Semiparametric Regression Models for Claims Reserving and Credibility: the Mixed Model Approach.

Katrien Antonio*† Jan Beirlant ‡

January 12, 2006

Abstract

Verrall (1996) and England & Verrall (2001) considered the use of smoothing methods in the context of claims reserving, by applying two smoothing procedures in a likelihood-based way, namely the locally weighted regression smoother ('loess') and the cubic smoothing spline smoother. Using the statistical methodology of semiparametric regression and its connection with mixed models (see e.g. Ruppert *et al.*, 2003), this paper revisits smoothing models for loss reserving and considers their use in an example from credibility. Next to the flexibility of a semiparametric regression model, advantages of the presented approach are threefold. Firstly, because the constructed semiparametric models have an interpretation as (generalized) linear mixed models ((G)LMMs), standard statistical theory and software for (G)LMMs can be used. Secondly, a Bayesian implementation of these smoothing models is relatively straightforward and allows simulation from the full predictive distribution of quantities of interest. Since actuaries are interested in predictions, this is a major advantage. Thirdly, more complicated statistical models, dealing for example with semicontinuous data or extensive longitudinal data, can be handled within the same framework. Throughout this work, data examples illustrate these different aspects. Evidently, the methodology is not restricted to the problems discussed in this paper, but is relevant for other kinds of actuarial regression problems.

Keywords: loss reserving, credibility, generalized additive mixed models, P-splines, Bayesian statistics.

*Corresponding author: Katrien.Antonio@econ.kuleuven.be (phone: +32 (0) 16 32 67 69)

†Ph. D. student, University Center for Statistics, KU Leuven, W. de Croylaan 54, 3001 Heverlee, Belgium.

‡Full professor, University Center for Statistics, KU Leuven, W. de Croylaan 54, 3001 Heverlee, Belgium.

1 Introduction

Claims originating in a particular year often can not be finalized in the same year. Many causes for delay of the payment process are possible, for example long-lasting juridical procedures are the rule with liability insurance. Alongside the *reported but not settled* (RBNS) claims, a company also needs to manage claims that *incurred* already *but* are *not* yet *reported* (IBNR) to the insurer. For both types of claims, provisions will be held to meet the future obligations of the insurer towards its policy holders.

A broad literature is available concerning deterministic and stochastic models used for loss reserving. We refer to England & Verrall (2002) for an overview. The methods discussed in that paper are framed within the context of a run-off triangle. Table 1 is a specific example of such a triangle. Antonio *et al.* (2006) analyze data with a different structure.

Continuing the earlier work by Verrall (1996) and England & Verrall (2001), the first part of this paper revisits the use of semiparametric regression models in a claims reserving exercise. In a semiparametric regression model, parametric as well as nonparametric functional relationships are allowed, where the latter have the advantage that they are able to model flexible relationships between a response and a covariate.

In the specific context of claims reserving, we explore the use of semiparametric models to capture the main trends in the data in the direction of arrival, development and calendar years (abbreviated in the sequel with 'AY', 'DY' and 'CY'). A widely used alternative approach is the specification of appropriate categorical variables to model such trends. However, the specification of the linear predictor in a lognormal or a generalized linear model often turns out to be a very difficult and time-consuming exercise, see for instance the discussion in Kaas *et al.* (2001, Chapter 9) or the quest for an appropriate trend model (De Vylder & Goovaerts, 1979). The intention of this paper is to reduce the work involved in the specification of the predictor in a model for loss reserving. To achieve this we rely on semiparametric regression models.

In the above mentioned papers by P. England and R. Verrall, cubic smoothing splines and locally weighted regression smoothers ('loess') were applied in a frequentist way, using the `gam()` function in SPlus [1]. The semiparametric models in this paper are implemented via the concept of penalized regression splines (also called P-splines or pseudo-splines) and their connection with mixed models (as discussed for example in Ruppert *et al.*, 2003). This (generalized) linear mixed model formulation of the smoothers opens many doors. Not only can we rely on software for GLMMs (like Proc Mixed and Proc Glimmix in SAS [2] and `lme()` in SPlus), also a Bayesian implementation of the models and, consequently, simulation from the predictive distribution of quantities of interest is rela-

[1]SPlus is a commercial statistical software package; see http://www.insightful.com

[2]SAS is a commercial software package; see http://www.sas.com

tively straightforward. Moreover, longitudinal and cross-sectional data can be modelled semiparametrically in the same framework.

Some recent papers, for instance Denuit & Lang (2004) and Fahrmeir *et al.* (2003), also used semiparametric regression models based on P-splines for ratemaking with actuarial data. Their approach is purely Bayesian and relies on an implementation with B-splines, as developed in the software package `BayesX` [3]. Our work is in the same line, but differs from their setup in the sense that we only consider the use of the mixed model methodology (likelihood-based and Bayesian) from Ruppert *et al.* (2003). Moreover, throughout this paper, emphasis is put on claims reserving, although an example in the context of longitudinal credibility data is also included.

The merits of Bayesian actuarial statistics have been discussed by many authors (see e.g. Verrall, 2005, page 149, for a recent opinion). Also in the statistical literature on semiparametric regression, *'going Bayesian'* is becoming very popular (see for instance Ruppert *et al.*, 2003, page xiv). In the specific problems discussed in this contribution, the use of Bayesian statistics and MCMC simulations allows us to obtain the predictive distribution of the reserves (in claims reserving) or future payments (in credibility). For the situation of claims reserving, this enables us to deal with more sophisticated statistical models, which for example include a stochastic discounting process (see Section 3.1.2), combine data on paid losses and claim counts (see Section 3.2) or model semicontinuous data consisting of exact zeros and strictly positive payments (see Section 3.2). In the likelihood-based approach in the England and Verrall (2001) paper, a 'standard' run-off triangle is considered.

The rest of the paper is organized as follows. In Section 1.1, the data are introduced that will be analyzed later on. Section 2 provides background on smoothing using penalized regression splines and the mixed model connection. An analysis of the presented data sets is given in Section 3 and Section 4 concludes. The reader should be familiar with basic concepts of (generalized) linear (mixed) models ((G)L(M)Ms). McCulloch & Searle (2001) offer a general overview and Antonio & Beirlant (2005) discuss applications of GLMMs in actuarial statistics. Ruppert *et al.* (2003) provide more details on smoothing with mixed models.

1.1 Description of the data sets

1.1.1 Aggregate data on claim intensities

In Table 1 we present the data set previously analyzed in England & Verrall (2001). It contains aggregate data on claim intensities, given as a classical run-off triangle with paid losses. Reserves obtained with the deterministic chain-ladder technique and the

[3]BayesX is a free software package for performing complex Bayesian inference; see http://www.stat.uni-muenchen.de/ bayesx/bayesx.html

chain-ladder development factors are displayed in Table 1. The reserves obtained with an over-dispersed Poisson model are included as benchmark results and are taken from England & Verrall (2001). Hereby the additive predictor consists of a smooth function of the logarithm of the development years, together with a parameter for each accident year.

Claim Payments										*Reserves Ch. Ladd.*	*Reserves Smooth o-P*
45,630	23,350	2,924	1,798	2,007	1,204	1,298	563	777	621	0	0
53,025	26,466	2,829	1,748	732	1,424	399	537	340		683	622
67,318	42,333	1,854	3,178	3,045	3,281	2,909	2,613			1,846	1,998
93,489	37,473	7,431	6,648	4,207	5,762	1,890				4,336	4,470
80,517	33,061	6,863	4,328	4,003	2,350					5,616	5,940
68,690	33,931	5,645	6,178	3,479						8,151	8,106
63,091	32,198	8,938	6,879							10,841	11,106
64,430	32,491	8,414								15,102	15,112
68,548	35,366									21,587	21,293
76,013										60,828	60,377
Dev. Fact.	1.491	1.052	1.042	1.027	1.025	1.015	1.013	1.007	1.008		

Table 1: *Run-off triangle with claim intensities, taken from England & Verrall (2001). The last two columns in the table display the reserves obtained with the deterministic chain-ladder ('Ch. Ladd.') and a smooth over-dispersed Poisson model ('Smooth o-P'), respectively.*

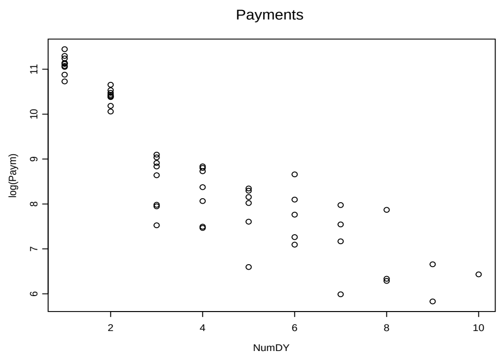

Figure 1: *Scatter plot of log-transformed data versus development year (DY); data from Table 1.*

In Section 3.1, the data in Table 1 will be analyzed by fitting a generalized additive model (GAM), using penalized regression splines. Categorical variables in the direction of arrival years and smoothing over the development years are used. Here we also consider the modelling of trends in the direction of calendar years, together with a Bayesian implementation of the constructed semiparametric regression model. The inclusion of a stochastic discounting process in the smoothing model is illustrated as well.

1.1.2 Aggregate data on claim intensities and claim counts

We illustrate how information on claim counts and claim amounts can be combined in a semiparametric regression model. Using a Bayesian implementation of the smoothers considered in this text, the data considered in de Alba (2002) are reanalyzed. These are displayed in Table 2 and 3 and illustrated in Figure 2. 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.

Claim Payments										*Reserves Ch. Ladd.*
357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948	0
352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046		94,634
290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405			469,511
310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286				709,638
443,160	693,190	991,983	769,488	504,851	470,639					984,889
396,132	937,085	847,498	805,037	705,960						1,419,460
440,832	847,631	1,131,398	1,063,296							2,177,641
359,480	1,061,648	1,443,370								3,920,301
376,686	986,608									4,278,972
344,014										4,625,811
Dev. Fact.	3.491	1.747	1.457	1.174	1.104	1.086	1.054	1.077	1.018	

Table 2: *Run-off triangle with claim intensities, taken from de Alba (2002). The last column in the table contains the reserves obtained with the deterministic chain-ladder ('Ch. Ladd.').*

Claim Numbers										*Reserves Ch. Ladd.*
40	124	157	93	141	22	14	10	3	2	0
37	186	130	239	61	26	23	6	6		2
35	158	243	153	48	26	14	5			7
41	155	218	100	67	17	6				13
30	187	166	120	55	13					25
33	121	204	87	37						39
32	115	146	103							89
43	111	83								155
17	92									239
22										333
Dev. Fact.	5.055	1.930	1.350	1.134	1.035	1.023	1.011	1.007	1.003	

Table 3: *Run-off triangle with claim numbers, taken from de Alba (2002). The last column in the table contains the reserves obtained with the deterministic chain-ladder ('Ch. Ladd.').*

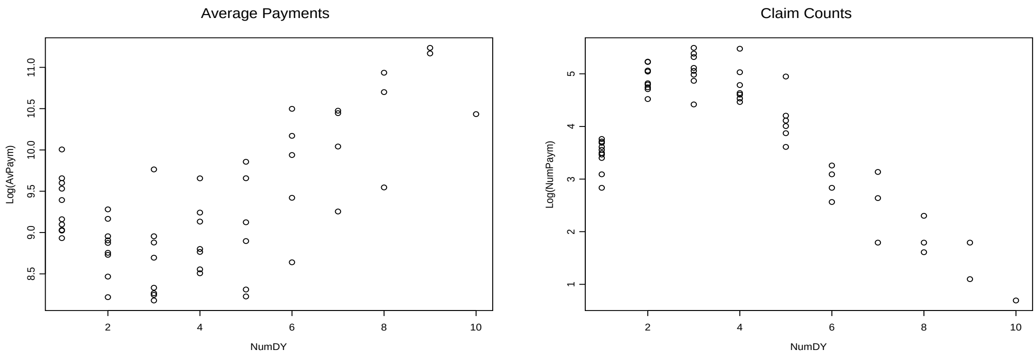

Figure 2: *Scatter plots of log-transformed data versus development year (DY); data from Table 2 (left) and Table 3 (right).*

1.1.3 A two-part semiparametric model for semicontinuous data

The run-off triangle in Table 4 consists of strictly positive payments and exact zeros. The modelling of such semicontinuous [4] data (with a hump at zero) requires specific attention. A two-part generalized additive mixed model is presented for these data. Hereby, a mixed model with semiparametric predictor is fitted to the binary data set which represents the occurrence of a payment. Given that a payment has occurred, its severity is modelled again with a GAM, but now in a different distributional framework (for example lognormal or gamma). Using a Bayesian analysis, the predictive distribution of the different reserves is obtained in this two-part model.

1.1.4 An example from credibility

To illustrate the use of semiparametric regression based on mixed models in a credibility context, the data from Frees & Wang (2005) are revisited. Automobile bodily injury liability claims from a sample of $n = 29$ Massachusetts towns are considered. Yearly data over a period of 6 years (1993-1998) are available. Thus, in contrast with our previous examples, these data are longitudinal. The response variable is average claims ('AC'), which is the total claim amount divided by the amount of exposure, for each town and each year. Two explanatory variables are available, namely the per capita income ('PCI') and population per square mile ('PPSM'). More details can be found in Frees & Wang (2005). These authors analyzed the data using a gamma generalized linear model with canonical link, such that $\theta_{it} = \beta_0 + \beta_1 \text{PCI}_{it} + \beta_2 \text{PPSM}_{it}$, for each town i and year t and θ_{it} the canonical parameter in the generalized linear model. The dependencies between observations on the same town are modelled by combining them with a $t-$copula. In

[4]'A semicontinuous random variable combines a continuous distribution with point masses at one or more locations.' (Olsen & Schafer, 2001, page 730)

Claim Payments													Reserves Ch. Ladd.
2,216	744	10	5	0	0	0	0	0	0	0	0	0	0
2,713	0	75	3	4	0	0	0	0	0	0	0		0
2,383	874	0	89	37	7	8	19	6	0	0			0
3,173	0	136	15	0	27	13	33	21	1				0
3,079	1,898	137	66	0	3	6	0	0					0.41
14,286	2,898	0	202	75	58	0	0						28
8,379	3,890	440	95	31	7	8							40
9,401	0	336	188	164	127								39
11,197	5,452	398	89	60									134
16,527	0	233	239										217
14,172	4,871	435											464
13,300	0												594
16,142													4,170
Dev. Fact.	1.205	1.020	1.011	1.005	1.004	1	1.001	1.002	1	1	1	1	

Table 4: *Run-off triangle with claim payments (exact zeros and strictly positive payments), data obtained from Belgian insurance company. The last column in the table contains the reserves obtained with the deterministic chain-ladder ('Ch. Ladd.').*

our analysis, we will investigate whether nonlinear effects of PCI and PPSM are suitable. The dependencies between observations on the same town are modelled using typical machinery from mixed models, i.e. the inclusion of random effects or the specification of a special structure for the covariance matrix of the residuals. A Bayesian analysis is used for prediction.

2 Generalized additive mixed models

This section describes generalized additive models (GAMs) for cross-sectional data and generalized additive mixed models (GAMMs) for longitudinal data, together with their specification using penalized regression splines. In this way, the GA(M)Ms can be rewritten as generalized linear mixed models (GLMMs). Likelihood-based and Bayesian inference for the smoothing models is described.

2.1 Observation model

Numerous illustrations of the use of generalized linear models (GLMs) in typical problems from actuarial statistics are available; see Haberman & Renshaw (1996) for an overview. Similar to a GLM, a GAM consists of three components: a random component, a systematic component and a link function. For the random component, let $Y_1, \ldots, Y_n$ be independent random variables with a density $f(.)$ from the exponential family, namely

$$f(y) = \exp\left(\frac{y\theta - \psi(\theta)}{\phi} + c(y, \phi)\right), \tag{1}$$

where $\psi(.)$ and $c(.)$ are known functions and θ is the natural parameter and ϕ the scale parameter. Distributions from this class are – for example – the normal, Bernoulli, gamma and Poisson distribution. The main difference between a GAM and a GLM lies in the specification of the systematic component. The linear predictor $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$ from a GLM is in a GAM replaced by an additive predictor,

$$\begin{aligned} \eta_i &= \sum_{h=1}^{l} f_h(x_{ih}), \\ \text{and } \mu_i &= g(\eta_i), \quad i = 1, \ldots, n, \end{aligned} \tag{2}$$

where $\mathrm{E}[Y_i] = \mu_i$ and $g(.)$ is called the link function. Hereby the functions f_h ($h = 1, \ldots, l$) are 'smooth' functions of covariates x_h ($h = 1, \ldots, l$). Instead of being fully nonparametric, the additive predictor in (2) possibly is a combination of parametric (like $f_h(x_{ih}) = x_{ih}\beta_h$) and nonparametric components. To estimate a GAM, some kind of smoother is used for the unknown functions $f_h(.)$. Possible smoothers are cubic smoothing splines, locally weighted regression (loess) or kernel smoothers, of which the first two were considered by Verrall (1996) and England & Verrall (2001) in the context of a claims reserving exercise. For the interested reader, Hastie & Tibshirani (1990) provide full details on the different aspects of GAMs. Instead of using the so-called local scoring algorithm for GAMs, we will rely on the inferential techniques developed for generalized linear mixed models (GLMMs), as discussed in Sections 2.2 and 2.3 below.

Concerning the observation model for the longitudinal data in Section 1.1.4, let Y_{ij} denote the j^{th} observation for subject i, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, N$. Thus, there are N subjects in the data set and n_i is the number of observations available for subject i. Similar to a GLMM (like in Antonio & Beirlant, 2005), conditional on the random effects $\boldsymbol{b}_i$ ($q \times 1$) for subject i, $Y_{i1}, \ldots, Y_{in_i}$ are assumed to be independent with a distribution from the exponential family in (1), thus

$$f(y_{ij}|\boldsymbol{b}_i) = \exp\left(\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{\phi} + c(y_{ij}, \phi)\right). \tag{3}$$

The predictor, η_{ij}, in a GAMM is then specified as

$$g(\mu_{ij}) = \eta_{ij} = \sum_{h=1}^{l} f_h(x_{ijh}) + \boldsymbol{z}_{ij}^{'}\boldsymbol{b}_i, \tag{4}$$

where $\mu_{ij} = \mathrm{E}[Y_{ij}|\boldsymbol{b}_i]$ and some of the functions $f_h(.)$ can simply be parametric. To complete the specification, the $\boldsymbol{b}_i$ ($i = 1, \ldots, N$) are assumed to be multivariate normally distributed with mean 0 and covariance matrix $\boldsymbol{D}$. In the specification given in (4), the nonparametric functions $f_h(.)$ apply on the population-level. This can be generalized further to subject-specific semiparametric functions, as in Durban *et al.* (2004).

In the sequel of this section, only the use of regression penalized splines to fit GA(M)Ms is considered. For the use of other types of smoothers, we refer to the literature. Following

Ruppert *et al.* (2003), we first describe how the use of regression penalized splines leads to a GLMM specification of the models discussed in this text.

2.2 Penalized splines and GLMM formulation

The idea behind regression penalized splines is to estimate the unknown nonparametric effect of a covariate, say x, on the response as a linear combination of some basis functions. To obtain a smooth fit, constraints are put on some of the coefficients used in this linear combination; they are *penalized*.

Clarifying this approach for unfamiliar readers, let us start from the simple example of scatterplot smoothing: data (x_i, y_i) $(i = 1, \dots, n)$ are given and the model $Y_i = f(x_i) + \epsilon_i$ $(i = 1, \dots, n)$ is fitted. To estimate the unknown function $f(.)$, 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 $\kappa_1, \dots, \kappa_K$ [5], define the design matrix $\boldsymbol{B}$ as

$$\boldsymbol{B} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^p & (x_1 - \kappa_1)_+^p & \dots & (x_1 - \kappa_K)_+^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^p & (x_n - \kappa_1)_+^p & \dots & (x_n - \kappa_K)_+^p \end{bmatrix}. \tag{5}$$

The unknown function $f(.)$ 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 a constraint $\sum_{k=1}^K \beta_{pk}^2 < C$ to obtain a smooth fit. Hereby, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p, \beta_{p1}, \dots, \beta_{pK})'$ and thus the penalized coefficients correspond with the truncated power functions. Using a Lagrange multiplier argument, this optimization problem is rewritten as

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n (y_i - \boldsymbol{B}(x_i)\boldsymbol{\beta})^2 + \alpha \boldsymbol{\beta}' \boldsymbol{P} \boldsymbol{\beta}, \tag{6}$$

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

$$\boldsymbol{P} = \begin{bmatrix} 0_{p+1 \times p+1} & 0_{p+1 \times K} \\ 0_{K \times p+1} & \mathbf{I}_{K \times K} \end{bmatrix}. \tag{7}$$

Ruppert *et al.* (2003) (among others) rewrite the argument of the optimization problem in (6), after dividing by σ_ϵ^2, as

$$\frac{1}{\sigma_\epsilon^2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u}\|^2 + \frac{1}{\sigma_u^2} \|\boldsymbol{u}\|^2, \tag{8}$$

[5] 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 $\{1, x, x^2, \dots, x^p, (x - \kappa_1)_+^p, \dots, (x - \kappa_K)_+^p\}$ span the vector space of piecewise functions of degree p with knots at $\kappa_1, \dots, \kappa_K$.

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 $1,\ x,\ x^2, \ldots,\ x^p$), $\boldsymbol{u} = (\beta_{p1}, \ldots, \beta_{pK})'$,

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^p \end{bmatrix} \text{ and } \boldsymbol{Z} = \begin{bmatrix} (x_1 - \kappa_1)_+^p & \ldots & (x_1 - \kappa_K)_+^p \\ \vdots & \vdots & \vdots \\ (x_n - \kappa_1)_+^p & \ldots & (x_n - \kappa_K)_+^p \end{bmatrix}. \tag{9}$$

By considering $\boldsymbol{u}$ as random effects with $\boldsymbol{u} \sim N(0, \sigma_u^2 \boldsymbol{I}_{K \times K})$, (8) 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 the GAM specified by (1) and (2). Construct the design matrix $\boldsymbol{X}$ as

$$\boldsymbol{X} = \left[\begin{array}{ccccc|c|cccc} 1 & x_{11} & x_{11}^2 & \ldots & x_{11}^p & \ldots & x_{1l} & x_{1l}^2 & \ldots & x_{1l}^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n1}^2 & \ldots & x_{n1}^p & \ldots & x_{nl} & x_{nl}^2 & \vdots & x_{nl}^p \end{array}\right]. \tag{10}$$

In the above specification the l blocks specify the unpenalized basis functions for estimation of the unknown functions $f_1(.), \ldots, f_l(.)$. 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

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

where K_i denotes the number of knots to estimate $f_i(.)$ $(i = 1, \ldots, l)$. In case of a GAM, the log-likelihood is considered as a function of the additive predictor $\boldsymbol{\eta}$ from (2) and, using penalized regression splines, $\widehat{\boldsymbol{\eta}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}\widehat{\boldsymbol{u}}$ is obtained as the solution of the following penalized log-likelihood

$$\max_{\boldsymbol{\beta}, \boldsymbol{u}} \{\boldsymbol{y}'(\boldsymbol{X\beta} + \boldsymbol{Zu}) - \boldsymbol{1}'\psi(\boldsymbol{X\beta} + \boldsymbol{Zu})\} - \frac{1}{2} \sum_{j=1}^{l} \alpha_j \boldsymbol{u}_j' \boldsymbol{u}_j, \tag{12}$$

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 (10) (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 $f_j(.)$ and say $\boldsymbol{u} = (\boldsymbol{u}_1', \ldots, \boldsymbol{u}_l')'$. The optimization problem in (12) is equivalent to the penalized quasi-likelihood optimization problem in a generalized linear mixed model (see Breslow & Clayton, 1993) with the GLMM specified as

$$\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{u}) &= \exp\left(\boldsymbol{y}'(\boldsymbol{X\beta}+\boldsymbol{Zu}) - \mathbf{1}'\psi(\boldsymbol{X\beta}+\boldsymbol{Zu}) + \mathbf{1}'c(\boldsymbol{y})\right), \\
\boldsymbol{u} &\sim N(\mathbf{0},\boldsymbol{\Lambda}), \\
\text{and } \boldsymbol{\Lambda} &= \begin{bmatrix} \sigma_1^2\mathbf{I}_{K_1\times K_1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_l^2\mathbf{I}_{K_l\times K_l} \end{bmatrix},
\end{aligned} \tag{13}$$

where $\sigma_j^2 = 1/\alpha_j$ $(j = 1,\dots,l)$ and – again – a canonical link is used in (13) for ease of notation. Both (12) and (13) are easily generalized to the case of a non-canonical link. In that situation, the relation $g\{\psi'(\boldsymbol{\theta})\} = \boldsymbol{\eta}$ is used.

In line with the previous specifications, the GAMM for longitudinal data, specified in (3) and (4), can be rewritten as a GLMM as well. Specify the design matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ for subject i $(i = 1,\dots,N)$ as

$$\boldsymbol{X}_i = \left[\begin{array}{ccccc|c|cccc} 1 & x_{i11} & x_{i11}^2 & \dots & x_{i11}^p & \dots & x_{i1l} & x_{i1l}^2 & \dots & x_{i1l}^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{in_i1} & x_{in_i1}^2 & \dots & x_{in_i1}^p & \dots & x_{in_il} & x_{in_il}^2 & \vdots & x_{in_il}^p \end{array}\right], \tag{14}$$

and

$$\mathbf{Z}_i^{pen} = \left[\begin{array}{ccc|c|ccc} (x_{i11}-\kappa_1^1)_+^p & \dots & (x_{i11}-\kappa_{K_1}^1)_+^p & \dots & (x_{i1l}-\kappa_1^l)_+^p & \dots & (x_{i1l}-\kappa_{K_l}^l)_+^p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (x_{in_i1}-\kappa_1^1)_+^p & \dots & (x_{in_i1}-\kappa_{K_1}^1)_+^p & \dots & (x_{in_il}-\kappa_1^l)_+^p & \dots & (x_{in_il}-\kappa_{K_l}^l)_+^p \end{array}\right]. \tag{15}$$

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

$$\mathbf{Z}_i^{ran} = \begin{bmatrix} z_{i11} & \dots & z_{i1q} \\ \vdots & \ddots & \vdots \\ z_{in_i1} & \dots & z_{in_iq} \end{bmatrix} \quad \text{and} \quad \mathbf{Z}_i = [\mathbf{Z}_i^{pen}|\mathbf{Z}_i^{ran}], \tag{16}$$

the contribution of subject i to the GLMM specification of the GAMM from (3) and (4) is given by

$$\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{r}_i) &= \exp\left(\boldsymbol{y}_i'(\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{r}_i) - \mathbf{1}'\psi(\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{r}_i) + \mathbf{1}'c(\boldsymbol{y}_i)\right), \\
\boldsymbol{r}_i &= (\boldsymbol{u}_i',\boldsymbol{b}_i')' \sim N(\mathbf{0},\boldsymbol{\Lambda}_i), \\
\text{and } \boldsymbol{\Lambda}_i &= \begin{bmatrix} \sigma_1^2\mathbf{I}_{K_1\times K_1} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_l^2\mathbf{I}_{K_l\times K_l} & 0 \\ 0 & 0 & \dots & 0 & \boldsymbol{D} \end{bmatrix}.
\end{aligned} \tag{17}$$

The assumption of independence among subjects completes the specification of the GLMM representation of the GAMM from (3) and (4).

2.3 Likelihood-based and Bayesian inference

In a likelihood-based context, penalized quasi-likelihood (PQL) is used to estimate the GLMMs, constructed for the above GA(M)Ms. Then, (restricted) maximum likelihood ((RE)ML) estimation of the variance components in (13) and (17) leads to an automatic choice of the smoothing parameters, namely $\hat{\alpha}_j = 1/\hat{\sigma}_j^2$ $(j = 1, \ldots, l)$. Note that, in the likelihood-based approach, all estimates for variance components reported in this paper are obtained with REML. Other inferential tools developed for GLMMs, such as a hypothesis test for the need of a random effect or the construction of confidence bands, can also be applied in the context of smoothing models. For a Gaussian response and normally distributed random effects, analytical expressions are available for the maximum likelihood estimators (MLEs) for the fixed effects parameters and the best linear unbiased predictors (BLUPs) for the random effects. In case of a non-Gaussian response the estimation in GLMMs is hindered by the presence of intractable multivariate integrals. To overcome this, Proc Nlmixed in SAS relies on the Gauss-Hermite quadrature formula for numerical integration, but can only deal with a limited number of random effects. Proc Glimmix in SAS relies on the Laplace approximation of the involved integrals and thus solves an approximate problem. Apart from these limitations of the likelihood approach, also note that they all rely on *'plugging-in'* the estimated variance components in formulas that are derived conditional on, or given the variance components. For more details, we refer to Ruppert *et al.* (2003) and Antonio & Beirlant (2005), for illustrations in actuarial statistics.

In the context of claims reserving or credibility, a Bayesian implementation of the GLMMs in (13) and (17) is especially useful, since this allows simulation from the full predictive distribution of the reserves or future payments. By specifying a prior distribution for the variance components, the Bayesian inferential tools take all sources of uncertainty into account, whereas the likelihood approach plugs in the estimated variance components, as noted above. Prior specifications for the unknown parameters are discussed in Section 3 and Antonio & Beirlant (2005) give more details on the full conditionals involved in Gibbs sampling for GLMMs. Zhao *et al.* (2004) compare different MCMC sampling schemes for estimation and inference in what they call *'general design Bayesian generalized linear mixed models'* (to which belong the GLMMs considered here). They consider the use of the Metropolis-Hasting algorithm for sampling from non-standard distributions, adaptive rejection sampling for sampling from univariate log concave distributions, the use of auxiliary variables and simple slice sampling. In the rest of this text, MCMC simulations are performed using the WinBUGS package.

3 An investigation in the context of claims reserving and credibility

3.1 A run-off triangle with aggregate data on claim intensities

3.1.1 A semiparametric Poisson model for claims reserving

In line with the analysis in England & Verrall (2001), an (overdispersed) Poisson model is used for the data in Table 1. Trends in the direction of development year and calendar year are modelled using penalized splines. The results obtained using semiparametric regression with mixed models are compared to those earlier reported in the literature.

Denote with Y_{ij} ($i, j = 1, \ldots, n$ and $n = 10$ in Table 1) the random variable corresponding with the amount paid out in arrival year i and development year j. Now start with the following model specification

$$\begin{aligned} \frac{Y_{ij}}{\phi} &\sim \text{Poisson}\left(\frac{\mu_{ij}}{\phi}\right), \\ \text{where } \log(\mu_{ij}) &= \alpha_1 I(i=1) + \ldots + \alpha_{10} I(i=10) + f(j). \end{aligned} \tag{18}$$

Thus, Y_{ij} follows an over-dispersed Poisson distribution, with $\mathrm{E}[Y_{ij}] = \mu_{ij}$ and $\mathrm{Var}[Y_{ij}] = \phi\mu_{ij}$. $f(.)$ is a smooth function over the development years. In a first stage of the analysis, we modelled $f(.)$ using truncated lines and $K = 3$ user-specified knots, namely $\kappa_1 = 3$, $\kappa_2 = 5$ and $\kappa_3 = 7$.

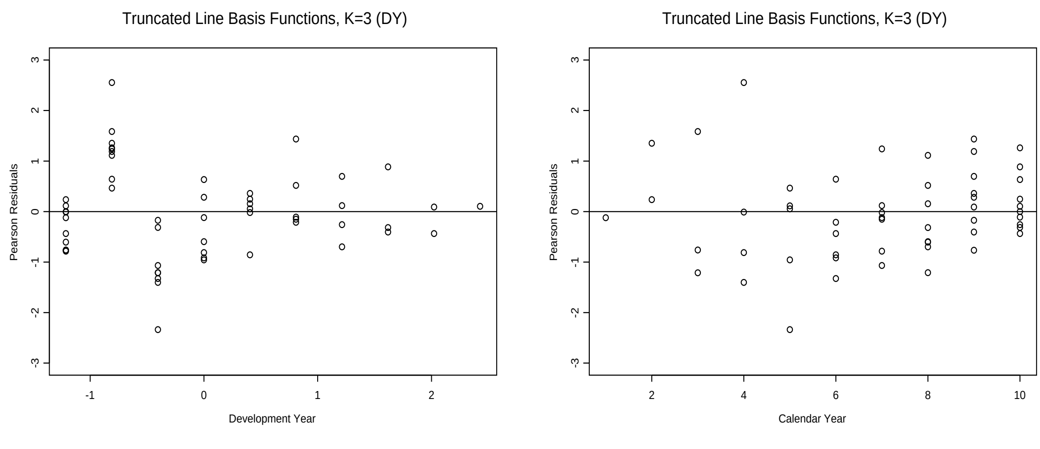

Figure 3: *Pearson-type residuals against development year (left) and calendar year (right): truncated line basis functions with $K = 3$ ($\kappa_1 = 3, \kappa_2 = 5, \kappa_3 = 7$) for smooth function in direction of development years.*

As illustrated by the residual plots in Figure 3, the model specification in (18) and its implementation given above, is not able to cope with all trends in the direction of development and calendar years. The initially suggested model is therefore refined by adapting the predictor in the following way

$$\log(\mu_{ij}) = \alpha_1 I(i=1) + \ldots + \alpha_9 I(i=9) + f(j) + g(i+j-1). \tag{19}$$

$f(.)$ is modelled by truncated line basis functions with 4 knots, namely $\kappa_{DY,1}=2$, $\kappa_{DY,2}=3$, $\kappa_{DY,3}=5$ and $\kappa_{DY,4}=7$. To model $g(.)$, truncated line basis functions are used as well, with 2 knots at positions $\kappa_{CY,1}=4$ and $\kappa_{CY,2}=6$. The over-dispersed Poisson model with a predictor as in (19) results in the residual plots given in Figure 4, which are clearly to be preferred above those in Figure 3. Other choices for the number and positions of the knots lead to similar results.

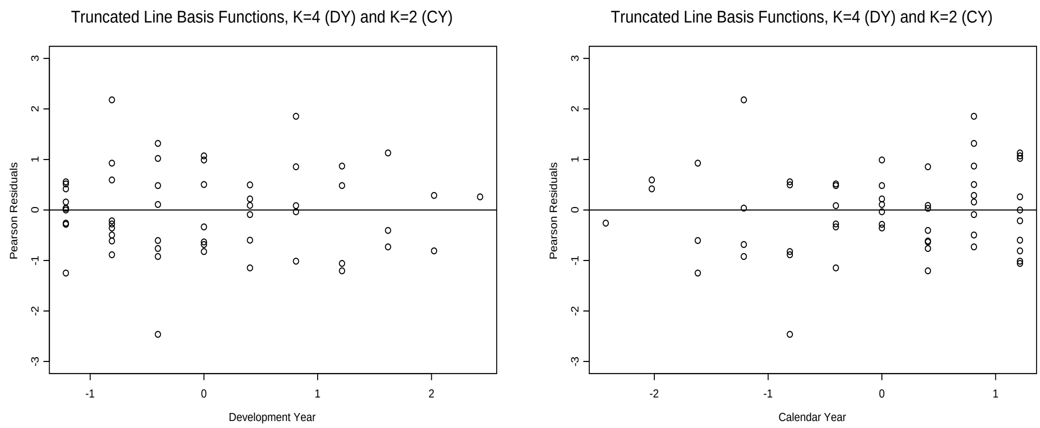

Figure 4: *Pearson-type residuals against development year (left) and calendar year (right): truncated line basis functions with* $K_{DY}=4$ $(\kappa_{DY,1}=2, \kappa_{DY,2}=3, \kappa_{DY,3}=5, \kappa_{DY,4}=7)$ *and* $K_{CY}=2$ $(\kappa_{CY,1}=4, \kappa_{CY,2}=6)$.

As a final remark regarding the specification of the additive predictor, consider the residual plot in Figure 5 which shows Pearson-type residuals against calendar year for a Poisson model with chain-ladder type structure for the linear predictor (thus $\log(\mu_{ij}) = \alpha_i + \beta_j$). This plot gives another motivation to model trends in the direction of calendar years.

To obtain the predictive distribution of the reserves, we consider a Bayesian implementation of the model in (19). For the over-dispersion factor ϕ the estimate obtained with a likelihood-based analysis is used, namely $\hat{\phi}=3.967$. The following prior specifications are used for the remaining parameters

$$\begin{aligned} \alpha_i &\sim N(0,10^5) \text{ with } i=1,\ldots,9, \\ \beta,\ \gamma &\sim N(0,10^5), \\ \sigma_b^2 &\sim \text{Inv-Gamma}(a,b), \\ \sigma_\gamma^2 &\sim \text{Inv-Gamma}(a,b). \end{aligned} \tag{20}$$

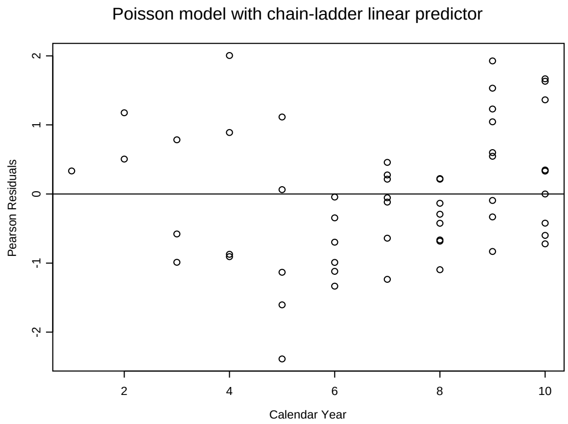

Figure 5: *Pearson-type residuals against calendar year: Poisson model with chain-ladder type structure for the mean.*

Table 5 contains the parameter estimates from a likelihood-based as well as from a Bayesian analysis of the over-dispersed Poisson model with predictor as in (19) ($(a,b) = (0.01, 0.01)$ in the prior specifications). Following the specification in Section 2, β and γ are 'fixed effects' parameters and σ_b^2 and σ_γ^2 denote the variance of the random effects used in the GLMM specification. Table 6 then displays the results for the different reserves.

Parameter	Mean (St. Err.) Lik.	Mean Bayes.	St. Dev. Bayes.	2.5% Bayes.	50% Bayes.	97.5% Bayes.
α_1	5.7402 (0.3935)	5.757	0.402	4.978	5.750	6.561
α_3	4.9293 (0.2465)	4.937	0.252	4.445	4.933	5.434
α_7	2.1875 (0.1259)	2.192	0.135	1.937	2.190	2.462
β	-3.4821 (0.1109)	-3.492	0.111	-3.712	-3.492	-3.275
γ	1.5734 (0.2059)	1.586	0.210	1.184	1.582	2.004
σ_β^2	6.2681 (5.2885)	9.485	22.833	1.533	5.541	40.495
σ_γ^2	0.1474 (0.1873)	1.719	54.6	0.02	0.206	6.362

Table 5: *Parameter estimates obtained with likelihood-based analysis (second column) and Bayesian analysis. 700,000 simulations used, to which a thinning factor of 10 is applied, after a burn-in of 50,000 simulations.*

To investigate the sensitivity of the results on the prior specifications for the variance components, Table 7 shows the posterior distributions for the fixed regression parameters and the total reserve, obtained via smoothing with truncated line basis functions and with

	Mean Lik.	Mean Bayes.	St.Dev. Bayes.	2.5% Bayes.	50% Bayes.	97.5% Bayes.
AY 2	717	672	620.18	0	390	2,340
AY 3	2,638	2,473	1,451	390	2,340	5,850
AY 4	6,323	6,037	2,579	1,950	5,850	12,090
AY 5	9,323	9,258	3,091	4,290	8,970	16,380
AY 6	13,285	13,387	3,872	7,020	12,870	22,230
AY 7	17,229	17,253	4,315	10,140	16,770	26,520
AY 8	22,463	22,503	4,918	14,040	22,230	33,540
AY 9	30,689	30,827	5,792	20,670	30,420	43,680
AY 10	75,163	75,390	10,009	57,330	74,880	96,720
Total	177,830	177,801	26,692	132,990	175,500	237,510

Table 6: *Results for different reserves obtained with likelihood-based analysis (second column) and Bayesian analysis. 700,000 simulations used, to which a thinning factor of 10 is applied, after a burn-in of 50,000 simulations.*

various prior specifications for the variance components, σ^2_β and σ^2_γ [6]:

$$\begin{aligned} \sigma^2_b,\ \sigma^2_\gamma &\sim \text{Inv-Gamma}(a,b) \text{ with } a,b = 0.1 \text{ or } 0.001, \\ \sigma_b,\ \sigma_\gamma &\sim \text{folded Cauchy with } s = 12 \text{ or } 25, \\ \sigma_b,\ \sigma_\gamma &\sim \text{Uniform}(0,50). \end{aligned} \tag{21}$$

3.1.2 Building in a stochastic discounting process

As another illustration of the flexibility of a Bayesian smoothing model, we build a stochastic discounting process into model (19). Assume that the reserve will be invested such that an amount of 1 at time $t-1$ becomes e^{Z_t} at time t. The discount factor for a payment of 1 at time t is then given by $e^{-(Z_1+\ldots+Z_t)} := e^{-Z(t)}$. Let us assume for instance that

$$Z(t) = \left(\mu - \frac{\delta^2}{2}\right) t + \delta B(t), \tag{22}$$

where $B(t)$ is the standard Brownian motion. The total discounted reserve – say R – reflects the time value of money and is then specified as

$$\begin{aligned} R &:= \sum_{i=2}^{n} \sum_{j=n+2-i}^{n} Y_{ij} e^{-Z(i+j-n-1)} \\ &= \sum_{i=2}^{n} \sum_{j=n+2-i}^{n} Y_{ij} \exp\{-(\mu-\delta^2/2)(i+j-n-1) - \delta B(i+j-n-1)\}, \end{aligned} \tag{23}$$

[6]folded Cauchy: $\sigma \propto (\sigma^2 + s^2)^{-1}$

Prior		Mean	St. Dev.	2.5%	50%	97.5%
Inv-Gamma(a, b)	α_1	5.762	0.415	4.968	5.758	6.577
$(a, b = 0.1)$	β	-3.492	0.117	-3.719	-3.493	-3.262
	γ	1.588	0.220	1.164	1.585	2.023
	Total	182,582	26,243	137,280	180,570	239,850
Inv-Gamma(a, b)	α_1	5.759	0.4012	4.954	5.758	6.548
$(a = b = 0.001)$	β	-3.49	0.1114	-3.71	-3.49	-3.271
	γ	1.586	0.2099	1.172	1.583	1.995
	Total	178,032	26,660	132,210	175,890	237,510
folded Cauchy	α_1	5.789	0.410	4.986	5.791	6.591
$s = 12$	β	-3.495	0.114	-3.717	-3.495	-3.268
	γ	1.601	0.216	1.170	1.603	2.026
	Total	182,229	27,272	136,110	179,790	242,970
Uniform	α_1	5.780	0.407	4.972	5.779	6.582
$(0, 50)$	β	-3.494	0.113	-3.717	-3.493	-3.272
	γ	1.596	0.214	1.178	1.598	2.019
	Total	181,973	27,195	136,110	179,400	242,970

Table 7: *Investigation of sensitivity with respect to the prior distribution of the variance components: posterior distributions for selection of parameters obtained with various choices of priors for σ^2_β and σ^2_γ. 700,000 simulations used, to which a thinning factor of 10 is applied, after a burn-in of 50,000 simulations.*

where the future payments Y_{ij} are modelled using an over-dispersed Poisson model as in (19). In any realistic model for the return process, R will be a sum of strongly dependent random variables. Because one can not rely on traditional risk theory, it becomes hard or even impossible to compute the cumulative distribution function ('cdf') of R analytically, though this cdf – and the calculation of different risk measures from it – is of interest in a decision making process. For a parametric claims reserving model, Antonio *et al.* (2005) illustrated how the predictive distribution of the discounted reserve can be obtained in a Bayesian way. Using Bayesian statistics and the implementation of the smoothing models in Section 2, simulations from the posterior predictive distribution of R, in case of a semiparametric model for the payments, can be obtained. With $\mu = 0.08$ and $\delta = 0.11$ the results in Table 8 follow. Figure 6 illustrates the mixing and convergence of the generated Markov Chains for some fixed effects parameters and the total discounted reserve.

	Mean Bayes.	St.Dev. Bayes.	5% Bayes.	50% Bayes.	97.5% Bayes.
AY 2	626	585	0	415	2,054
AY 3	2,240	1,338	325	2,027	5,426
AY 4	5,334	2,350	1,775	5,002	10,870
AY 5	7,952	2,854	3,481	7,578	14,610
AY 6	11,220	3,602	5,532	10,740	19,430
AY 7	14,200	4,118	7,627	13,710	23,680
AY 8	18,160	4,796	10,390	17,620	29,010
AY 9	24,490	5,918	14,840	23,820	37,840
AY 10	63,780	11,460	44,070	62,780	88,720
Total	147,997	29,133	99,703	144,944	213,675

Table 8: *Over-dispersed Poisson model with discounting process included. Bayesian results for arrival year and total reserves. 700,000 simulations used, to which a thinning factor of 10 is applied, after a burn-in of 50,000 simulations.*

3.2 Combining data on claim intensities and claim counts

Denote with Y_{ij} the aggregate payment for cell (i, j), as shown in Table 2, and let N_{ij} be the corresponding number of claims, as displayed in Table 3. 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

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

Furthermore, the Z_{ij}'s and N_{ij}'s are assumed to be independent. Thus, in contrast with de Alba (2002), a semiparametric regression model is fitted, which models the additive predictor for the average payments and the number of claims as a sum of smooth functions over the development years, together with categorical variables in the direction of arrival years. The plots in Figure 2 illustrate that appropriate modelling of the trends over the development period is necessary.

With the assumption of independence, the likelihood corresponding with (24) is the product of the likelihood expressions for the average payments and the claim numbers.

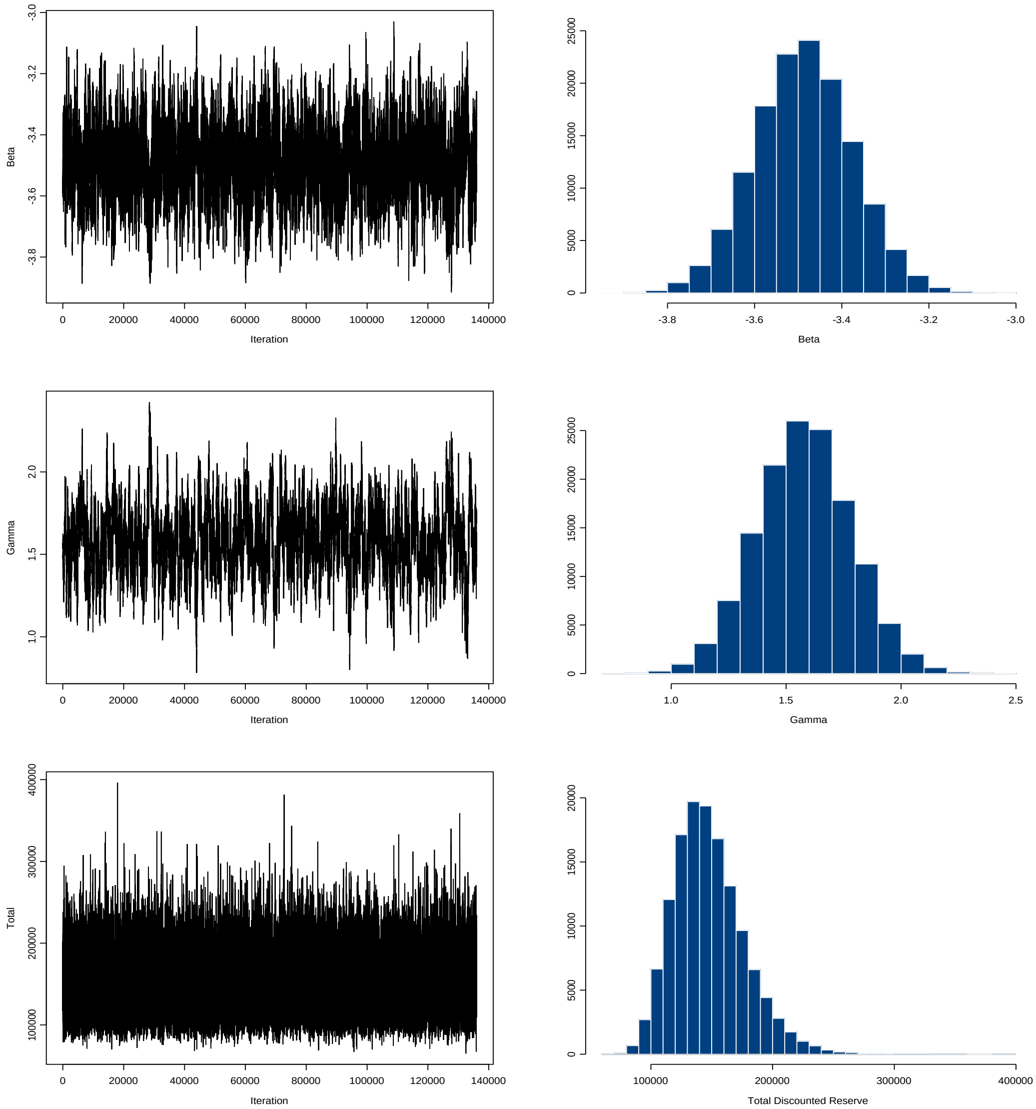

Figure 6: *Trace plots and histograms of the generated chains for β, γ and the total discounted reserve, over-dispersed Poisson model.*

The various reserves are then obtained by appropriately summing up the fitted values $\hat{Z}_{ij}\hat{N}_{ij}$, or the simulated values from the predictive distribution of $Z_{ij}N_{ij}$.

In an initial stage of the analysis, we also experimented with trends in the direction of calendar years. However, the specification in (24) is to be preferred. This is in line with Antonio *et al.* (2005), where a trend model for these data is used which does not contain parameters in the direction of calendar years either. Truncated line and quadratic basis functions, as well as radial basis functions, are used to model $f^{Av}(.)$ and $f^{Num}(.)$ [7]. Based

[7] Truncated basis functions of degree p: $\alpha_1 * I(i=1) + \ldots + \alpha_{10} * I(i=10) + \beta_1 * j + \ldots + \beta_p * j^p + \sum_{i=1}^{K} b_k (j-\kappa_k)_+^p$ and radial basis functions: $\alpha_1 * I(i=1) + \ldots + \alpha_{10} * I(i=10) + \beta_1 * j + \sum_{i=1}^{k} b_i * |j-\kappa_i|^3$ and $(b_1, \ldots, b_k)^t \sim N(\mathbf{0}, \sigma_b^2(\mathbf{\Omega}^{-1/2})(\mathbf{\Omega}^{-1/2})^t)$ where $\mathbf{\Omega} = \left[|\kappa_k - \kappa_l|^3_{1\leq k,l\leq K}\right]$.

on an inspection of the scatterplots in Figure 2 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. The resulting fits for $f^{Num}(j)$, as obtained with the different types of basis functions, are illustrated in Figure 6.

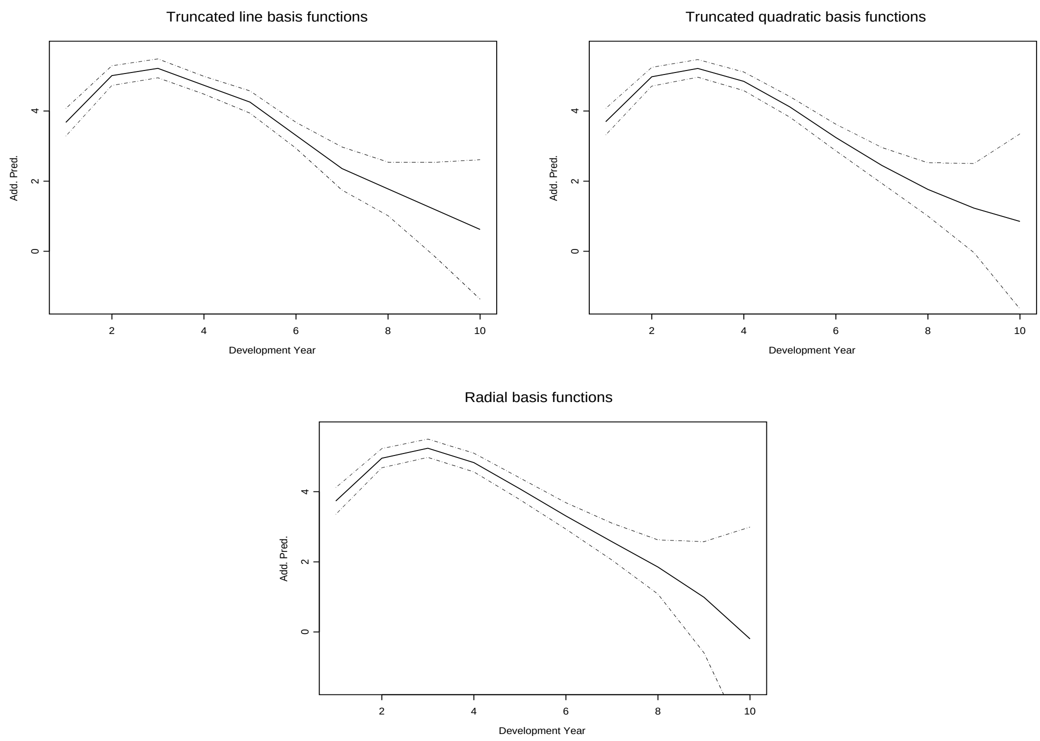

Figure 7: *Trend over development period for the claim numbers. Truncated line, truncated quadratic and radial basis functions with 4 knots, over-dispersed Poisson model. Fitted functions, together with pointwise 95% confidende intervals for* $f^{Num}(j)$.

Priors for the Bayesian analysis (with similar notations as in Section 3.1.1) are given by

$$\begin{aligned}
\beta^{Av},\ \beta^{Num} &\sim \text{dunif}(-10, 10)\\
\alpha_i^{Av},\ \alpha_i^{Num}\ (i = 1, \ldots, 10) &\sim \text{dunif}(-10, 10)\\
\nu &\sim \text{dunif}(0, 100)\\
\sigma^2_{b,Av} &\sim \text{Inv-Gamma}(0.01, 0.01)\\
\sigma^2_{b,Num} &\sim \text{Inv-Gamma}(0.01, 0.01). \qquad (25)
\end{aligned}$$

β^{Av} and β^{Num} are the fixed effects used when fitting the smooth functions. α_i^{Av} and α_i^{Num} are parameters in the direction of arrival years. $\sigma^2_{\beta,Av}$ and $\sigma^2_{\beta,Num}$ are the variance components used for the smooth functions $f^{Av}(.)$ and $f^{Num}(.)$, respectively.

95% credible intervals for some of the parameters used in the specification in (25) are given in Table 9. Results for the reserves from this model are summarized in Table 10 (claim counts) and Table 11 (total payments, obtained by multiplying claim numbers and average payments). In Table 10 the results obtained with a regular as well as an over-dispersed Poisson model are shown, both with an additive predictor as in (25). Note that the former are close to the results from a deterministic chain-ladder, as shown in Table 3.

	Mean Bayes.	St.Dev. Bayes.	2.5% Bayes.	50% Bayes.	97.5% Bayes.
α_1^{Num}	7.685	0.405	6.889	7.684	8.48
α_3^{Num}	7.818	0.402	7.033	7.817	8.608
β^{Num}	1.344	0.1739	1.009	1.342	1.69
$\sigma^2_{b,Num}$	1.177	2.587	0.188	0.696	4.905
α_1^{Av}	7.615	0.4719	6.636	7.631	8.494
α_3^{Av}	7.892	0.477	6.91	7.908	8.771
β^{Av}	-0.403	0.181	-0.782	-0.393	-0.077
$\sigma^2_{b,Av}$	0.258	0.621	0.028	0.139	1.176
ν	0.218	0.049	0.142	0.212	0.331

Table 9: *95% credible intervals for parameters in model (25): results from a Bayesian analysis with truncated line basis functions, using a smooth function over the development years (for both claim counts and average payments). 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 A two-part semiparametric model for semicontinuous data

Actuaries often have to deal with data sets containing an inflated number of zeros, in case of claim counts, or semicontinuous data. The latter combine a continuous distribution with point masses at one or more locations. For an example, consider the data in Table 4 which consist of a mixture of zeros and strictly positive values. Following (among others) Olsen & Schafer (2001) and Kunkler (2004), for the specific context of a run-off triangle, two extra random variables are introduced to describe such data, namely

$$\delta_{ij} = \begin{cases} 0 & \text{if } Y_{ij} = 0 \\ 1 & \text{if } Y_{ij} > 0, \end{cases} \quad \text{and} \quad Y^{\star}_{ij} = \begin{cases} \text{irrelevant} & \text{if } Y_{ij} = 0 \\ Y_{ij} & \text{if } Y_{ij} > 0, \end{cases} \tag{26}$$

where Y_{ij} represents the aggregate amount paid out for cell (i, j) in the run-off triangle. For the data example in Section 1.1.3, for instance, the δ_{ij} variables $(i, j = 1, \ldots, 13)$

	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 10: *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.*

denote whether at least one claim for cell (i, j) occurs. Given that a claim has occurred, $Y^{\star}_{ij}$ records the total severity for cell (i, j). Whereas the construction of the linear predictor in Kunkler (2004) was not explained in detail, we consider here a semiparametric approach for the specification of the additive predictor in each part of the two-part model in (26). As mentioned before, this is an easy to implement and very flexible way of working.

In the sequel of this analysis, both parts of the two-part model are modelled using a smooth function over the development period. Firstly, the GAM for the binary data set is described. For every cell (i, j),

$$\begin{aligned} \delta_{ij} &\sim \text{Binomial}(1, \pi_{ij}), \\ \text{where logit}(\pi_{ij}) &= f^{Bin}(j), \end{aligned} \tag{27}$$

and $f^{Bin}(.)$ is a smooth function over the development years. Secondly, for the random variables $Y^{\star}_{ij}$, a lognormal model (as in (28) below) or a model from the class of generalized linear models is fitted. Let us assume for instance that the effect of arrival years can be described using categorical variables and that a smooth function over development years applies. Thus,

$$\begin{aligned} \log(Y^{\star}_{ij}) &= a^{Sev} + \alpha_2 * I(i = 2) + \ldots + \alpha_9 * I(i = 9) + \alpha_{10} * I(i \geq 10) + f^{Sev}(j) + \epsilon_{ij}, \\ \text{where } \epsilon_{ij} &\sim N(0, \sigma^2_{\epsilon}). \end{aligned} \tag{28}$$

Denote with $\boldsymbol{\beta}^{Bin}$ the fixed effects and let $\boldsymbol{b}^{Bin}$ be the vector with the random effects that are used to model $f^{Bin}(.)$. The corresponding variance parameter is $\sigma^2_{b,Bin}$. Similarly,

	Mean Bayes.	St.Dev. Bayes.	2.5% Bayes.	50% Bayes.	97.5% Bayes.
AY 2	165,258	499,602	0	0	1,549,391
AY 3	371,707	742,439	0	0	2,407,384
AY 4	605,817	909,093	0	312,485	3,059,008
AY 5	1,038,426	1,126,787	0	726,448	3,963,113
AY 6	1,562,111	1,305,660	111,057	1,239,298	4,908,412
AY 7	2,472,569	1,611,940	522,865	2,103,166	6,510,212
AY 8	3,801,502	2,328,059	947,493	3,288,440	9,693,514
AY 9	5,503,211	3,522,317	1,343,632	4,673,287	14,506,980
AY 10	5,982,886	5,937,080	494,505	4,242,313	21,772,130
Total	21,503,490	8,990,044	9,512,551	19,753,450	43,902,620

Table 11: *Predictive distribution of the reserves obtained with model (24): results from a Bayesian analysis with truncated line basis functions for smooth functions over development. A burn-in of 50,000 simulations was used, followed by another 450,000 simulations to which a thinning factor of 10 was applied.*

$\boldsymbol{\beta}^{Sev}$ are fixed effects parameters in the direction of development years. $\boldsymbol{b}^{Sev}$ then denotes the random effects used for this smooth function, with corresponding variance parameter $\sigma^2_{b,Sev}$. Independence between $\boldsymbol{b}^{Bin}$ and $\boldsymbol{b}^{Sev}$ is assumed.

The likelihood for this two-part model then becomes

$$\begin{aligned}
& L(\boldsymbol{\beta}^{Bin}, \boldsymbol{\beta}^{Sev}, \sigma_\epsilon, \sigma_{b,Bin}, \sigma_{b,Sev} | y_{ij}, i = 1, \ldots, n, j = 1, \ldots, n-i+1) \\
= & \prod_{i,j} \int f(y_{ij} | \boldsymbol{\beta}^{Bin}, \boldsymbol{\beta}^{Sev}, \boldsymbol{b}^{Bin}, \boldsymbol{b}^{Sev}, \sigma_\epsilon) \\
& \times \quad f(\boldsymbol{b}^{Bin}, \boldsymbol{b}^{Sev} | \sigma_{b,Bin}, \sigma_{b,Sev}) d\boldsymbol{b}^{bin} d\boldsymbol{b}^{Sev} \\
= & \int \prod_Z (1-\pi_{ij}) \prod_{NZ} \pi_{ij} f(y^\star_{ij} | \boldsymbol{\beta}^{Sev}, \boldsymbol{b}^{Sev}, \sigma_\epsilon) \\
& \times \quad f(\boldsymbol{b}^{Bin}, \boldsymbol{b}^{Sev} | \sigma_{b,Bin}, \sigma_{b,Sev}) d\boldsymbol{b}^{bin} d\boldsymbol{b}^{Sev} \\
= & \int \exp\left(\sum_{NZ,Z} l_{\delta_{ij}}\right) \exp\left(\sum_{NZ} l_{Y^\star_{ij}}\right) f(\boldsymbol{b}^{Bin}, \boldsymbol{b}^{Sev} | \sigma_{b,Bin}, \sigma_{b,Sev}) d\boldsymbol{b}^{bin} d\boldsymbol{b}^{Sev} \\
= & \int \exp\left(\sum_{NZ,Z} l_{\delta_{ij}}\right) f(\boldsymbol{b}^{Bin} | \sigma_{b,Bin}) d\boldsymbol{b}^{Bin} \int \exp\left(\sum_{NZ} l_{Y^\star_{ij}}\right) f(\boldsymbol{b}^{Sev} | \sigma_{b,Sev}) d\boldsymbol{b}^{Sev}. \quad (29)
\end{aligned}$$

Here, $\sum_{NZ}$ (NZ: Non-Zero) denotes summation over all $y_{ij} > 0$ and $\sum_Z$ (Z: Zero) summation over all $y_{ij} = 0$. $l_{Y^\star_{ij}}$ is the part of the log-likelihood related to a strict positive claim, conditional on the random effects $\boldsymbol{b}^{Sev}$. $l_{\delta_{ij}} = \delta_{ij}\text{logit}(\pi_{ij}) + \log(1-\pi_{ij})$, conditional on the random effects $\boldsymbol{b}^{Bin}$. The last equation in (29) illustrates that both parts of the likelihood have to be maximized separately when maximizing the complete two-part likelihood.

In a Bayesian approach using Gibbs sampling, two separate sampling schemes are set up; one for the binary model and one for the lognormal model. To obtain simulated values from the posterior predictive distribution of the Y_{ij}, both are combined via $Y_{ij} = \delta_{ij} Y^{\star}_{ij} + (1-\delta_{ij}) * 0$. Table 12 contains 95% credible intervals for the parameters used in the models specified by (27) and (28). In this table, a_{Bin} is the intercept used to model $f^{Bin}(.)$. Figure 8 displays the fitted additive predictors, obtained by using truncated line basis functions. The predictive distributions of the various reserves are summarized in Table 13.

	Mean Bayes.	St.Dev. Bayes.	2.5% Bayes.	50% Bayes.	97.5% Bayes.
a_{Bin}	0.764	0.778	-0.8989	0.8236	1.993
β^{Bin}	-0.3909	0.2471	-0.8931	-0.3786	0.03485
$\sigma^2_{b,Bin}$	0.0512	0.2129	0.0028	0.0153	0.3084
a_{Sev}	1.391	1.131	-0.799	1.382	3.64
α_3^{Sev}	1.556	0.489	0.6	1.556	2.516
β^{Sev}	-1.374	0.3042	-1.967	-1.376	-0.7716
$\sigma^2_{b,Sev}$	1.811	3.647	0.1844	1.005	8.173
σ^2_{ϵ}	0.5637	0.1242	0.3716	0.5469	0.8546

Table 12: *95% credible intervals for parameters used in models (27) and (28): results from a Bayesian analysis with truncated line basis functions. 300,000 simulations used, to which a thinning factor of 5 is applied, after a burn-in of 50,000 simulations.*

3.4 A semiparametric model for longitudinal credibility data

Credibility rate making is a technique for predicting future expected claims of a risk class, given past claims of the given and related risk classes. Recently, the use of (generalized) linear mixed models as a statistical tool for credibility rate making has been discussed in Frees *et al.* (1999,2002) and Antonio & Beirlant (2005).

To illustrate the use of semiparametric regression models for credibility, the data introduced in Section 1.1.4, originally from Frees & Wang (2005), are analyzed. Since the data are longitudinal, the dependencies over time between observations on the same town should be taken into account appropriately. Whereas Frees & Wang (2005) used a t-copula to achieve this, we consider typical concepts of mixed models, namely the inclusion of random effects or the use of a special structure for the covariance matrix of

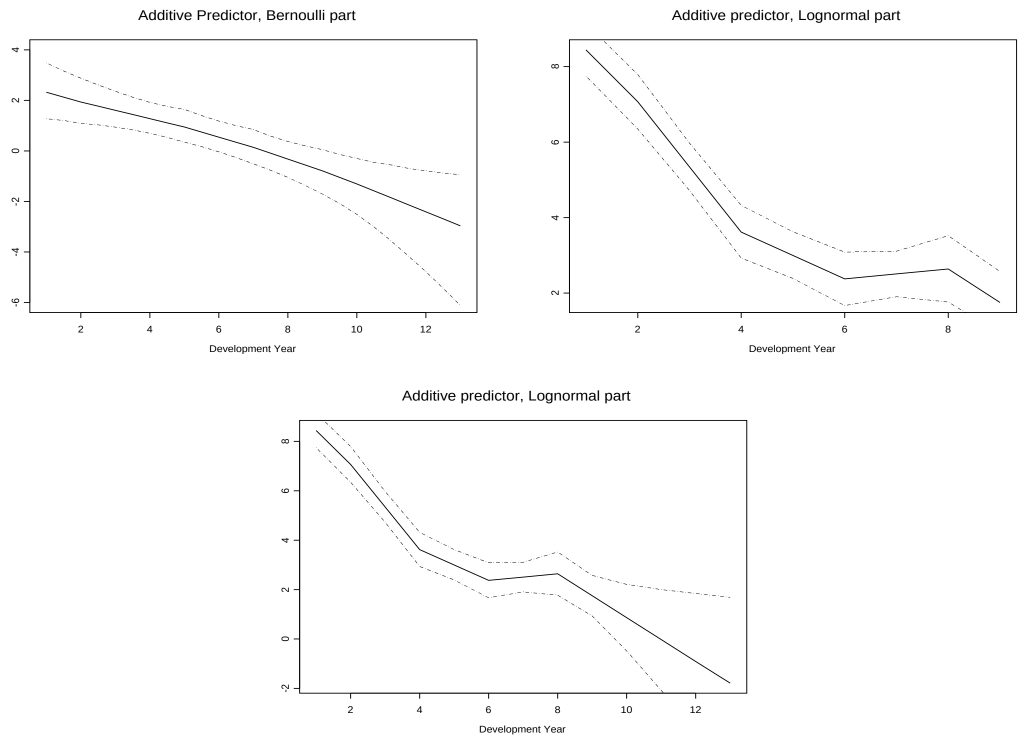

Figure 8: *Fitted additive predictors for models (27) and (28): results from a Bayesian analysis with truncated line basis functions. Posterior mean of $f^{Bin}(j)$ and $f^{Sev}(j)$, together with 95% pointwise credible bands. 300,000 simulations used, to which a thinning factor of 5 is applied, after a burn-in of 50,000 simulations.*

the residual terms (in a linear mixed model). The effects of the explanatory variables, PCI and PPSM, are modelled semiparametrically.

Firstly, the marginal effects of both explanatory variables, PCI and PPSM, on the response variable, AC, are illustrated. The lognormal distribution is used for the distribution of average claims (AC). The qqplot in Frees & Wang (2005, page 36) illustrates that this is a plausible choice. Ignoring the longitudinal structure of the data, we fit (with n the total number of observations)

$$\log{(AC)_i} = f(PCI_i) + \epsilon_i, \ i = 1, \ldots, n, \tag{30}$$

$$\log{(AC)_i} = g(PPSM_i) + \epsilon_i \ i = 1, \ldots, n. \tag{31}$$

Both $f(.)$ and $g(.)$ are estimated using mixed models, for instance with truncated line basis functions. We used 15 knots for both functions, which were automatically chosen with the procedure from Ruppert *et al.* (2003, page 125). Following Frees & Wang (2005), a rescaled version of PCI is used, namely PCI/1000, together with the logarithm of PPSM. The observations for year 1998 are reserved as the 'hold-out' sample, to validate predictions in a later stage of the analysis. Parameter estimates are in Table 14.

	Mean Bayes.	St.Dev. Bayes.	2.5% Bayes.	50% Bayes.	97.5% Bayes.
AY 2	0.28	1.965	0	0.041	1.843
AY 3	1.191	13.66	0.003	0.17	7.75
AY 4	3.272	8.32	0.111	1.367	18.01
AY 5	2.036	11.29	0.018	0.4817	12.93
AY 6	69.6	105	8.984	46.04	269
AY 7	58	51	11.86	44.53	183
AY 8	43.7	89.87	0	7.71	271
AY 9	38.45	60.38	0	19.66	186.9
AY 10	86.14	106.8	0.2163	55.78	356.4
AY 11	153.5	151.4	12.43	112.3	538
AY 12	385	411	58,87	280	1,407
AY 13	5,302	5,151	339.6	4,012	18,210
Total	6,132	5,193	976.7	4,848	19,100

Table 13: *Two-part model for semicontinuous data: results from a Bayesian analysis with truncated line basis functions. 300,000 simulations used, to which a thinning factor of 5 is applied, after a burn-in of 50,000 simulations.*

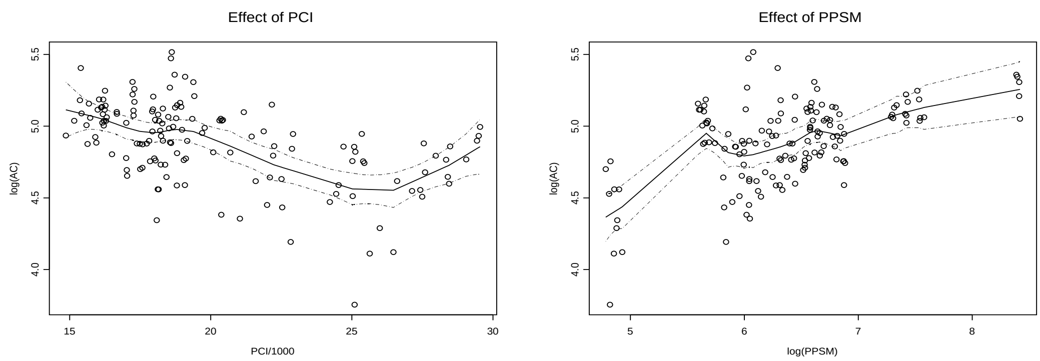

Figure 9: *Marginal effect of PCI (left) and PPSM (right) on log(AC), together with 95% pointwise confidence intervals for $f(PCI_i)$ and $g(PPSM_i)$; truncated line basis functions with 15 knots. Results obtained with Proc Mixed in SAS.*

Secondly, a lognormal mixed model is fitted to the data, in which nonlinear effects of both PCI and PPSM are allowed. To take the dependencies over time into account, two strategies are considered: the inclusion of a random intercept per town on the one hand and the specification of a special structure for the covariance matrix of the residual terms

Marginal Effect	β_0	β_1	σ_b^2	σ_ϵ^2
PCI	5.8497 (1.2641)	-0.04946 (0.07916)	0.0056	0.05225
PPSM	1.4537 (2.6472)	0.6075 (0.5429)	0.4373	0.04864

Table 14: *Scatterplot smoothing of* $\log(AC)$ *versus PCI and PPSM, respectively. Results obtained with Proc Mixed in SAS.*

on the other hand. Thus, in its most general specification, models of the form

$$\begin{aligned}
\log(AC_{it}) &= f(PCI_{it}) + g(PPSM_{it}) + b_i + \epsilon_{it}, \quad i = 1, \dots, N \text{ and } j = 1, \dots, n_i, \\
b_i &\sim N(0, \sigma_b^2) \\
\boldsymbol{\epsilon}_i &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_i), \text{ where } \boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})'
\end{aligned} \tag{32}$$

are considered. Modeling nonlinear effects of PCI and PPSM, the combination of random intercepts and a non-diagonal $\boldsymbol{\Sigma}_i$ did not lead to convergence of Proc Mixed in SAS. Models with nonlinear effects of PCI and PPSM, together with a non-diagonal $\boldsymbol{\Sigma}_i$, did not lead to convergence of the procedure either. The results for different – convergent – model specifications are summarized in Table 15.

Parameter	Model I	Model II	Model III	Model IV
Intercept	4.1913	4.2301	4.334	4.3729
β_{PCI}	-0.02852	-0.02844	-0.03216	-0.03198
β_{PPSM}	0.1998	0.1928	0.1869	0.1816
$\sigma^2_{b,PCI}$	0	/	/	/
$\sigma^2_{b,PPSM}$	0.0001	/	/	/
σ_b^2	0.0208	0.0209	/	/
σ_ϵ^2	0.022	0.022	0.03874	0.04107
AR(1)	/	/	/	0.4335
Toep(2)	/	/	0.01056	/
Toep(3)	/	/	0.01618	/

Table 15: *Model 1: nonlinear effects of PCI and PPSM, together with random intercepts and Toep(1) structure for covariance of residual terms. Model II: linear effects of PCI and PPSM, together with random intercepts. Model III: linear effects of PCI and PPSM, Toep(3) structure for covariance of residual terms. Model IV: linear effects of PCI and PPSM, AR(1) structure for covariance of residual terms.*

For prediction purposes, Models II, III and IV were considered. Apart from these, a fifth model is also considered which has the same specifications as Model II, but assumes a gamma distribution for the data (cfr. the qqplots in Frees & Wang, 2005). The predicted values for the hold-out sample our calculated for the different models. As in Frees & Wang (2005), the sum of squared prediction errors (SSPE) is used to compare the predictive performance of the different models. The SSPE is tabulated in Table 16. The SSPE for full credibility (thus, $\hat{y}_{i,n_i+1} = \bar{y}_i = \frac{1}{n_i}\sum_{t=1}^{n_i} y_{it}$) is 15,701 and for the Bühlmann model (thus, $\hat{y}_{i,n_i+1,B} = \zeta\bar{y}_i + (1-\zeta)\bar{y}$ where ζ is the credibility factor and $\bar{y}$ the overall mean) is 14,916. Except for Model III, all the reported SSPE are lower than those in Frees & Wang (2005). Although a visual inspection of the plots in Figure 8 might suggest nonlinear effects of PCI and PPSM, an analysis using mixed models (both in a lognormal and a gamma framework) revealed that – as in Frees & Wang (2005) – linear effects of PCI and PPSM are sufficient.

	Model II	Model III	Model IV	Model V
SSPE	14,244	15,436	13,775	14,117

Table 16: *Comparisons of Sum of Squared Prediction Error.*

To obtain the full predictive distributions of AC for year 1998 and for the different towns, a Bayesian implementation of one of the suggested should be used.

4 Conclusions

This paper revisits the use of semiparametric regression models in the context of claims reserving and credibility. Penalized splines and their connection with mixed models are used, both in a likelihood-based and in a Bayesian way. In an actuarial context, a first advantage of this approach is the fact that the results from the likelihood approach can be obtained efficiently, using a standard statistical software package like SAS. This is an important advantage, for instance in the search for an appropriate predictor structure when analyzing a run-off triangle. Moreover, the use of semiparametric regression models allows to specify such a predictor structure in a flexible way and under various distributional assumptions, using standard software packages. Since actuaries are mainly interested in predictions, the straightforward Bayesian implementation of these models is a second main advantage of the semiparametric regression models based on mixed models. Three examples from claims reserving – which have the intention to be representative for actuarial applications – were analyzed and discussed, as well as an example dealing with longitudinal credibility data. Future research in this area will concentrate on the use of different model selection criteria and the implementation of semiparametric regression models for cross-sectional and longitudinal data within the class of heavy-tailed distributions.

Acknowledgements

Interested readers can contact the first author for illustrative SAS, SPlus and `WinBUGS` code.

References

[1] Antonio, Katrien, and Jan Beirlant. 2005. Applications of Generalized Linear Mixed Models in Actuarial Statistics. *Working Paper. Online at http://www.econ.kuleuven.be/insurance.*

[2] Antonio, Katrien, Jan Beirlant, and Tom Hoedemakers. 2005. Discussion of 'A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving.' *North American Actuarial Journal* 9(3): 143-145.

[3] Antonio, Katrien, Jan Beirlant, Tom Hoedemakers and Robert Verlaak. 2006. Lognormal mixed models for reported claims reserving. *North American Actuarial Journal*, 10(1): 1-19.

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

[5] Crainiceanu, C., D. Ruppert and M.P. Wand. 2004. Bayesian analysis for penalized spline regression using WinBUGS. *Working paper.*

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

[7] Denuit, M. and S. Lang. 2004. Non-life ratemaking with Bayesian GAMs. *Insurance: Mathematics and Economics* 35(3): 627-647.

[8] De Vylder, F. and M.J. Goovaerts. 1979. *Proceedings of the first meeting of the contact group "Actuarial Sciences".* KU Leuven, nr 7904B, wettelijk depot: D/1979/23761/5.

[9] Durban, M., J. Harezlak, M.P. Wand and R.J. Carroll. 2005. Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine.*

[10] England, P.D. and R.J. Verrall. 2002. Stochastic claims reserving in general insurance. *British Actuarial Journal* 8: 443-544.

[11] England, P.D. and R.J. Verrall. 2001. A flexible framework for stochastic claims reserving. *Proceedings of the casualty actuarial society forum.*

[12] Fahrmeir, L., S. Lang and F. Spies. 2003. Generalized geoadditive models for insurance claims data. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* 26: 7-23.

[13] Frees, E.W., V.R. Young, and Y. Luo. 1999. *A longitudinal data analysis interpretation of credibility models.* Insurance: Mathematics and Economics, 24(3), 229-247.

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

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

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

[17] Hastie, T.J. and R.J. Tibshirani. 1990. *Generalized additive models.* London: Chapman & Hall.

[18] Kaas, R., M.J. Goovaerts, J. Dhaene, and M. Denuit. 2001. *Modern actuarial risk theory.* Kluwer, The Netherlands.

[19] Kunkler, M. 2004. Modelling zeros in stochastic reserving models. *Insurance: Mathematics and Economics* 34(1): 23-35.

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

[21] Olsen, M.K. and J.L. Schafer. 2001. A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association* 96 (454): 730-745.

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

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

[24] Verrall, R.J. 1996. Claims reserving and generalized additive models. *Insurance: Mathematics and Economics* 19: 31-43.

[25] Verrall, R.J. 2005. Discussion of 'A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving.' *North American Actuarial Journal* 9(3): 149.

[26] Zhao, Y., J. Staudenmayer, B.A. Coull and M.P. Wand. 2004. General design Bayesian generalized linear mixed models. *Working paper.*