

Issues in Claims Reserving and Credibility: a Semiparametric Approach with Mixed Models

Katrien Antonio^{*†} Jan Beirlant[‡]

January 16, 2007

Abstract

Verrall (1996) and England & Verrall (2001) first considered the use of smoothing methods in the context of claims reserving. They applied 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 credibility. Apart from the flexibility inherent to all semiparametric methods, advantages of the semiparametric approach developed here are threefold. Firstly, a Bayesian implementation of these smoothing models is relatively straightforward and allows simulation from the full predictive distribution of quantities of interest. Since the main interest of actuaries lies in prediction, this is a major advantage. Secondly, because the constructed models have an interpretation as (generalized) linear mixed models ((G)LMMs), standard statistical theory and software for (G)LMMs can be used. Thirdly, more complicated data sets, dealing for example with quarterly development in a reserving context, heavy-tails, semicontinuous data, or extensive longitudinal data, can be modelled within this framework. Throughout this article, data examples illustrate these different aspects. Several comments are included regarding model specification, estimation and selection.

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, Statistics Section and AFI Research Center, KU Leuven, W. de Croylaan 54, 3001 Heverlee, Belgium.

[‡]Full professor, Statistics Section, 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. For these claims, provisions will be held to meet 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 by these authors are framed within the context of a run-off triangle like the one in Table 1. Its design is typical for a claims reserving problem.

<i>Arrival</i> <i>Year</i>	<i>Development Year</i>						
	1	2	...	<i>j</i>	...	<i>n</i> - 1	<i>n</i>
1	Y_{11}	Y_{12}	...	Y_{1j}	...	$Y_{1,n-1}$	Y_{1n}
2	Y_{21}	Y_{22}	...	Y_{2j}	...	$Y_{2,n-1}$	
⋮		
<i>i</i>	Y_{i1}	Y_{ij}			
⋮				
<i>n</i>	Y_{n1}						

Table 1: *Random variables in a run-off triangle.*

The random variable Y_{ij} (for $i, j = 1, \dots, n$) denotes the claim figure for year of origin (arrival or incurral year) i and development year j . It represents for instance claim counts, incremental or cumulative payments or loss ratios, aggregated per (i, j) combination. Random variables on the (i, j) diagonal correspond with payments made in the same calendar year, namely calendar year $i + j - 1$. The purpose of loss reserving techniques is to complete this run-off triangle to a square or a rectangle. To achieve this, stochastic reserving techniques will be used. Current state-of-the-art models are loglinear location-scale (Doray, 1994), lognormal or generalized linear models with the mean or predictor for Y_{ij} (say, η_{ij} , for $i, j = 1, \dots, n$) specified in a parametric way. Well-known and widely used specifications are

$$\eta_{ij} = \alpha_i + \beta_j \text{ ('chain-ladder')}; \quad (1)$$

$$\eta_{ij} = \alpha_i + \sum_{k=1}^{j-1} \beta_k + \sum_{t=1}^{i+j-2} \gamma_t \text{ ('probabilistic trend family')}; \quad (2)$$

$$\eta_{ij} = \alpha_i + \beta_i \log(j) + \gamma_i j \text{ ('Hoerl curve')}. \quad (3)$$

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 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 (structure (2)) (De Vylder & Goovaerts, 1979). The intention of this paper is to reduce the amount of work involved in the specification of the predictor by relying on semiparametric regression models. The benefits of a semiparametric approach in reserving become even more obvious when more extensive data are considered, such as the ‘triangles’ with quarterly development described in Example 1.1.3.

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 SPlus¹. 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 SAS²), also a Bayesian implementation of the models and, consequently, simulation from the predictive distribution of quantities of interest is relatively straightforward. In this way, we extend the work of England and Verrall to include predictive distributions.

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 3.4). In the likelihood-based approach in England and Verrall (2001), only a ‘standard’ run-off triangle could be considered. Moreover, using P-splines and their Bayesian implementation, it

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

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

is illustrated in Section 3.3 how semiparametric Burr reserving models (and loglinear location-scale models) can be constructed. This extends the parametric approach in Doray (1994) and Beirlant *et al.* (1998).

Using penalized splines and their connection with mixed models, longitudinal and cross-sectional data can be modelled semiparametrically in the same framework. From an actuarial point of view, this feature is very appealing since it offers a natural machinery to deal with both claims reserving and credibility problems. A quote taken from the discussion of the seminal paper by England & Verrall (2002, page 529) puts further light on this issue: “*If you look within the general insurance part of the actuarial profession, there is a body of thinking that has grown up around premium rating and a body of thinking that has grown up around reserving. Are we getting ‘over-siloed’?*” Indeed, extensive data sets from reserving problems, where for instance data on individual claims – instead of data aggregated in cells as in Table 1 – with quarterly development could be available, can be modelled by combining the ideas from Example 1.1.3 and Example 1.1.5.

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 2 a data set previously analyzed in England & Verrall (2001) is shown. 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, as well as the chain-ladder development factors, are given in Table 2 as benchmark results. For actuaries, the chain-ladder is a simple, yet widely used, technique to construct reserve estimates. Its development factors are calculated in the following way (for the case that Y_{ij} represents incremental payments)

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}},$$

where $D_{ij} = \sum_{k=1}^j Y_{ik}$ (the ‘cumulative claims’), (4)

and the predictions for future values of the cumulative claims are then obtained via

$$\begin{aligned}\hat{D}_{i,n-i+2} &= D_{i,n-i+1}\hat{\lambda}_{n-i+2}, \\ \hat{D}_{i,k} &= \hat{D}_{i,k-1}\hat{\lambda}_k, \quad k = n - i + 3, \dots, n.\end{aligned}\tag{5}$$

The reserves in the last column of Table 2 are from the smooth over-dispersed Poisson model in England & Verrall (2001). They will also serve as benchmark results. Hereby the additive predictor consists of a smooth function of the logarithm of the development years, together with a parameter for each accident year. Calculations were done with the SPlus function `gam()`.

<i>Claim Payments</i>										<i>Reserves</i>	<i>Reserves</i>
										<i>Ch. Ladd.</i>	<i>Smooth o-P</i>
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 2: *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.*

In Section 3.1, the data in Table 2 will be analyzed by fitting a generalized additive model (GAM), using penalized regression splines and their connection with mixed models. Categorical variables in the direction of arrival years and smoothing over the development years will be used. We will also consider the modelling of trends in the direction of calendar years, together with a Bayesian implementation of the constructed semiparametric regression model. Recall that the approach in England & Verrall (2001) did not include predictive distributions and relied on heavy analytical calculations to obtain prediction error estimates. The inclusion of a stochastic discounting process in this statistical model is illustrated in Section 3.1.2, an easy to obtain by-product of the Bayesian implementation.

1.1.2 Aggregate data on claim intensities and claim counts

With this example we want to illustrate how information on claim counts and claim amounts can be combined in a semiparametric regression model. Using a Bayesian implementation of the smoothers used in this article, the data considered in de Alba (2002) are

reanalyzed. Ntzoufras and Dellaportas (2002) discuss a similar problem in a parametric way. The data are displayed in Table 3 and 4 and illustrated in Figure 1. A generalized additive model will be 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.

<i>Claim Payments</i>										<i>Reserves</i>
										<i>Ch. Ladd.</i>
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 3: *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.').*

<i>Claim Numbers</i>										<i>Reserves</i>
										<i>Ch. Ladd.</i>
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 4: *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.').*

1.1.3 Incremental claims with quarterly development

The benefits of smoothing techniques become more obvious when extensive run-off triangles are considered. Assume that the development of aggregate data is followed per

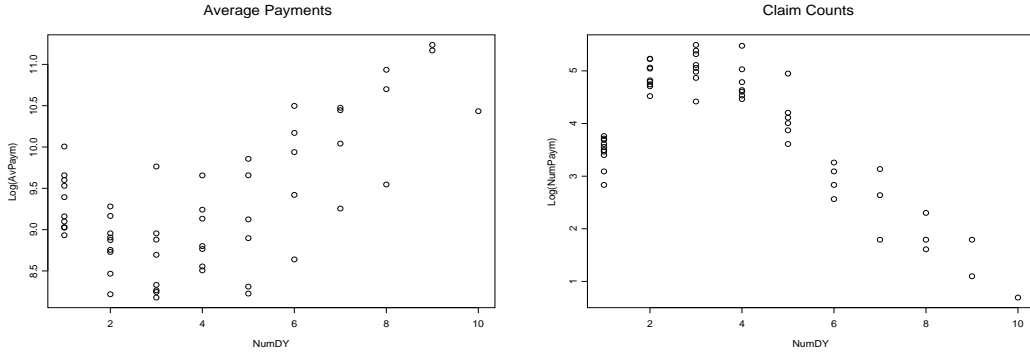


Figure 1: *Scatter plots of log-transformed data versus development year (DY); data from Table 3 (left) and Table 4 (right).*

quarter, instead of per year. Thus, the Y_{ij} , with $i = 1, \dots, n_{years}$ and $j = 1, \dots, n_{quarters}$, denote claim figures corresponding to arrival year i and development quarter j . To illustrate this, first, a real data set with quarterly development from an insurance company is considered. For reasons of confidentiality, the run-off data can not be displayed. A histogram is shown in Figure 2 (right), together with a plot of the claim payments versus their development quarter (left).

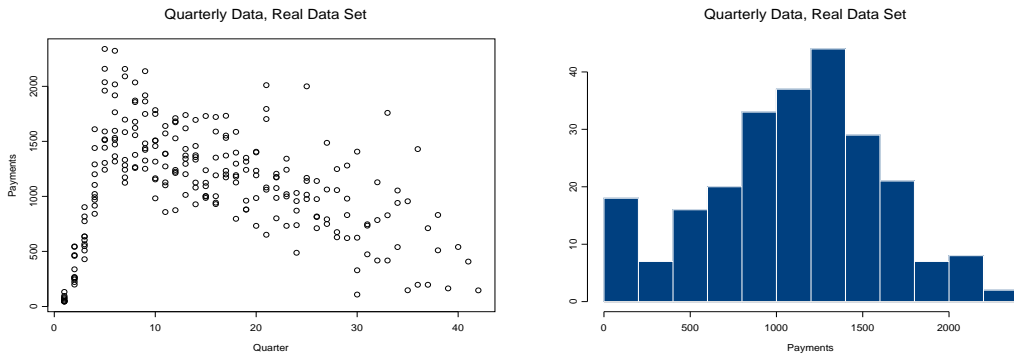


Figure 2: *(Left) Trend in the direction of development quarter. (Right) Histogram of claim payments, quarterly real data from an insurance company.*

Secondly, a similar extended triangle is simulated from a Burr distribution where

$$\begin{aligned} Y_{ij} &\sim \text{Burr}(\beta_{ij}, \lambda, \tau) \\ \beta_{ij} &= \exp(\tau \mu_{ij}), \quad i = 1, \dots, n_{years}, \quad j = 1, \dots, n_{quarters}, \end{aligned} \quad (6)$$

($\lambda = 1$, $\tau = 3$, for our simulated data). Recall that the Burr distribution is heavy-tailed, with extreme value index $\frac{1}{\lambda\tau} > 0$ (see Beirlant *et al.*, 2004). The development pattern constructed for this example is illustrated in Figure 3 (left). A histogram of the simulated data is in Figure 3 (right).

The above examples illustrate two important issues. On the one hand, specifying one of the state-of-the-art predictor structures (like those in (1)-(3)) and reducing the number of parameters through hypothesis testing will become a cumbersome job for these extensive data. Smoothing techniques then provide an elegant alternative. On the other hand, through our Bayesian approach, we are not restricted to the class of generalized linear models, but can deal for instance with heavy-tailed semiparametric regression models (as in (6)). The loglinear location-scale regression models from Doray (1994) constitute another class for which a Bayesian semiparametric approach could be considered, using the approach discussed in this article.

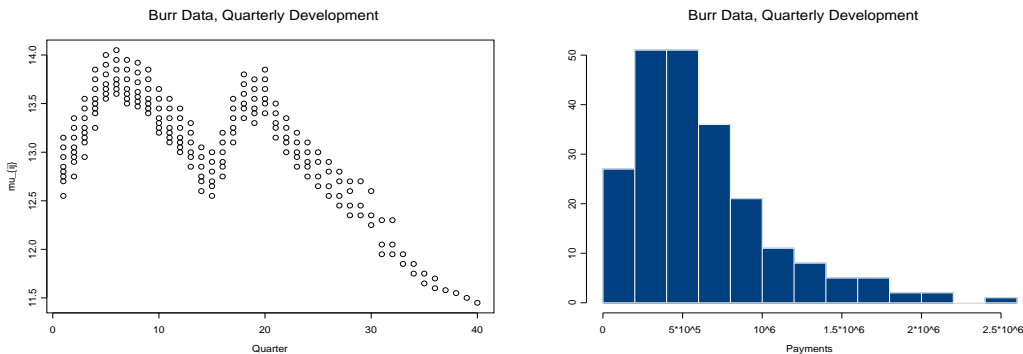


Figure 3: (Left) Scatter plot of μ_{ij} versus development quarter. (Right) Histogram of data simulated from a Burr model.

1.1.4 A two-part semiparametric model for semicontinuous data

The run-off triangle in Table 5 consists of strictly positive payments and exact zeros. The modelling of such semicontinuous ³ data (with a hump at zero) requires specific attention. A two-part generalized additive mixed model is presented for these data. Hereby, a regression 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. Again, when more extensive data become available (which often goes together with more zeros), the use of smoothing techniques is an obvious alternative for the (awkward) specification of categorical variables (for both the binary and the positive part of the data).

³A semicontinuous random variable combines a continuous distribution with point masses at one or more locations.’ (Olsen & Schafer, 2001, page 730)

<i>Claim Payments</i>											<i>Reserves</i>	
												<i>Ch. Ladd.</i>
2,216	744	10	5	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
<i>Dev. Fact.</i>	1.205	1.020	1.011	1.005	1.004	1	1.001	1.002	1	1	1	1

Table 5: *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.').*

1.1.5 An example from credibility

To illustrate the use of semiparametric regression (through 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. Whereas our previous examples were cross-sectional, these data are longitudinal. The response variable is average claims ('AC'), which is the total claim amount divided by a certain 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 , with θ_{it} the canonical parameter in the generalized linear model. In our analysis, we will investigate whether nonlinear effects of PCI and PPSM are suitable.

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 are 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, \dots, Y_n be independent random variables with a density $f(\cdot)$ from the exponential family, namely

$$f(y) = \exp\left(\frac{y\theta - \psi(\theta)}{\phi} + c(y, \phi)\right), \quad (7)$$

where $\psi(\cdot)$ and $c(\cdot)$ are known functions, 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 $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ from a GLM is in a GAM replaced by an additive predictor,

$$\begin{aligned} \eta_i &= \sum_{h=1}^l f_h(x_{ih}), \\ \text{and } \mu_i &= g(\eta_i), \quad i = 1, \dots, n, \end{aligned} \quad (8)$$

where $E[Y_i] = \mu_i$ and $g(\cdot)$ is the link function. Hereby the functions f_h ($h = 1, \dots, l$) are ‘smooth’ functions of covariates x_h ($h = 1, \dots, l$). Instead of being fully nonparametric, the additive predictor in (8) 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 $f_h(\cdot)$. 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 1.1.5, let Y_{ij} denote the j^{th} observation for subject i , where $j = 1, \dots, n_i$ and $i = 1, \dots, N$. 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 \mathbf{b}_i ($q \times 1$) for subject i , Y_{i1}, \dots, Y_{in_i} are assumed to be independent with a distribution from the exponential family in (7), thus

$$f(y_{ij}|\mathbf{b}_i) = \exp\left(\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{\phi} + c(y_{ij}, \phi)\right). \quad (9)$$

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

$$g(\mu_{ij}) = \eta_{ij} = \sum_{h=1}^l f_h(x_{ijh}) + \mathbf{z}'_{ij}\mathbf{b}_i, \quad (10)$$

where $\mu_{ij} = E[Y_{ij}|\mathbf{b}_i]$ and some of the functions $f_h(\cdot)$ can simply be parametric. To complete the specification, the \mathbf{b}_i ($i = 1, \dots, N$) are assumed to be multivariate normally distributed with mean $\mathbf{0}$ and covariance matrix \mathbf{D} . In (10) the nonparametric functions $f_h(\cdot)$ apply on the population-level. This can be generalized further to subject-specific semiparametric functions.

In the sequel of this section, only the use of penalized regression 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 the GLMM specification of the models discussed in this paper.

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*.

In order to clarify 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 $f(\cdot)$, 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 $\kappa_1, \dots, \kappa_K$ ⁴, define the design matrix \mathbf{B} as

$$\mathbf{B} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^p & (x_1 - \kappa_1)_+^p & \dots & (x_1 - \kappa_K)_+^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^p & (x_n - \kappa_1)_+^p & \dots & (x_n - \kappa_K)_+^p \end{bmatrix}. \quad (11)$$

The unknown function $f(\cdot)$ is then estimated as $\hat{f}(x) = \mathbf{B}(x)\hat{\boldsymbol{\beta}}$ where $\mathbf{B}(x)$ is a row vector, similar to a row from \mathbf{B} , and $\hat{\boldsymbol{\beta}}$ is the solution of the least-squares problem $\min_{\boldsymbol{\beta}} \sum_{i=1}^n (y_i - \mathbf{B}(x_i)\boldsymbol{\beta})^2$, 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

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n (y_i - \mathbf{B}(x_i)\boldsymbol{\beta})^2 + \alpha \boldsymbol{\beta}' \mathbf{P} \boldsymbol{\beta}, \quad (12)$$

⁴The 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 α is the so-called smoothing parameter and \mathbf{P} a penalty matrix given by

$$\mathbf{P} = \begin{bmatrix} 0_{p+1 \times p+1} & 0_{p+1 \times K} \\ 0_{K \times p+1} & \mathbf{I}_{K \times K} \end{bmatrix}. \quad (13)$$

Ruppert *et al.* (2003) (among others) rewrite the argument of the optimization problem in (12), after dividing by σ_ϵ^2 , as

$$\frac{1}{\sigma_\epsilon^2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}\|^2 + \frac{1}{\sigma_u^2} \|\mathbf{u}\|^2, \quad (14)$$

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

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^p \end{bmatrix} \text{ and } \mathbf{Z} = \begin{bmatrix} (x_1 - \kappa_1)_+^p & \dots & (x_1 - \kappa_K)_+^p \\ \vdots & \vdots & \vdots \\ (x_n - \kappa_1)_+^p & \dots & (x_n - \kappa_K)_+^p \end{bmatrix}. \quad (15)$$

By considering \mathbf{u} as random effects with $\mathbf{u} \sim N(0, \sigma_u^2 \mathbf{I}_{K \times K})$, (14) reduces to minus two times the log-likelihood of (\mathbf{Y}, \mathbf{u}) in the linear mixed model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$, under the assumptions $\mathbf{Y}|\mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma_\epsilon^2 \mathbf{I})$, $\mathbf{u} \sim N(0, \sigma_u^2 \mathbf{I})$ and $\boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \mathbf{I})$.

A similar reasoning leads to the penalized splines formulation of the GAM specified by (7) and (8). Construct the design matrix \mathbf{X} as

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & \dots & x_{11}^p & \dots & x_{1l} & x_{1l}^2 & \dots & x_{1l}^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n1}^2 & \dots & x_{n1}^p & \dots & x_{nl} & x_{nl}^2 & \vdots & x_{nl}^p \end{bmatrix}. \quad (16)$$

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

$$\mathbf{Z}^{pen} = \begin{bmatrix} (x_{11} - \kappa_1^1)_+^p & \dots & (x_{11} - \kappa_{K_1}^1)_+^p & \dots & (x_{1l} - \kappa_1^l)_+^p & \dots & (x_{1l} - \kappa_{K_l}^l)_+^p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (x_{n1} - \kappa_1^1)_+^p & \dots & (x_{n1} - \kappa_{K_1}^1)_+^p & \dots & (x_{nl} - \kappa_1^l)_+^p & \dots & (x_{nl} - \kappa_{K_l}^l)_+^p \end{bmatrix}, \quad (17)$$

where K_i denotes the number of knots to estimate $f_i(\cdot)$ ($i = 1, \dots, l$). In case of a GAM, the log-likelihood is considered as a function of the additive predictor $\boldsymbol{\eta}$ from (8) and, using penalized regression splines, $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$ is obtained as the solution of the following penalized log-likelihood

$$\max_{\boldsymbol{\beta}, \mathbf{u}} \{ \mathbf{y}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}'\psi(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) \} - \frac{1}{2} \sum_{j=1}^l \alpha_j \mathbf{u}'_j \mathbf{u}_j, \quad (18)$$

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 (16) (one parameter per column of \mathbf{X}). $\mathbf{u}_j = (u_{j1}, \dots, u_{jK_j})'$ ($j = 1, \dots, l$), α_j ($j = 1, \dots, l$) is the smoothing parameter for function $f_j(\cdot)$ and say $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_l)'$. The optimization problem in (18) is equivalent to the penalized quasi-likelihood optimization problem in a generalized linear mixed model (see Breslow & Clayton, 1993) with the GLMM specified as

$$\begin{aligned} f(\mathbf{y}|\mathbf{u}) &= \exp(\mathbf{y}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}'\psi(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) + \mathbf{1}'c(\mathbf{y})), \\ \mathbf{u} &\sim N(\mathbf{0}, \boldsymbol{\Lambda}), \\ \text{and } \boldsymbol{\Lambda} &= \begin{bmatrix} \sigma_1^2 \mathbf{I}_{K_1 \times K_1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_l^2 \mathbf{I}_{K_l \times K_l} \end{bmatrix}, \end{aligned} \quad (19)$$

where $\sigma_j^2 = 1/\alpha_j$ ($j = 1, \dots, l$) and – again – a canonical link is used in (19) for ease of notation. Both (18) and (19) are easily generalized to the case of a non-canonical link. In that situation, the relation $g\{\psi'(\boldsymbol{\theta})\} = \boldsymbol{\eta}$ (using the notation from (7) and (8)) is used.

In line with the previous specifications, the GAMM for longitudinal data, specified in (9) and (10), can be rewritten as a GLMM as well. Specify the design matrices \mathbf{X}_i and \mathbf{Z}_i for subject i ($i = 1, \dots, N$) as

$$\mathbf{X}_i = \left[\begin{array}{cccc|cccc} 1 & x_{i11} & x_{i11}^2 & \dots & x_{i11}^p & \dots & x_{i1l} & x_{i1l}^2 & \dots & x_{i1l}^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{in_i1} & x_{in_i1}^2 & \dots & x_{in_i1}^p & \dots & x_{in_il} & x_{in_il}^2 & \vdots & x_{in_il}^p \end{array} \right], \quad (20)$$

and

$$\mathbf{Z}_i^{pen} = \left[\begin{array}{cccc|cccc} (x_{i11} - \kappa_1^1)_+^p & \dots & (x_{i11} - \kappa_{K_1}^1)_+^p & \dots & (x_{i1l} - \kappa_1^l)_+^p & \dots & (x_{i1l} - \kappa_{K_l}^l)_+^p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (x_{in_i1} - \kappa_1^1)_+^p & \dots & (x_{in_i1} - \kappa_{K_1}^1)_+^p & \dots & (x_{in_il} - \kappa_1^l)_+^p & \dots & (x_{in_il} - \kappa_{K_l}^l)_+^p \end{array} \right]. \quad (21)$$

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

$$\mathbf{Z}_i^{ran} = \begin{bmatrix} z_{i11} & \dots & z_{i1q} \\ \vdots & \ddots & \vdots \\ z_{in_i1} & \dots & z_{in_iq} \end{bmatrix} \quad \text{and} \quad \mathbf{Z}_i = [\mathbf{Z}_i^{pen} | \mathbf{Z}_i^{ran}], \quad (22)$$

the contribution of subject i to the GLMM specification of the GAMM from (9) and (10) is given by

$$\begin{aligned} f(\mathbf{y}_i|\mathbf{r}_i) &= \exp(\mathbf{y}'_i(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{r}_i) - \mathbf{1}'\psi(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{r}_i) + \mathbf{1}'c(\mathbf{y}_i)), \\ \mathbf{r}_i &= (\mathbf{u}', \mathbf{b}'_i)' \sim N(\mathbf{0}, \boldsymbol{\Lambda}_i), \\ \text{and } \boldsymbol{\Lambda}_i &= \begin{bmatrix} \sigma_1^2 \mathbf{I}_{K_1 \times K_1} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_l^2 \mathbf{I}_{K_l \times K_l} & 0 \\ 0 & 0 & \dots & 0 & \mathbf{D} \end{bmatrix}. \end{aligned} \quad (23)$$

The assumption of independence among subjects completes the specification of the GLMM representation of the GAMM from (9) and (10).

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 (19) and (23) leads to an automatic choice of the smoothing parameters, namely $\hat{\alpha}_j = 1/\hat{\sigma}_j^2$ ($j = 1, \dots, l$). 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 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 (19) and (23) 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. Prior specifications for the unknown parameters are discussed in Section 3. In the rest of this article, MCMC simulations are performed using the WinBUGS package.

3 An investigation in the context of claims reserving and credibility

3.1 Aggregate data on claim severities in a run-off triangle

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 2. Trends in the direction of development year and – at a second stage – calendar year are modelled using penalized splines. The results obtained in this way are compared to those reported earlier in the literature. These state-of-the-art results were displayed in Table 2. In the example discussed in Section 3.2, some comments are included regarding the choice of the error distribution for Bayesian claims reserving. However, for this example we solely rely on the distribution suggested in England & Verrall (2001).

Denote by Y_{ij} ($i, j = 1, \dots, n$ and $n = 10$ in Table 2) the random variable corresponding to the amount paid out in arrival year i and development year j . Now start with the following model specification

$$\frac{Y_{ij}}{\phi} \sim \text{Poisson} \left(\frac{\mu_{ij}}{\phi} \right),$$
$$\text{where } \log(\mu_{ij}) = \alpha_1 I(i = 1) + \dots + \alpha_{10} I(i = 10) + f(j). \quad (24)$$

Thus, Y_{ij} follows an over-dispersed Poisson distribution, with $E[Y_{ij}] = \mu_{ij}$ and $\text{Var}[Y_{ij}] = \phi\mu_{ij}$. $f(\cdot)$ is a smooth function over the development years. In a first stage of the analysis, we modelled $f(\cdot)$ using truncated lines and $K = 4$ user-specified knots, namely $\kappa_1 = 2$, $\kappa_2 = 3$, $\kappa_2 = 5$ and $\kappa_3 = 7$. Columns 5-7 in Table 6 summarize the posterior distribution for the arrival year and total reserves. To enable comparison with a standard specification of the linear predictor in a generalized linear model, the predictive distributions of the reserves obtained with chain-ladder structure (as in (1)) are reported as well. Both specifications lead to very similar posterior distributions. The same observation holds when the means of the predictive distributions are compared with the estimates given in Table 2.

The first two models in Table 6 are empirically Bayesian in the sense that the overdispersion factor, ϕ , is estimated beforehand, using the SAS procedure (Proc Glimmix) for likelihood-based estimation in a generalized linear mixed model. For the semiparametric model and a rescaled response (namely ‘Claim Payments / 500’), $\hat{\phi} = 1.0635$, and for the model with chain-ladder structure, $\hat{\phi} = 1.16$. A fully Bayesian implementation of model (24) leads to the reserves reported in the last three columns of Table 6. Hereby a gamma distribution with mean $\hat{\phi}$ and variance 0.01 is used as the prior for the overdispersion factor (see Skollnik, 2006). Model complexity and fit of the Bayesian regression model are

summarized by the number of equivalent parameters p_D and the DIC (see Spiegelhalter *et al.*, 2002), displayed in Table 9.

	Emp. Bayes.			Emp. Bayes.			Full Bayes.		
	Ch. Ladd. o-P.			Smooth o-P.			Smooth o-P.		
	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
AY 2	0	578	2,890	0	532	2,127	0	0	2,663
AY 3	0	1,734	5,780	0	1,595	5,318	0	1,574	6,109
AY 4	1,156	4,046	9,826	1,064	4,254	9,572	710	3,937	10,460
AY 5	1,734	5,202	10,982	2,127	5,849	11,167	1,523	5,613	12,130
AY 6	3,468	8,092	13,872	3,722	7,976	13,294	2,876	7,713	14,760
AY 7	5,202	10,404	17,340	5,849	10,635	17,016	4,758	10,570	18,420
AY 8	8,670	15,028	22,542	8,508	14,889	22,334	7,776	14,720	23,710
AY 9	13,872	21,386	30,634	14,357	21,802	30,310	12,920	21,680	32,630
AY 10	45,084	60,690	78,608	45,731	60,088	78,167	42,730	60,410	81,300
Total	100,572	127,738	164,730	101,564	128,152	162,184	97,090	128,100	169,300

Table 6: *Predictive distributions for various reserves: overdispersed Poisson model with chain-ladder structure for the linear predictor versus semiparametric specification of predictor.*

As illustrated by the residual plot in Figure 4, the model with chain-ladder structure for the linear predictor does not seem to be able to remove all trends in the direction of calendar years. A similar observation holds for the semiparametric model. We therefore consider a refinement of the previous models and include calendar year effects:

$$\log(\mu_{ij}) = \alpha_1 I(i = 1) + \dots + \alpha_9 I(i = 9) + f(j) + g(i + j - 1). \quad (25)$$

$f(\cdot)$ is modelled by truncated line basis functions with 4 knots, namely $\kappa_{DY,1} = 2$, $\kappa_{DY,2} = 3$, $\kappa_{DY,3} = 5$ and $\kappa_{DY,4} = 7$. For $g(\cdot)$ truncated line basis functions are used as well, with 4 knots at positions $\kappa_{CY,1} = 3$, $\kappa_{CY,2} = 5$, $\kappa_{CY,3} = 7$ and $\kappa_{CY,4} = 9$.

To obtain the predictive distribution of the reserves, a Bayesian implementation of (25) is considered. For the over-dispersion factor ϕ again the estimate obtained with a likelihood-based implementation is used. Priors for the remaining parameters are

$$\begin{aligned} \alpha_i &\sim N(0, 10^5) \text{ with } i = 1, \dots, 10, \\ \beta, \gamma &\sim N(0, 10^5), \\ \sigma_b^2 &\sim \text{Inv-Gamma}(a, b), \\ \sigma_\gamma^2 &\sim \text{Inv-Gamma}(a, b). \end{aligned} \quad (26)$$

Table 7 contains the parameter estimates obtained with a likelihood-based as well as with a Bayesian analysis of the over-dispersed Poisson model with predictor structure (25)

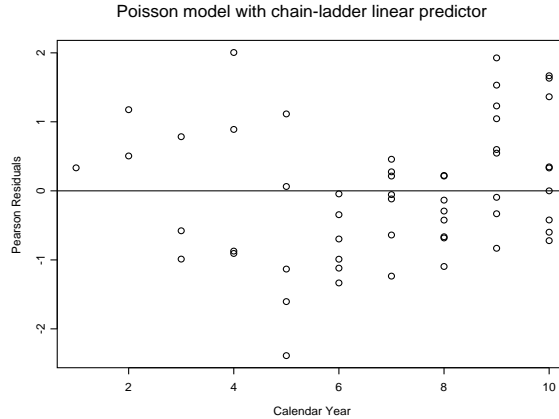


Figure 4: *Pearson-type residuals against calendar year: Poisson model with chain-ladder type structure for the mean, likelihood-based implementation.*

((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. p_D and DIC for this model are added in Table 9. Because the semiparametric model in (24) leads to a lower DIC and does not require extrapolation (as models with calendar year effects do), this model is preferred above the one in (25). However, a simplification of model (25), where σ_γ^2 is put equal to zero (as suggested for instance by the crude variance component test reported by SAS Proc Glimmix) and with γ significantly different from zero, leads to predictive distributions which are very similar to those reported in Table 6.

Parameter	Mean (St. Err.) Lik.	Mean Bayes.	St. Dev. Bayes.	2.5% Bayes.	50% Bayes.	97.5% Bayes.
α_1	5.554 (0.499)	5.556	0.536	4.458	5.574	6.562
α_3	4.816 (0.319)	4.816	0.341	4.121	4.826	5.461
α_7	2.143 (0.163)	2.138	0.175	1.795	2.137	2.488
β	-3.44 (0.138)	-3.445	0.15	-3.735	-3.446	-3.148
γ	1.47 (0.268)	1.474	0.291	0.881	1.483	2.025
σ_β^2	4.716 (3.49)	9.383	23.08	1.518	5.526	39.41
σ_γ^2	0.056 (0.061)	0.127	0.37	0.01	0.064	0.6

Table 7: *Parameter estimates obtained with likelihood-based analysis (second column) and Bayesian analysis. 700,000 simulations used, after a burn-in of 50,000 simulations.*

To investigate the sensitivity of the results on the prior specification for the variance

component, Table 8 shows the posterior distributions for the fixed regression parameters and the total reserve, obtained via smoothing with truncated line basis functions and with various prior specifications for the variance component, σ_β^2 , in our preferred model ⁵:

$$\begin{aligned}
\sigma_b^2 &\sim \text{Inv-Gamma}(a, b) \text{ with } a, b = 0.1 \text{ or } 0.001, \\
\sigma_b &\sim \text{folded Cauchy with } s = 12 \text{ or } 25, \\
\sigma_b &\sim \text{Uniform}(0, 50).
\end{aligned}
\tag{27}$$

Prior		Mean	St. Dev.	2.5%	50%	97.5%
Inv-Gamma(a, b) ($a, b = 0.1$)	α_1	2.301	0.162	1.994	2.299	2.626
	β	-1.803	0.13	-2.048	-1.806	-1.54
	Total	128,985	15,296	101,564	128,152	161,120
Inv-Gamma(a, b) ($a = b = 0.001$)	α_1	2.309	0.166	1.973	2.309	2.627
	β	-1.795	0.133	-2.067	-1.795	-1.542
	Total	129,112	15,322	101,564	128,152	161,652
folded Cauchy $s = 12$	α_1	2.338	0.164	2.016	2.337	2.659
	β	-1.77	0.13	-2.025	-1.772	-1.517
	Total	129,461	15,356	101,564	128,684	162,184
Uniform (0, 50)	α_1	2.332	0.162	2.014	2.334	2.641
	β	-1.775	0.129	-2.034	-1.774	-1.535
	Total	129,446	15,352	102,096	128,634	162,184

Table 8: *Investigation of sensitivity with respect to the prior distribution of the variance component in (25): posterior distributions for selection of parameters obtained with various choices of priors for σ_b^2 or σ_b . 700,000 simulations used, after a burn-in of 50,000 simulations.*

Ch. Ladd.		Smooth DY		Smooth DY+CY	
p_D	DIC	p_D	DIC	p_D	DIC
18.753	313.629	14.92	314.589	16.868	341.675

Table 9: *Model complexity and fit, as summarized by p_D and DIC: over-dispersed Poisson model with chain-ladder structure (columns 1-2), model (24) (columns 3-4) and model (25) (columns 5-6).*

⁵folded Cauchy: $\sigma \propto (\sigma^2 + s^2)^{-1}$

3.1.2 Building in a stochastic discounting process

To further illustrate the flexibility of a Bayesian smoothing model, we build a stochastic discounting process into model (24). 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 $e^{-(Z_1+\dots+Z_t)} := e^{-Z(t)}$. Here we use the classical model

$$Z(t) = \left(\mu - \frac{\delta^2}{2} \right) t + \delta B(t), \quad (28)$$

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

$$\begin{aligned} R &:= \sum_{i=2}^n \sum_{j=n+2-i}^n Y_{ij} e^{-Z(i+j-n-1)} \\ &= \sum_{i=2}^n \sum_{j=n+2-i}^n Y_{ij} \exp \{ -(\mu - \delta^2/2)(i+j-n-1) - \delta B(i+j-n-1) \}, \quad (29) \end{aligned}$$

where the future payments Y_{ij} are modelled using an over-dispersed Poisson model as in (24). 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 10 follow. Figure 5 illustrates the mixing and convergence of the generated Markov chains for some fixed effects parameters and the total discounted reserve.

3.2 Combining data on claim intensities and claim counts

Denote by Y_{ij} the aggregate payment for cell (i, j) , as shown in Table 3, and let N_{ij} be the corresponding number of claims, as displayed in Table 4. 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.

	Mean	St.Dev.	2.5%	50%	97.5%
	Bayes.	Bayes.	Bayes.	Bayes.	Bayes.
AY 2	547	622	0	472	2,134
AY 3	1,777	1,278	0	1,539	4,878
AY 4	3,820	1,970	865	3,537	8,465
AY 5	5,117	2,096	1,746	4,867	9,918
AY 6	6,782	2,372	2,884	6,530	12,120
AY 7	9,066	2,771	4,457	8,782	15,280
AY 8	12,350	3,366	6,733	12,030	19,860
AY 9	17,870	4,365	10,560	17,460	27,570
AY 10	52,680	9,649	35,960	51,930	73,700
Total	110,015	19,166	77,421	108,275	152,415

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

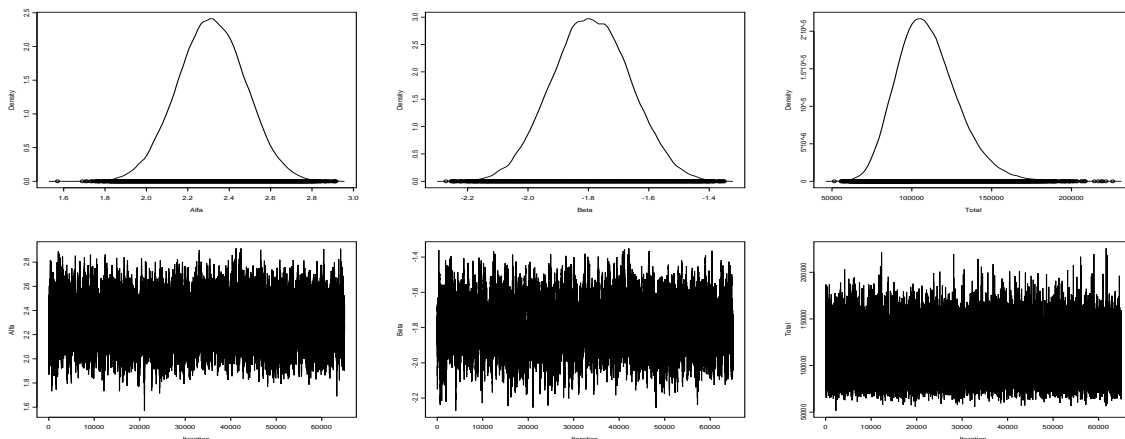


Figure 5: *Density and trace plots of the generated chains for α_1 , β and the total discounted reserve, over-dispersed Poisson model.*

Let $Z_{ij} := Y_{ij}/N_{ij}$ be the average payment for cell (i, j) and specify

$$\begin{aligned}
Z_{ij} &\sim \Gamma(\nu, \mu_{ij}^{Av}/\nu), \\
\text{where } \log(\mu_{ij}^{Av}) &= \alpha_1 * I(i = 1) + \dots + \alpha_{10} * I(i = 10) + f^{Av}(j) \\
\text{and } \frac{N_{ij}}{\phi} &\sim \text{Poisson}\left(\frac{\mu_{ij}^{Num}}{\phi}\right), \\
\text{where } \log(\mu_{ij}^{Num}) &= \alpha_1 * I(i = 1) + \dots + \alpha_{10} * I(i = 10) + f^{Num}(j). \quad (30)
\end{aligned}$$

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 1 illustrate that appropriate modelling of the trends over the development period is necessary.

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

In an initial stage of the analysis, we also experimented with trends in the direction of calendar years. However, the specification in (30) 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 $f^{Av}(\cdot)$ and $f^{Num}(\cdot)$.⁶ 4 knots in the direction of development years, with positions (2, 3, 5, 7) (for claim counts and average payments), are used, though similar results were obtained with other choices for the number and positions of the knots. The resulting fits for $f^{Num}(j)$, as obtained with the different types of basis functions, are illustrated in Figure 6.

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

$$\begin{aligned}
\beta^{Av}, \beta^{Num} &\sim \text{dunif}(-10, 10) \\
\alpha_i^{Av}, \alpha_i^{Num} \quad (i = 1, \dots, 10) &\sim \text{dunif}(-10, 10) \\
\nu &\sim \text{dunif}(0, 100) \\
\sigma_{b,Av}^2 &\sim \text{Inv-Gamma}(0.01, 0.01) \\
\sigma_{b,Num}^2 &\sim \text{Inv-Gamma}(0.01, 0.01). \tag{31}
\end{aligned}$$

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. $\sigma_{\beta,Av}^2$ and $\sigma_{\beta,Num}^2$ are the variance components used for the smooth functions $f^{Av}(\cdot)$ and $f^{Num}(\cdot)$, respectively.

95% credible intervals for some of the parameters used in (30) are given in Table 11. The reserves obtained with this model are summarized in Table 12 (claim counts) and Table 13 (total payments, obtained by multiplying claim numbers and average payments). In Table 12 the results obtained with a regular as well as an over-dispersed Poisson model are shown, both with additive predictor (30). Note that the former are close to the results from a deterministic chain-ladder, as shown in Table 4. Regarding the choice of the error distribution for Z_{ij} , Table 14 reports p_D and DIC for both a gamma and a lognormal model. Both specifications lead to very similar results.

⁶Truncated basis functions of degree p : $\alpha_1 * I(i = 1) + \dots + \alpha_{10} * I(i = 10) + \beta_1 * j + \dots + \beta_p * j^p + \sum_{i=1}^K b_k(j - \kappa_k)_+^p$ and radial basis functions: $\alpha_1 * I(i = 1) + \dots + \alpha_{10} * I(i = 10) + \beta_1 * j + \sum_{i=1}^k b_i * |j - \kappa_i|^3$ and $(b_1, \dots, b_k)^t \sim N(\mathbf{0}, \sigma_b^2(\mathbf{\Omega}^{-1/2})(\mathbf{\Omega}^{-1/2})^t)$ where $\mathbf{\Omega} = \left[|\kappa_k - \kappa_l|_{1 \leq k, l \leq K}^3 \right]$.

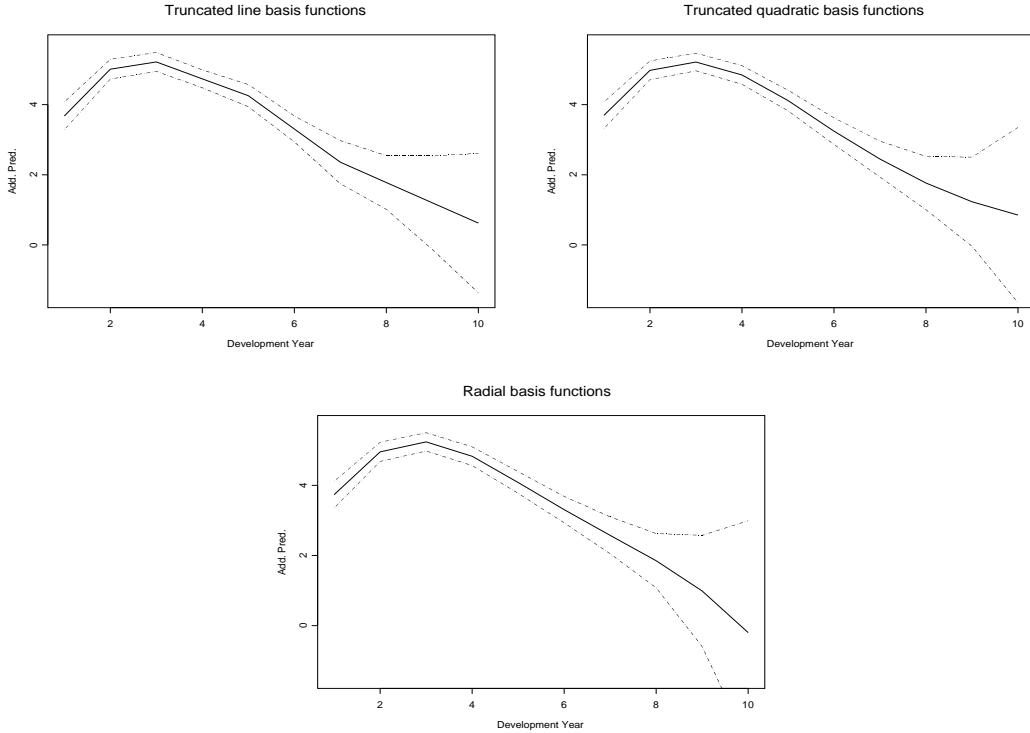


Figure 6: *Trend over development period for the claim numbers. Truncated line, truncated quadratic and radial basis functions with 4 knots, over-dispersed Poisson model. Fitted functions, together with pointwise 95% confidence intervals for $f^{Num}(j)$.*

3.3 Incremental claims with quarterly development

The merits of smoothing techniques for claims reserving become more obvious when data with extensive development periods are available. For the Belgian insurance data in Figure 2, both a normal and a Student-t regression model were considered with mean

$$\mu_{ij} = \alpha_1 * I(i = 1) + \dots + \alpha_{10} * I(i = 10) + \alpha_{11} * I(i = 11) + f(j). \quad (32)$$

We use Bayesian truncated line basis and radial basis functions for the trend in the direction of development quarters. Results included here are obtained with 11 knots, namely $\kappa_1 = 2$, $\kappa_2 = 3$, $\kappa_3 = 5$, $\kappa_4 = 9$, $\kappa_5 = 14$, $\kappa_6 = 16$, $\kappa_7 = 18$, $\kappa_8 = 22$, $\kappa_9 = 26$, $\kappa_{10} = 30$ and $\kappa_{11} = 32$, though similar results follow for a different number and positioning of the knots. A uniform discrete prior was specified for the degrees of freedom parameter (ν) in the Student-t model. Possible values for ν were 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 30 and 50, with equal probability. This resulted in the posterior distribution shown in Figure 9 (upper-left corner).

To estimate the Burr regression model in (6), observe that $Z_{ij} = \log(Y_{ij})$ follows a

	Mean	St.Dev.	2.5%	50%	97.5%
	Bayes.	Bayes.	Bayes.	Bayes.	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
$\sigma_{b,Num}^2$	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
$\sigma_{b,Av}^2$	0.258	0.621	0.028	0.139	1.176
ν	0.218	0.049	0.142	0.212	0.331

Table 11: 95% credible intervals for parameters in model (30): 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.

generalized logistic distribution with density given by

$$f_{Z_{ij}}(z_{ij}) = \frac{\lambda \tau \exp(\tau(z_{ij} - \mu_{ij}))}{[1 + \exp(\tau(z_{ij} - \mu_{ij}))]^{\lambda+1}}. \quad (33)$$

For trends in the location parameter, a specification equal to (32) is considered. Again a Bayesian implementation with truncated line basis functions as well as radial basis functions is used (with 7 knots, namely $\kappa_1 = 3$, $\kappa_2 = 6$, $\kappa_3 = 9$, $\kappa_4 = 15$, $\kappa_5 = 18$, $\kappa_6 = 20$, $\kappa_7 = 28$). Priors are the same as in Section 3.1.1, together with

$$\begin{aligned} \lambda &\sim \text{Gamma}(0.01, 0.01) \\ \tau &\sim \text{Unif}(0, 5) \text{ for } \tau, \lambda \text{ in (33)}. \end{aligned} \quad (34)$$

This results in estimated development trends ($\hat{\alpha}_1 + \hat{f}(j)$) as displayed in Figure 7 (upper: real insurance data and lower: Burr model; left: truncated line basis functions and right: radial basis functions). Posterior distributions for a selection of parameters from the Burr regression model are shown in Figure 8.

The fit of these regression models can be assessed using residual plots. Figure 9 shows qqplots for the Belgian insurance data (Student-t and normal model). The Student-t model is slightly better than the normal regression model, though deviations in the tail are still present. For the residuals, the mean of the posterior distribution of $(Y_{ij} - \mu_{ij})/\sigma$ is used. For the Burr regression model, the residuals $R_{ij} = \log(Y_{ij}) - \mu_{ij}$ have a distribution

$$F_{R_{ij}}(r_{ij}) = 1 - \frac{1}{[1 + \exp(\tau r_{ij})]^\lambda}. \quad (35)$$

	Mean	Mean	St.Dev.	5%	50%	97.5%
	Poisson	o-Poisson	Bayes.	Bayes.	Bayes.	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 12: *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.*

Figure 9 shows the residual quantile plot that corresponds with (35). For the residuals, the mean of the posterior distribution of $\log(Y_{ij}) - \mu_{ij}$ is used. The straight line pattern in this plot indicates a good fit.

3.4 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 5 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

$$\delta_{ij} = \begin{cases} 0 & \text{if } Y_{ij} = 0 \\ 1 & \text{if } Y_{ij} > 0, \end{cases} \quad \text{and} \quad Y_{ij}^* = \begin{cases} \text{irrelevant} & \text{if } Y_{ij} = 0 \\ Y_{ij} & \text{if } Y_{ij} > 0, \end{cases} \quad (36)$$

where Y_{ij} represents the aggregate amount paid out for cell (i, j) in the run-off triangle. For the data example in Section 1.1.4, for instance, the δ_{ij} variables ($i, j = 1, \dots, 13$) denote whether at least one claim for cell (i, j) has occurred. Given that a claim has occurred, Y_{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

	Mean	St.Dev.	2.5%	50%	97.5%
	Bayes.	Bayes.	Bayes.	Bayes.	Bayes.
AY 2	169,508	480,862	0	0	1,502,417
AY 3	378,139	724,771	0	0	2,346,970
AY 4	610,770	867,159	0	342,363	2,914,468
AY 5	1,041,495	1,081,350	0	749,758	3,851,428
AY 6	1,555,307	1,232,978	153,324	1,256,105	4,706,811
AY 7	2,461,469	1,553,093	558,292	2,118,104	6,337,255
AY 8	3,802,244	2,265,758	1,011,863	3,304,696	9,419,485
AY 9	5,539,692	3,460,206	1,433,985	4,733,528	14,410,460
AY 10	5,974,638	5,647,918	601,500	4,351,736	20,846,550
Total	21,533,260	8,682,724	9,883,250	19,883,880	42,867,290

Table 13: *Predictive distribution of the reserves obtained with model (30): 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.*

for the specification of the additive predictor in each part of the two-part model in (36). 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 specification 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) ,

$$\begin{aligned} \delta_{ij} &\sim \text{Binomial}(1, \pi_{ij}), \\ \text{where } \text{logit}(\pi_{ij}) &= f^{Bin}(j), \end{aligned} \quad (37)$$

and $f^{Bin}(\cdot)$ is a smooth function over the development years. Secondly, for the random variables Y_{ij}^* , a lognormal model (as in (38) below) 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,

$$\begin{aligned} \log(Y_{ij}^*) &= a^{Sev} + \alpha_2 * I(i = 2) + \dots + \alpha_9 * I(i = 9) + \alpha_{10} * I(i \geq 10) + f^{Sev}(j) + \epsilon_{ij}, \\ \text{where } \epsilon_{ij} &\sim N(0, \sigma_\epsilon^2). \end{aligned} \quad (38)$$

Denote by $\boldsymbol{\beta}^{Bin}$ the fixed effects and let \mathbf{b}^{Bin} be the vector with the random effects that are used to model $f^{Bin}(\cdot)$. The corresponding variance parameter is $\sigma_{b, Bin}^2$. Similarly, $\boldsymbol{\beta}^{Sev}$ are fixed effects parameters in the direction of development years. \mathbf{b}^{Sev} then denotes the random effects used for this smooth function, with corresponding variance parameter $\sigma_{b, Sev}^2$. Independence between \mathbf{b}^{Bin} and \mathbf{b}^{Sev} is assumed.

	Gamma (smooth)		Lognormal (smooth)	
	p_D	DIC	p_D	DIC
Count	14.488	276.156	14.533	276.217
Severity	14.107	1110.58	15.1	1112.42

Table 14: *Model complexity and fit, as summarized by p_D and DIC: over-dispersed Poisson for claim counts, gamma and lognormal model for average payments, smoothing over development years.*

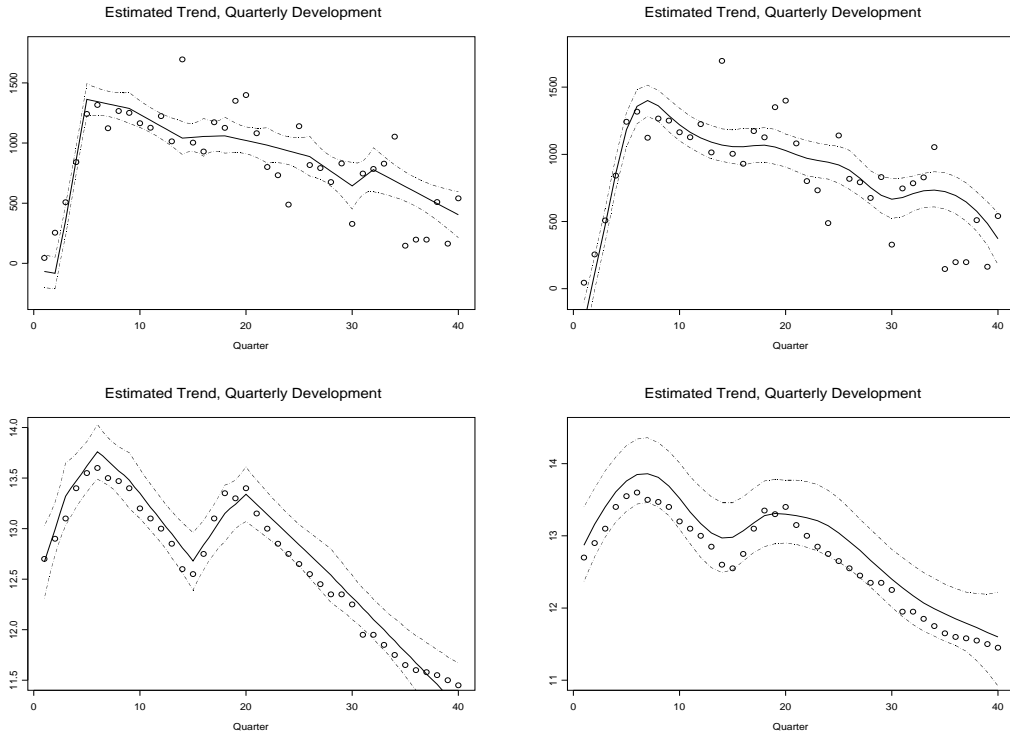


Figure 7: *Posterior mean of $\alpha_1 + f(j)$, together with 95% pointwise credible bands. 300,000 simulations used, to which a thinning factor of 5 is applied, after a burn-in of 50,000 simulations. Upper: Real Data Set, the real observed data for arrival year 1 are added. Lower: Burr regression model, the simulated values for μ_{1j} are added. Left: semiparametric regression with truncated line basis functions. Right: semiparametric regression with radial basis functions.*

The likelihood for this two-part model is

$$\begin{aligned}
& L(\boldsymbol{\beta}^{Bin}, \boldsymbol{\beta}^{Sev}, \sigma_\epsilon, \sigma_{b,Bin}, \sigma_{b,Sev} | y_{ij}, i = 1, \dots, n, j = 1, \dots, n - i + 1) \\
&= \prod_{i,j} \int f(y_{ij} | \boldsymbol{\beta}^{Bin}, \boldsymbol{\beta}^{Sev}, \mathbf{b}^{Bin}, \mathbf{b}^{Sev}, \sigma_\epsilon) \\
&\quad \times f(\mathbf{b}^{Bin}, \mathbf{b}^{Sev} | \sigma_{b,Bin}, \sigma_{b,Sev}) d\mathbf{b}^{bin} d\mathbf{b}^{Sev} \\
&= \int \prod_Z (1 - \pi_{ij}) \prod_{NZ} \pi_{ij} f(y_{ij}^* | \boldsymbol{\beta}^{Sev}, \mathbf{b}^{Sev}, \sigma_\epsilon) \\
&\quad \times f(\mathbf{b}^{Bin}, \mathbf{b}^{Sev} | \sigma_{b,Bin}, \sigma_{b,Sev}) d\mathbf{b}^{bin} d\mathbf{b}^{Sev} \\
&= \int \exp\left(\sum l_{\delta_{ij}}\right) \exp\left(\sum l_{Y_{ij}^*}\right) f(\mathbf{b}^{Bin}, \mathbf{b}^{Sev} | \sigma_{b,Bin}, \sigma_{b,Sev}) d\mathbf{b}^{bin} d\mathbf{b}^{Sev}
\end{aligned}$$

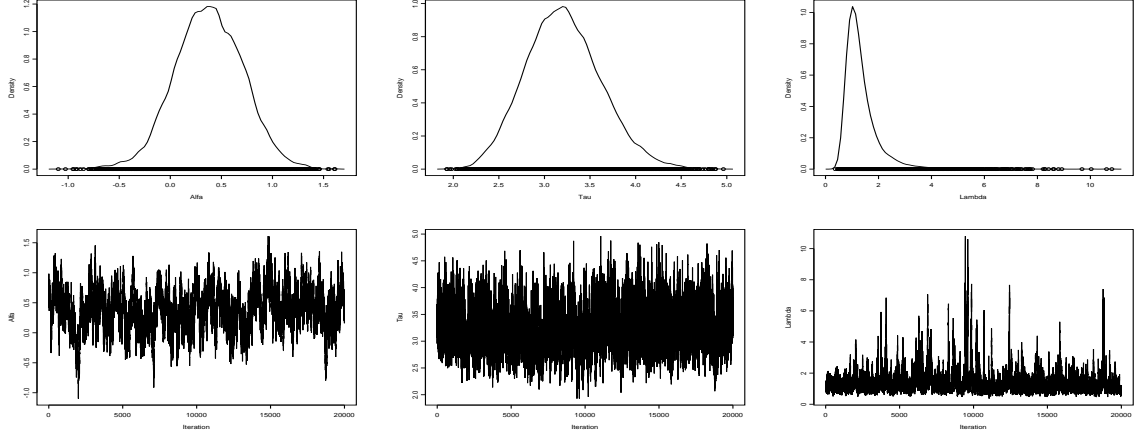


Figure 8: *Posterior densities and trace plots for parameters used in the Burr regression (from left to right: α_1 , β and λ). 100,000 simulations, after a burn-in of 20,000 iterations.*

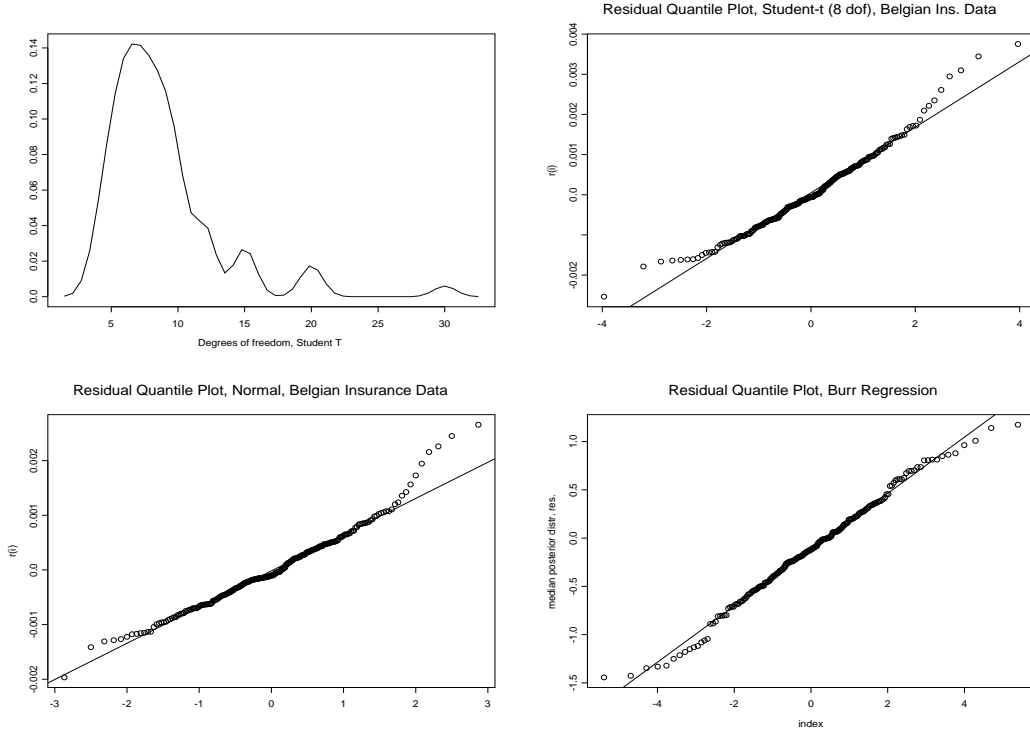


Figure 9: *(Real Data Set) Posterior distribution of degrees of freedom parameter in Student-t regression model. Residual quantile plot, standard Student-t with 8 dof and standard normal distribution. (Burr data) Residual quantile plot, Burr regression.*

$$= \int \exp \left(\sum_{NZ,Z} l_{\delta_{ij}} \right) f(\mathbf{b}^{Bin} | \sigma_{b,Bin}) d\mathbf{b}^{Bin} \int \exp \left(\sum_{NZ} l_{Y_{ij}^*} \right) f(\mathbf{b}^{Sev} | \sigma_{b,Sev}) d\mathbf{b}^{Sev}. \quad (39)$$

Here, \sum_{NZ} (NZ: Non-Zero) denotes summation over all $y_{ij} > 0$ and \sum_Z (Z: Zero) sum-

mation over all $y_{ij} = 0$. $l_{Y_{ij}^*}$ is the part of the log-likelihood related to a strict positive claim, conditional on the random effects \mathbf{b}^{Sev} . $l_{\delta_{ij}} = \delta_{ij} \text{logit}(\pi_{ij}) + \log(1 - \pi_{ij})$, conditional on the random effects \mathbf{b}^{Bin} . The last equation in (39) 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 $Y_{ij} = \delta_{ij} Y_{ij}^* + (1 - \delta_{ij}) * 0$. Table 15 contains 95% credible intervals for the parameters used in the models specified by (37) and (38). In this table, a_{Bin} is the intercept used to model $f^{Bin}(\cdot)$. Figure 10 shows the fitted additive predictors, obtained with truncated line basis functions. The predictive distributions of the various reserves are summarized in Table 16.

	Mean	St.Dev.	2.5%	50%	97.5%
	Bayes.	Bayes.	Bayes.	Bayes.	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
$\sigma_{b, Bin}^2$	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
$\sigma_{b, Sev}^2$	1.811	3.647	0.1844	1.005	8.173
σ_ϵ^2	0.5637	0.1242	0.3716	0.5469	0.8546

Table 15: 95% credible intervals for parameters used in models (37) and (38): 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.

3.5 A semiparametric model for longitudinal credibility data

Credibility ratemaking 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 ratemaking has been discussed in Frees *et al.* (1999, 2001) and Antonio & Beirlant (2005).

To illustrate the use of semiparametric regression models for credibility, the data introduced in Section 1.1.5, 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

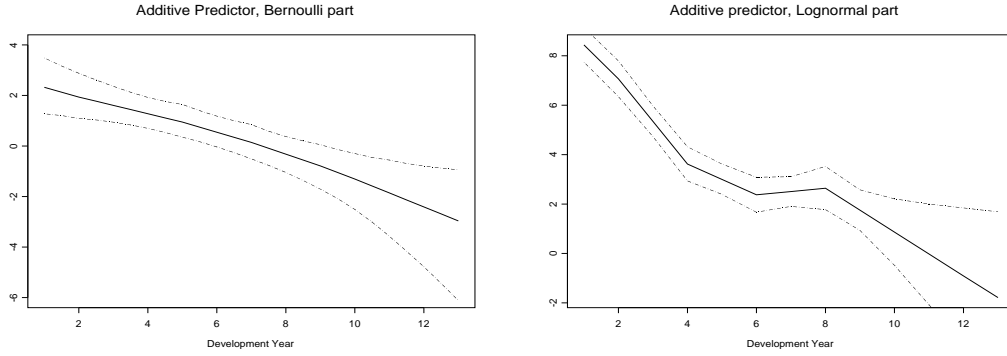


Figure 10: *Fitted additive predictors for models (37) and (38): results from a Bayesian analysis with truncated line basis functions. Posterior mean of $f^{Bin}(j)$ and $f^{Sev}(j)$, together with 95% pointwise credible bands. 300,000 simulations used, to which a thinning factor of 5 is applied, after a burn-in of 50,000 simulations.*

t-copula to achieve this, typical concepts of mixed models, namely the inclusion of random effects or the use of a special structure for the covariance matrix of the residual terms (in a linear mixed model), are considered here. The effects of the available explanatory variables, ‘PCI’ and ‘PPSM’, are modelled semiparametrically.

Firstly, the marginal effects of both explanatory variables on the response variable, ‘AC’, are illustrated. The lognormal distribution is used for the distribution of average claims (‘AC’). Frees & Wang (2005, page 36) already indicated that this is a reasonable choice. Ignoring the longitudinal structure of the data, we fit (with n the total number of observations)

$$\log(AC)_i = f(PCI_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (40)$$

$$\log(AC)_i = g(PPSM_i) + \epsilon_i \quad i = 1, \dots, n. \quad (41)$$

Both $f(\cdot)$ and $g(\cdot)$ 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 17.

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

	Mean	St.Dev.	2.5%	50%	97.5%
	Bayes.	Bayes.	Bayes.	Bayes.	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 16: *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.*

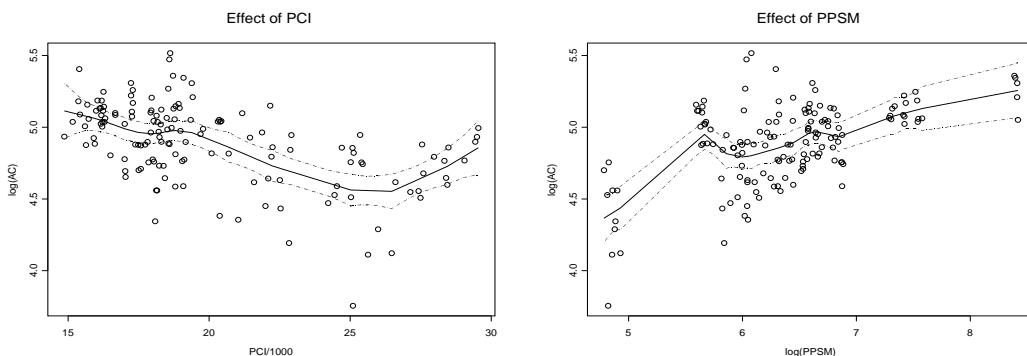


Figure 11: *Marginal effect of PCI (left) and PPSM (right) on $\log(AC)$, together with 95% pointwise confidence intervals for $f(PCI_i)$ and $g(PPSM_i)$; truncated line basis functions with 15 knots. Results obtained with Proc Mixed in SAS.*

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

$$\begin{aligned}
 \log(AC_{it}) &= f(PCI_{it}) + g(PPSM_{it}) + b_i + \epsilon_{it}, \quad i = 1, \dots, N \text{ and } j = 1, \dots, n_i, \\
 b_i &\sim N(0, \sigma_b^2) \\
 \epsilon_i &\sim N(\mathbf{0}, \Sigma_i), \text{ where } \epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})'
 \end{aligned} \tag{42}$$

are considered. Modeling nonlinear effects of ‘PCI’ and ‘PPSM’, the combination of

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

random intercepts and a non-diagonal Σ_i did not lead to convergence of Proc Mixed in SAS. Models with nonlinear effects of ‘PCI’ and ‘PPSM’, together with a non-diagonal Σ_i , did not lead to convergence of the procedure either. The results for different – convergent – model specifications are summarized in Table 18.

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
$\sigma_{b,PCI}^2$	0	/	/	/
$\sigma_{b,PPSM}^2$	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 18: *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 19. 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 11 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 19: *Comparisons of Sum of Squared Prediction Error.*

As mentioned in the introduction, this example is included just to illustrate that within the framework of smoothing with mixed models the statistical approaches for reserving and credibility can be unified. When dealing with extensive data in a claims reserving exercise (think of quarterly development of individual claims), this feature is appealing.

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. Important characteristics and advantages of our approach are summarized below:

- Trends in run-off triangles for claims reserving are modelled in a semiparametric way. This is especially useful in more extensive ‘triangles’ where for instance quarterly development is reported.
- Cross-sectional as well as longitudinal data are analyzed in the same framework. This helps further unifying the actuarial ‘reserving’ and ‘ratemaking’ methodology. Moreover, in a reserving exercise, the same techniques will allow to consider individual development instead of data aggregated in cells.
- Both a likelihood-based and a Bayesian implementation of the models are illustrated.
- More complicated data structures are considered. It is illustrated how to incorporate a stochastic discounting factor, combine data on claim counts and severities and model semicontinuous data.
- Apart from the class of generalized linear models, an example with a heavy-tailed regression model is also included.

References

- [1] Antonio, Katrien, and Jan Beirlant. 2005. Applications of Generalized Linear Mixed Models in Actuarial Statistics. *Insurance: Mathematics and Economics*, in press.
- [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. Log-normal mixed models for reported claims reserving. *North American Actuarial Journal*, 10(1): 30-48.
- [4] Beirlant, Jan, Yuri Goegebeur, Robert Verlaak and Petra Vynckier. 1998. Burr regression and portfolio segmentation. *Insurance: Mathematics and Economics*, 23, 231-250.
- [5] Beirlant, Jan, Yuri Goegebeur, Johan Segers and Jef Teugels. 2004. *Statistics of Extremes*. Wiley.
- [6] 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.
- [7] Crainiceanu, C., D. Ruppert and M.P. Wand. 2004. Bayesian analysis for penalized spline regression using WinBUGS. *Working paper*.
- [8] de Alba, E. 2002. Bayesian estimation of outstanding claims reserves. *North American Actuarial Journal* 6(4): 1-20.
- [9] 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.
- [10] Doray, L. IBNR reserve under a loglinear location-scale regression model. *Casualty Actuarial Science Forum 1994*, 2, 607-652.
- [11] England, P.D. and R.J. Verrall. 2002. Stochastic claims reserving in general insurance. *British Actuarial Journal* 8: 443-544.
- [12] England, P.D. and R.J. Verrall. 2001. A flexible framework for stochastic claims reserving. *Proceedings of the casualty actuarial society forum*.
- [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] Ntzoufras, N. and P. Dellaportas. 2002. *Bayesian modelling of outstanding liabilities incorporating claim count uncertainty*. *North American Actuarial Journal* 6, 113-128.
- [22] 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.
- [23] Ruppert, D., M.P. Wand and R.J. Carroll. 2003. *Semiparametric regression*. Cambridge: Cambridge University Press.
- [24] Skollnik, D. 2006. Discussion of ‘A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving.’ *North American Actuarial Journal* 9(3): 146-149.
- [25] Spiegelhalter, D.J., N.G. Best, B.P. Carlin and A. Van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64: 583-639.
- [26] Verrall, R.J. 1996. Claims reserving and generalized additive models. *Insurance: Mathematics and Economics* 19: 31-43.
- [27] 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.