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 Lånsförsäkringar Alliance, SE-106 50 Stockholm, Sweden
c Cass Business School, City University, 106 Bunhill Row, London EC1Y 8TZ, United Kingdom

A R T I C L E   I N F O

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

A B S T R A C T

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 fast 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.
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 smoothing 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 \( \{C_{ij}; i, j \in \Delta \} \) denote the incremental observations of paid claims, which are assumed to be available in a development triangle \( \Delta = \{(i, j); i = 1, \ldots , t; j = 1, \ldots , t - i + 1\} \). 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 \( \Delta \). 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 \( R = \sum_{j=1}^{t} C_{ij} \) 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 \( \hat{R}_i = \sum_{j=1}^{t-i+1} C_{ij} \), where \( \Delta_i \) denotes the row corresponding to origin year \( i \) in \( \Delta \).

Estimators of the outstanding claims per origin year and the grand total are obtained by \( \hat{R}_i = \sum_{j=1}^{t-i+1} \hat{C}_{ij} \) and \( \hat{R} = \sum_{i=1}^{t} \hat{C}_{ij} \), 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.

<table>
<thead>
<tr>
<th>Table 2.1</th>
<th>The triangle ( \nabla ) of observed incremental payments.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin year</td>
<td>Development year</td>
</tr>
<tr>
<td>1</td>
<td>1 2 3 ( \cdots ) t - 1 t</td>
</tr>
<tr>
<td>1</td>
<td>( C_{11} ) ( C_{12} ) ( C_{13} ) ( \cdots ) ( C_{1,t-1} ) ( C_{1,t} )</td>
</tr>
<tr>
<td>2</td>
<td>( C_{21} ) ( C_{22} ) ( C_{23} ) ( \cdots ) ( C_{2,t-1} ) ( C_{2,t} )</td>
</tr>
<tr>
<td>3</td>
<td>( C_{31} ) ( C_{32} ) ( C_{33} ) ( \cdots ) ( C_{3,t-1} ) ( C_{3,t} )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( \vdots ) ( \vdots ) ( \vdots ) ( \vdots ) ( \vdots ) ( \vdots )</td>
</tr>
<tr>
<td>t - 1</td>
<td>( C_{t-1,1} ) ( C_{t-1,2} ) ( \vdots ) ( \vdots ) ( \vdots )</td>
</tr>
<tr>
<td>t</td>
<td>( C_{t,1} ) ( C_{t,2} ) ( \vdots ) ( \vdots ) ( \vdots )</td>
</tr>
</tbody>
</table>

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

\[
D_{ij} = \sum_{t=i}^{C_{ij}} C_{it}.
\] (2.1)

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

\[
\hat{f}_j = \frac{\sum_{i=1}^{t-j} \mu_{ij}}{\sum_{i=1}^{t} \mu_{ij}}.
\] (2.2)

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

\[
\hat{f}_j = \frac{\sum_{i=1}^{t-j} D_{ij}}{\sum_{i=1}^{t} D_{ij}}.
\] (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 \( j \geq 4 \). Let \( \hat{f}_j \) denote the new sequence of development factors. Estimates \( \hat{\mu}_j \) for \( \Delta \) can now be computed as in the standard chain-ladder method yielding

\[
\hat{\mu}_j = D_{i,t-i} \hat{f}_{t-i+1} \hat{f}_{t-i+2} \cdots \hat{f}_{t-j+1}
\] (2.4)

and

\[
\hat{C}_{ij} = \hat{\mu}_{ij} - \hat{\mu}_{i,j-1}.
\] (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 make 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

\[
\ln(C_{ij}) = \eta_{ij} + \epsilon_{ij},
\] (2.6)

where \( \eta_{ij} = E(\ln(C_{ij})) \) and \( \epsilon_{ij} \sim N(0, \sigma^2) \). Hence, \( C_{ij} \sim LN(\eta_{ij}, \sigma^2) \), where \( N \) and \( LN \) denote normal and lognormal distributions,
respectively. Moreover, two models,
\[ n_{ij} = c + \alpha_i + \beta_j \]  
and
\[ n_{ij} = c + \alpha_i + \beta_i \ln(j + \gamma_i j), \]  
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 \( \beta_i \) and \( \gamma_i \) 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 \( \alpha_i \) and \( \beta_i \) 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
\[ \eta_{ij} = \alpha_i + \sum_{l=1}^{j-1} \beta_i + \sum_{k=2}^{\infty} \gamma_k \]  
in (2.6). Here the \( \beta_i \)'s and \( \gamma_k \)’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
\[ E(C_{ij}) = m_{ij} \quad \text{and} \quad \text{Var}(C_{ij}) = \phi_j m_{ij}^p \]

\[ \ln(m_{ij}) = \eta_{ij}, \]

where
\[ \eta_{ij} = u_{ij} + c + \alpha_i + \beta_j \ln(j + \gamma_j j + \delta_k), \]  
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), where \( \delta_k \) 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. \( \phi_j = \phi \), and, hence, we will stick to this assumption from now on.

Eq. (2.7) can be extended to include a calendar year parameter according to
\[ \eta_{ij} = c + \alpha_i + \beta_j + \gamma_k, \quad k = 2, \ldots, 2t. \]  

(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.
\[ \alpha_1 = \beta_1 = \gamma_2 = 0 \]  
(2.13)
is needed to estimate the \( v = 3t - 2 \) remaining model parameters \( v \), \( \beta_i \) and \( \gamma_k \), 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 \( \gamma_k \) 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 \( \alpha_i \). 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 \( \beta_i \), 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
\[ \eta = \mathbf{X}_{\text{full}} \boldsymbol{\theta}_{\text{full}} \]  
using \( \boldsymbol{\theta}_{\text{full}} = (c \quad \alpha \quad \beta \quad \gamma)^T \) and \( \eta = (\eta_{11} \ldots \eta_{1t} \eta_{21} \ldots \eta_{2,t-1} \ldots \eta_{t1})^T \), where \( \alpha = (\alpha_2 \ldots \alpha_t) \), \( \beta = (\beta_2 \ldots \beta_t) \), \( \gamma = (\gamma_3 \ldots \gamma_{t+1}) \). Recall that the number of observations in \( \mathcal{V} \) is \( n = t (t + 1)/2 \), which is also the length of \( \eta \). 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 \( v = 3t - 2 \) original parameters \( \alpha_i, \beta_i \) and \( \gamma_k \) and, hence, reparameterize the system (3.1) we introduce a new set of parameters \( \mathbf{a} = (a_1 \ldots a_q) \), \( \mathbf{b} = (b_1 \ldots b_r) \) and \( \mathbf{g} = (g_1 \ldots g_s) \), where \( 0 \leq q, r, s \leq t - 1 \). Let
\[ \mathbf{a} \mathbf{B} = \alpha \]
\[ \mathbf{b} \mathbf{B} = \beta \]
\[ \mathbf{g} \mathbf{I} = \gamma. \]  

(3.2)

which corresponds to
\[ \begin{pmatrix} a_1 & \ldots & a_q \end{pmatrix} \begin{pmatrix} A_{12} & \ldots & A_{1t} \\ \vdots & \ddots & \vdots \\ A_{q2} & \ldots & A_{qt} \end{pmatrix} = \begin{pmatrix} \alpha_2 & \ldots & \alpha_t \end{pmatrix} . \]  

(3.3)

\[ \begin{pmatrix} b_1 & \ldots & b_r \end{pmatrix} \begin{pmatrix} B_{12} & \ldots & B_{1t} \\ \vdots & \ddots & \vdots \\ B_{r2} & \ldots & B_{rt} \end{pmatrix} = \begin{pmatrix} \beta_2 & \ldots & \beta_t \end{pmatrix} . \]  

(3.4)

\[ \begin{pmatrix} g_1 & \ldots & g_s \end{pmatrix} \begin{pmatrix} \Gamma_{13} & \ldots & \Gamma_{1,t+1} \\ \vdots & \ddots & \vdots \\ \Gamma_{s3} & \ldots & \Gamma_{s,t+1} \end{pmatrix} = \begin{pmatrix} \gamma_3 & \ldots & \gamma_{t+1} \end{pmatrix} . \]  

(3.5)

Moreover, let \( \theta = (c \quad \alpha \quad \beta \quad \mathbf{g}) \), containing \( w = 1 + q + r + s \) parameters. We can now express \( \theta_{\text{full}} \) as
\[ \theta_{\text{full}} = \mathbf{D} \theta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \mathbf{A} & 0 & 0 \\ 0 & 0 & \mathbf{B} & 0 \\ 0 & 0 & 0 & \mathbf{I} \end{pmatrix} \theta \]  

(3.6)
using the blocks $A$, $B$, and $\Gamma$ in (3.2)-(3.5). The matrix $D$ is of dimension $r \times w$.

Finally (3.1) can be rewritten using the new parameters
\[ \eta = X_{\text{full}} \hat{\theta}_{\text{full}} = X_{\text{full}} D \theta = X \theta, \]
(3.7)
where $X$ is the new design matrix.

**Example 1 (The Full GLM).** If $q = r = s = t - 1$ and $A = B = \Gamma = I$, where $I$ is the identity matrix, we get the full GLM. However, if we choose other values of $q$, $r$, $s$ and $A$, $B$, $\Gamma$, 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 $\beta = (\beta \quad \gamma \gamma)^T$ and
\[ D = \begin{pmatrix} 1 & 0 & 0 & \cdots & \alpha_l & \beta & \gamma \gamma \end{pmatrix} \]
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:
\[ \alpha_i = a_{i-1}; \quad 2 \leq i \leq q \]
\[ \alpha_i = a_{q-1} + a_q (i-q) ; \quad q + 1 \leq i \leq t \]
\[ \beta_j = b_{j-1}; \quad 2 \leq j \leq r \]
\[ \beta_j = b_{j-1} + b_r (j-r); \quad r + 1 \leq j \leq t \]
\[ \gamma_k = g_{k-1}; \quad 2 \leq k \leq s \]
\[ \gamma_k = g_{s-1} + g_s (k-s); \quad s + 1 \leq k \leq t, \]
where $1 \leq q, r, s \leq t - 1$. 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 $a_k, b_r$ 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 $\beta_j$ beyond $t$ for a tail. It might not make sense to do the same for $\alpha_i$, however, the model could be useful in order to forecast calendar year effects by extrapolation of $\gamma_k$. 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 $\Gamma = 0$ in (3.8). Thus, $D$ will be of size $(2t - 1) \times (t + r)$,
\[ D = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 1 & t-r \end{pmatrix}. \]
(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
\[ f_j = \frac{\sum_{l=1}^{t-j+1} m_{jl}}{\sum_{l=1}^{t-j} m_{jl}}, \]
(3.10)
where $m_{jl}$ is defined by the underlying GLM structure in (2.10), (3.1) and (3.8). Recall that the observed development factors
\[ \hat{f}_j = \frac{\sum_{l=1}^{t-j+1} \hat{m}_{jl}}{\sum_{l=1}^{t-j} \hat{m}_{jl}}, \]
(3.11)
equal the chain-ladder development factors $f_j$ when we let $r = t - 1$ (and $p = 1$ in (2.10)).

Thus,
\[ \ln (f_j - 1) = \ln \left( \sum_{l=1}^{t-j+1} e^{a_l + \beta_l} \right) - 1 = \ln \left( \sum_{l=1}^{t-j+1} e^{\hat{a}_l} \right) - 1 = \beta_{j+1} - \ln \left( \sum_{l=1}^{j} e^{\hat{a}_l} \right). \]
(3.12)
For $j \geq r$ Eq. (3.12) can be written as
\[ \ln (f_j - 1) = b_p + b_{p-1} (j + 1 - q + 1) - \ln \left( \sum_{l=1}^{j} e^{\hat{a}_l} \right). \]
(3.13)
Since $r$ is supposed to be so large that the linear extrapolation captures the tail only small values of $\beta_j$ will remain to the right of $r$. Therefore the linear parametrization of $\beta_j$ 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 $\hat{\theta}_{qs}$ denote the estimated parameter vector for a model with fixed $(q, r, s)$. We can choose the model
\[ (\hat{q}, \hat{r}, \hat{s}) = \arg\min_{(q,r,s)\in I} \text{Crit}(\hat{\theta}_{qs}), \]
(4.1)
that minimizes a model selection criterion $\text{Crit}(\hat{\theta}_{qs})$ among a pre-chosen set $I$ of candidate models. We then take $\hat{\theta}_{(\hat{q},\hat{r},\hat{s})}$ 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

$$I = \{(t - 1, 1, 0), \ldots, (t - 1, t - 1, 0)\}$$  \hspace{1cm} (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

$$I = \{(t - 2, 1, 0), \ldots, (t - 2, t - 1, 0), (t - 1, 1, 0), \ldots, (t - 1, t - 1, 0)\}.$$  \hspace{1cm} (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

$$\text{Crit} = \text{AIC}(\hat{\theta}_{qrs}) = 2w - 2l(\hat{\theta}_{qrs})$$  \hspace{1cm} (4.4)

and

$$\text{Crit} = \text{BIC}(\hat{\theta}_{qrs}) = \ln(n)w - 2l(\hat{\theta}_{qrs}),$$  \hspace{1cm} (4.5)

see e.g. Miller (2002). Here $w = 1 + q + r + s$ is the number of parameters and $l(\hat{\theta}_{qrs})$ 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 $\text{MSEP} (\hat{R}) = E \left( (R - \hat{R})^2 \right)$, where the expected value is with respect to $R$ and $\hat{R}$. Estimating MSEP by bootstrapping, we get a model selection criterion

$$\text{Crit} = \text{MSEP}(\hat{\theta}_{qrs}) = E \left( (R^* - \hat{R}^*)^2 \right),$$  \hspace{1cm} (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 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 $Q(\theta)$ of Wedderburn (1974). Pan (2001) suggested a quasi-likelihood information criterion

$$\text{QIC}(\hat{\theta}_{qrs}) = 2 \text{trace}(\hat{\Sigma} \hat{\Sigma}) - 2Q(\hat{\theta}_{qrs}),$$  \hspace{1cm} (4.7)

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

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 $\text{MSEP}$ 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 $\phi$ 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 $m_{ij} - m_{ij}$, which accounts for systematic effects not captured by the smaller model. Hence, estimating $\phi$ for each model $(t - 1, r, 0)$ yields a higher value compared to estimating $\phi$ 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 $\phi$ 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 $\phi$ for AIC and BIC we use the Pearson residuals

$$r_{ij} = \frac{C_{ij} - \hat{m}_{ij}^\text{full}}{\sqrt{\hat{m}_{ij}^p}},$$  \hspace{1cm} (5.1)

and for MSEP we use

$$r_{ij} = \frac{C_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}^p}},$$  \hspace{1cm} (5.2)

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

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

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

$$\hat{\phi} = \frac{1}{n - w} \sum_{\nu} (r_{ij})^2.$$  \hspace{1cm} (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 $X_{\text{full}}$ 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 \(A\) and \(B\), and hence, the block matrix \(D\).

4. Calculate \(X = X_{\text{adj}} D\).

5. Set up the new GLM \(\eta = X \theta\). Then use some standard software to compute an estimate \(\hat{\theta}_{t-1,r,0}\) of \(\theta = \theta_{t-1,r,0}\) by maximizing e.g. the likelihood or quasi-likelihood.

6. Evaluate the chosen model selection criterion \(\text{Crit}(\hat{\theta}_{t-1,r,0})\) 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 \(M_{\text{SEP}}\) in (4.6).

END.

7. Select model \((t - 1, \hat{r}, 0)\) as in (4.1).

8. Obtain estimators \(\hat{E}(C_{ij}) = \hat{m}_{ij}\) and \(\text{Var}(C_{ij}) = \hat{\phi} \hat{m}_{ij}^2\) from (2.10) and (2.7), with \(c, \alpha\) and \(\beta_i\) replaced by estimates, computed from the first four equations of (3.8) and \(\hat{\theta}_{t-1,i,0}\). 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 \(\hat{\theta}_{t-1,i,0}\) selected by the reserving algorithm in Section 5.2 or for the full model \(\hat{\theta}_{t-1,i,1,0}\) 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 \(F = F(\hat{m}_{ij}, \phi \hat{m}_{ij}^2)\). Typically we assume that \(F\) is either an ODP or a gamma distribution parameterized by the mean \(\hat{\mu}_i\) and variance, so that we may write \(F(\hat{\mu}_i, \phi \hat{\mu}_i^2)\). We analyze the dataset from Taylor and Aske (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 \(VC\) is presented in Table 7.1.

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 Aske (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 \(VC\) is presented in Table 7.1.
Table 7.1
Observations of paid claims \( \nabla C \) from Taylor and Ashe (1983).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>357848</td>
<td>766940</td>
<td>610542</td>
<td>482940</td>
<td>527326</td>
<td>574398</td>
<td>146342</td>
<td>139950</td>
<td>227229</td>
<td>67948</td>
</tr>
<tr>
<td>2</td>
<td>352118</td>
<td>884021</td>
<td>933894</td>
<td>118328</td>
<td>445745</td>
<td>320996</td>
<td>527804</td>
<td>266172</td>
<td>425046</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>396132</td>
<td>937085</td>
<td>847498</td>
<td>805037</td>
<td>504851</td>
<td>470639</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>359480</td>
<td>106164</td>
<td>144337</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>376686</td>
<td>986608</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. The \( \ln(\hat{\phi} - 1) \) 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 MSEP in (4.6) are presented in Table 7.2. Recall that \( \phi \) 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 \( \phi m_{ij}^2 \), i.e.
The ln(\hat{j}_j - 1) curves for the full GLM (\ldots), the smoothed GLM (\ldots) and the smoothed chain-ladder (\ldots) under the assumption of a gamma distribution (p = 2). On the x-axis we have j.

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

<table>
<thead>
<tr>
<th>r</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sqrt{MSEP}) (10^3)</td>
<td>3039</td>
<td>3222</td>
<td>3247</td>
<td>3114</td>
<td>3000</td>
<td>3051</td>
<td>3234</td>
<td>3659</td>
<td>4921</td>
</tr>
<tr>
<td>Deviance (10^3)</td>
<td>1903.0</td>
<td>2073.0</td>
<td>2077.5</td>
<td>2079.2</td>
<td>2108.1</td>
<td>2402.0</td>
<td>2607.2</td>
<td>3161.3</td>
<td>7807.9</td>
</tr>
</tbody>
</table>

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

\[
\ell(\hat{\theta}, \hat{\phi}_{\text{full}}) = \frac{1}{\phi_{\text{full}}} \sum_v \left( -\frac{C_{ij}}{m_{ij}} - \log(m_{ij}) \right) + \sum_v \left( \frac{1}{\phi_{\text{full}}} \log\left(\frac{C_{ij}}{\phi_{\text{full}}}\right) - \log(C_{ij}) - \log\Gamma\left(\frac{1}{\phi}\right) \right).
\]

The values of AIC and BIC are presented in Table 7.3 together with the square root of \(MSEP\) and the deviance.

As can be seen, \(MSEP\) selects model \(r = 5\) or \(r = 9\) for the assumption of an ODP distribution. For the gamma assumption AIC and \(MSEP\) select model \(r = 9\) or \(r = 5\), while BIC selects model \(r = 3\) or \(r = 2\). Hence, for the triangle under analysis AIC and \(MSEP\) 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, \(\sqrt{MSEP}\) in (e) is slightly larger than the standard deviation of the bootstrap samples in (d). The 95

![Fig. 2](image-url)
Table 7.3
The values of AIC, BIC, the square root of $\sqrt{\text{MSEP}}$ in (4.6) and the deviance when a gamma distribution is assumed.

<table>
<thead>
<tr>
<th>$r$</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>1502.3</td>
<td>1508.9</td>
<td>1506.9</td>
<td>1505.0</td>
<td>1503.1</td>
<td>1505.1</td>
<td>1504.6</td>
<td>1508.6</td>
<td>1578.3</td>
</tr>
<tr>
<td>BIC</td>
<td>1540.5</td>
<td>1545.1</td>
<td>1541.1</td>
<td>1537.1</td>
<td>1533.2</td>
<td>1533.2</td>
<td>1530.7</td>
<td>1532.6</td>
<td>1600.4</td>
</tr>
<tr>
<td>$\sqrt{\text{MSEP}}$ ($10^3$)</td>
<td>2736</td>
<td>3047</td>
<td>3012</td>
<td>2944</td>
<td>2915</td>
<td>2974</td>
<td>3024</td>
<td>3160</td>
<td>3733</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Estimated reserve</th>
<th>Bootstrap mean</th>
<th>Bootstrap mean − Est. res.</th>
<th>Bootstrap stand. dev.</th>
<th>$\sqrt{\text{MSEP}}$</th>
<th>95 percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC 18 085 773</td>
<td>17 911 099</td>
<td>−174 674</td>
<td>2735 238</td>
<td>2740 673</td>
<td>22 082 887</td>
</tr>
<tr>
<td>BIC 18 071 392</td>
<td>17 969 537</td>
<td>−101 855</td>
<td>3031 233</td>
<td>3033 233</td>
<td>22 602 603</td>
</tr>
</tbody>
</table>

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 $B = 10,000$.

<table>
<thead>
<tr>
<th>$r$</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>7010</td>
<td>24</td>
<td>85</td>
<td>166</td>
<td>1240</td>
<td>454</td>
<td>801</td>
<td>220</td>
<td>0</td>
</tr>
<tr>
<td>BIC</td>
<td>9</td>
<td>10</td>
<td>32</td>
<td>47</td>
<td>117</td>
<td>368</td>
<td>5394</td>
<td>4023</td>
<td>0</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Year</th>
<th>( r = 9 )</th>
<th>( r = 8 )</th>
<th>( r = 7 )</th>
<th>( r = 6 )</th>
<th>( r = 5 )</th>
<th>( r = 4 )</th>
<th>( r = 3 )</th>
<th>( r = 2 )</th>
<th>( r = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>94634</td>
<td>240187</td>
<td>221536</td>
<td>230156</td>
<td>202906</td>
<td>142453</td>
<td>178698</td>
<td>234891</td>
<td>397438</td>
</tr>
<tr>
<td>3</td>
<td>469511</td>
<td>490001</td>
<td>486650</td>
<td>481624</td>
<td>435577</td>
<td>322911</td>
<td>394173</td>
<td>505214</td>
<td>826224</td>
</tr>
<tr>
<td>4</td>
<td>709638</td>
<td>769423</td>
<td>767511</td>
<td>778890</td>
<td>725379</td>
<td>571929</td>
<td>677310</td>
<td>844415</td>
<td>1330160</td>
</tr>
<tr>
<td>5</td>
<td>984889</td>
<td>1039712</td>
<td>1029537</td>
<td>1029175</td>
<td>992396</td>
<td>840830</td>
<td>961998</td>
<td>1162612</td>
<td>1755176</td>
</tr>
<tr>
<td>6</td>
<td>1419459</td>
<td>1477137</td>
<td>1466432</td>
<td>1471334</td>
<td>1483356</td>
<td>1373765</td>
<td>1508001</td>
<td>1756918</td>
<td>252484</td>
</tr>
<tr>
<td>7</td>
<td>2177641</td>
<td>2241521</td>
<td>2229665</td>
<td>2237087</td>
<td>2208130</td>
<td>2310842</td>
<td>2399913</td>
<td>2667064</td>
<td>3585818</td>
</tr>
<tr>
<td>8</td>
<td>3920301</td>
<td>3996865</td>
<td>3982652</td>
<td>3991552</td>
<td>3956845</td>
<td>3864518</td>
<td>3655170</td>
<td>3776692</td>
<td>4609461</td>
</tr>
<tr>
<td>9</td>
<td>4278972</td>
<td>4342643</td>
<td>4330826</td>
<td>4338225</td>
<td>4309362</td>
<td>4232583</td>
<td>4278865</td>
<td>3735382</td>
<td>3778396</td>
</tr>
<tr>
<td>10</td>
<td>4625811</td>
<td>4681894</td>
<td>4671485</td>
<td>4678001</td>
<td>4652579</td>
<td>4584950</td>
<td>4625716</td>
<td>4690075</td>
<td>2135908</td>
</tr>
<tr>
<td>Total</td>
<td>18680856</td>
<td>19279383</td>
<td>1916297</td>
<td>19237844</td>
<td>18966529</td>
<td>1824781</td>
<td>1867943</td>
<td>19373942</td>
<td>2096067</td>
</tr>
</tbody>
</table>

### Table A.2
Bootstrap results for the total under the assumption of a gamma distribution.

<table>
<thead>
<tr>
<th>Year</th>
<th>( r = 9 )</th>
<th>( r = 8 )</th>
<th>( r = 7 )</th>
<th>( r = 6 )</th>
<th>( r = 5 )</th>
<th>( r = 4 )</th>
<th>( r = 3 )</th>
<th>( r = 2 )</th>
<th>( r = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>93316</td>
<td>202152</td>
<td>203511</td>
<td>207892</td>
<td>199638</td>
<td>172114</td>
<td>184757</td>
<td>211549</td>
<td>309558</td>
</tr>
<tr>
<td>3</td>
<td>464505</td>
<td>408480</td>
<td>409958</td>
<td>416799</td>
<td>404635</td>
<td>351964</td>
<td>376550</td>
<td>429348</td>
<td>639118</td>
</tr>
<tr>
<td>4</td>
<td>611145</td>
<td>629353</td>
<td>629086</td>
<td>632612</td>
<td>618547</td>
<td>574367</td>
<td>611511</td>
<td>688582</td>
<td>1018712</td>
</tr>
<tr>
<td>5</td>
<td>992023</td>
<td>100479</td>
<td>1008970</td>
<td>1003960</td>
<td>996081</td>
<td>934022</td>
<td>967584</td>
<td>1070345</td>
<td>1415607</td>
</tr>
<tr>
<td>6</td>
<td>1453085</td>
<td>1470922</td>
<td>1471480</td>
<td>1473661</td>
<td>1497377</td>
<td>1481516</td>
<td>1511006</td>
<td>1616061</td>
<td>2080145</td>
</tr>
<tr>
<td>7</td>
<td>2186161</td>
<td>2208097</td>
<td>2210383</td>
<td>2212695</td>
<td>2205657</td>
<td>2416362</td>
<td>238537</td>
<td>2497200</td>
<td>3015061</td>
</tr>
<tr>
<td>8</td>
<td>3656065</td>
<td>3687154</td>
<td>3689803</td>
<td>3684859</td>
<td>3651627</td>
<td>3402313</td>
<td>3351061</td>
<td>3906657</td>
<td>3906657</td>
</tr>
<tr>
<td>9</td>
<td>4122398</td>
<td>4139309</td>
<td>4139332</td>
<td>4141344</td>
<td>4136502</td>
<td>4107371</td>
<td>4118985</td>
<td>3557488</td>
<td>348913</td>
</tr>
<tr>
<td>10</td>
<td>4516073</td>
<td>4532001</td>
<td>4532447</td>
<td>4532966</td>
<td>4528998</td>
<td>4502114</td>
<td>4512328</td>
<td>4524778</td>
<td>1755546</td>
</tr>
<tr>
<td>Total</td>
<td>18085773</td>
<td>18287657</td>
<td>18293470</td>
<td>18317784</td>
<td>18293470</td>
<td>18293470</td>
<td>18293470</td>
<td>18293470</td>
<td>18293470</td>
</tr>
</tbody>
</table>

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

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

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 \( \phi \) 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


References


