ACTUARIAL STATISTICS AND MIXED MODELS: APPLICATIONS AND OPPORTUNITIES

Katrien Antonio and Jan Beirlant

University Center for Statistics, KULeuven, W. de Croylaan 54, 3001 Leuven, Belgium
Email: katrien.antonio@econ.kuleuven.be, jan.beirlant@wis.kuleuven.be

Abstract

The purpose of this paper is twofold. On the one hand, it is a short overview of our recent work on the use of mixed model methodology in actuarial statistics, which covers topics from credibility, claims reserving and non-life ratemaking. On the other hand, opportunities and challenges for future research are sketched.

1. INTRODUCTION

We discuss how mixed models can be applied in the analysis of insurance data and the decision making process following it. Starting point for the use of mixed models in actuarial statistics are traditional credibility models and their connection with linear mixed models. The credibility ratemaking problem concerns the prediction of future claims of a risk class, given past claims of that and related risk classes. Traditional credibility formulas can be reconstructed using the explicit expressions for the maximum likelihood estimations (MLE) of the fixed effects and the best linear unbiased predictor (BLUP) for the random effects in a linear mixed model. This appealing analogy was presented in Frees et al. (1999) and is a first step towards the interpretation of traditional credibility schemes in the framework of generalized linear models, using the methodology of generalized linear mixed models.

Next to the credibility ratemaking problem, examples from loss reserving and non-life ratemaking with mixed models are discussed. Using the concept of mixed models, their connection with smoothing methods and their implementation with Bayesian statistics, we present some new and promising alternatives for the techniques that are currently in use.

Section 2 contains a brief overview of the statistical concepts that are involved. In Section 3 some concrete examples are discussed and possibilities for further research are sketched. More details regarding the material presented here, are given in Antonio et al. (2006), Antonio and Beirlant (2006a) and Antonio and Beirlant (2006b).

2. STATISTICAL DETAILS

2.1. Linear mixed models (LMMs): specification and estimation

Linear mixed models extend classical linear regression models by incorporating random effects in the structure for the mean. Assume the data set at hand consists of N subjects. Let n_i denote the number of observations for subject i and $\boldsymbol{Y}_i$ its vector of observations ($1 \leq i \leq N$). The general linear mixed model is given by

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i. \tag{1}$$

$\boldsymbol{\beta}$ $(p \times 1)$ contains the parameters for the p fixed effects in the model; these are fixed, but unknown, regression parameters, common to all subjects. $\boldsymbol{b}_i$ $(q \times 1)$ is the vector with the random effects for the i^{th} subject in the data set. The use of random effects reflects the belief that there is heterogeneity among subjects for a subset of the regression coefficients in $\boldsymbol{\beta}$. $\boldsymbol{X}_i$ $(n_i \times p)$ and $\boldsymbol{Z}_i$ $(n_i \times q)$ are the design matrices for the p fixed and q random effects. $\boldsymbol{\epsilon}_i$ $(n_i \times 1)$ contains the residual components for subject i. Independence between subjects is assumed. $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ are also assumed to be independent and we follow the traditional assumption that they are normally distributed with mean vector $\boldsymbol{0}$ and covariance matrices, say $\boldsymbol{D}$ $(q \times q)$ and $\boldsymbol{\Sigma}_i$ $(n_i \times n_i)$, respectively. Different structures for these covariance matrices are possible; an overview of some frequently used ones can be found in Verbeke and Molenberghs (2000). It is easy to see that $\boldsymbol{Y}_i$ then has a marginal normal distribution with mean $\boldsymbol{X}_i\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{V}_i = \text{Var}(\boldsymbol{Y}_i)$, given by

$$\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^{'} + \boldsymbol{\Sigma}_i. \tag{2}$$

In this interpretation it becomes clear that the fixed effects enter only the mean $\text{E}[Y_{ij}]$, whereas the inclusion of subject-specific effects specifies the structure of the covariance between observations on the same unit.

Denote the unknown parameters in the covariance matrix $\boldsymbol{V}_i$ with $\boldsymbol{\alpha}$. Conditional on $\boldsymbol{\alpha}$, a closed form expression for the maximum likelihood estimator of $\boldsymbol{\beta}$ exists, namely

$$\hat{\boldsymbol{\beta}} = (\sum_{i=1}^{N} \boldsymbol{X}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{X}_i)^{-1}\sum_{i=1}^{N} \boldsymbol{X}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{Y}_i. \tag{3}$$

To predict the random effects, the mean of the posterior distribution of the random effects given the data, $\boldsymbol{b}_i|\boldsymbol{Y}_i$, is used. Conditional on $\boldsymbol{\alpha}$, we have

$$\hat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^{'}\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}), \tag{4}$$

which can be proven to be the Best Linear Unbiased Predictor (BLUP) of $\boldsymbol{b}_i$ (where 'best' is in the sense of minimal mean squared error). For estimation of $\boldsymbol{\alpha}$ maximum likelihood (ML) or restricted maximum likelihood (REML) is used. The expression maximized by the ML (L_1), respectively REML (L_2), estimates is given by

$$L_1(\boldsymbol{\alpha}; \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N) = c_1 - \frac{1}{2}\sum_{i=1}^{N} \log|\boldsymbol{V}_i| - \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{r}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i \tag{5}$$

$$L_2(\boldsymbol{\alpha}; \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N) = c_2 - \frac{1}{2}\sum_{i=1}^{N} \log|\boldsymbol{V}_i| - \frac{1}{2}\sum_{i=1}^{N} \log|\boldsymbol{X}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{X}_i| - \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{r}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \tag{6}$$

where $\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{X}_i \left(\sum_{i=1}^{N} \boldsymbol{X}_i^{'}\boldsymbol{V}_i\boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{y}_i\right)$ and c_1, c_2 are appropriate constants. Equations (5) and (6) are maximized using iterative numerical techniques such as Fisher scoring or Newton-Raphson. In (3) and (4) the unknown $\boldsymbol{\alpha}$ is then replaced with $\hat{\boldsymbol{\alpha}}_{ML}$ or $\hat{\boldsymbol{\alpha}}_{REML}$, leading to the empirical BLUE for $\boldsymbol{\beta}$ and the empirical BLUP for $\boldsymbol{b}_i$. For inference regarding the fixed and random effects and the variance components, appropriate likelihood ratio and Wald tests are explained in Verbeke and Molenberghs (2000).

2.2. Generalized linear mixed models (GLMMs): specification and estimation

GLMMs extend generalized linear models (GLMs) by allowing for random, or subject-specific, effects in the linear predictor. These models are useful when the interest of the analyst lies in the individual response profiles rather than the marginal mean $\mathrm{E}[Y_{ij}]$. The inclusion of random effects in the linear predictor reflects the idea that there is natural heterogeneity across subjects in (some of) their regression coefficients. Diggle et al. (2002) and Molenberghs and Verbeke (2005) are useful references for full details on GLMMs.

Say we have a data set at hand consisting of N subjects. For each subject i ($1 \leq i \leq N$), n_i observations are available. Given the vector $\boldsymbol{b}_i$ with the random effects for subject (or cluster) i, the repeated measurements $Y_{i1}, \ldots, Y_{in_i}$ are assumed to be independent with a density from the exponential family

$$f(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) = \exp\left(\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{\phi} + c(y_{ij}, \phi)\right), \; j = 1, \ldots, n_i. \tag{7}$$

Similar to a GLM, the following (conditional) relations hold

$$\mu_{ij} = \mathrm{E}[Y_{ij}|\boldsymbol{b}_i] = \psi^{'}(\theta_{ij}) \text{ and } \mathrm{Var}[Y_{ij}|\boldsymbol{b}_i] = \phi\psi^{''}(\theta_{ij}) = \phi V(\mu_{ij}) \tag{8}$$

where $g(\mu_{ij}) = \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{'}\boldsymbol{b}_i$. As before, $g(\cdot)$ is called the link and $V(\cdot)$ the variance function. $\boldsymbol{\beta}$ ($p \times 1$) denotes the fixed effects parameter vector and $\boldsymbol{b}_i$ ($q \times 1$) the random effects vector. $\boldsymbol{x}_{ij}$ ($p \times 1$) and $\boldsymbol{z}_{ij}$ ($q \times 1$) contain subject i's covariate information for the fixed and random effects, respectively. The specification of the GLMM is completed by assuming that the random effects, $\boldsymbol{b}_i$ ($i = 1, \ldots, N$), are mutually independent and identically distributed with density function $f(\boldsymbol{b}_i|\boldsymbol{\alpha})$. Hereby $\boldsymbol{\alpha}$ denotes (again) the unknown parameters in the density. Traditionally, one works under the assumption of (multivariate) normally distributed random effects with zero mean and covariance matrix determined by $\boldsymbol{\alpha}$. Correlation between observations on the same subject arises because they share the same random effects $\boldsymbol{b}_i$.

The likelihood function for the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and ϕ then becomes (with $\boldsymbol{y} = (\boldsymbol{y}_1^{'}, \ldots, \boldsymbol{y}_N^{'})^{'}$)

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi; \boldsymbol{y}) &= \prod_{i=1}^{N} f(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi) \\ &= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) f(\boldsymbol{b}_i|\boldsymbol{\alpha}) d\boldsymbol{b}_i, \end{aligned} \tag{9}$$

where the integral is with respect to the q dimensional vector $\boldsymbol{b}_i$. When both the data and the random effects are normally distributed (as in the linear mixed model), the integral can be worked out analytically and closed-form expressions exist for the maximum likelihood estimator of $\boldsymbol{\beta}$ and the BLUP for $\boldsymbol{b}_i$ (see (3) and (4)). For general GLMMs, however, approximations to the likelihood or numerical integration techniques are required to maximize equation (9) with respect to the unknown parameters. Restricted pseudo-likelihood ((RE)PL) (Wolfinger and O'Connell (1993)) and (adaptive) Gauss-Hermite quadrature (Liu and Pierce (1994)) are two widely used techniques to perform the maximum likelihood estimation. Both techniques are available in the commercial software package SAS and their use will be illustrated in Section 3. The pseudo-likelihood technique corresponds with the penalized quasi-likelihood (PQL) method of Breslow and Clayton (1993). Since maximum likelihood techniques are hindered by the integration over the q-dimensional vector of random effects, a Bayesian implementation of GLMMs is considered as well. Hereby random numbers are drawn from the relevant posterior and predictive distributions using Markov Chain Monte Carlo (MCMC) techniques. WINBUGS allows easy implementation of these models. Illustrative code for both SAS and WINBUGS is available on the web [1].

2.3. Smoothing with mixed models

To provide some background for smoothing with mixed model methodology, let us start from the simple example of scatterplot smoothing. Data (x_i, y_i) $(i = 1, \ldots, n)$ are given and the model $Y_i = f(x_i) + \epsilon_i$ $(i = 1, \ldots, n)$ is fitted. To estimate the unknown function $f(\cdot)$, a linear combination of some basis functions is used. Possible basis functions are *truncated power basis functions*, *B-splines* or *radial basis functions*, among others. For truncated power basis functions of degree p with K knots $\kappa_1, \ldots, \kappa_K$ [2], define the design matrix $\boldsymbol{B}$ as

$$\boldsymbol{B} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^p & (x_1 - \kappa_1)_+^p & \ldots & (x_1 - \kappa_K)_+^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^p & (x_n - \kappa_1)_+^p & \ldots & (x_n - \kappa_K)_+^p \end{bmatrix}. \tag{10}$$

The unknown function $f(\cdot)$ is then estimated as $\widehat{f}(x) = \boldsymbol{B}(x)\widehat{\boldsymbol{\beta}}$ where $\boldsymbol{B}(x)$ is a row vector, similar to a row from $\boldsymbol{B}$, and $\widehat{\boldsymbol{\beta}}$ is the solution of the least-squares problem $\min_{\boldsymbol{\beta}} \sum_{i=1}^n (y_i - \boldsymbol{B}(x_i)\boldsymbol{\beta})^2$, subject to the constraint $\sum_{k=1}^K \beta_{pk}^2 < C$ to obtain a smooth fit. Hereby, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p, \beta_{p1}, \ldots, \beta_{pK})'$ and thus the penalized coefficients correspond with the truncated power functions. Using a Lagrange multiplier argument, this optimization problem is rewritten as

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n (y_i - \boldsymbol{B}(x_i)\boldsymbol{\beta})^2 + \alpha \boldsymbol{\beta}' \boldsymbol{P} \boldsymbol{\beta}, \tag{11}$$

where α is the so-called smoothing parameter and $\boldsymbol{P}$ a penalty matrix given by

$$\boldsymbol{P} = \begin{bmatrix} 0_{p+1 \times p+1} & 0_{p+1 \times K} \\ 0_{K \times p+1} & \mathbf{I}_{K \times K} \end{bmatrix}. \tag{12}$$

[1] see http://www.econ.kuleuven.be/katrien.antonio

[2] The truncated line $(x - \kappa_k)_+$ is zero, when $x < \kappa_k$ and equals $x - \kappa_k$ elsewhere. $(x - \kappa_k)_+^p$ has to be interpreted as $\{(x - \kappa_k)_+\}^p$. The basis functions $\{1, x, x^2, \ldots, x^p, (x - \kappa_1)_+^p, \ldots, (x - \kappa_K)_+^p\}$ span the vector space of piecewise functions of degree p with knots at $\kappa_1, \ldots, \kappa_K$.

Ruppert et al. (2003) (among others) rewrite the argument of the optimization problem in (11), after dividing by σ_ϵ^2, as

$$\frac{1}{\sigma_\epsilon^2}\|\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Zu}\|^2+\frac{1}{\sigma_u^2}\|\boldsymbol{u}\|^2, \tag{13}$$

where $\sigma_u^2=\sigma_\epsilon^2/\alpha$, $\boldsymbol{y}=(y_1,\ldots,y_n)'$, $\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_p)'$ (i.e. the regression parameters for the basis functions $1,\ x,\ x^2,\ldots,\ x^p$), $\boldsymbol{u}=(\beta_{p1},\ldots,\beta_{pK})'$,

$$\boldsymbol{X}=\begin{bmatrix}1 & x_1 & x_1^2 & \ldots & x_1^p\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_n & x_n^2 & \ldots & x_n^p\end{bmatrix} \text{ and } \boldsymbol{Z}=\begin{bmatrix}(x_1-\kappa_1)_+^p & \ldots & (x_1-\kappa_K)_+^p\\ \vdots & \vdots & \vdots\\ (x_n-\kappa_1)_+^p & \ldots & (x_n-\kappa_K)_+^p\end{bmatrix}. \tag{14}$$

By considering $\boldsymbol{u}$ as random effects with $\boldsymbol{u}\sim N(0,\sigma_u^2\boldsymbol{I}_{K\times K})$, (13) reduces to minus two times the log-likelihood of $(\boldsymbol{Y},\boldsymbol{u})$ in the linear mixed model $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{Zu}+\boldsymbol{\epsilon}$, under the assumptions $\boldsymbol{Y}|\boldsymbol{u}\sim N(\boldsymbol{X\beta}+\boldsymbol{Zu},\sigma_\epsilon^2\boldsymbol{I})$, $\boldsymbol{u}\sim N(\boldsymbol{0},\sigma_u^2\boldsymbol{I})$ and $\boldsymbol{\epsilon}\sim N(\boldsymbol{0},\sigma_\epsilon^2\boldsymbol{I})$.

A similar reasoning leads to the penalized splines formulation of a GAM, where $Y_1,\ldots,Y_n$ are independent random variables with a density $f(\cdot)$ from the exponential family and an additive predictor $\eta_i=\sum_{h=1}^{l}f_h(x_{ih})$ $(i=1,\ldots,n)$. Construct the design matrix $\boldsymbol{X}$ as

$$\boldsymbol{X}=\left[\begin{array}{ccccc|c|cccc}1 & x_{11} & x_{11}^2 & \ldots & x_{11}^p & \ldots & x_{1l} & x_{1l}^2 & \ldots & x_{1l}^p\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_{n1} & x_{n1}^2 & \ldots & x_{n1}^p & \ldots & x_{nl} & x_{nl}^2 & \vdots & x_{nl}^p\end{array}\right]. \tag{15}$$

In the above specification the l blocks specify the unpenalized basis functions for estimation of the unknown functions $f_1(\cdot),\ldots,f_l(\cdot)$. As in the scatterplot smoothing example, a smooth fit results by putting constraints on the coefficients of the truncated basis functions. This is done by treating them as random effects in a mixed model formulation. Define

$$\mathbf{Z}^{pen}=\left[\begin{array}{ccc|c|ccc}(x_{11}-\kappa_1^1)_+^p & \ldots & (x_{11}-\kappa_{K_1}^1)_+^p & \ldots & (x_{1l}-\kappa_1^l)_+^p & \ldots & (x_{1l}-\kappa_{K_l}^l)_+^p\\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ (x_{n1}-\kappa_1^1)_+^p & \ldots & (x_{n1}-\kappa_{K_1}^1)_+^p & \ldots & (x_{nl}-\kappa_1^l)_+^p & \ldots & (x_{nl}-\kappa_{K_l}^l)_+^p\end{array}\right], \tag{16}$$

where K_i denotes the number of knots to estimate $f_i(\cdot)$ $(i=1,\ldots,l)$. In case of a GAM, the log-likelihood is considered as a function of the additive predictor $\boldsymbol{\eta}$ and, using penalized regression splines, $\widehat{\boldsymbol{\eta}}=\boldsymbol{X}\widehat{\boldsymbol{\beta}}+\boldsymbol{Z}\widehat{\boldsymbol{u}}$, where $\widehat{\boldsymbol{\beta}}$ is obtained from the following penalized log-likelihood

$$\max_{\boldsymbol{\beta}}\{\boldsymbol{y}'(\boldsymbol{X\beta}+\boldsymbol{Zu})-\boldsymbol{1}'\psi(\boldsymbol{X\beta}+\boldsymbol{Zu})\}-\frac{1}{2}\sum_{j=1}^{l}\alpha_j\boldsymbol{u}_j'\boldsymbol{u}_j, \tag{17}$$

and $\widehat{\boldsymbol{u}}$ from $E[\boldsymbol{u}|\boldsymbol{y}]$ where – for ease of notation – a canonical link is assumed. $\boldsymbol{\beta}$ is the column vector with the parameters for the unpenalized basis functions in (15) (one parameter per column of $\boldsymbol{X}$). $\boldsymbol{u}_j=(u_{j1},\ldots,u_{jK_j})'$ $(j=1,\ldots,l)$, α_j $(j=1,\ldots,l)$ is the smoothing parameter for function $f_j(\cdot)$ and say $\boldsymbol{u}=(\boldsymbol{u}_1',\ldots,\boldsymbol{u}_l')'$. The optimization problem in (17) is equivalent to the optimization problem in a generalized linear mixed model (see Breslow and Clayton (1993)) with the GLMM specified as

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{u}) &= \exp\left(\boldsymbol{y}^{'}(\boldsymbol{X\beta}+\boldsymbol{Zu}) - \mathbf{1}^{'}\psi(\boldsymbol{X\beta}+\boldsymbol{Zu}) + \mathbf{1}^{'}c(\boldsymbol{y})\right), \\ \boldsymbol{u} &\sim N(\mathbf{0},\boldsymbol{\Lambda}), \\ \text{and } \boldsymbol{\Lambda} &= \begin{bmatrix} \sigma_1^2\mathbf{I}_{K_1\times K_1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_l^2\mathbf{I}_{K_l\times K_l} \end{bmatrix}, \end{aligned} \tag{18}$$

where $\sigma_j^2 = 1/\alpha_j$ ($j = 1,\dots,l$) and – again – a canonical link is used in (18) for ease of notation. Both (17) and (18) are easily generalized to the case of a non-canonical link.

In line with the previous specifications, a GAMM for longitudinal data can be rewritten as a GLMM as well. Let Y_{ij} denote the j^{th} observation for subject i, where $i = 1,\dots,N$ and $j = 1,\dots,n_i$. Conditional on the random effects $\boldsymbol{b}_i$ ($q\times 1$) for subject i (and $\boldsymbol{b}_i \sim N(\mathbf{0},\boldsymbol{D})$), $Y_{i1},\dots,Y_{in_i}$ are independent with a density from the exponential family and a predictor $\eta_{ij} = \sum_{h=1}^{l} f_h(x_{ijh}) + \boldsymbol{z}_{ij}^{'}\boldsymbol{b}_i$. Specify the design matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ for subject i ($i = 1,\dots,N$) as

$$\boldsymbol{X}_i = \left[\begin{array}{ccccc|c|cccc} 1 & x_{i11} & x_{i11}^2 & \dots & x_{i11}^p & \dots & x_{i1l} & x_{i1l}^2 & \dots & x_{i1l}^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{in_i1} & x_{in_i1}^2 & \dots & x_{in_i1}^p & \dots & x_{in_il} & x_{in_il}^2 & \vdots & x_{in_il}^p \end{array}\right], \tag{19}$$

and

$$\mathbf{Z}_i^{pen} = \left[\begin{array}{ccc|c|ccc} (x_{i11}-\kappa_1^1)_+^p & \dots & (x_{i11}-\kappa_{K_1}^1)_+^p & \dots & (x_{i1l}-\kappa_1^l)_+^p & \dots & (x_{i1l}-\kappa_{K_l}^l)_+^p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (x_{in_i1}-\kappa_1^1)_+^p & \dots & (x_{in_i1}-\kappa_{K_1}^1)_+^p & \dots & (x_{in_il}-\kappa_1^l)_+^p & \dots & (x_{in_il}-\kappa_{K_l}^l)_+^p \end{array}\right]. \tag{20}$$

Together with the ‘classical’ design matrix for the random effects for $\boldsymbol{b}_i$ ($i = 1,\dots,N$),

$$\mathbf{Z}_i^{ran} = \begin{bmatrix} z_{i11} & \dots & z_{i1q} \\ \vdots & \ddots & \vdots \\ z_{in_i1} & \dots & z_{in_iq} \end{bmatrix} \text{ and } \mathbf{Z}_i = [\mathbf{Z}_i^{pen}|\mathbf{Z}_i^{ran}], \tag{21}$$

the contribution of subject i to the GLMM specification of the GAMM is given by

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{r}_i) &= \exp\left(\boldsymbol{y}_i^{'}(\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{r}_i) - \mathbf{1}^{'}\psi(\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{r}_i) + \mathbf{1}^{'}c(\boldsymbol{y}_i)\right), \\ \boldsymbol{r}_i &= (\boldsymbol{u}^{'},\boldsymbol{b}_i^{'})^{'} \sim N(\mathbf{0},\boldsymbol{\Lambda}_i), \\ \text{and } \boldsymbol{\Lambda}_i &= \begin{bmatrix} \sigma_1^2\mathbf{I}_{K_1\times K_1} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_l^2\mathbf{I}_{K_l\times K_l} & 0 \\ 0 & 0 & \dots & 0 & \boldsymbol{D} \end{bmatrix}. \end{aligned} \tag{22}$$

The assumption of independence among subjects completes the specification of the GLMM representation of the GAMM.

3. APPLICATIONS AND OPPORTUNITIES

3.1. Credibility

Using linear mixed models Frees et al. (1999) already gave a longitudinal data analysis interpretation of the well-known credibility models of Bühlmann (1967), Bühlmann (1969), Bühlmann and Straub (1970), Hachemeister (1975) and Jewell (1975). They explained how to specify the fixed and random effects for every subject or risk class i ($i = 1, \ldots, N$) and used $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}_i$ (as in (3) and (4)) to derive the Best Linear Unbiased Predictor for the conditional mean of a future observation ($\mathrm{E}[Y_{i,n_i+1}|\boldsymbol{b}_i]$). For the above mentioned credibility models, this BLUP corresponds with the classical credibility formulas.

However, the normal-normal model (normality for both responses and random effects) will not always be plausible for the data at hand (which can be, for instance, counts, binary or skewed data). Therefore it is useful to revisit the credibility models in the context of GLMs and to consider their specification as a GLMM. In this way, estimators and predictors will be used that take the distributional features of the data into account.

Interpreting traditional credibility models in the context of GLMMs implies that the additive regression structure in terms of fixed and subject-specific (or risk class specific) effects is specified on the scale of the linear predictor, namely

$$g(\mu_{ij}) = \eta_{ij} = \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{'}\boldsymbol{b}_i. \tag{23}$$

Hereby i ($i = 1, \ldots, N$) denotes the subject, for instance a policy(holder) or risk class, and j refers to its j^{th} measurement, unless it is stated otherwise. The link function $g(\cdot)$ and variance function $V(\cdot)$ are determined by the chosen GLM. More details are given in Antonio and Beirlant (2006a).

3.2. Claims reserving

We illustrate how information on claim counts and claim amounts can be combined in a semiparametric regression model for claims reserving. Using a Bayesian implementation of the smoothers from Section 2, the data considered in de Alba (2002) are reanalyzed. A generalized additive model is constructed that combines data on claim numbers and claim intensities. We illustrate that, by using Bayesian statistics, simulation from the predictive distributions in this more complicated model is possible without many additional efforts. Full details are in Antonio and Beirlant (2006b).

Denote with Y_{ij} the aggregate payment for cell (i, j) and let N_{ij} be the corresponding number of claims. Thus, $Y_{ij} = \sum_{k=1}^{N_{ij}} Y_{ijk}$, with Y_{ijk} the payments composing the aggregate claim Y_{ij}. Following de Alba (2002), a model is considered which combines information on the number of claims registered and the total amount paid out for these claims, per arrival/development year combination. Let $Z_{ij} := Y_{ij}/N_{ij}$ be the average payment for cell (i,j) and model

$$\begin{aligned} Z_{ij} &\sim \Gamma(\nu, \mu_{ij}^{Av}/\nu), \\ \text{where } \log(\mu_{ij}^{Av}) &= \alpha_1^{Av} * I(i=1) + \cdots + \alpha_{10}^{Av} * I(i=10) + f^{Av}(j) \\ \text{and } \frac{N_{ij}}{\phi} &\sim \text{Poisson}\left(\frac{\mu_{ij}^{Num}}{\phi}\right), \\ \text{where } \log(\mu_{ij}^{Num}) &= \alpha_1^{Num} * I(i=1) + \cdots + \alpha_{10}^{Num} * I(i=10) + f^{Num}(j). \end{aligned} \quad (24)$$

Furthermore, the Z_{ij}'s and N_{ij}'s are assumed to be independent.

Based on an inspection of the scatterplots and residual plots from an analysis with Proc Glimmix in SAS (not shown), 4 knots in the direction of development years, with positions $(2, 3, 5, 7)$ (for claim counts and average payments), are used. Results for the reserves from this model are summarized in Table 1 (claim counts) and Table 2 (total payments, obtained by multiplying claim numbers and average payments).

	Mean Poisson	Mean o-Poisson	St.Dev. Bayes.	5% Bayes.	50% Bayes.	97.5% Bayes.
AY 2	2	2	4.36	0	0	17
AY 3	7	5	7.424	0	0	25
AY 4	13	9	10.372	0	8	34
AY 5	22	19	14.418	0	17	51
AY 6	41	40	21.06	8	34	85
AY 7	97	96	33.702	34	93	169
AY 8	149	147	47.275	68	144	246
AY 9	240	240	84.071	102	229	432
AY 10	332	322	215.339	42	279	855
Total	902	879	248.871	500	847	1,465

Table 1: *Predictive distribution for the number of claims: results from a Bayesian analysis with truncated line basis functions for smooth function over development years. A burn-in of 50,000 simulations was used, followed by another 450,000 simulations to which a thinning factor of 10 was applied.*

3.3. Non-life ratemaking

We consider a data set from Frees et al. (2001). These authors focused on the longitudinal character of the data and modelled the logarithmic transformation of ‘PP=Loss/Payroll’, using linear mixed models. Our analysis as well takes the longitudinal character of the data into account and considers inference and prediction regarding individual risk classes. Use is made, however, of a gamma GLMM; in this way no transformation of the data is required. ‘Loss’ is the response variable and

	Mean Bayes.	St.Dev. Bayes.	2.5% Bayes.	50% Bayes.	97.5% Bayes.
AY 2	165	500	0	0	2
AY 3	372	742	0	0	2
AY 4	606	909	0	312	3
AY 5	1,038	1,127	0	726	3,963
AY 6	1,562	1,306	111	1,239	4,908
AY 7	2,473	1,612	523	2,103	6,510
AY 8	3,802	2,328	947	3,288	9,694
AY 9	5,503	3,522	1,344	4,673	14,507
AY 10	5,983	5,937	495	4,242	21,772
Total	21,503	8,990	9,513	19,753	43,903

Table 2: *Predictive distribution of the reserves (data displayed in thousands): results from a Bayesian analysis with truncated line basis functions for smooth functions over development period. A burn-in of 50,000 simulations was used, followed by another 450,000 simulations to which a thinning factor of 10 was applied.*

'Payroll' is used as an offset. The following models are considered

$$\begin{aligned} Y_{ij}|\boldsymbol{b}_i &\sim \Gamma(\nu, \mu_{ij}/\nu) \\ \text{where } \log(\mu_{ij}) &= \log(Payroll_{ij}) + \beta_0 + b_{i,0} \qquad (25) \\ \text{versus } \log(\mu_{ij}) &= \log(Payroll_{ij}) + \beta_0 + \beta_1 Year_{ij} + b_{i,0} \qquad (26) \\ \text{and } \log(\mu_{ij}) &= \log(Payroll_{ij}) + \beta_0 + \beta_1 Year_{ij} + b_{i,0} + b_{i,1} Year_{ij}. \qquad (27) \end{aligned}$$

The gamma density function is specified as $f(y) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu y}{\mu}\right)^{\nu} \exp\left(\frac{-\nu y}{\mu}\right)\frac{1}{y}$. The specification in (27) did not lead to convergence of the SAS procedures. Structure (26) is the preferred choice for the linear predictor. Table 3 contains the results of a maximum-likelihood and Bayesian analysis, where non-informative priors were used. Fitted values against real observations are plotted in Figure 1. More details and related examples are in Antonio and Beirlant (2006a).

	PQL		*adaptive G-H*		*Bayesian*	
	Est.	SE	Est.	SE	Mean	90% Cred. Int.
β_0	-4.172	0.091	-4.148	0.091	-4.147	(-4.298, -3.996)
β_1	0.042	0.012	0.042	0.012	0.042	(0.022, 0.062)
δ_1	0.915	0.128	0.912	0.127	0.938	(0.741, 1.174)

Table 3: *Workers' compensation data (Losses): results of maximum likelihood and Bayesian analysis. REML is used in PQL.*

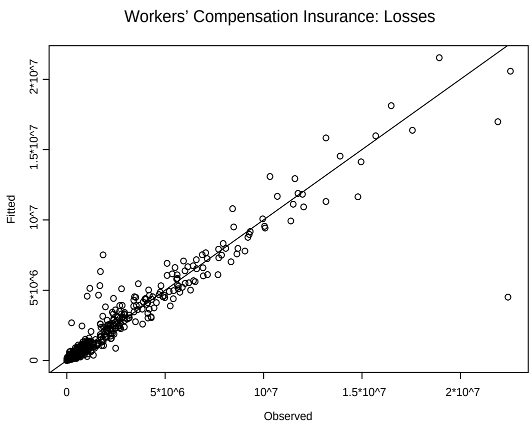

Figure 1: *Workers' compensation data (Losses): observed versus fitted values.*

3.4. Discussion

We presented some new statistical approaches for the analysis of actuarial data related to claims reserving and credibility. To illustrate further possibilities in this framework, we mention three interesting topics of our current research. Firstly, it is interesting to compare the mixed model approach with a copula construction to model the dynamics in panel data (as in Frees and Wang (2005)). Secondly, the joint modelling of longitudinal data on claim numbers and claim amounts through a mixed model, can be considered and contrasted with – again – a copula construction. Thirdly, instead of working in the framework of the exponential distribution, regression models for heavy-tailed data are of interest for actuaries. In this way, a combination of the models discussed above with heavy-tailed regression models, can be useful for actuarial applications.

References

K. Antonio and J. Beirlant. Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, 2006a. to be published.

K. Antonio and J. Beirlant. Semiparametric regression models for claims reserving and credibility: the mixed model approach. *Working Paper (online at www.econ.kuleuven.be/katrien.antonio)*, 2006b.

K. Antonio, J. Beirlant, T. Hoedemakers, and R. Verlaak. Lognormal mixed models for reported claims reserves. *North American Actuarial Journal*, 10:30–48, 2006.

N.E. Breslow and D.G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Society*, 88(421):9–25, 1993.

H. Bühlmann. Experience rating and credibility I. *ASTIN Bulletin*, 4:199–207, 1967.

H. Bühlmann. Experience rating and credibility II. *ASTIN Bulletin*, 5:157–165, 1969.

H. Bühlmann and E. Straub. Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, pages 111–133, 1970.

E. de Alba. Bayesian estimation of outstanding claims reserves. *North American Actuarial Journal*, 6(4):1–20, 2002.

P. Diggle, P. Heagerty, K.Y. Liang, and S.L. Zeger. *Analysis of longitudinal data.* Oxford University Press, Oxford, 2002.

E.W. Frees and P. Wang. Credibility using copulas. *North American Actuarial Journal*, 9(2): 31–48, 2005.

E.W. Frees, V.R. Young, and Y. Luo. A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, 24:229–247, 1999.

E.W. Frees, V.R. Young, and Y. Luo. Case studies using panel data models. *North American Actuarial Journal*, 5(4):24–42, 2001.

C.A. Hachemeister. Credibility for regression models with application to trend. In *Proceedings of the Berkeley Actuarial Research Conference on Credibility*, pages 129–163, New York, 1975. Adamic Press.

W.S. Jewell. The use of collateral data in credibility theory. *Giornale dell' Instituto Italiano degli Actuary*, 38:1–16, 1975.

Q. Liu and D.A. Pierce. A note on Gauss-Hermite quadrature. *Biometrika*, 81(3):624–629, 1994.

G. Molenberghs and G. Verbeke. *Models for discrete longitudinal data.* Springer series in statistics. Springer, New York, 2005.

D. Ruppert, M.P. Wand, and R.J. Carroll. *Semiparametric regression.* Cambridge University Press, Cambridge, 2003.

G. Verbeke and G. Molenberghs. *Linear mixed models for longitudinal data.* Springer, New York, 2000.

R. Wolfinger and M. O'Connell. Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48:233–243, 1993.