Insurance: Mathematics and Economics 49 (2011) 27–37

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

A generalized linear model with smoothing effects for claims reserving

Susanna Björkwall[a,*], Ola Hössjer[a], Esbjörn Ohlsson[b], Richard Verrall[c]

[a] Department of Mathematics, Division of Mathematical Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

[b] Länsförsäkringar Alliance, SE-106 50 Stockholm, Sweden

[c] Cass Business School, City University, 106 Bunhill Row, London EC1Y 8TZ, United Kingdom

ARTICLE INFO

Article history:
Received September 2010
Received in revised form January 2011
Accepted 28 January 2011

Keywords:
Bootstrap
Generalized linear model
Model selection
Smoothing
Stochastic claims reserving

ABSTRACT

In this paper, we continue the development of the ideas introduced in England and Verrall (2001) by suggesting the use of a reparameterized version of the generalized linear model (GLM) which is frequently used in stochastic claims reserving. This model enables us to smooth the origin, development and calendar year parameters in a similar way as is often done in practice, but still keep the GLM structure. Specifically, we use this model structure in order to obtain reserve estimates and to systemize the model selection procedure that arises in the smoothing process. Moreover, we provide a bootstrap procedure to achieve a full predictive distribution.

1. Introduction

Many stochastic claims reserving methods have now been established, see, for example, Wütrich and Merz (2008) and England and Verrall (2002). However, many actuaries use intuitive approaches in conjunction with a standard reserving method, and it is necessary for stochastic methods to also enable these to be used. One such intuitive method is the subject of this paper, which is to allow some smoothing to be applied to the shape of the development pattern.

In Björkwall et al. (2009) an example of a deterministic development factor scheme, which is frequently used in practice, was provided. This approach could be varied in many different ways by the actuary depending on the particular data set under analysis. This includes smoothing the development factors, perhaps after excluding the oldest observations and outliers. In this way the impact of irregular observations in the data set is reduced and more reliable reserve estimates are obtained. A common approach is to use exponential smoothing by log-linear regression of the development factors, but other curves could be applied too, see Sherman (1984). These could also be carried over to a bootstrap procedure in order to obtain the corresponding prediction error as well as the predictive distribution, whose size and width, respectively, can be changed according to the amount of smoothing and extrapolation.

Despite the intuitiveness and transparency of this approach it is certainly accompanied by some statistical drawbacks. For instance, the statistical quality of the reserve estimates is not optimal since they are not maximum likelihood estimators. Moreover, bootstrapping requires some stochastic model assumptions anyway. When this model is defined at the resampling stage, rather than at the original data generation stage the resulting reserving exercise leads to somewhat ad hoc decisions and more subjectiveness compared to a more systematic methodology, where a stochastic model is defined from the start.

England and Verrall (2001) presented a Generalized Additive Model (GAM) framework of stochastic claims reserving, which has the flexibility to include several well-known reserving models as special cases as well as to incorporate smoothing and extrapolation in the model-fitting procedure. Using this framework implies that the actuary simply would have to choose one parameter corresponding to the amount of smoothing, the error distribution and how far to extrapolate, then the fitted model automatically provides statistics of interest, e.g. reserve estimates, measures of precision and tests for goodness-of-fit. Such an approach is appealing, partly due to its statistical qualities and partly in order to obtain a tool for selection and comparison of models, which then could be systemized.

Recently Antonio and Beirlant (2008) applied a similar approach using a semi-parametric regression model which is based on a generalized linear mixed model (GLMM) approach. However, both GAMs and GLMMs might be considered as too sophisticated in order to become popular in reserving practice.

* Corresponding author. Tel.: +46 70 434 7680.
E-mail addresses: susanna@math.su.se, sbj@student.su.se (S. Björkwall), ola@math.su.se (O. Hössjer), esbjorn.ohlsson@lansforsakringar.se (E. Ohlsson), R.J.Verrall@city.ac.uk (R. Verrall).

0167-6687/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2011.01.012

Table 2.1
The triangle ∇ of observed incremental payments.

Origin year	Development year					
	1	2	3	$\cdots$	$t-1$	t
1	C_{11}	C_{12}	C_{13}	$\cdots$	$C_{1,t-1}$	$C_{1,t}$
2	C_{21}	C_{22}	C_{23}	$\cdots$	$C_{2,t-1}$	
3	C_{31}	C_{32}	C_{33}	$\cdots$		
$\vdots$	$\vdots$	$\vdots$	$\vdots$			
$t-1$	$C_{t-1,1}$	$C_{t-1,2}$				
t	$C_{t,1}$					

Table 2.2
The triangle Δ of unobserved future payments.

Origin year	Development year					
	1	2	3	$\cdots$	$t-1$	t
1						
2						$C_{2,t}$
3					$C_{3,t-1}$	$C_{3,t}$
$\vdots$					$\vdots$	$\vdots$
$t-1$			$C_{t-1,3}$	$\cdots$	$C_{t-1,t-1}$	$C_{t-1,t}$
t		$C_{t,2}$	$C_{t,3}$	$\cdots$	$C_{t,t-1}$	$C_{t,t}$

In this paper, we instead suggest the use of a reparameterized version of the popular Generalized Linear Model (GLM) introduced in a claims reserving context by Renshaw and Verrall (1998). This model enables us to smooth origin, development and calendar year parameters in a similar way as is often done in practice, but still keep a GLM structure which we can use to obtain reserve estimates and to systemize the model selection procedure that arises in the smoothing process. While England and Verrall (2001) used the GAM in order to analytically compute prediction errors we instead implement a bootstrap procedure to achieve a full predictive distribution for the suggested GLM in accordance with the method developed in Björkwall et al. (2009).

The paper is set out as follows. The notation and a short summary of existing smoothing approaches, which to a large extent is based on the work of England and Verrall (2001, 2002), are given in Section 2. The suggested model is introduced in Section 3, which also contains three examples of how it can be used; one of them is the main topic of this paper. In Section 4 we discuss model selection and some criteria are provided. The estimation of the parameters and the reserving algorithm are described in Section 5, while Section 6 contains the bootstrap procedure. The theory is numerically studied in Section 7 and, finally, Section 8 contains a discussion.

2. Smoothing models in claims reserving

2.1. Notation

Let $\{C_{ij}; i,j \in \nabla\}$ denote the incremental observations of paid claims, which are assumed to be available in a development triangle $\nabla = \{(i,j); i = 1,\ldots,t; j = 1,\ldots,t-i+1\}$. The suffixes i and j refer to the origin year and the development year, respectively, see Table 2.1. In addition, the suffix $k = i + j$ is used for the calendar years, i.e. the diagonals of ∇. Let $n = t(t+1)/2$ denote the number of observations.

The purpose of a claims reserving exercise is to predict the sum of the delayed claim amounts in the lower, unobserved future triangle $\{C_{ij}; i,j \in \Delta\}$, where $\Delta = \{(i,j); i = 2,\ldots,t; j = t-i+2,\ldots,t\}$, see Table 2.2. We write $R = \sum_{\Delta} C_{ij}$ for this sum, which is the outstanding claims for which the insurance company must hold a reserve. The outstanding claims per origin year are specified by summing per origin year $R_i = \sum_{j\in\Delta_i} C_{ij}$, where Δ_i denotes the row corresponding to origin year i in Δ.

Estimators of the outstanding claims per origin year and the grand total are obtained by $\hat{R}_i = \sum_{j\in\Delta_i} \hat{C}_{ij}$ and $\hat{R} = \sum_{\Delta} \hat{C}_{ij}$, respectively, where $\hat{C}_{ij}$ is a prediction of C_{ij}. With an underlying stochastic reserving model, $\hat{C}_{ij}$ is a function of the estimated parameters of that model, typically chosen to make it an (asymptotically) unbiased predictor of C_{ij}.

2.2. A deterministic development factor method

This section contains an example of how the chain-ladder development factors might be smoothed and extrapolated into a tail according to a deterministic algorithm.

The chain-ladder and other development factor methods operate on cumulative claim amounts

$$D_{ij} = \sum_{\ell=1}^{j} C_{i\ell}. \tag{2.1}$$

Let $\mu_{ij} = E(D_{ij})$. Development factors

$$f_j = \frac{\sum_{i=1}^{t-j} \mu_{i,j+1}}{\sum_{i=1}^{t-j} \mu_{ij}}, \tag{2.2}$$

where $j = 1,\ldots,t-1$, are estimated for a fully non-parametric model without any smoothing of parameters by

$$\hat{f}_j = \frac{\sum_{i=1}^{t-j} D_{i,j+1}}{\sum_{i=1}^{t-j} D_{ij}}. \tag{2.3}$$

By examining a graph of the sequence of $\hat{f}_j$'s the actuary might decide to smooth them, for instance, for $j \geq 4$. Exponential smoothing could be used for that purpose, i.e. the $\hat{f}_j$'s are replaced by estimators obtained from a linear regression of $\ln(\hat{f}_j - 1)$ on j. By extrapolation in the linear regression this also yields development factors for a tail $j = t, t+1 \ldots, t+u$. The original $\hat{f}_j$'s are kept for $j < 4$ and the smoothed ones used for all $j \geq 4$. Let $\hat{f}_j^s$ denote the new sequence of development factors. Estimates $\hat{\mu}_{ij}$ for Δ can now be computed as in the standard chain-ladder method yielding

$$\hat{\mu}_{ij} = D_{i,t-i}\hat{f}^s_{t-i}\hat{f}^s_{t-i+1}\ldots\hat{f}^s_{j-1} \tag{2.4}$$

and

$$\hat{C}_{i,j} = \hat{\mu}_{i,j} - \hat{\mu}_{i,j-1}. \tag{2.5}$$

Note that the truncation point $j = 4$ of the unsmoothed development factors has to be decided by eye. Moreover, this approach could be varied, the actuary might choose, for example, to disregard some of the latest development factors for the regression procedure and then more decisions have to be made. Hence, this approach is quite ad hoc, and a more stringent methodology requires a stochastic model for the claims; see Verrall and England (2005) and Verrall (2007) for further discussion.

2.3. Lognormal models

Early smoothing models applied to claims reserving were parametric and, for simplicity, normal distributions were assumed. The usual assumptions were that C_{ij} are independent with

$$\ln(C_{ij}) = \eta_{ij} + \epsilon_{ij}, \tag{2.6}$$

where $\eta_{ij} = E(\ln(C_{ij}))$ and $\epsilon_{ij} \sim \mathrm{N}(0,\sigma^2)$. Hence, $C_{ij} \sim \mathrm{LN}(\eta_{ij},\sigma^2)$, where N and LN denote normal and lognormal distributions, respectively. Moreover, two models,

$$\eta_{ij} = c + \alpha_i + \beta_j \tag{2.7}$$

and

$$\eta_{ij} = c + \alpha_i + \beta_i \ln j + \gamma_i j, \tag{2.8}$$

were suggested.

Model (2.7) was introduced by Kremer (1982). Model (2.8), which is referred to as the Hoerl curve, can be ascribed to Zehnwirth (1985), see e.g. England and Verrall (2001). The original document does no longer exist according to Insureware even though it is frequently referred to in the literature. However, the Hoerl curve is also mentioned in the conference paper Zehnwirth (1989). This model was popular since it often provides a reasonable approximation to the shape of the payment pattern; it starts with a rapidly increasing peak and then decays exponentially. Moreover, it can be used for extrapolation of a tail of payments beyond t.

De Jong and Zehnwirth (1983) used the Kalman filter in order to smooth the estimates of the parameters β_i and γ_i in (2.8) according to a framework for a family of models. Verrall (1989) also used the Kalman filter in order to smooth the estimates of α_i and β_j in (2.7).

Barnett and Zehnwirth (2000) introduced a model which is referred to as the probabilistic trend family (PTF), and this model could be expressed using

$$\eta_{ij} = \alpha_i + \sum_{l=1}^{j-1} \beta_l + \sum_{k=2}^{i+j-1} \gamma_k \tag{2.9}$$

in (2.6). Here the β_l's and γ_k's account for linear trends between the development years and calendar years, respectively.

2.4. Generalized linear models

Wright (1990) introduced a parametric alternative which is based on a risk theoretic model including the assumption of Poisson distributed claim numbers and gamma distributed claim amounts. The model can be expressed as a GLM

$$E(C_{ij}) = m_{ij} \quad \text{and} \quad \text{Var}(C_{ij}) = \phi_{ij}\, m_{ij}^p$$

$$\ln(m_{ij}) = \eta_{ij}, \tag{2.10}$$

where

$$\eta_{ij} = u_{ij} + c + \alpha_i + \beta_i \ln j + \gamma_i j + \delta k, \tag{2.11}$$

and $p = 1$, see e.g. England and Verrall (2001) for the derivation. The term u_{ij} is a known *offset*, a function of an exposure and a known adjustment term, see Wright (1990), while δk allows for claims inflation. It is easy to see that (2.11) is similar to the Hoerl curve in (2.8), but the relation between the responses and the predictor differs. Moreover, the error distribution no longer has to be normal. Wright (1990) used the Kalman filter to produce smoothed estimates of the parameters.

Renshaw and Verrall (1998) used (2.7) in (2.10) and related the model to the chain-ladder method for $p = 1$. Note that the scale parameter is usually assumed to be constant in this context, i.e. $\phi_{ij} = \phi$, and, hence, we will stick to this assumption from now on.

Eq. (2.7) can be extended to include a calender year parameter according to

$$\eta_{ij} = c + \alpha_i + \beta_j + \gamma_k, \quad k = 2, \ldots, 2t. \tag{2.12}$$

However, the number of parameters is then usually too large compared to the small data set of aggregated individual paid claims in Table 2.1. In any case, a side constraint, e.g.

$$\alpha_1 = \beta_1 = \gamma_2 = 0 \tag{2.13}$$

is needed to estimate the $\upsilon = 3t - 2$ remaining model parameters c, α_i, β_j and γ_k, typically under the assumption $p = 1$ or $p = 2$, corresponding to an over-dispersed Poisson (ODP) distribution or a gamma distribution, respectively. Note that it is only possible to estimate γ_k for $k = 3, \ldots, t$, while a further assumption is needed regarding the future $k = t + 1, \ldots, 2t$.

2.5. Generalized additive models

GAMs, which include non-parametric smoothers, are an alternative in order to obtain more flexibility than parametric smoothing models can provide. Using this approach, Verrall (1996) extended the model in (2.10) and (2.7) to incorporate smoothing of the origin year parameters α_i. England and Verrall (2001) extended that idea further by creating a general framework which can express several previous reserving models as special cases. The framework was then used, among other things, to allow for smoothing over the development year parameter β_j, in (2.10) and (2.7), too.

2.6. Generalized linear mixed models

Antonio and Beirlant (2008) presented a semi-parametric regression model which is based on a GLMM approach. Using a Bayesian implementation they extended the work of England and Verrall (2001) to include simulation of predictive distributions. In addition, the suggested model could be used for more complicated data sets involving e.g. quarterly development or longitudinal data.

3. GLM with log-linear smoothing

3.1. A general parametrization

In this section we introduce a matrix representation of Eq. (2.12) according to

$$\boldsymbol{\eta} = \mathbf{X}_{\text{full}} \boldsymbol{\theta}_{\text{full}} \tag{3.1}$$

using $\boldsymbol{\theta}_{\text{full}} = \begin{pmatrix} c & \boldsymbol{\alpha} & \boldsymbol{\beta} & \boldsymbol{\gamma} \end{pmatrix}^T$ and $\boldsymbol{\eta} = (\eta_{11} \cdots \eta_{1t}\, \eta_{21} \cdots \eta_{2,t-1} \cdots \eta_{t1})^T$, where $\boldsymbol{\alpha} = \begin{pmatrix} \alpha_2 & \cdots & \alpha_t \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_2 & \cdots & \beta_t \end{pmatrix}$, $\boldsymbol{\gamma} = \begin{pmatrix} \gamma_3 & \cdots & \gamma_{t+1} \end{pmatrix}$. Recall that the number of observations in ∇C is $n = t\,(t+1)/2$, which is also the length of $\boldsymbol{\eta}$. Moreover, $\mathbf{X}_{\text{full}}$ is the design matrix of the system. The index '*full*' refers to the GLM in (2.10) and (2.12) which from now on will be considered as *full* in contrast to the subsequent smoothed version.

In order to smooth the $\upsilon = 3t - 2$ original parameters α_i, β_j and γ_k and, hence, reparametrize the system (3.1) we introduce a new set of parameters $\boldsymbol{a} = \begin{pmatrix} a_1 & \cdots & a_q \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} b_1 & \cdots & b_r \end{pmatrix}$ and $\boldsymbol{g} = \begin{pmatrix} g_1 & \cdots & g_s \end{pmatrix}$, where $0 \le q, r, s \le t - 1$. Let

$$\begin{aligned} \boldsymbol{a}\,\boldsymbol{A} &= \boldsymbol{\alpha} \\ \boldsymbol{b}\,\boldsymbol{B} &= \boldsymbol{\beta} \\ \boldsymbol{g}\,\boldsymbol{\Gamma} &= \boldsymbol{\gamma}, \end{aligned} \tag{3.2}$$

which corresponds to

$$\begin{pmatrix} a_1 & \cdots & a_q \end{pmatrix} \begin{pmatrix} A_{12} & \cdots & A_{1t} \\ \vdots & \vdots & \vdots \\ A_{q2} & \cdots & A_{qt} \end{pmatrix} = \begin{pmatrix} \alpha_2 & \cdots & \alpha_t \end{pmatrix} \tag{3.3}$$

$$\begin{pmatrix} b_1 & \cdots & b_r \end{pmatrix} \begin{pmatrix} B_{12} & \cdots & B_{1t} \\ \vdots & \vdots & \vdots \\ B_{r2} & \cdots & B_{rt} \end{pmatrix} = \begin{pmatrix} \beta_2 & \cdots & \beta_t \end{pmatrix} \tag{3.4}$$

$$\begin{pmatrix} g_1 & \cdots & g_s \end{pmatrix} \begin{pmatrix} \Gamma_{13} & \cdots & \Gamma_{1,t+1} \\ \vdots & \vdots & \vdots \\ \Gamma_{s3} & \cdots & \Gamma_{s,t+1} \end{pmatrix} = \begin{pmatrix} \gamma_3 & \cdots & \gamma_{t+1} \end{pmatrix}. \tag{3.5}$$

Moreover, let $\boldsymbol{\theta} = \begin{pmatrix} c & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{g} \end{pmatrix}^T$, containing $w = 1 + q + r + s$ parameters. We can now express $\boldsymbol{\theta}_{\text{full}}$ as

$$\boldsymbol{\theta}_{\text{full}} = \boldsymbol{D}\,\boldsymbol{\theta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \boldsymbol{A} & 0 & 0 \\ 0 & 0 & \boldsymbol{B} & 0 \\ 0 & 0 & 0 & \boldsymbol{\Gamma} \end{pmatrix}^T \boldsymbol{\theta} \tag{3.6}$$

using the blocks $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{\Gamma}$ in (3.2)–(3.5). The matrix $\boldsymbol{D}$ is of dimension $v \times w$.

Finally (3.1) can be rewritten using the new parameters

$$\boldsymbol{\eta} = \boldsymbol{X}_{\text{full}}\, \boldsymbol{\theta}_{\text{full}} = \boldsymbol{X}_{\text{full}}\, \boldsymbol{D}\, \boldsymbol{\theta} = \boldsymbol{X}\, \boldsymbol{\theta}, \tag{3.7}$$

where $\boldsymbol{X}$ is the new design matrix.

Example 1 (*The Full GLM*). If $q = r = s = t - 1$ and $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{\Gamma} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix, we get the full GLM. However, if we choose other values of q, r, s and $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Gamma}$, respectively, we will still keep a GLM structure.

Example 2 (*A Hoerl Curve*). A special case of the Hoerl curve in (2.8), where $\beta_i = \beta$ and $\gamma_i = \gamma$, can be expressed according to (3.7) using $\boldsymbol{\beta} = \begin{pmatrix} \beta & \gamma \end{pmatrix}$, $\boldsymbol{\Gamma} = \mathbf{0}$,

$$\boldsymbol{\theta} = \begin{pmatrix} c & \alpha_2 & \cdots & \alpha_t & \beta & \gamma \end{pmatrix}^T$$

and

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & \ln(1) & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \ln(t) & t \end{pmatrix}.$$

Hence, $q = t - 1$, $r = 2$ and $s = 0$. Note that there is an overlap in the naming of parameters in (2.8) and (3.1).

3.2. The log-linear smoothing model

In this section, we consider the GLM which is studied in the remainder of this paper. For smoothing purposes, curves such as the following are of interest:

$$\begin{aligned} &\alpha_i = a_{i-1}; \quad 2 \le i \le q \\ &\alpha_i = a_{q-1} + a_q\,(i - q); \quad q + 1 \le i \le t \\ &\beta_j = b_{j-1}; \quad 2 \le j \le r \\ &\beta_j = b_{r-1} + b_r\,(j - r); \quad r + 1 \le j \le t \\ &\gamma_k = g_{k-1}; \quad 2 \le k \le s \\ &\gamma_k = g_{s-1} + g_s(k - s); \quad s + 1 \le k \le t, \end{aligned} \tag{3.8}$$

where $1 \le q, r, s \le t - 1$. For definiteness, $a_0 = 0, b_0 = 0$ and $g_0 = 0$ in the second, fourth and sixth equations when $q = 1, r = 1$ and $s = 1$, respectively. Moreover, note that $a_1, \ldots, a_{q-1}, b_1, \ldots, b_{r-1}$ and $g_1, \ldots, g_{s-1}$ are varying intercept parameters, while a_q, b_r and g_s are slope parameters.

The amount of smoothing can now be set by the choice of q, r and s. Note that this has some similarities to the ad hoc procedure described in Section 2.2. For instance, model (3.8) could be used in order to smooth the later part of the run-off pattern and perhaps extend β_j beyond t for a tail. It might not make sense to do the same for α_i, however, the model could be useful in order to forecast calendar year effects by extrapolation of γ_k. The key to this is the choice of q, r and s, and we will set out an automatic way to do this, to replace the ad hoc procedure often used.

3.2.1. A special case: Smoothing of the run-off pattern

From now on we will stick to the assumption $q = t - 1, s = 0$ and $\boldsymbol{\Gamma} = \mathbf{0}$ in (3.8). Thus, $\boldsymbol{D}$ will be of size $(2t - 1) \times (t + r)$,

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & t - r \end{pmatrix}. \tag{3.9}$$

This special case is particularly interesting since it can be related to the log-linear smoothing of chain-ladder development factors in Section 2.2.

The theoretical development factors in Eq. (2.2) can be rewritten as

$$f_j = \frac{\sum_{i=1}^{t-j} \sum_{l=1}^{j+1} m_{il}}{\sum_{i=1}^{t-j} \sum_{l=1}^{j} m_{il}}, \tag{3.10}$$

where m_{ij} is defined by the underlying GLM structure in (2.10), (3.1) and (3.8). Recall that the observed development factors

$$\hat{f}_j = \frac{\sum_{i=1}^{t-j} \sum_{l=1}^{j+1} \hat{m}_{il}}{\sum_{i=1}^{t-j} \sum_{l=1}^{j} \hat{m}_{il}} \tag{3.11}$$

equal the chain-ladder development factors in (2.3) when we let $r = t - 1$ (and $p = 1$ in (2.10)).

Thus,

$$\begin{aligned} \ln\left(f_j - 1\right) &= \ln\left(\frac{\sum_{i=1}^{t-j} \sum_{l=1}^{j+1} e^{c+\alpha_i+\beta_l}}{\sum_{i=1}^{t-j} \sum_{l=1}^{j} e^{c+\alpha_i+\beta_l}} - 1\right) \\ &= \ln\left(\frac{\sum_{l=1}^{j+1} e^{\beta_l}}{\sum_{l=1}^{j} e^{\beta_l}} - 1\right) \\ &= \beta_{j+1} - \ln\left(\sum_{l=1}^{j} e^{\beta_l}\right). \end{aligned} \tag{3.12}$$

For $j \ge r$ Eq. (3.12) can be written as

$$\ln\left(f_j - 1\right) = b_p + b_{p-1}\,(j + 1 - q + 1) - \ln\left(\sum_{l=1}^{j} e^{\beta_l}\right). \tag{3.13}$$

Since r is supposed to be so large that the linear extrapolation captures the tail only small values of β_j will remain to the right of r. Therefore the linear parametrization of β_j approximately leads to log-linear smoothing of theoretical development factors f_j, analogous to the log-linear smoothing of true development factors accounted for in Section 2.2. A numerical comparison of the two approaches is provided in Section 7.2.

4. Model selection

In practice, we would like to select the truncation points q, r and s of Section 3.2 from data. To this end, we let $\hat{\boldsymbol{\theta}}_{\text{qrs}}$ denote the estimated parameter vector for a model with fixed (q, r, s). We can choose the model

$$(\hat{q}, \hat{r}, \hat{s}) = \operatorname{argmin}_{(q,r,s)\in I} \operatorname{Crit}(\hat{\boldsymbol{\theta}}_{\text{qrs}}), \tag{4.1}$$

that minimizes a model selection criterion $\operatorname{Crit}(\hat{\boldsymbol{\theta}}_{\text{qrs}})$ among a pre-chosen set I of candidate models. We then take $\hat{\boldsymbol{\theta}}_{(\hat{q},\hat{r},\hat{s})}$ as the final parameter estimators on which to base reserves.

An ad hoc GLM claims reserving approach would correspond to using a single model $I = \{(q, r, s)\}$. This is often what is done in practice, but it requires prior knowledge of the truncation points q, r and s, which is often not realistic. Indeed, there are many situations when several candidate models should be allowed for. For instance, suppose smoothing of development years is of concern, whereas accident years are not smoothed and inflation not included in the model, then

$$I = \{(t-1, 1, 0), \dots, (t-1, t-1, 0)\} \tag{4.2}$$

is of interest. If we also allow the possibility of the last accident year parameter being linearly interpolated from the previous two, we put

$$I = \{(t-2, 1, 0), \dots, (t-2, t-1, 0), (t-1, 1, 0), \dots, (t-1, t-1, 0)\}. \tag{4.3}$$

A linear inflation trend can be incorporated into either of (4.2) and (4.3) by simply replacing $s = 0$ with $s = 1$ everywhere, and so on.

Note that low values of r, where the most extreme choice would be $r = 1$, are not of any actual interest in (4.2) and (4.3), however, we still choose to include them for completeness and illustration, see Section 7.2.

In connection with (4.2), the classical exponential smoothing looks at the chain-ladder estimated development factors and then chooses I to contain one single model $(t-1, r, 0)$, where r is the visually determined break point for a linear trend on the log scale.

Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) could be used as model selection criteria when working with likelihood functions. The idea is to add a penalty term to the log-likelihood in order to avoid choosing too large models that result in over-fitting, using criterion functions. The definitions are

$$\text{Crit} = \text{AIC}(\hat{\boldsymbol{\theta}}_{\text{qrs}}) = 2w - 2l(\hat{\boldsymbol{\theta}}_{\text{qrs}}) \tag{4.4}$$

and

$$\text{Crit} = \text{BIC}(\hat{\boldsymbol{\theta}}_{\text{qrs}}) = \ln(n)w - 2l(\hat{\boldsymbol{\theta}}_{\text{qrs}}), \tag{4.5}$$

see e.g. Miller (2002). Here $w = 1 + q + r + s$ is the number of parameters and $l(\hat{\boldsymbol{\theta}}_{\text{qrs}})$ is the maximized log-likelihood function with respect to model (q, r, s). It can be seen that AIC and BIC differ in that they penalize the fitted log-likelihood of large models in different ways.

Bootstrapping could also be used to estimate the mean squared error of prediction $MSEP(\hat{R}) = E\left((R - \hat{R})^2\right)$, where the expected value is with respect to R and $\hat{R}$. Estimating MSEP by bootstrapping, we get a model selection criterion

$$\text{Crit} = \widehat{MSEP}(\hat{\boldsymbol{\theta}}_{\text{qrs}}) = E\left((R^{**} - \hat{R}^*)^2\right), \tag{4.6}$$

where R^{**} and $\hat{R}^*$ are resampled reserves and resampled estimated reserves, respectively. The resampled data are created by a parametric bootstrap from model (q, r, s). This corresponds to the criterion BOOT in Pan and Le (2001). Note that $\widehat{MSEP}$ measures the deviation between the future claims and the predicted ones under a certain model, while AIC and BIC measure the deviation between the observations and their expected values.

Another possibility, pursued by Pan (2001), is to define an analogue of AIC, where the likelihood is replaced by the quasi-likelihood function $Q(\boldsymbol{\theta})$ of Wedderburn (1974). Pan (2001) suggested a quasi-likelihood information criterion

$$\text{QIC}(\hat{\boldsymbol{\theta}}_{\text{qrs}}) = 2\,\text{trace}(\hat{\boldsymbol{\Omega}}\,\hat{\boldsymbol{V}}) - 2\,Q(\hat{\boldsymbol{\theta}}_{\text{qrs}}), \tag{4.7}$$

where $\hat{\boldsymbol{\Omega}} = -\frac{\partial^2}{\partial \boldsymbol{\theta}^2} Q(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{qrs}}}$ and $\hat{\boldsymbol{V}}$ is an estimator of $\boldsymbol{V} = \text{Cov}(\hat{\boldsymbol{\theta}}_{\text{qrs}})$, see Liang and Zeger (1986) for an example. Note that QIC reduces to AIC when the quasi and true likelihoods coincide, since then $\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{V}}^{-1}$.

England and Verrall (2001) discuss the use of the deviance for model comparison in their GAM framework. They remark that it is not obvious how many degrees of freedom should be used since the GAM smoothers are non-parametric. Here, on the other hand, the deviance is inappropriate since it does not penalize large models and, hence, it will by definition be lower for larger models.

Note that the outcome of the model selection is sensitive to the chosen criterion. For instance, AIC usually tends to select large models. Moreover, the suggested criteria do not help us choose among the underlying distributional assumptions. In Section 7 we will illustrate model selection numerically using AIC, BIC and $\widehat{MSEP}$ as criteria for the special case in Section 3.2.1.

5. Estimating the model parameters

Here the estimation procedure for the special case in Section 3.2.1 is described, which could be extended to consider the general model in Section 3.2.

5.1. Estimation of ϕ for model selection

Care must be taken regarding the choice of estimator $\hat{\phi}$, since it will strongly affect the outcome of the model selection. The error terms $C_{ij} - m_{ij}$ of a smaller model with some degree of smoothing, i.e. $(t-1, r, 0)$ where $r < t-1$, can be expressed as a sum of two terms; the random errors of the full model $(t-1, t-1, 0)$ and a second term $m_{ij}^{\text{full}} - m_{ij}$, which accounts for systematic effects not captured by the smaller model. Hence, estimating ϕ for each model $(t-1, r, 0)$ yields a higher value compared to estimating ϕ based on the full model $(t-1, t-1, 0)$.

For model selection using AIC and BIC in (4.4) and (4.5) we are interested in comparing the random errors of the models. Therefore the estimator of ϕ should be based on the full model, see e.g. Pan (2001). However, if we use bootstrapping, as in (4.6), the systematic effects of model $(t-1, r, 0)$ should be included, otherwise smaller models will benefit since $MSEP$ would be underestimated.

Hence, in order to estimate ϕ for AIC and BIC we use the Pearson residuals

$$r_{ij}^{\text{full}} = \frac{C_{ij} - \hat{m}_{ij}^{\text{full}}}{\sqrt{\left(\hat{m}_{ij}^{\text{full}}\right)^p}} \tag{5.1}$$

and for $\widehat{MSEP}$ we use

$$r_{ij} = \frac{C_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}^p}}, \tag{5.2}$$

where $\hat{m}_{ij}$ is the estimator of m_{ij} for the particular model under analysis. This yields

$$\hat{\phi}^{\text{full}} = \frac{1}{n - 2t + 1} \sum_{\nabla} (r_{ij}^{\text{full}})^2, \tag{5.3}$$

and, since $w = 1 + q + r + s = 2t - 1$ is the number of estimated parameters (excluding $\hat{\phi}$) in the full model $(t-1, t-1, 0)$,

$$\hat{\phi} = \frac{1}{n - w} \sum_{\nabla} (r_{ij})^2. \tag{5.4}$$

Here, the index '*full*' is only used for clarity.

5.2. The reserving algorithm

We can now implement a log-linear smoothing procedure for the reserving exercise according to the following scheme.

1. Define the design matrix $\mathbf{X}_{\text{full}}$ of the full model in (3.1) when the GLM in (2.10) and (2.7) is assumed.

2. Define a family I of models, based on truncation points $(t-1, r, 0)$, from which we would like to select a model.
   FOR all $(t-1, r, 0) \in I$ DO:
3. Create $\mathbf{A}$ and $\mathbf{B}$, and hence, the block matrix $\mathbf{D}$.
4. Calculate $\mathbf{X} = \mathbf{X}_{\text{full}}\, \mathbf{D}$.
5. Set up the new GLM $\boldsymbol{\eta} = \boldsymbol{X}\,\boldsymbol{\theta}$. Then use some standard software to compute an estimate $\hat{\boldsymbol{\theta}}_{t-1,r,0}$ of $\boldsymbol{\theta} = \boldsymbol{\theta}_{t-1,r,0}$ by maximizing e.g. the likelihood or quasi likelihood.
6. Evaluate the chosen model selection criterion $\text{Crit}(\hat{\boldsymbol{\theta}}_{t-1,r,0})$ for model $(t-1, r, 0)$ using either (5.3) for AIC, BIC and QIC in (4.4), (4.5) and (4.7) or (5.4) for $\widehat{MSEP}$ in (4.6).
   END.
7. Select model $(t-1, \hat{r}, 0)$ as in (4.1).
8. Obtain estimators $\widehat{E}(C_{ij}) = \hat{m}_{ij}$ and $\widehat{\text{Var}}(C_{ij}) = \hat{\phi}\,\hat{m}_{ij}^{p}$ from (2.10) and (2.7), with c, α_i and β_j replaced by estimates, computed from the first four equations of (3.8) and $\hat{\boldsymbol{\theta}}_{t-1,\hat{r},0}$. Here $\hat{\phi}$ is obtained from (5.4).

6. Bootstrap and quantile prediction

It is now straightforward to implement a bootstrap procedure for either the model $\hat{\boldsymbol{\theta}}_{t-1,\hat{r},0}$ selected by the reserving algorithm in Section 5.2 or for the full model $\hat{\boldsymbol{\theta}}_{t-1,t-1,0}$ corresponding to the GLM in (2.10) and (2.7). Including the model selection part in the bootstrap procedure implies that each resampled pseudo-triangle is being individually analyzed in the bootstrap world in a similar way as it had been by the actuary if it was an observed triangle in the real world. Hence, if there are outliers in the data set the prediction error could decrease compared to the situation when we less accurately apply the same reserving algorithm to all pseudo-triangles. It is difficult to implement such a procedure for the deterministic approach in Section 2.2 since the truncation points for the smoothing of the development factors usually is based on an ad hoc decision instead of a systematic one.

We will use the bootstrap technique provided in Björkwall et al. (2009) for the implementation. However, since it is beyond the scope of this paper to compare bootstrap algorithms we will focus only on the parametric approach due to its robustness even for small data sets.

Hence, in addition to the assumptions in (2.10) and (2.7) we assume a full distribution F for C_{ij}, parameterized by the mean and variance, so that we may write $F = F(m_{ij}, \phi m_{ij}^{p})$. Typically we assume that F is either an ODP or a gamma distribution corresponding to $p = 1$ and $p = 2$, respectively. We draw C_{ij}^{*} from $F(\hat{m}_{ij}, \hat{\phi}\hat{m}_{ij}^{p})$, where $\hat{m}_{ij}$ is calculated according to the model $(t-1, \hat{r}, 0)$ selected as in (4.1), for all $i, j \in \nabla$ and $\hat{\phi}$ is obtained from (5.4). This produces the pseudo-triangles ∇C^{*}. Steps 1–8 in Section 5.2 are then repeated for each ∇C^{*} and the bootstrap estimates $\hat{R}_i^{*} = \sum_{j\in\Delta_i} \hat{m}_{ij}^{*}$ and $\hat{R}^{*} = \sum_{\Delta} \hat{m}_{ij}^{*}$ are finally calculated for each pseudo-triangle.

In order to get $R_i^{**} = \sum_{j\in\Delta_i} C_{ij}^{**}$ and $R^{**} = \sum_{\Delta} C_{ij}^{**}$, corresponding to the process error, we sample once again from $F(\hat{m}_{ij}, \hat{\phi}\hat{m}_{ij}^{p})$ to get C_{ij}^{**} for all $i, j \in \Delta$. Finally, the B observations of the standardized or unstandardized prediction errors given in Björkwall et al. (2009) are plotted to yield the predictive distribution.

If we let $I = \{(t-1, t-1, 0)\}$ in step 2 in Section 5.2 and then implement the resampling described above, we get a bootstrap algorithm for the full model $\hat{\boldsymbol{\theta}}_{t-1,t-1,0}$. Hence, assuming $p = 1$ in (2.10) and an ODP distribution for F yields a bootstrap algorithm for the chain-ladder method.

7. Numerical study

The purpose of this numerical study is to illustrate the smoothing effect of the GLM parametrization in Section 3.2.1; the special case corresponding to smoothing of the run-off pattern. Moreover, we will show how the model selection can be carried out and how the predictive distribution can be simulated by the suggested bootstrap approach.

In the subsequent sections, we work under the assumption of ODP ($p = 1$) and gamma ($p = 2$) distributions. We use the unstandardized prediction errors for the bootstrap procedure since they are always defined, see Björkwall et al. (2009).

7.1. Data

We analyze the data set from Taylor and Ashe (1983), even though the chain-ladder development factors of this particular triangle are already quite smooth, since it is useful to compare with previous studies. The triangle of paid claims ∇C is presented in Table 7.1.

7.2. A comparison of the smoothing effects

In order to illustrate the smoothing effect of the model Figs. 1 and 2 show the $\ln(\hat{f}_j - 1)$ curves under the assumption of an ODP and a gamma distribution, respectively. Here $\hat{f}_j$ are the estimated development factors before and after the smoothing. A sequence of deterministically smoothed chain-ladder development factors has also been included in the figures for comparison.

It is difficult to compare the reparameterized GLM with the deterministic approach since there are no rules of how to smooth the latter. As mentioned previously, the deterministic approach could be varied in many different ways based on ad hoc decisions made by the actuary. However, in order to construct a consistent and repeatable algorithm we have first applied a log-linear regression to the *two* last chain-ladder development factors (this does of course not change the estimates, since all we do is draw the same line as before between the two factors). This would correspond to $r = 9$. Secondly, we have applied a log-linear regression to the *three* last chain-ladder development factors and created a new smoothed sequence using the first seven original factors and the three new ones. This would correspond to $r = 8$. In this way we have proceeded through the sequences until we included all the original factors in the log-linear regression and then created a smoothed sequence which completely consists of new smoothed factors. This would correspond to $r = 2$. Finally, we have fitted a straight line through the origin, which would correspond to only one model parameter and, hence, $r = 1$.

Note that the model $r = 1$ is of hardly any actual interest for reserving purposes, but we still choose to include it here for completeness and illustration. In practice the first original factors are often kept, while only the tail is smoothed and, hence, low values of r should be excluded from I in a real application.

Recall that the factors for $r = 9$ under the assumption of an ODP in Fig. 1 equal the pure chain-ladder development factors, while the corresponding estimates under the assumption of a gamma distribution in Fig. 2 slightly differ. As can be seen, the reparameterized GLM yields a smoother curve than the deterministic approach when it is used in the tail of the factor sequence. Also note that the smoothed part of the curves is almost a straight line for large values of r, just as would be expected according to Eq. (3.13). For lower values of r the curves are convex in comparison with the straight line obtained by the deterministic approach. The difference between the assumptions of an ODP and a gamma distribution is the slope of the curve, which tends to

Table 7.1
Observations of paid claims ∇C from Taylor and Ashe (1983).

	1	2	3	4	5	6	7	8	9	10
1	357 848	766 940	610 542	482 940	527 326	574 398	146 342	139 950	227 229	67 948
2	352 118	884 021	933 894	1 183 289	445 745	320 996	527 804	266 172	425 046	
3	290 507	1 001 799	926 219	1 016 654	750 816	146 923	495 992	280 405		
4	310 608	1 108 250	776 189	1 562 400	272 482	352 053	206 286			
5	443 160	693 190	991 983	769 488	504 851	470 639				
6	396 132	937 085	847 498	805 037	705 960					
7	440 832	847 631	1 131 398	1 063 269						
8	359 480	1 061 648	1 443 370							
9	376 686	986 608								
10	344 014									

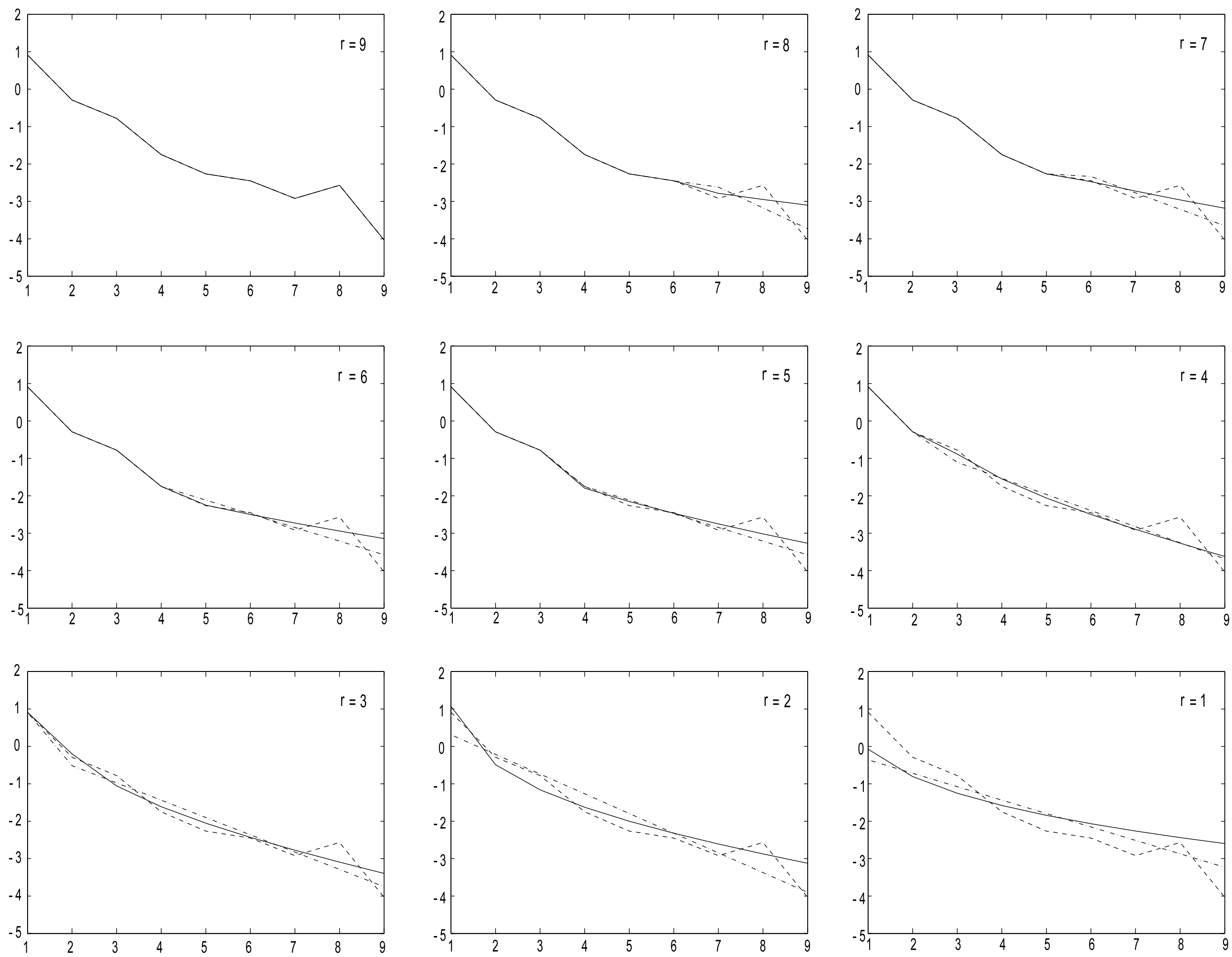


Fig. 1. The $\ln(\hat{f}_j - 1)$ curves for the full GLM (– – –), the smoothed GLM (—) and the smoothed chain-ladder (· — ·) under the assumption of an ODP ($p = 1$) distribution. On the x-axis we have j.

be lower for the latter one. However, for $r = 4$ the two curves coincide quite well. Note that the reparameterized GLM yields a heavier tail than the deterministic approach. If the curve were to be extrapolated for a tail this might lead to an unrealistically large reserve according to the actuary's judgement.

A decision by eye would probably result in the use of one of the models $r = 5, 6$ for the ODP distribution and $r = 5, 6$, or maybe $r = 7, 8$, for the gamma distribution. However, Figs. 1 and 2 do not reveal how to choose between the two distributional assumptions. For the deterministic approach we would choose the amount of smoothing corresponding to $r = 5, 6$. However, this approach does not yield maximum likelihood reserve estimates and we will instead focus on the difference between the ODP and gamma assumptions for the reparameterized GLM.

7.3. The model selection

AIC and BIC cannot be used as model selection criteria in (4.1) for the assumption of an ODP distribution since we do not have a likelihood function due to the over-dispersion. However, the square roots of $\widehat{MSEP}$ in (4.6) are presented in Table 7.2. Recall that $\hat{\phi}$ in (5.4) is used for the resampling in the bootstrap procedure. We also present the deviance computed in the statistical software used for the modeling (MATLAB); here it is defined as the sum of the squared deviance residuals. As remarked in Section 4 the deviance does not penalize large models and, hence, it will by definition be lower for larger models.

AIC and BIC can be used for the assumption of a full gamma distribution with expected value m_{ij} and variance ϕm_{ij}^2, i.e.

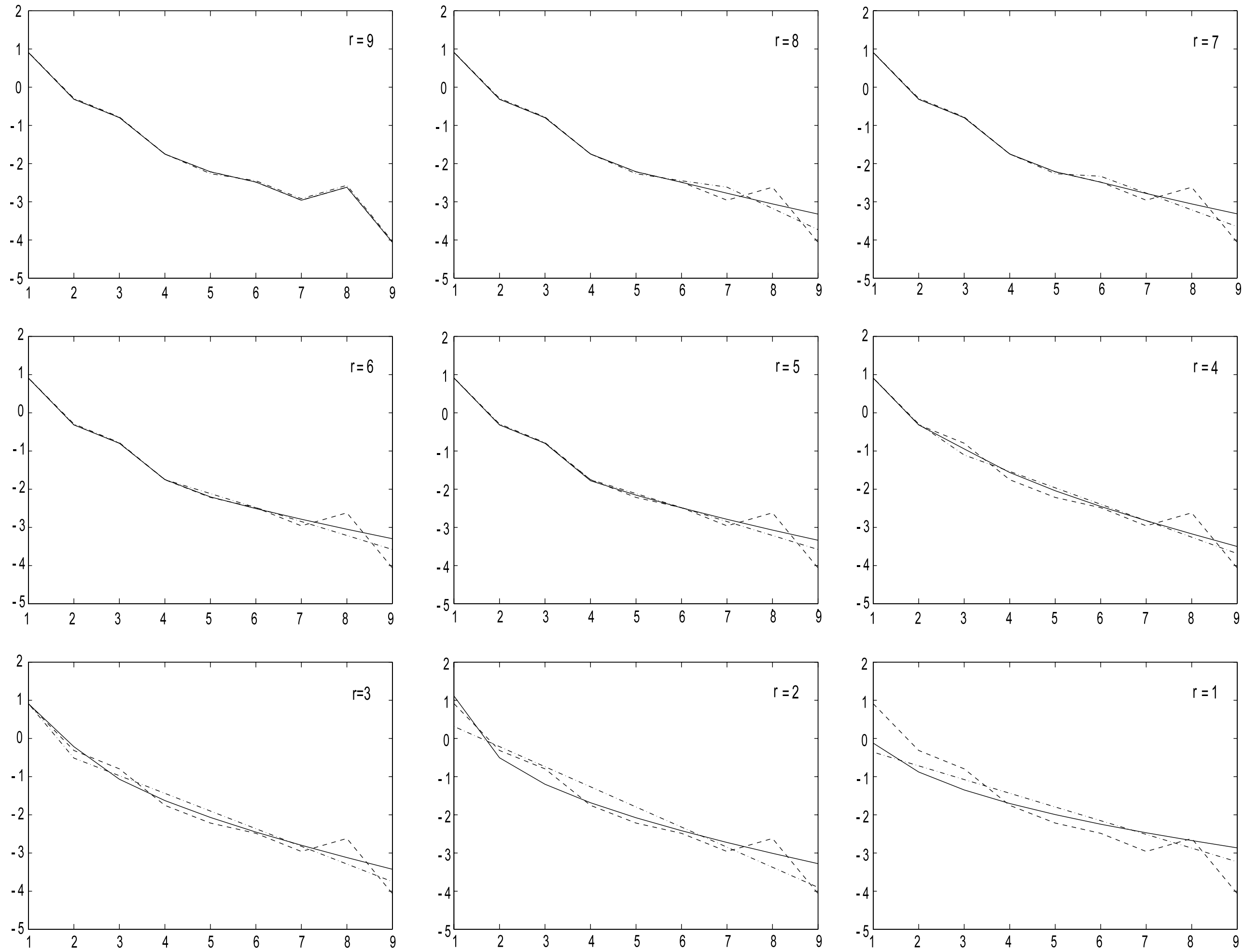


Fig. 2. The $\ln(\hat{f}_j - 1)$ curves for the full GLM (– – –), the smoothed GLM (—) and the smoothed chain-ladder (· — ·) under the assumption of a gamma distribution ($p = 2$). On the x-axis we have j.

Table 7.2
The square root of $\widehat{MSEP}$ in (4.6) and the deviance when an ODP distribution is assumed.

r	9	8	7	6	5	4	3	2	1
$\sqrt{\widehat{MSEP}}$ (10^3)	3039	3222	3247	3114	3000	3051	3234	3659	4921
Deviance (10^3)	1903.0	2073.0	2077.5	2079.2	2108.1	2402.0	2607.2	3161.3	7807.9

$C_{ij} \sim \Gamma\left(\frac{1}{\phi}, \phi\, m_{ij}\right)$. Summing over the n observations in ∇C, which are assumed to be independent, yields

$$l(\hat{\boldsymbol{\theta}}, \hat{\phi}^{\text{full}}) = \frac{1}{\hat{\phi}^{\text{full}}} \sum_{\nabla} \left(-\frac{C_{ij}}{\hat{m}_{ij}} - \log(\hat{m}_{ij}) \right) + \sum_{\nabla} \left(\frac{1}{\hat{\phi}^{\text{full}}} \log\left(\frac{C_{ij}}{\hat{\phi}^{\text{full}}} \right) - \log(C_{ij}) - \log \Gamma\left(\frac{1}{\hat{\phi}^{\text{full}}} \right) \right).$$

The values of AIC and BIC are presented in Table 7.3 together with the square root of $\widehat{MSEP}$ and the deviance.

As can be seen, $\widehat{MSEP}$ selects model $r = 5$ or $r = 9$ for the assumption of an ODP distribution. For the gamma assumption AIC and $\widehat{MSEP}$ select model $r = 9$ or $r = 5$, while BIC selects model $r = 3$ or $r = 2$. Hence, for the triangle under analysis AIC and $\widehat{MSEP}$ seem to be better selection criteria than BIC, since model $r = 5$ was one of the models chosen by eye in Section 7.2 too.

The number of parameters w has a large impact on the value of AIC in (4.4), since the log-likelihood is quite constant for $5 \leq r \leq 8$ for the triangle under analysis. This explains why AIC selects model $r = 9$ or $r = 5$.

7.4. Reserve estimates and quantiles

The reserve estimates and bootstrap statistics under the assumption of an ODP and a gamma distribution, respectively, are shown in detail in Appendix. In Fig. 3 we present the results graphically in order to get an overview for different choices of r. Here $B = 10\,000$ iterations were used in the bootstrap procedure.

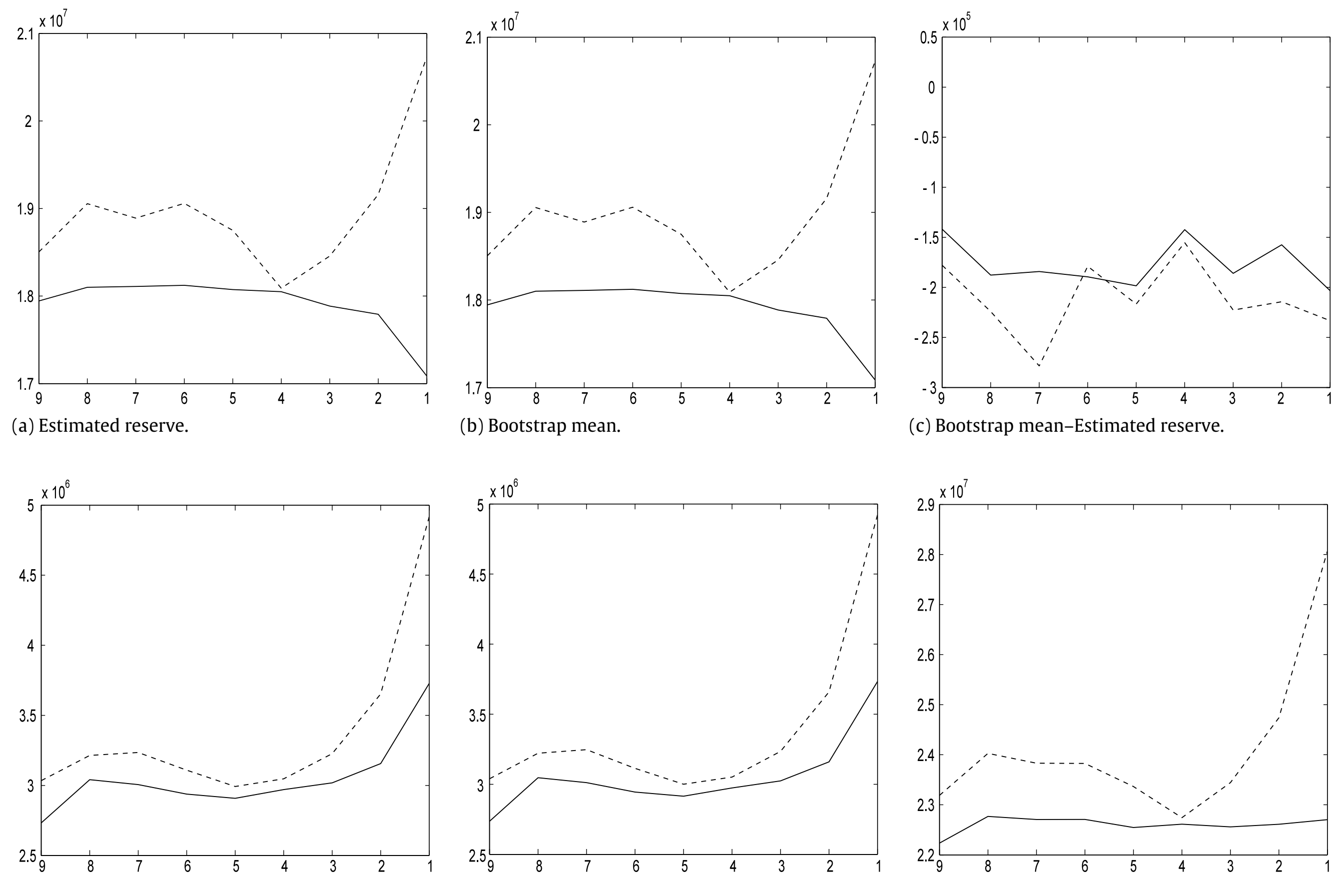
As can be seen from (a), smoothing affects the reserve estimates in different ways for the assumption of an ODP and a gamma distribution. For $r = 4$ the reserve estimates coincide, as was already concluded in Section 7.2. The means of the bootstrap samples in (b) follow the reserve estimates, with the latter ones being slightly larger since the difference in (c) is negative (but random). Therefore, $\sqrt{\widehat{MSEP}}$ in (e) is slightly larger than the standard deviation of the bootstrap samples in (d). The 95

Table 7.3
The values of AIC, BIC, the square root of $\widehat{MSEP}$ in (4.6) and the deviance when a gamma distribution is assumed.

r	9	8	7	6	5	4	3	2	1
AIC	1502.3	1508.9	1506.9	1505.0	1503.1	1505.1	1504.6	1508.6	1578.3
BIC	1540.5	1545.1	1541.1	1537.1	1533.2	1533.2	1530.7	1532.6	1600.4
$\sqrt{\widehat{MSEP}}$ (10^3)	2736	3047	3012	2944	2915	2974	3024	3160	3733
Deviance (10^3)	4023.5	4931.9	4932.0	4934.3	4951.3	5372.0	5526.8	6155.5	13 717.8

Table 7.4
Bootstrap results when the model selection has been implemented in the bootstrap procedure.

	Estimated reserve	Bootstrap mean	Boot. mean − Est. res.	Bootstrap stand. dev.	$\sqrt{\widehat{MSEP}}$	95 percentile
AIC	18 085 773	17 911 099	−174 674	2 735 238	2 740 673	22 082 887
BIC	18 071 392	17 969 537	−101 855	3 031 674	3 033 233	22 602 603



(a) Estimated reserve. (b) Bootstrap mean. (c) Bootstrap mean–Estimated reserve.

(d) Bootstrap standard deviation. (e) $\sqrt{\widehat{MSEP}}$. (e) 95th percentile.

Fig. 3. Reserve estimates and bootstrap statistics for the assumption of an ODP and a gamma distribution, respectively. Here we use (– – –) for the ODP distribution and (—) for the gamma distribution. On the x-axis we have r.

Table 7.5
The frequency of chosen models in a bootstrap simulation where $B = 10\,000$.

r	9	8	7	6	5	4	3	2	1
AIC	7010	24	85	166	1240	454	801	220	0
BIC	9	10	32	47	117	368	5394	4023	0

percentiles in (f) mainly follow the shape of the reserve estimates, since the standard deviations are relatively constant (except for the lowest values of r).

In general, the first development factors of a sequence should be kept unsmoothed since they are supposed to provide significant information regarding the data set. Hence, it is only relevant to smooth the tail of the development factor sequence. However, Fig. 3 shows that, for the particular triangle under analysis, smoothing of the tail only has a minor effect on the results, while the distributional assumption seems more important. To illustrate this, we investigate the reserve estimate and the risk corresponding to the 95th percentile. Suppose that we start with the chain-ladder method, i.e. $r = 9$ and the assumption of an ODP distribution, and, moreover, that we decide to smooth the development factors using model $r = 5$ in order to eliminate the shakiness appearing in Fig. 1. According to Tables A.1 and A.3 in the Appendix we would then see a 1.5% increase in the reserve estimate and a 0.8% increase in the 95th percentile. If we instead would stick to $r = 9$, but assume a gamma distribution, we would see a 3.3% decrease in the reserve estimate and a 4.3% decrease in the 95th percentile according to Tables A.2 and A.4.

Table A.1
The estimated reserves under the assumption of an ODP distribution.

Year	$r = 9$	$r = 8$	$r = 7$	$r = 6$	$r = 5$	$r = 4$	$r = 3$	$r = 2$	$r = 1$
2	94 634	240 187	221 536	230 156	202 906	142 453	178 698	234 891	397 438
3	469 511	490 001	468 650	481 624	435 577	322 911	394 173	505 214	826 224
4	709 638	769 423	767 511	778 890	725 379	571 929	677 310	844 415	1 330 160
5	984 889	1 039 712	1 029 537	1 029 175	992 396	840 830	961 998	1 162 612	1 755 176
6	1 419 459	1 477 137	1 466 432	1 473 134	1 483 356	1 373 765	1 508 001	1 756 918	2 521 484
7	2 177 641	2 241 521	2 229 665	2 237 087	2 208 130	2 310 842	2 399 913	2 667 064	3 585 818
8	3 920 301	3 996 865	3 982 655	3 991 552	3 956 845	3 864 518	3 655 170	3 776 692	4 609 461
9	4 278 972	4 342 643	4 330 826	4 338 225	4 309 362	4 232 583	4 278 865	3 735 382	3 778 936
10	4 625 811	4 681 894	4 671 485	4 678 001	4 652 579	4 584 950	4 625 716	4 690 754	2 155 908
Total	18 680 856	19 279 383	19 168 297	19 237 844	18 966 529	18 244 781	18 679 843	19 373 942	20 960 607

Table A.2
The estimated reserves under the assumption of a gamma distribution.

Year	$r = 9$	$r = 8$	$r = 7$	$r = 6$	$r = 5$	$r = 4$	$r = 3$	$r = 2$	$r = 1$
2	93 316	202 152	203 511	207 892	199 638	172 114	184 757	211 549	309 558
3	446 505	408 480	409 958	416 799	404 635	351 964	376 550	429 348	639 118
4	611 145	629 353	629 086	632 612	618 547	574 367	611 511	688 582	1 018 712
5	992 023	1 008 479	1 008 970	1 003 960	996 081	934 022	967 584	1 073 045	1 415 607
6	1 453 085	1 470 922	1 471 480	1 473 661	1 497 337	1 481 516	1 511 006	1 616 061	2 080 145
7	2 186 161	2 209 807	2 210 383	2 212 695	2 206 567	2 416 362	2 386 357	2 497 200	3 015 061
8	3 665 066	3 687 154	3 687 803	3 689 857	3 684 059	3 651 627	3 402 313	3 351 061	3 906 657
9	4 122 398	4 139 309	4 139 832	4 141 344	4 136 502	4 107 371	4 118 985	3 557 488	3 149 813
10	4 516 073	4 532 001	4 532 447	4 532 966	4 528 998	4 502 114	4 512 328	4 524 778	1 755 546
Total	18 085 773	18 287 657	18 293 470	18 311 784	18 272 364	18 191 456	18 071 392	17 949 111	17 290 218

Table A.3
Bootstrap results for the total under the assumption of an ODP distribution.

r	Estimated reserve	Bootstrap mean	Boot. mean — Est. res.	Bootstrap stand. dev.	$\sqrt{\widehat{MSEP}}$	95 percentile
9	18 680 856	18 502 852	−178 004	3 034 174	3 039 240	23 187 718
8	19 279 383	19 055 285	−224 098	3 214 046	3 221 689	24 027 278
7	19 168 297	18 889 878	−278 419	3 235 449	3 247 245	23 830 275
6	19 237 844	19 058 883	−178 961	3 108 771	3 113 762	23 825 274
5	18 966 529	18 749 895	−216 634	2 992 653	3 000 334	23 365 377
4	18 244 781	18 089 415	−155 366	3 047 533	3 051 339	22 742 333
3	18 679 843	18 457 164	−222 679	3 226 904	3 234 417	23 442 951
2	19 373 942	19 159 588	−214 354	3 652 746	3 658 848	24 738 289
1	20 960 607	20 727 316	−233 291	4 915 800	4 921 087	28 065 538

Table A.4
Bootstrap results for the total under the assumption of a gamma distribution.

r	Estimated reserve	Bootstrap mean	Boot. mean — Est. res.	Bootstrap stand. dev.	$\sqrt{\widehat{MSEP}}$	95 percentile
9	18 085 773	17 943 796	−141 977	2 732 628	2 736 177	22 233 262
8	18 287 657	18 099 941	−187 716	3 041 109	3 046 746	22 767 497
7	18 293 470	18 109 290	−184 180	3 006 144	3 011 631	22 706 271
6	18 311 784	18 122 332	−189 452	2 938 433	2 944 387	22 707 132
5	18 272 364	18 073 957	−198 407	2 908 233	2 914 848	22 545 380
4	18 191 456	18 049 088	−142 368	2 970 365	2 973 626	22 612 284
3	18 071 392	17 885 372	−186 020	3 018 708	3 024 283	22 558 091
2	17 949 111	17 791 615	−157 496	3 156 046	3 159 815	22 611 167
1	17 290 218	17 086 896	−203 322	3 727 575	3 732 930	22 702 370

7.5. Implementing model selection in the bootstrap procedure

In this section, model selection is included in the bootstrap procedure, as described in Section 6, for the assumption of a gamma distribution. The results are presented in Table 7.4. Recall that $\hat{\phi}^{\text{full}}$ is always used for the calculation of AIC and BIC, but for the resampling of pseudo-triangles in the bootstrap procedure $\hat{\phi}$ is used. Also recall that the first choice of model for AIC and BIC was $r = 9$ and $r = 3$, respectively, see Table 7.3. It is not possible to distinguish these results from the ones for $r = 9$ and $r = 3$, respectively, in Table A.4. Table 7.5 shows the frequency of the selection of each model in the 10 000 iterations of the bootstrap, where it can be seen that AIC strongly prefers high values of r, while BIC does the opposite.

8. Discussion

In this paper, a model has been described which allows for smoothing of origin, development and calendar year parameters. This smoothing model for the run-off pattern has some similarities to the somewhat ad hoc log-linear smoothing of chain-ladder development factors which is often used in practice. The suggested model is much simpler, but less flexible, than the GAM framework presented in England and Verrall (2001) and the GLMM approach provided by Antonio and Beirlant (2008). It can be used as a stochastic foundation for a claims reserving exercise including smoothing, model selection and bootstrapping for either prediction errors or a full predictive distribution.

While it is difficult to make any final conclusions from the single data set which has been analyzed in this paper, it is interesting to note that the distributional assumption of the model had a larger impact on the results than the smoothing effect. Hence, it seems important to first find an appropriate model, which then possibly could be adjusted by smoothing of the model parameters.

The main weakness of a GLM with a log-link is that the model cannot be used for data sets including negative increments when a full distribution is assumed. A future development would be to use pure quasi-likelihood and the resulting estimating equations. In that case resampling of residuals would be required for the bootstrap procedure.

Acknowledgements

Ola Hössjer's work was supported by the Swedish Research Council, grant 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine.

Appendix

See Tables A.1–A.4.

References

Antonio, K., Beirlant, B., 2008. Issues in claims reserving and credibility: a semi-parametric approach with mixed models. Journal of Risk and Insurance 75 (3), 643–676.
Barnett, G., Zehnwirth, B., 2000. Best estimates for reserves. Proceedings of the Casualty Actuarial Society LXXXVII, 245–321.
Björkwall, S., Hössjer, O., Ohlsson, E., 2009. Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving. Scandinavian Actuarial Journal 4, 306–331.
De Jong, P., Zehnwirth, B., 1983. Claims reserving, state-space models and the Kalman filter. Journal of the Institute of Actuaries 110, 157–182.
England, P., Verrall, R., 2001. A flexible framework for stochastic claims reserving. Proceedings of the Casualty Actuarial Society LXXXVIII, 1–38.
England, P., Verrall, R., 2002. Stochastic claims reserving in general insurance. British Actuarial Journal 8 (3), 443–518.
Kremer, E., 1982. IBNR claims and the two way model of ANOVA. Scandinavian Actuarial Journal 47–55.
Liang, K.-Y., Zeger, S., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.
Miller, A., 2002. Subset Selection in Regression. Chapman and Hall, Boca Raton.
Pan, W., 2001. Akaike's information criterion in generalized estimating equations. Biometrics 57 (1), 120–125.
Pan, W., Le, C.T., 2001. Bootstrap model selection in generalized linear models. Journal of Agricultural, Biological, and Environmental Statistics 6 (1), 49–61.
Renshaw, A., Verrall, R., 1998. A stochastic model underlying the chain ladder technique. British Actuarial Journal 4, 903–923.
Sherman, R., 1984. Extrapolating, smoothing and interpolating development factors. Proceedings of the Casualty Actuarial Society LXXI, 122–192.
Taylor, G., Ashe, F.R., 1983. Second moments of estimates of outstanding claims. Journal of Econometrics 23, 37–61.
Verrall, R., 1989. A state space representation of the chain ladder linear model. Journal of the Institute of Actuaries 116, 589–609.
Verrall, R., 1996. Claims reserving and generalized additive models. Insurance: Mathematics and Economics 19, 31–43.
Verrall, R., 2007. Obtaining predictive distributions for reserves which incorporate expert opinion. Variance 1, 53–80.
Verrall, R., England, P., 2005. Incorporating expert opinion into a stochastic model for the chain-ladder technique. Insurance: Mathematics and Economics 37, 355–370.
Wedderburn, R.W.M., 1974. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. Biometrika 61, 439–447.
Wright, T., 1990. A stochastic method for claims reserving in general insurance. Journal of the Institute of Actuaries 117, 677–731.
Wütrich, M., Merz, M., 2008. Stochastic Claims Reserving Methods in Insurance. John Wiley & Sons Ltd.
Zehnwirth, B., 1985. ICRFS Version 4 Manual and Users Guide. Benhar Nominees Pty. Ltd., Tumurra, NSW, Australia.
Zehnwirth, B., 1989. Regression methods—applications. In: Conference Paper. 1989 Casualty Loss Reserve Seminar. Available online at: http://www.casact.org/pubs/CLRSTrans/1989/825.pdf.