} \\ &= \mathbb{P}\{g[(1-r)\mu_{ij}] - g(\mu_{ij}) \le g(\hat{\mu}_{ij}) - g(\mu_{ij}) \le g[(1+r)\mu_{ij}] - g(\mu_{ij})\} \\ &= \mathbb{P}\{g[(1-r)\mu_{ij}] - \boldsymbol{X}_{ij}'\boldsymbol{\beta} - T_{ij}'\boldsymbol{u}_i \le \boldsymbol{X}_{ij}'\widehat{\boldsymbol{\beta}} + T_{ij}'\widehat{\boldsymbol{u}}_i - \boldsymbol{X}_{ij}'\boldsymbol{\beta} - T_{ij}'\boldsymbol{u}_i \\ &\qquad \le g[(1+r)\mu_{ij}] - \boldsymbol{X}_{ij}'\boldsymbol{\beta} - T_{ij}'\boldsymbol{u}_i\}.\end{aligned} \tag{3.16}$$

Using the same idea as in Theorem 3.2 (see Liang and Zeger, 1986 for the asymptotic normal distribution of the GLMM estimators), we obtain the following result for GLMMs.

Theorem 3.3. *For any link function g, let $s_{ij}^2 = \mathbb{V}(\boldsymbol{X}_{ij}'\boldsymbol{\beta} + T_{ij}'\boldsymbol{u}_i)$ and Q_{1j}, Q_{2j} be defined as*

$$Q_{1j} = g[(1-r)\mu_{ij}] - g(\mu_{ij}) \quad \textit{and} \quad Q_{2j} = g[(1+r)\mu_{ij}] - g(\mu_{ij}), \tag{3.17}$$

then

$$\pi_{ij} \doteq \Phi\left(\frac{Q_{2j}}{s_{ij}}\right) - \Phi\left(\frac{Q_{1j}}{s_{ij}}\right), \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i. \tag{3.18}$$

3.2. The choice of link function

As shown in the previous section, the main idea here is to transfer the full credibility condition (3.1) to an equivalent form that is easier to implement, as in Theorems 3.1-3.2. Expression (3.13) 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 3.1. *Rescaling or shifting a given link function g, such as in* $h(x) = cg(x) + \tau$, *does not affect the approximate confidence probabilities* π_i in (3.13).

Proof: For a link function g, (2.3) can be rewritten as $g(\mu_i) = \beta_0^{(g)} + \boldsymbol{X}_i'\boldsymbol{\beta}^{(g)}$, where $\beta_0^{(g)}$ is the intercept. Let the new link function be $h(x) = cg(x) + \tau$. Then $h(\mu_i) = \beta_0^{(h)} + \boldsymbol{X}_i'\boldsymbol{\beta}^{(h)} = cg(\mu_i) + \tau$ and hence $g(\mu_i) = \frac{\beta_0^{(h)} - \tau}{c} + \boldsymbol{X}_i'\frac{\boldsymbol{\beta}^{(h)}}{c}$. It follows that $\beta_0^{(g)} = \frac{\beta_0^{(h)} - \tau}{c}$ and $\boldsymbol{\beta}^{(g)} = \frac{\boldsymbol{\beta}^{(h)}}{c}$.

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

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

for $i = 1, 2$. Refer to (3.13) and substitute $Q_i^{(h)}$ and $s_i^{(h)}$ above, to see that

$$\pi_i^{(g)} \doteq \Phi\left(\frac{Q_2^{(g)}}{s_i^{(g)}}\right) - \Phi\left(\frac{Q_1^{(g)}}{s_i^{(g)}}\right) = \Phi\left(\frac{cQ_2^{(g)}}{cs_i^{(g)}}\right) - \Phi\left(\frac{cQ_1^{(g)}}{cs_i^{(g)}}\right) \doteq \pi_i^{(h)}.$$

□

Example 3.4 gives a numerical illustration of Lemma 3.1. It shows how the estimated probabilities π_i, in (3.13), but where s_i is estimated with $\widehat{s}_i$ 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, $\widehat{\boldsymbol{\beta}}$, $\widehat{\mu}_i = g^{-1}(\boldsymbol{X}_i'\widehat{\boldsymbol{\beta}})$ and in our estimated $\widehat{Q}_1$, $\widehat{Q}_2$ in (3.14). 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 3.2. ***(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*

$$\mathbb{E}[\varphi(X)] \geq (resp. \leq)\ \varphi(\mathbb{E}[X]). \tag{3.19}$$

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 3.4. *$\widehat{Q}_1$ and $\widehat{Q}_2$ in* (3.14) *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 $\widehat{Q}_1 = g[(1-r)\widehat{\mu}_i] - g(\widehat{\mu}_i)$ and $Q_1 = g[(1-r)\mu_i] - g(\mu_i)$, where $g(\mu_i) = \boldsymbol{X}_i'\boldsymbol{\beta}$ and $g(\widehat{\mu}_i) = \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}}$. Then

$$\begin{aligned}
\text{bias}(\widehat{Q}_1) &= \mathbb{E}(\widehat{Q}_1) - Q_1 \\
&= \mathbb{E}\{g[(1-r)\widehat{\mu}_i]\} - \mathbb{E}[\boldsymbol{X}_i'\widehat{\boldsymbol{\beta}}] - g[(1-r)\mu_i] + \boldsymbol{X}_i'\boldsymbol{\beta} \\
&= \mathbb{E}\{g[(1-r)\widehat{\mu}_i]\} - g[(1-r)\mu_i] - \boldsymbol{X}_i'\text{bias}(\widehat{\beta}).
\end{aligned} \tag{3.20}$$

Three cases need to be distinguished:

1. If g is linear then $\mathbb{E}\{g[(1-r)\widehat{\mu}_i]\} - g[(1-r)\mu_i] = 0$ and $\widehat{\beta}$ is unbiased, hence so is $\widehat{Q}_1$.
2. If g is a convex upward decreasing function, then by Jensen's inequality in (3.19)

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

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

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

   and $\widehat{\boldsymbol{\beta}}$ is asymptotically unbiased, then asymptotically $\mathbb{E}(\widehat{Q}_1) - Q_1 \geq 0$. Hence $\widehat{Q}_1$ 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 $\mathbb{E}(\widehat{Q}_1) - Q_1 \leq 0$ and $\widehat{Q}_1$ is an asymptotically downward-biased estimator.

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

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 criterium for the choice of the link function.

Theorem 3.5. *For a GLM problem,* $\hat{\pi}_i$ *given by* (3.15) *can be used as a criterium 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*.*

3.3. Numerical examples

Example 3.2. *Car Insurance Claims Data (GLM)*

TABLE 2

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

The SAS Technical Report P-243 (1993) gives the illustrative dataset in Table 6 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).

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

$$\mathrm{x}_{i1} = 1,$$

$$x_{i2} = \begin{cases} 1 & \text{if } \textit{car type} \text{ is } \textit{large} \\ 0 & \text{otherwise,} \end{cases}$$

$$x_{i3} = \begin{cases} 1 & \text{if } \textit{car type} \text{ is } \textit{medium} \\ 0 & \text{otherwise,} \end{cases}$$

$$x_{i4} = \begin{cases} 1 & \text{if } \textit{age group} \text{ is } \textit{1} \\ 0 & \text{otherwise.} \end{cases}$$

In this notation $\boldsymbol{X}_4 = (1,0,0,0)'$ 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 $\boldsymbol{\Sigma}$ in Proposition 3.3 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 2.1-2).

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

Let the tolerance level $r = 0.1$ and $\boldsymbol{X}_3 = (1,1,0,1)'$ for the third class of drivers, i.e. with a *large* car type in age group *1*. Then the asymptotic value in (3.10) for $s_3^2 = \boldsymbol{X}_3'\boldsymbol{\Sigma}\boldsymbol{X}_3 = 0.082236$ and from (3.11) we get $\pi_3 = \Phi\left(\frac{\ln(1+r)}{s_3}\right) - \Phi\left(\frac{\ln(1-r)}{s_3}\right) = 0.273533$. 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 $\boldsymbol{X}_1 = (1,0,0,1)'$ 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 3 reports the asymptotic variances $s_i^2 = \mathbb{V}(\boldsymbol{X}_{ij}'\boldsymbol{\beta}) \to \boldsymbol{X}_i'\boldsymbol{\Sigma}\boldsymbol{X}_i$ and the credibility probabilities π_i for all 6 classes.

TABLE 3

ASYMPTOTIC VARIANCES $\mathbf{X}_i'\boldsymbol{\Sigma}\boldsymbol{X}_i$ AND CONFIDENCE PROBABILITIES π_i

Class	$\boldsymbol{i = 1}$	**2**	**3**	**4**	**5**	**6**
$\boldsymbol{X}_i'$	(1,0,0,1)	(1,0,1,1)	(1,1,0,1)	(1,0,0,0)	(1,0,1,0)	(1,1,0,0)
$\boldsymbol{X}_i'\boldsymbol{\Sigma}\boldsymbol{X}_i$	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

Example 3.3. *Effect of Sample Size (GLM)*

Furthermore, if we modify Example 3.2 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 3.4. *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 6, but rearrange the claim counts in each group as in Table 4.

TABLE 4

MODIFIED CAR INSURANCE DATA

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

Then for $X_3 = (1,1,0,1)'$ 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 3.2. Clearly the credibility of GLM estimates depends on the distribution of the covariates.

Example 3.5. *Effect of the Link Function (GLM)*

Let the link function $g(x) = c\ln(x) + \tau$. Lemma 3.1 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 5 shows that the estimated confidence values π_i in (3.13), for classes $i = 1$ and 3 remain essentially the same.

TABLE 5

RESCALED CAR INSURANCE DATA

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

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

Example 3.6. *GLMM with Territory as Random Effect*

This example illustrates the credibility results for GLMMs. We add one more variable, called territory, to Example 3.2. It takes two values, "rural" and "urban", which will illustrate the random effect of a GLMM. The fixed-effects parameters estimates, $\widehat{\boldsymbol{\beta}}$, those for the random-effects, $\widehat{\boldsymbol{u}}$, as well as their variance-covariance matrix were obtained with the GLIMMIX procedure in SAS (see SAS/STAT (2006)).

TABLE 6

CAR INSURANCE DATA WITH TERRITORY RANDOM EFFECTS

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

As in Example 3.2, let y_i be Poisson and g be a log-link function. Furthermore, let the covariates $\mathbf{X}_i = (x_{i1}, \ldots, x_{i5})'$ and $\mathbf{T}_j = (t_{1j}, t_{2j})'$, be coded as:

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

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

$$
\boldsymbol{\Sigma} = \begin{pmatrix} 0.016500 & -0.007440 & -0.007600 & -0.005230 \\ -0.007440 & 0.074590 & 0.006032 & 0.003173 \\ -0.007600 & 0.006032 & 0.017680 & -0.001150 \\ -0.005230 & 0.003173 & -0.001150 & 0.018250 \end{pmatrix}
$$

$$\mathbf{D} = \begin{pmatrix} 0.101262 & 0 \\ 0 & 0.101262 \end{pmatrix} \tag{3.21}$$

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 3.3 is $s_{13}^2 = \mathbf{X}_3'\mathbf{\Sigma}\mathbf{X}_3 + \mathbf{T}_1'\mathbf{D}\mathbf{T}_1 = 0.100608$. This gives $\pi_3 = \Phi\left(\frac{\ln(1+r)}{s_{13}}\right) - \Phi\left(\frac{\ln(1-r)}{s_{13}}\right) = 0.248217$. 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.

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 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.

Acknowledgments

The authors are grateful to the anonymous referees and to the Editor, Prof. Michel Denuit, for their constructive comments.

References

AITKIN, M., ANDERSON, D., FRANCIS, B. and HINDE, J. (1989) *Statistical Modelling in GLIM*. Oxford Science Publications, Oxford.

ANTONIO, K. and BEIRLANT, J. (2007) Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, **40(1)**, 58-76.

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).

BÜHLMANN, H. (1967) Experience rating and credibility. *ASTIN Bulletin*, **4(3)**, 199-207.

BÜHLMANN, H. (1969) Experience rating and credibility. *ASTIN Bulletin*, **5(2)**, 157-165.

BÜHLMANN, H. and STRAUB, E. (1970) Glaubwürdigkeit für Schadensätze. *Bulletin of the Swiss Association of Actuaries*, 111-133.

CORDEIRO, G.M. and MCCULLAGH, P. (1991) Bias correction in generalized linear models. *Journal of the Royal Statistical Society, B* **53(3)**, 629-643.

DEMINDENKO, E. (2004) *Mixed Models: Theory and Applications*. Hoboken, New Jersey.
DOBSON, A. (1990) *An Introduction to Generalized Linear Models*. Chapman and Hall, London.
FAHRMEIR, L. and KAUFMANN, H. (2003) Consistency and asymptotic normalty of the maximum likelihood estimator in generalized linear models. *The Annuals of Statistics*, **13(1)**, 342-368.
HABERMAN, S. and RENSHAW, A.E. (1996) Generalized linear models and actuarial science. *The Statistician*, **45(4)**, 407-436.
HACHEMEISTER, C.A. (1975) Credibility for regression models with application to trend. In *Credibility, Theory and Application*. Academic Press, New York, 129-163.
HARBIN, J.W. and HILBE, J.M. (2007) *Generalized Linear Models and Extensions*, 2nd Edition. Chapman and Hall/CRC, Boca Raton.
JEWELL, W.S. (1975) The use of collateral data in credibility theory: A hierarchical model. *Giornale dell'Instituto Italiano degli Attuari*, **38**, 1-16.
LEE, Y., NELDER, J.A. and PAWITAN, Y. (2007) *Generalized Linear Models With Random Effects*. Chapman and Hall/CRC, Boca Raton.
LIANG, K.Y. and ZEGER, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73(1)**, 13-22.
MCCULLAGH, P. and NELDER, J.A. (1989) *Generalized Linear Models*. Chapman and Hall, New-York.
MCCULLOCH, C.E. and SEARLE, S.R. (2001) *Generalized, Linear and Mixed Models*. Wiley, New York.
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.
NELDER, J.A. and VERRALL, R.J. (1997) Credibility theory and generalized linear models. *ASTIN Bulletin*, **27(1)**, 71-82.
SAS Technical Report P-243 (1993) *SAS/STAT Software: The GENMOD Procedure, Release 6.09*, SAS Institute Inc., Cary, NC.
SAS/STAT Software (2006) *GLIMMIX Procedure, Release 9.1*, June 2006, SAS Institute Inc., Cary, NC.
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.
SCHMITTER, H. (2004) The sample size needed for the calculation of a GLM tariff. *ASTIN Bulletin*, **34(1)**, 249-262.
WHITNEY, A.W. (1918) The theory of experience rating. *Proc. of the Casualty Actuarial Society*, **4**, 274-292.
WOLFINGER, R. and O'CONNELL, M. (1993) Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, **4**, 233-243.

JOSE GARRIDO (corresponding author)
Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd. West
Montreal, Qc
H3G 1M8, Canada
E-Mail: garrido@mathstat.concordia.ca

JUN ZHOU
Research and Development
Aviva Canada Inc.
8th floor, 630 René Lévesque W.
Montreal, QC
H3B 1S6, Canada
E-Mail: jun_zhou@avivacanada.com