Projection for Claims Triangles by Affine Age-to-Age Development

by Thomas Müller

ABSTRACT

Actuaries have always had the impression that the chain-ladder reserving method applied to real data has some kind of "upward" bias. This bias will be explained by the newly reported claims (true IBNR) and taken into account with an additive part in the age-to-age development. The multiplicative part in the development is understood to be restricted to the changes in the already reported claims (IBNER, "incurred but not enough reserved"). Based on regression theory the reserve as well as error formulae are generalized from the purely multiplicative chain-ladder model to our considerably more stable "affine" models.

KEYWORDS

Claims reserving, chain-ladder method, IBNR and IBNER, linear regression, standard error, weighted least squares estimators, chain-ladder bias

1. Multiplicative and additive reserving methods

Halliwell (2007) noticed that the age-to-age development on incurred real data mostly yields a regression line with a positive intercept, which he comprehends as a "bias" of the chain-ladder projection. In this paper, we propose that this positive intercept is caused by newly reported claims, i.e., claims which were not reported in the previous year, the so-called IBNYR claims (incurred but not *yet* reported). These newly reported claims amounts, also denoted as "true IBNR," are usually considered together with the changes in the amounts of the *already* reported claims, denoted as IBNER (incurred but not *enough* reported claims). It seems obvious that the amounts of newly reported claims depend on the volume of the business, e.g., the premium, rather than on the previously reported claims. The volume of the considered business is usually rather stable, at least compared with the reported claims. Assuming this volume to be constant, the observed positive intercept may well be considered an estimate of the average amount of newly reported claims, because within the database, i.e., the loss triangles of incurred claims, the newly reported claims amounts are, as already mentioned, commonly summed up with the newly estimated amounts of the previously reported claims. Therefore, an affine model, i.e., a model with an additive as well as a multiplicative part for the age-to-age development, better corresponds to the real circumstances of this development than, for instance, the purely multiplicative chain-ladder model.

In general, one distinguishes between additive or multiplicative reserving models: additive models refer the estimated claims development to the volume of the considered business, which in practice mostly is considered to be proportional to the premium volume, whereas multiplicative methods refer to the reported claims amounts.

Multiplicative methods usually are much more unstable, in particular if claims are reported rather late, such that for some smaller portfolios there may in fact be no reported claims at all for the latest accident years, as is the case in the example of Brosius (1993). In this case, the multiplicative chain-ladder model yields, for these accident years, the amount of zero as estimation for all development years. Positive reserve estimation demands some reported claims; therefore, the estimated reserve may dramatically change as soon as some claims are reported. In the Brosius case, the prediction error for the chain-ladder model even becomes infinitely large: indeed, the multiplicative model cannot cope with the age-to-age development of zero claims amounts into any positive claims amounts, and the usual formulae for the prediction error are not defined in this case. However, for a series of preceding-year claims amounts converging to zero with a given positive next-year claims amount, the corresponding prediction errors will be arbitrarily large. This can be interpreted as an infinite prediction error for the chain-ladder model.

Compared to multiplicative methods, additive methods are commonly much more stable and may also be applied in the case of triangles with a rather unsmooth, inhomogeneous pattern such as the Brosius example. As mentioned above, this is essentially due to the fact that the premium volume for the different accident years is much more stable than the reported claims amounts. With the proposed affine method, which considers a multiplicative as well as an additive relation, one gains the advantages of both model types: affine models are more stable than the chain-ladder models and may therefore also be applied to more fragmented triangle patterns. Moreover, some highly developed theoretical tools such as the famous Mack formulae for the prediction error of chain-ladder reserve estimates may be generalized to the proposed affine model. And, above all, the affine age-to-age development corresponds rather well with the reality of the data collected. Finally, the additive part of the affine models enables us to explain the aforementioned upward bias of the chain-ladder projection. By ignoring this bias and by using purely multiplicative methods, one overestimates the leverage of the reported claims on the estimated claims reserve instead of considering a supplementary additive component depending on a constant, the development of premiums, or any other appropriate notion of exposure.

In this paper, we propose various Gauss-Markov predictors for the reserves, including the chain-ladder model. Our models may be understood as generalizations of the chain-ladder model, rendering reserve estimation more stable in practice, especially for recent accident years with little or no experience. We develop these predictors as well as their standard error within one single framework known from multivariate statistics and obtain the prediction error derived by Mack (1993) for the chain-ladder model. Although, as in the well-known purely multiplicative chain-ladder method, no a priori assumptions such as a presumed claim ratio are required, these affine methods are considerably more stable with regard to fluctuations in recent accident years.

Venter and Zehnwirth (1998) and Barnett and Zehnwirth (2000) proposed similar affine methods to the ones discussed in this paper. They realized the importance of the additive component in comparison with pure chain-ladder projections by analyzing real incurred data taken from several business sectors. Ludwig and Schmidt (2010) suggested several Gauss-Markov predictors, with their so-called "combined model" integrating additive and multiplicative components as we do here with our affine model. For this reason some technical aspects may seem similar to Ludwig and Schmidt (2010). However, there is a fundamental difference between these approaches. The combined models described in Ludwig and Schmidt (2010) do not specifically include the chain-ladder method, because their multiplicative age-to-age development always depends on the claims in the first year and not—as in the chain-ladder as well as in our models—on the actual previous year.

2. Model structure

2.1. Age-to-age development

Let the vectors X_j and X'_{j+1} denote the incurred claims of development year j and $j+1$, respectively, for the accident years 1 to $n-j$, and let v_j be a volume function, e.g., the written premium, in the corresponding accident year 1 to $n-j$,

$$X_j = \begin{pmatrix} X_{1,j} \\ \vdots \\ X_{n-j,j} \end{pmatrix}, X'_{j+1} = \begin{pmatrix} X_{1,j+1} \\ \vdots \\ X_{n-j,j+1} \end{pmatrix}, v_j = \begin{pmatrix} V_1 \\ \vdots \\ V_{n-j} \end{pmatrix}.$$

We assume the volume function v_j to be given and we consider two different types of age-to-age developments. In the first case, the development only depends on the incurred claims of the previous year, whereas in the second case it simultaneously depends on both previous-year claims and the volume function. For the description of the latter case, we introduce the matrices

$$X_j^* = \begin{pmatrix} V_1 & X_{1,j} \\ \vdots & \vdots \\ V_{n-j} & X_{n-j,j} \end{pmatrix},$$ composed of previous-year claims and the volume function.

Let the age-to-age development be defined by a deterministic part f_j^* and a stochastic part ε_j, where

$$f_j^* = \begin{pmatrix} c_j \\ f_j \end{pmatrix}$$ is the deterministic part and

$$\varepsilon_j = \begin{pmatrix} e_{1j} \\ \vdots \\ e_{n-j\,j} \end{pmatrix} \text{ with } \begin{pmatrix} E[e_{1j}] \\ \vdots \\ E[e_{n-j\,j}] \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$ is the random part.

Depending on the different assumptions for the relevant parameters, we will consider three different types of models for the age-to-age development:

First, we have $X'_{j+1} = X_j^* \cdot f_j^* + \varepsilon_j$, the *affine model,* with the two development parameters of f_j^*, where the additive parameter c_j defines a development proportional to the volume of accident year j. The parameter f_j describes a development proportional to the previous-year claims and is therefore conceived as a multiplicative component of the age-to-age development.

Within the theory of general linear models, f_j^* is called design matrix. In the affine model, the design matrix has two parameters, the multiplicative f_j and the additive c_j.

In addition, we also consider the cases with only one parameter:

$X'_{j+1} = X_j \cdot f_j + \varepsilon_j$, the *multiplicative model,* and
$X'_{j+1} = V_j \cdot c_j + \varepsilon_j$, the *additive model,* hereinafter referred to as "incremental loss ratio method."

Since the expected values of the random part are supposed to be 0, the non-random part defines the age-to-age development of the conditional expected value of claims:

$E[X'_{j+1} \mid X_j] = E[X_j^*] \cdot f_j^*$	affine model
$E[X'_{j+1} \mid X_j] = E[X_j] \cdot f_j$	multiplicative model
$E[X'_{j+1} \mid X_j] = v_j \cdot c_j$	incremental loss ratio method

2.2. The random part of the age-to-age development

The random part ε_j defines the covariance matrix, which is composed of a scalar component and a matrix W specifying the structure,

$$\operatorname{cov}(\varepsilon_j \cdot \varepsilon_j^t) = \sigma_j^2 \cdot W = \sigma_j^2 \cdot \begin{pmatrix} W_{11} & \cdots & W_{1\ n-j} \\ \vdots & & \vdots \\ W_{n-j\,1} & \cdots & W_{n-j\ n-j} \end{pmatrix}.$$

Here, we only consider diagonal matrices W, i.e.,

$$\begin{pmatrix} W_{11} & \cdots & W_{1\ n-j} \\ \vdots & & \vdots \\ W_{n-j\,1} & \cdots & W_{n-j\ n-j} \end{pmatrix} = \begin{pmatrix} W_{11} & 0 & \cdots & 0 \\ 0 & W_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{n-j\ n-j} \end{pmatrix}.$$

In addition, we assume two different types of diagonals:

– the constant diagonal with $W_{ii} = 1$, meaning that $W = I_{n-j}$, i.e., the unit matrix of size $n - j$, and
– the diagonal proportional to the previous-year claims $W_{ii} = X_{i,j}$, i.e., the claims of development year j, for accident years $1 \le i \le j$.

The diagonality of the matrix W means that the claims developments of different development years are independent.

The variance of the conditional random variable of the claims in development year j given the previous-year amount will then lead to

$Var[X'_{j+1} \mid X_j] = \sigma_j^2$, if W has a constant diagonal and
$Var[X'_{j+1} \mid X_j] = \sigma_j^2 \cdot X_j$, if W has a diagonal proportional to the previous-year claims.

The different assumptions about the covariance matrix correspond to different risk-measure models and result in different estimations even beyond the non-random age-to-age development parameters. With the diagonal of W being constant, all variations of the subsequent year will be valued equally. Alternatively, if the diagonal is proportional to previous-year claims amounts, variations in the subsequent year are more probable for higher previous-year claims, and these data have therefore less weight than the data for smaller claims amounts. This is justified by risk-theoretical reflections. Higher previous-year claims are expected to vary more in their subsequent year development than smaller ones. By assuming that the variance is proportional to the previous-year amount itself and not to the square of it, one even takes a diversification effect into account. This diversification is based on the assumption that higher claims amounts for a given accident year are expected to be composed of a higher number of single claims. Assuming the single claims composing the claims amounts to be equally distributed for all accident years, the number of single claims is expected to be proportional to the claims amounts. In most common models for the distribution of the number of claims, such as in the

Poisson distribution, the variance is proportional to the expected value itself. This justifies the assumption that the variance in the age-to-age development is proportional to the previous-year claims.

2.3. The "memorylessness" assumption for the age-to-age development process

We assume that the age-to-age development process is "memoryless" in the following sense: the probability distribution of the claims development at age $j+1$, $X_{k,j+1}$, only depends on the claims amounts at the preceding age j, $X_{k,j}$, and does not depend on former claims developments $X_{k,i}$ at age i, $i < j$. Within the additive and affine models, $X_{k,j+1}$ may of course also depend on the volume V_k.

Typical examples of such "memorylessness" are Markov processes, which for instance may be applied to construct life contingencies. We need this assumption in order to be able to link in a feasible way the insights into the single-year developments with the required multi-year development and thus to complete the triangle for the entire period. The memorylessness assumption ensures that the expected value and the prediction error of the ultimate claims development $X_{k,n}$ behave as we would expect them to do:

- The expected value of $X_{k,n}$ corresponds to the iterated application of the estimated affine age-to-age development from $X_{k,n-k+1}$, the latest of the known claims amounts $X_{k,j}$, $j \le n-k+1$.
- The prediction error $\hat{M}SEP$ of the estimated reserve corresponds to the sum of the prediction errors $\hat{M}SEP_j$ of the estimated claims amounts of the development year $j+1$, given the claims amounts at development year j, multiplied by $\hat{f}^2_{j:n}$, where $\hat{f}_{j:n} = \prod_{k=j+1}^{n-1} \hat{f}_k$ denotes the product of the remaining estimated multiplicative projection factors from year $j+1$ up to year n. Thus the prediction error for the entire development corresponds to the sum of the prediction errors of the age-to-age developments, scaled up to the level of the claims amounts of the final development year.

Thus, this paper discusses two rather different topics:

- First, the age-to-age development based on the widely used techniques of multivariate statistics. These methods are very useful to gain all kinds of insights into some given data; they rely on a powerful mathematical theory and may be expressed in a concise way using matrix calculus. Up to and including Section 6, only this one-year case will be considered.
- Second, we need to link these age-to-age developments to the required multi-year development. The kind of reflection used here is much less common than the multivariate statistics discussed under the first topic. As mentioned before, these considerations essentially only prove that this linking may be done in a rather obvious way. Nevertheless, some subtle reflections based among others on the law of total expectations as introduced by Mack (1993) are needed. Sections 7 and 8 study this conclusion from the one-year to the multi-year case.

3. Consideration of the different models

We will not treat all six possible models in detail in this paper. As an additive development we will only consider the case where the constant variance assumption holds, i.e., $W_{ii} = 1$, and as a multiplicative development the chain-ladder case with the proportional assumption, i.e., $W_{ii} = X_{i,j}$. Instead we will look more closely at the affine models. In doing so, we will also consider affine models with constant volume, e.g., $V \equiv 1$ for all accident years. This model is particularly interesting in case no volume function is available within the data. In such a case, we strongly recommend comparing the affine models with a constant volume assumption to the results of a purely multiplicative method, as chain ladder is, especially if no volume function is available. If the results of these two models differ significantly, the multiplicative method—that is, the chain-ladder method—might not be appropriate for the given triangles. Table 1 shows the main assumptions in the different models.

Table 1. Main assumptions in the different models

Model	# of Parameters in the Design Matrix 2, multiplicative and additive (affine), or only one of them	Volume Function i.e., the reference for the additive parameter	Risk Assumption/ Risk Measure i.e., the diagonal W_{ii} of the diagonal matrix W
generalized chain ladder	2, affine model	general	proportional to X, $W_{ii} = X_{i,j}$
generalized chain ladder with constant volume	2, affine model	special (constant e.g., $V \equiv 1$)	proportional to X, $W_{ii} = X_{i,j}$
chain ladder	1, only multiplicative one	none	proportional to X, $W_{ii} = X_{i,j}$
generalized linear regression	2, affine model	general	constant, $W_{ii} = 1$
linear regression with intercept	2, affine model	special (constant e.g., $V \equiv 1$)	constant, $W_{ii} = 1$
incremental loss ratio method	1, only additive one	general	constant, $W_{ii} = 1$

4. The weighted least squares estimators of the parameters $\hat{f}_j^*$

Usually, one distinguishes between two cases. In either case, W is a diagonal matrix, i.e., all off-diagonal entries of W are 0. In the special case corresponding to our constant-risk-measure assumption, all diagonal elements are equal to 1, whereas in the general case, there is no such restriction.

First of all, the special case is treated within regression analysis and the parameters are estimated by the least squares estimator. Then the theorem of Gauss-Markov shows that these estimators are the best linear unbiased estimators (BLUE). The special case presuming the diagonal entries of W to be constant is called the homoscedasticity model, as opposed to the more general heteroskedastic model with different diagonal entries. The diagonal entries of the heteroskedastic model can be understood as different weights. Therefore, one takes into account these different weights $W_{ii}^{-1} = X_{i,j}^{-1}$ within the least squares estimators and evaluates a weighted least square estimator. Hence the general heteroskedastic case may be reduced to the special case, meaning that in the general case the theorem of Gauss-Markov still holds. This is why the weighted least squares estimator is also the best linear unbiased estimator.

Let us introduce the notation $\hat{X}'_{j+1} = X_j^* \cdot \hat{f}_j^*$ for the estimated claims amounts of the subsequent year. We seek $\hat{f}_j^*$ minimizing the weighted squared differences WSQ between the estimated and the observed claims development. WSQ may be expressed by matrix calculation, taking into account the weights by the inverse matrix of W. Thus the components with high diagonal entries $W_{ii} = X_{i,j}$ will have less impact on parameter estimation than those with low diagonal entries:

$$\begin{aligned}\text{WSQ} &= \left(\hat{X}'_{j+1} - X'_{j+1}\right)^t W^{-1}\left(\hat{X}'_{j+1} - X'_{j+1}\right)\\ &= \left(X_j^* \cdot \hat{f}_j^* - X'_{j+1}\right)^t W^{-1}\left(X_j^* \cdot \hat{f}_j^* - X'_{j+1}\right).\end{aligned}$$

We now determine $\hat{f}_j^*$ such that WSQ becomes minimal. Hence the two partial derivatives of WSQ with respect to the two parameters must be set equal to zero. Since WSQ is a quadratic expression, the derivatives have a linear and a constant term, which, in matrix terms, leads to the following equation:

$$\left(X_j^*\right)^t \cdot W^{-1} \cdot X_j^* \cdot \hat{f}_j^* = \left(X_j^*\right)^t \cdot W^{-1} X'_{j+1}.$$

This results in the weighted least squares estimators

$$\hat{f}_j^* = \left(\left(X_j^*\right)^t \cdot W^{-1} \cdot X_j^*\right)^{-1} \cdot \left(X_j^*\right)^t \cdot W^{-1} \cdot X'_{j+1},$$

where the matrices in question are presumed to be invertible, as is assumed for the remainder of this paper, even if not specifically mentioned. The relations used in this article are mostly well known in multivariate statistics and have been described in textbooks

(Fahrmeir, Hammerle and Tutz 1996, for instance, or Halliwell 2007).

A table with the specific formulae for the weighted least squares estimators for the parameters in the different models, as well as the derivation of these formulae, can be found in Appendix A.

5. The error of the weighted least squares estimator

In Sections 5 and 6, as in the two previous sections, we only consider one single development year, i.e., the development from year j to year $j+1$. We suppose the claims up to the development year j to be known and therefore consider the conditional probabilities given the history

$$H_j = \{X_{i,l}, i+l \le n+1, l \le j\}.$$

The volumes V_j are not relevant for the history considered, since they are not regarded as stochastic variables in this model. Since in these sections we always study conditional probabilities given the history H_j, this may be omitted to improve readability, especially for the matrix formulae.

To simplify the notation even further, we set $A = ((X_j^*)^t \cdot W^{-1} \cdot X_j^*)^{-1}$. Then

$$\begin{aligned}
\operatorname{cov}\left(\hat{f}_j^*, \hat{f}_j^*\right) &= \operatorname{cov}\left(A\left(X_j^*\right)^t W^{-1}\varepsilon_j, A\left(X_j^*\right)^t W^{-1}\varepsilon_j\right) \\
&= \operatorname{cov}\left(A\left(X_j^*\right)^t W^{-1}\varepsilon_j \left(A\left(X_j^*\right)^t W^{-1}\varepsilon_j\right)^t\right) \\
&= \operatorname{cov}\left(A \cdot \left(X_j^*\right)^t W^{-1}\varepsilon_j \cdot \varepsilon_j^t \left(W^{-1}\right)^t \cdot X_j^* A^t\right) \\
&= \sigma^2 \cdot A \cdot \left(X_j^*\right)^t \left(W^{-1}\right)^t \cdot X_j^* \cdot A^t \\
&= \sigma^2 \cdot A \cdot \left(A^{-1}\right)^t \cdot A^t = \sigma^2 \cdot A \\
&= \sigma^2 \cdot \left(\left(X_j^*\right)^t \cdot W^{-1} \cdot X_j^*\right)^{-1}.
\end{aligned}$$

We are now interested in the impact of the parameter estimation error on the estimation of the age-to-age claims development. This development starts with the sum $\sum_{k=n-j+1}^{n} \hat{X}_{k,j}$ of the estimated previous-year claims for all accident years which have not yet been developed and therefore have to be estimated. This sum corresponds to the sum of the rows beyond the diagonal and includes the diagonal itself, i.e., $\hat{X}_{n-j+1,j} = X_{n-j+1,j}$ because these claims are already included in the data.

$$\text{We set } Z_j = j \cdot \begin{pmatrix} \overline{\hat{V}_j} \\ \overline{\hat{X}_j} \end{pmatrix} = \begin{pmatrix} \sum_{k=n-j+1}^{n} V_j \\ X_{n-j+1,j} + \sum_{k=n-j+2}^{n} \hat{X}_{k,j} \end{pmatrix} = \begin{pmatrix} \sum_{k=n-j+1}^{n} V_j \\ \sum_{k=n-j+1}^{n} \hat{X}_{k,j} \end{pmatrix},$$

thus defining Z_j, $\overline{\hat{V}_j}$ and $\overline{\hat{X}_j}$. The notation $\overline{\hat{V}_j}$ with the hat may be misinterpreted: we do not estimate new volumes here; this notation merely suggests that we take the average over the remaining accident years for the claims in $\overline{\hat{X}_j}$.

The so-called parameter error

$$Err_j^{Param} = E\left(\sum_{k=n-j+1}^{n} \hat{X}_{k,j+1} - E\left[\sum_{k=n-j+1}^{n} X_{k,j+1}\right]\right)^2$$

may be estimated using the theory of generalized linear models. This estimator is calculable as a product of the matrices in question, that is,

$$\begin{aligned}
\hat{Err}_j^{Param} &= \operatorname{cov}\left(Z_j \cdot \hat{f}_j^*, Z_j \cdot \hat{f}_j^*\right) \\
&= Z_j^t \cdot \operatorname{cov}\left(\hat{f}_j^*, \hat{f}_j^*\right) \cdot Z_j \\
&= \hat{\sigma}_j^2 \cdot Z_j^t \cdot \left(\left(X_j^*\right)^t \cdot W^{-1} \cdot X_j^*\right)^{-1} \cdot Z_j.
\end{aligned}$$

The process variance describes the randomness of the process itself, and here we look directly at the effect of this process randomness on the development of the sum of claims of the accident years not yet given in the claims triangle. Therefore, we capture the effect of the process randomness of the development from year j to $j+1$ on the reserve

$$Var_j^{Process} = E\left[\left(\sum_{k=n-j+1}^{n} X_{k,j+1} - E\left[\sum_{k=n-j+1}^{n} X_{k,j+1}\right]\right)^2\right]$$

with the estimation

$$\hat{Var}_j^{Process} = \hat{\sigma}_j^2 \cdot \begin{cases} j & \text{for constant risk measure} \\ j \cdot \overline{\hat{X}_j} & \text{for proportional risk measure} \end{cases},$$

since the above sum is composed of the claims of j accident years which are supposed to be independent as the covariance matrix is diagonal. The history H_j up to development year j defines the estimates $j \cdot \overline{\hat{X}_j}$.

Putting both together, we obtain the mean squared error of the predicted sum of claims for the development from year j to year $j+1$, given the history H_j defining again the required estimates $j \cdot \overline{\hat{X}_j}$:

$$\begin{aligned} MSEP_j &= E\left[\left(\sum_{k=n-j+1}^{n} \hat{X}_{k,j+1} - \sum_{k=n-j+1}^{n} X_{k,j+1}\right)^2\right] \\ &= Var_j^{Process} + Err_j^{Param} \end{aligned}$$

with estimator

$$\hat{MSEP}_j = \hat{Var}_j^{Process} + \hat{Err}_j^{Param} = \hat{\tau}_j \cdot \hat{\sigma}_j^2, \text{ where}$$

$$\hat{\tau}_j = j \cdot \left\{\begin{matrix} 1 & \text{for constant risk measure} \\ \overline{\hat{X}_j} & \text{for proportional risk measure} \end{matrix}\right\} + j^2 \cdot \begin{pmatrix} \overline{\hat{V}_j} \\ \overline{\hat{X}_j} \end{pmatrix}^t \cdot \left(\left(X_j^*\right)^t \cdot W^{-1} \cdot X_j^*\right)^{-1} \cdot \begin{pmatrix} \overline{\hat{V}_j} \\ \overline{\hat{X}_j} \end{pmatrix}.$$

Note that the last factor $\hat{\tau}_{n-1}$ is generally not defined in the models with two parameters. In the numerical examples below, we therefore set $\hat{\tau}_{n-1} = \hat{\tau}_{n-2}^2/\hat{\tau}_{n-3}$.

A table with the specific formulae for the prediction error within the different models, as well as the derivation of these formulae, can be found in Appendix B. The well-known calculation method for the estimator $\hat{\sigma}_j^2$ of the process randomness is described in Appendix D.

6. A simple calculation method provided by standard spreadsheet applications

Estimates based on linear regression are much more common and accessible in computational tools such as spreadsheets. Hence practical calculations may be facilitated if it is possible to link the computations in the chain-ladder models to those in the linear regression models. In fact, this can be achieved by a simple transformation of the data, which we here regard as coordinates. Thus the previous-year claims amounts and the volume function represent the independent variables and are interpreted as x-coordinates, whereas the next-year claims amounts represent the dependent variables viewed as y-coordinates. More specifically, the data in the generalized chain-ladder model,

$$X_j^* = \begin{pmatrix} V_1 & X_{1,j} \\ \vdots & \vdots \\ V_{n-j} & X_{n-j,j} \end{pmatrix} \text{ and } X'_{j+1} = \begin{pmatrix} X_{1,j+1} \\ \vdots \\ X_{n-j,j+1} \end{pmatrix},$$

are transformed, by dividing each row by $X_{k,j}^{1/2}$, to the independent variables

$$\tilde{X}_j^* = \begin{pmatrix} \tilde{V}_1 & \tilde{X}_{1,j} \\ \vdots & \vdots \\ \tilde{V}_{n-j} & \tilde{X}_{n-j,j} \end{pmatrix} = \begin{pmatrix} V_1/X_{1,j}^{1/2} & X_{1,j}^{1/2} \\ \vdots & \vdots \\ V_{n-j}/X_{n-j,j}^{1/2} & X_{n-j,j}^{1/2} \end{pmatrix},$$

considered as x-coordinates in the linear regression model. By analogy, in the traditional chain-ladder model without an additive component, we get

$$\tilde{X}_j = \begin{pmatrix} \tilde{X}_{1,j} \\ \vdots \\ \tilde{X}_{n-j,j} \end{pmatrix} = \begin{pmatrix} X_{1,j}^{1/2} \\ \vdots \\ X_{n-j,j}^{1/2} \end{pmatrix}.$$

For both the generalized and the traditional chain-ladder models, the dependent variables then are

$$\tilde{X}'_{j+1} = \begin{pmatrix} X_{1,j}^{1/2} \cdot X_{1,j+1}/X_{1,j} \\ \vdots \\ X_{n-j,j}^{1/2} \cdot X_{n-j,j+1}/X_{n-j,j} \end{pmatrix} = \begin{pmatrix} X_{1,j+1}/X_{1,j}^{1/2} \\ \vdots \\ X_{n-j,j+1}/X_{n-j,j}^{1/2} \end{pmatrix},$$

which are now considered as y-coordinates in the linear regression model.

The weighted squared differences WSQ in the generalized chain-ladder (CL) model equal the WSQ in the linear regression (LR) model based on the transformed coordinates. Therefore, the estimation problems in the chain-ladder models may be reduced to a linear regression problem and solved accordingly by the multiple tools available:

$$\begin{aligned} &\mathrm{WSQ}\left(X_j^*, X'_{j+1}, W(CL)\right) \\ &= \left(X_j^* \cdot \hat{f}_j^* - X'_{j+1}\right)^t W(CL)^{-1} \left(X_j^* \cdot \hat{f}_j^* - X'_{j+1}\right) \\ &= \left(X_j^* \cdot \hat{f}_j^* - X'_{j+1}\right)^t \begin{pmatrix} X_{1,j}^{-1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & X_{n-j,n-j}^{-1} \end{pmatrix} \left(X_j^* \cdot \hat{f}_j^* - X'_{j+1}\right) \\ &= \left(\tilde{X}_j^* \cdot \hat{f}_j^* - \tilde{X}'_{j+1}\right)^t \begin{pmatrix} 1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & 1 \end{pmatrix} \left(\tilde{X}_j^* \cdot \hat{f}_j^* - \tilde{X}'_{j+1}\right) \\ &= \mathrm{WSQ}\left(\tilde{X}_j^*, \tilde{X}'_{j+1}, W(LR) = I_{n-j}\right). \end{aligned}$$

Panning (2005) proposed to introduce additional "dummy" variables: primarily you augment the dependent variable with a zero such that you have the same number of rows as there are independent variables, including the last row with the x-coordinates you wish to estimate. Thereafter you introduce a "dummy" column with all entries zero except the last one, which will be set to −1. This leads to the following matrices and vectors, respectively, for our main models:

Model	x-coordinate (independent variable)	y-coordinate (dependent variable)
generalized linear regression	$\begin{pmatrix} V_1 & X_{1,j} & 0 \\ \vdots & \vdots & \vdots \\ V_{n-j} & X_{n-j,j} & 0 \\ j \cdot \overline{\hat{V}_j} & j \cdot \overline{\hat{X}_j} & -1 \end{pmatrix}$	$\begin{pmatrix} X_{1,j+1} \\ \vdots \\ X_{n-j,j+1} \\ 0 \end{pmatrix}$
	Transformed x-coordinates, to be used in the linear regression tool:	Transformed y-coordinates, to be used in the linear regression tool:
chain ladder	$\begin{pmatrix} \tilde{X}_{1,j} & 0 \\ \vdots & \vdots \\ \tilde{X}_{n-j,j} & 0 \\ j \cdot \overline{\hat{\tilde{X}}} & -1 \end{pmatrix} = \begin{pmatrix} X_{1,j}^{1/2} & 0 \\ \vdots & \vdots \\ X_{n-j,j}^{1/2} & 0 \\ \left(j \cdot \overline{\hat{X}_j}\right)^{1/2} & -1 \end{pmatrix}$	$\begin{pmatrix} X_{1,j}^{1/2} \cdot X_{1,j+1}/X_{1,j} \\ \vdots \\ X_{n-j,j}^{1/2} \cdot X_{n-j,j+1}/X_{n-j,j} \\ 0 \end{pmatrix}$

Transformed *x*-coordinates, to be used in the linear regression tool:	Transformed *y*-coordinates, to be used in the linear regression tool:
generalized chain ladder	
$\begin{pmatrix} \tilde{V}_1 & \tilde{X}_{1,j} & 0 \\ \vdots & \vdots & \vdots \\ \tilde{V}_{n-j} & \tilde{X}_{n-j,j} & 0 \\ j\cdot\overline{\hat{\tilde{V}}_j} & j\cdot\overline{\hat{\tilde{X}}_j} & -1 \end{pmatrix} = \begin{pmatrix} V_1/X_{1,j}^{1/2} & X_{1,j}^{1/2} & 0 \\ \vdots & \vdots & \vdots \\ V_{n-j}/X_{n-j,j}^{1/2} & X_{n-j,j}^{1/2} & 0 \\ j\cdot\overline{\hat{V}_j}/\left(j\cdot\overline{\hat{X}_j}\right)^{1/2} & \left(j\cdot\overline{\hat{X}_j}\right)^{1/2} & -1 \end{pmatrix}$	as above for the chain-ladder model

Panning (2005) introduced a "dummy" column for each accident year $k = n - j + 1$ up to n, thus estimating each accident year separately. Here, we are only interested in the prediction error of the sum of the claims amounts at age $j + 1$ to catch the entire prediction error $\hat{M}SEP_j$ of the development from age j to age $j + 1$. The "LINEST" function in Microsoft Excel recommended by Panning (2005) calculates the linear regression coefficients as well as their standard error, and therefore this function particularly lends itself to computing the prediction error $\hat{M}SEP_j$. Please note that the "LINEST" function arranges the columns in reverse order (cf. the remarks in Panning (2005) on this subject). Placing the regression coefficients on the first row and their standard error on the second row leads to the matrix

$$\begin{pmatrix} \hat{c}_j & \hat{f}_j & \sum_{k=n-j+1}^{n} \dfrac{\hat{X}_{k,j+1}}{\left(j\cdot\overline{\hat{X}_j}\right)^{1/2}} \\ \hat{E}rr^{Par}\left(\hat{c}_j\right)^{1/2} & \hat{E}rr^{Par}\left(\hat{f}_j\right)^{1/2} & \hat{M}SEP_j^{1/2}/\left(j\cdot\overline{\hat{X}_j}\right)^{1/2} \end{pmatrix} \tag{GCL}$$

in the generalized chain-ladder model, and for the chain-ladder models without an additive component we have

$$\begin{pmatrix} \hat{f}_j & \sum_{k=n-j+1}^{n} \hat{X}_{k,j+1}/\left(j\cdot\overline{\hat{X}_j}\right)^{1/2} \\ \hat{E}rr^{Par}\left(\hat{f}_j\right)^{1/2} & \hat{M}SEP_j^{1/2}/\left(j\cdot\overline{\hat{X}_j}\right)^{1/2} \end{pmatrix}. \tag{CL}$$

This procedure therefore enables us to compute the prediction error for the chain-ladder models in a particularly simple way.

For the linear regression models, there is only one more small step to do: here the "LINEST" function applied to only one "dummy" variable does not take into account the entire process variance for the j accident years to be predicted. Thus this corresponds to a prediction error based on

$$\frac{\hat{M}SEP_j}{\hat{\sigma}_j^2} = j(\textit{instead of } 1 \textit{ provided by } \text{"LINEST"}) + \frac{j^2}{n-j}\,\frac{\left(\overline{\hat{V}_j}\right)^2\cdot\overline{X_j^2} - 2\,\overline{\hat{V}_j}\cdot\overline{\hat{X}_j}\cdot\left(\overline{V_j\cdot X_j}\right) + \left(\overline{\hat{X}_j}\right)^2\cdot\overline{V_j^2}}{\overline{X_j^2}\cdot\overline{V_j^2} - \left(\overline{V_j\cdot X_j}\right)^2}$$

(see Appendix B). As the "LINEST" function in Microsoft Excel also lists the process error in the second column of the third row, $\hat{M}SEP_j$ can easily be calculated with the two terms $(\hat{M}SEP_j - (j-1)\hat{\sigma}_j^2)^{1/2}$ and $\hat{\sigma}_j$. These terms are returned by the "LINEST" function, used—as in the chain-ladder cases—as an array function with the following output:

$$\begin{pmatrix} \hat{c}_j & \hat{f}_j & \sum_{k=n-j+1}^{n} \hat{X}_{k,j+1} \\ Err^{Par}\left(\hat{c}_j\right)^{1/2} & Err^{Par}\left(\hat{f}_j\right)^{1/2} & \left(\hat{M}SEP_j - (j-1)\hat{\sigma}_j^2\right)^{1/2} \\ * & \hat{\sigma}_j & * \end{pmatrix}. \tag{GLR}$$

The reasons that there is no need for the correction term $\hat{\sigma}_j$ in the chain-ladder models using the array function returned by the "LINEST" function are explained in Appendix C.

7. The completion of the claims triangle: Expected claims development

The usual task when projecting claims triangles is to complete these triangles to a square. Therefore we predict the development of the latest of the known claims amounts, $X_{k,n-k+1}$ for the development from year $n-k+1$ to year n by recursively defined estimators for $j = n-k+1, \dots, n-1$, with the notation $\hat{X}_{k,n-k+1} = X_{k,n-k+1}$ for the initial term.

Model	$\hat{X}_{k,j+1} =$
Affine models: generalized chain ladder and generalized linear regression	$\hat{f}_j \cdot \hat{X}_{k,j} + \hat{c}_j \cdot V_k$
1-parameter models: multiplicative models, e.g., chain ladder	$\hat{f}_j \cdot \hat{X}_{k,j}$
Additive models, e.g., incremental loss ratio method	$\hat{c}_j \cdot V_k$

The recursively defined $\hat{X}_{k,n}$ provides an unbiased estimator for the projected claims development for the accident year k, $1 \le k < n$, based on the data given by the claims triangle $D = \{X_{i,j},\ i + j \le n + 1\}$. Thus, we have $E[\hat{X}_{k,j+1} | D] = E[X_{k,j+1} | D]$, which can be proved—as Mack (1993) did for the chain-ladder model—using the law of total expectations:

$$
\begin{aligned}
E\left[\hat{X}_{k,j+1}\middle| D\right] &= E\left[\hat{X}_{k,j+1}\middle| X_{k,n+1-k}\right] \\
&= E\left[E\left[\hat{X}_{k,j+1}|X_{k,j}|X_{k,n+1-k}\right]\right] \\
&= E\left[E\left[\hat{f}_j X_{k,j} + \hat{c}_j V_k|X_{k,j}|X_{k,n+1-k}\right]\right] \\
&= E\left[\hat{f}_j|X_{k,j}\right]\cdot E\left[X_{k,j}|X_{k,n+1-k}\right] \\
&\quad + E\left[\hat{c}_j|X_{k,j}\right]\cdot V_k \\
&\overset{(*)}{=} f_j \cdot E\left[X_{k,j}|X_{k,n+1-k}\right] + c_j \cdot V_k = f_j \cdot f_{j-1} \\
&\quad \cdot E\left[X_{k,j-1}|X_{k,n+1-k}\right] + \left(f_j \cdot c_{j-1} + c_j\right)\cdot V_k \\
&= \cdots \overset{(**)}{=} f_{n-k:j+1} X_{k,n+1-k} + V_k \sum_{s=1}^{j+k-n} f_{n-k+s:j+1} \\
&\quad \cdot c_{n-k+s} = E\left[X_{k,j+1}|D\right],
\end{aligned}
$$

where $f_{a:b} = \prod_{k=a+1}^{b-1} f_k$ or the corresponding estimators $\hat{f}_{a:b} = \prod_{k=a+1}^{b-1} \hat{f}_k$, respectively, denote the projection factor from development year $a+1$ to development year b. Equation (*) is valid by the theorem of Gauss-Markov, stating that the weighted least squares estimators $\hat{f}_j$ and $\hat{c}_j$ are unbiased estimators, i.e., $E[\hat{f}_j | X_{k,j}] = f_j$ and $E[\hat{c}_j | X_{k,j}] = c_j$.

Setting $j + 1 = n$ in equation (**), we obtain an estimate for the ultimate claims development $\hat{X}_{k,n}$ from the latest available claims data for accident year k, thus the diagonal element $X_{k,n+1-k}$ is

$$\hat{X}_{k,n} = \hat{f}_{n-k:n} X_{k,n+1-k} + V_k \sum_{s=1}^{k-1} \hat{f}_{n-k+s:n} \cdot \hat{c}_{n-k+s}.$$

This formula shows the importance of the additive part $\hat{c}_j$ for the affine models. These additive terms, in particular those for the more recent accident years, i.e., those for higher values of k, may have an important impact on the estimate of the ultimate claim development. The reason for this is that the additive parts have to be considered for each of the $k-1$ remaining development years, since the formula for $\hat{X}_{k,n}$ involves the sum of $k-1$ additive terms in the estimate of the ultimate $\hat{X}_{k,n}$. Our numerical examples at the end of this paper will underline the potential importance of these additive parts for claims triangles based on real data.

8. The completion of the claims triangle: The prediction error of the reserves

In this section we combine the prediction errors for the age-to-age development in order to obtain the prediction error for the entire projection. The reserve itself depends on all these subsequent estimations, whereas the development from year j to year $j+1$ only depends on the sum of the estimated previous-year claims $\overline{\hat{X}}_j = (1/j) \cdot \sum_{k=n-j+1}^{n} \hat{X}_{k,j}$, as defined in Section 5.

We have already computed the mean squared error of prediction $MSEP_j$ for the development from year j to year $j+1$ of the sum of the claims given the estimation $\hat{X}_j$, i.e., the sum of the predicted previous-year claims. This amounts to

$$MSEP_j = Var_j^{Process} + Err_j^{Param} = \sigma_j^2 \cdot \tau_j,$$

with estimator

$$\hat{M}SEP_j = \hat{\sigma}_j^2 \cdot \hat{\tau}_j, \text{ where } \hat{\tau}_j$$
and $\hat{\sigma}_j$ depend on the chosen model.

To obtain the error of the predicted reserve from the age-to-age development errors, we have to take into account the multiplicative development from year $j+1$ to year n, because the estimation of the reserve is based on the estimated projection up to the ultimate development year n. Therefore the prediction error for the development year j has to be multiplied by the square of the previously defined projection factor $f_{j:n} = \prod_{k=j+1}^{n-1} f_k$. Due to our memorylessness assumption, neither the parameter error nor the variances of the process are correlated for the different development years. Therefore, the age-to-age development errors $\hat{M}SEP_j$, multiplied by the square of the estimated projection factor to scale them up to the level of the final development year, may simply be summed up to get the prediction error of the reserves:

$$MSEP = E\left[\left(\sum\nolimits_{k=2}^{n} \hat{X}_{k,n} - \sum\nolimits_{k=2}^{n} X_{k,n}\right)^2\right]$$
$$= \sum_{j=1}^{n-1} MSEP_j \cdot f_{j:n}^2.$$

The derivation of $MSEP$ as the sum of the scaled-up age-to-age prediction errors can be found in Appendix E.

Thus, we get a similarly constructed estimator,

$$\hat{M}SEP = \sum_{j=1}^{n-1} \hat{M}SEP_j \cdot \hat{f}_{j:n}^2 = \sum_{j=1}^{n-1} \left(\hat{\sigma}_j \cdot \hat{f}_{j:n}\right)^2 \cdot \hat{\tau}_j.$$

In case of the chain-ladder model, this results in

$$\hat{M}SEP(CL) = \sum_{j=1}^{n-1} \left(\overline{\hat{X}_j} \cdot \hat{\sigma}_j \cdot \hat{f}_{j:n}\right)^2 \left(\frac{j}{\overline{\hat{X}_j}} + \frac{j^2}{(n-j) \cdot \overline{X_j}}\right),$$

and for the generalized chain-ladder model (GCL) with an additive component depending on a volume, we have

$$\hat{M}SEP(GCL) = \sum_{j=1}^{n-1} \left(\hat{\sigma}_j \cdot \hat{f}_{j:n}\right)^2$$
$$\left(j \cdot \overline{\hat{X}_j} + j^2 \cdot \frac{\left(\begin{array}{c}\left(\overline{\hat{V}_j}\right)^2 \cdot \overline{X_j} - 2\, \overline{\hat{V}_j} \cdot \overline{V_j} \cdot \overline{\hat{X}_j} \\ + \left(\overline{\hat{X}_j}\right)^2 \cdot \overline{V_j^2 / X_j}\end{array}\right)}{(n-j) \cdot \left(\overline{V_j^2 / X_j} \cdot \overline{X_j} - \left(\overline{V_j}\right)^2\right)} \right).$$

Mack (1993) derived the formula for the mean squared error of prediction in the chain-ladder case, $\hat{M}SEP(CL)$, and thus stimulated extended research into the reliability of reserve calculation based on claims triangles. His approach was different in so far as all accident years were considered separately and not together as we do here. Therefore Mack's prediction error for the entire reserve is composed of the prediction errors for the different accident years. Because these prediction errors rely on the same parameter estimation, Mack has to take these dependencies into account. Thus, the structures of the equivalent formulae are different due to the different approach in handling the estimation error. In the development by Mack, our term for the estimation error of $(j \cdot \hat{X}_j)^2$ appears twice, once as the sum of the squares corresponding to the estimation error for the different accident years, and again as a mixed product considering the dependencies for the different accident years. Therefore, as noted below, the squared sum of our case splits into two terms in the Mack formulae and the mixed term appears when calculating the prediction error for the entire reserve from those of the individual accident years:

$$\left(j \cdot \hat{X}_j\right)^2 = \left(\sum\nolimits_{k=n-j}^{n} \hat{X}_{k,j}\right)^2$$
$$= \sum\nolimits_{k=n-j}^{n} \hat{X}_{k,j}^2 + 2 \cdot \sum\nolimits_{k=n-j}^{n-1} \hat{X}_{k,j} \cdot \sum\nolimits_{l=k+1}^{n} \hat{X}_{l,j}.$$

Table 2. Data of claims triangle of Mack (1993)

Accident Year	Volume	Development Year j								
		1	2	3	4	5	6	7	8	9
1	1	58	128	477	1,028	1,360	1,647	1,819	1,907	1,950
2	1	24	142	984	2,143	2,962	3,684	4,049	4,116	
3	1	33	275	1,523	3,203	4,446	5,159	5,343		
4	1	21	530	2,900	4,999	6,460	6,854			
5	1	40	763	2,921	4,990	5,649				
6	1	91	952	4,211	5,866					
7	1	62	868	1,955						
8	1	25	284							
9	1	13								

Note: Figures are rounded, and our calculations are also based on these rounded figures.

9. Numerical examples

In this section, we apply three of our methods—the two considered affine methods (that is, the generalized chain-ladder and the generalized linear regression), and the traditional chain-ladder method—to three given claims triangles. One of these triangles corresponds to the less regular example used by Mack (1993) to illustrate his newly discovered determination of the prediction error in the chain-ladder case. The second one was published by Schnieper (1991) in a case where he explicitly gathered the data of the newly reported claims in the corresponding accident year. Within our affine methods, these amounts are estimated by the additive component of the age-to-age development. Applied to real claims data, our affine model usually yields positive additive parameters, and only exceptionally does it produce negative parameters. This shows that the affine model fits rather well to the reality of the way in which claims are handled within the triangles. The third triangle was published by Brosius (1993) and reconsidered by Halliwell (2007).

9.1. The example of Mack (1993)

There is no volume function quoted in Mack's paper. We therefore assumed the volume to be constant for all accident years. The projection does not depend on the constant itself. By the choice of "1" for the volume in Table 2, the estimated additive development parameter may be directly interpreted as the estimated amount of newly reported claims of the corresponding development year. This example also shows that assuming a constant volume function in those cases where no such function is available within the data may lead to a more stable and more reliable projection result. Table 3 shows the

Table 3. Estimated parameters of the age-to-age development for the triangle of Mack (1993)

Model	Parameter	$\hat{f}_j^*$	j: Development from Year j to Year $j+1$							
			1	2	3	4	5	6	7	8
generalized linear regression	additive	$\hat{c}_j$	124	501	865	396	478	209	105.0	0
	multiplicative	$\hat{f}_j$	8.34	3.13	1.31	1.15	1.01	1.01	0.99	1.02
generalized chain ladder	additive	$\hat{c}_j$	156	335	526	221	299	154	105	0
	multiplicative	$\hat{f}_j$	7.61	3.45	1.47	1.21	1.06	1.02	0.99	1.02
chain ladder	multiplicative	$\hat{f}_j$	11.14	4.09	1.71	1.28	1.14	1.07	1.03	1.02

Table 4. Estimated reserves and their standard errors for the triangle of Mack (1993)

	IBNR Reserves as Projected from the Claims Triangle: Model		
Accident Year	Generalized Linear Regression	Generalized Chain Ladder	Chain Ladder
1	0	0	0
2	93	93	93
3	177	177	265
4	470	524	834
5	1,009	1,142	1,568
6	2,368	2,752	3,696
7	3,359	3,372	3,487
8	4,146	3,796	2,952
9	4,162	3,871	1,636
Total reserve	15,784	15,727	14,530
Estimated standard error = $\hat{M}SEP^{1/2}$	3,862	3,526	3,731
Error % = estimated standard error/total reserve	**24%**	**22%**	**26%**

estimated additive and multiplicative parameters for the triangle of Mack (1993).

Looking in particular at the ninth accident year in Table 4, huge differences in the estimated IBNR are found between the traditional chain-ladder model and the two affine models. We suggest that the chain-ladder method is not appropriate for the given data, because this method neglects the IBNYR-estimation and therefore seems not suitable, especially for the most recent accident year with a particularly small claims amount of just 13 (cf. Table 2).

In Table 5, $\hat{M}SEP_j^{1/2} \cdot \hat{f}_{j:n}$ estimates the standard error of the development from age j to age $j+1$ scaled up to the level of the final-year claims amounts, and $\hat{M}SEP^{1/2} = (\sum_{j=1}^{n-1} \hat{M}SEP_j \cdot \hat{f}_{j:n}^{2})^{1/2}$, the entire standard error of the estimated reserve.

9.2. The example of Schnieper (1991)

Schnieper (1991) proposed to explicitly separate the incurred data into newly reported claims (true IBNR) and changes in the amounts of reported claims (IBNER). He assumed the expected values of the true IBNR claims to depend on a volume function and changes in IBNER claims on the incurred claims of the previous year, as in the chain-ladder case. The example in Schnieper (1991) shows that these additional data may have a considerable impact on the estimation of the reserve, although these data are not regularly collected. Assuming an affine age-to-age development of incurred claims, we propose an estimate of these two components based solely on the incurred claims available in the usual claims triangle. The additive part provides an estimate for

Table 5. Estimated standard error for the triangle of Mack

		$\hat{M}SEP_j^{1/2} \cdot \hat{f}_{j:n}$ j: Development from Year j to Year $j+1$							
Model	$\hat{M}SEP^{1/2}$	1	2	3	4	5	6	7	8
generalized linear regression	3,862	1,626	1,626	1,180	1,180	991	991	1,196	1,196
generalized chain ladder	3,526	1,444	1,582	1,117	1,219	1,234	1,104	1,105	1,071
chain ladder	3,731	1,788	1,901	1,567	1,220	1,313	868	622	550

Table 6. Data of claims triangle of Schnieper (1991), based on a practical third-party-liability excess-of-loss pricing problem from motor insurance

Accident Year	Volume	Development Year j						
		1	2	3	4	5	6	7
1	10,224/15,000	7.5	28.9	52.6	84.5	80.1	76.9	79.5
2	12,752/15,000	1.6	14.8	32.1	39.6	55.0	60.0	
3	14,875/15,000	13.8	42.4	36.3	53.3	96.5		
4	17,365/15,000	2.9	14.0	32.5	46.9			
5	19,410/15,000	2.9	9.8	52.7				
6	17,617/15,000	1.9	29.4					
7	18,129/15,000	19.1						

the newly reported claims, the so-called true IBNR claims, and the multiplicative part gives an estimate for the changes in the reported claims (IBNER).

In his paper, Schnieper (1991) regards the premiums per accident year of the entire portfolio as a volume function. To facilitate interpretation, we normalize the volume via dividing the values in Table 6 by 15,000 to attain a size near to "1."

One may interpret the product of the volume V_j and the additive development parameter $\hat{c}_j$ as an estimation of the newly reported claims. In fact, with a volume function near to "1", the parameters $\hat{c}_j$ may already be regarded as a rough estimate of the newly reported claims.

Looking at Table 7, the negative values of the additive parameter for the development of year 3 to year 4 do not fit in with our interpretation that the said parameter models the newly reported claims. Nevertheless, the observation of such negative additive parameters should motivate us to look more closely and, perhaps, to collect more data, as was done in this case with the additional data of the newly reported claims shown in Table 8. Estimated reserves are shown in Table 9, and estimated standard errors in Table 10.

9.3. The example of Brosius (1993), reconsidered by Halliwell (2007)

The data of Brosius in Table 11 are based on a small book of business and consist of a claims trapezoid of seven accident and five development years. The data were completed to a triangle in order to apply the formulae mentioned above. Again, the volume is normalized to attain a size near to "1" via dividing by 10,000 in order to scale the parameters $\hat{c}_j$ up to the order of magnitude of the newly reported claims. Table 12 shows the estimated parameters.

The risk metric in the chain-ladder model, as well as in the generalized chain-ladder model, entails an infinite probability for the age-to-age development

Table 7. Estimated parameters of the age-to-age development for the triangle of Schnieper (1991)

Model	Parameter	$\hat{f}_j^*$	j: Development from Year j to Year $j+1$					
			1	2	3	4	5	6
generalized linear regression	additive	$\hat{c}_j$	10.1	31.7	–10.3	57.0	18.8	0.0
	multiplicative	$\hat{f}_j$	2.42	0.39	1.71	0.51	0.80	1.03
generalized chain ladder	additive	$\hat{c}_j$	12.3	32.8	–9.5	52.0	18.8	0.0
	multiplicative	$\hat{f}_j$	2.09	0.39	1.69	0.57	0.80	1.03
chain ladder	multiplicative	$\hat{f}_j$	4.55	1.88	1.46	1.31	1.01	1.03
from additional data of newly reported claims (cf. Schnieper 1991)	additive	$\hat{c}_j$	15.9	21.0	17.3	17.7	7.4	7.5
	multiplicative	$\hat{f}_j$	1.36	0.93	1.05	1.05	0.93	0.97

Table 8. Newly reported claims per development year: *N* triangle of genuine IBNR claims for the triangle of Schnieper (1991)

Accident Year	Collected Data "*N*", cf. Schnieper (1991) Development Year *j*						IBNYR-Estimation: $\hat{c}_{j-1} \cdot V_k$ in the Generalized Chain-Ladder Model					
k	2	3	4	5	6	7	2	3	4	5	6	7
1	18.3	28.5	23.4	18.4	0.7	5.1	6.9	21.6	–7.0	38.9	12.8	0.0
2	12.6	18.2	16.1	14.0	10.6		8.6	27.0	–8.7	48.5	16.0	
3	22.7	4.0	12.4	12.1			10.0	31.5	–10.2	56.5		
4	9.7	16.4	11.6				11.7	36.7	–11.9			
5	6.9	37.1					13.1	41.1				
6	27.5						11.9					

Table 9. Estimated reserves and their standard errors for the triangle of Schnieper (1991)

	IBNR Reserves as Projected from the Claims Triangle: Model			
Accident Year	From Additional Data (Schnieper 1991)	Generalized Linear Regression	Generalized Chain Ladder	Chain Ladder
1	0	0	0	0
2	4	2	2	2
3	5	3	3	5
4	33	50	47	17
5	60	66	64	53
6	77	79	78	81
7	104	100	99	307
Total reserve	284	300	294	464
Estimated standard error = $\hat{MSEP}^{1/2}$	122, see [L]	74	93	302
Error % = estimated standard error/ total reserve	43%	25%	32%	65%

Table 10. Estimated standard errors for the triangle of Schnieper (1991)

		$\hat{MSEP}_j^{1/2} \cdot \hat{f}_{j:n}$					
		j: Development from Year *j* to Year *j* + 1					
Model	$\hat{MSEP}^{1/2}$	1	2	3	4	5	6
generalized linear regression	74	2	23	6	61	32	13
generalized chain ladder	93	11	32	8	66	48	27
chain ladder	302	162	201	36	145	49	14

Table 11. Data of claims triangle of Brosius (1993)

Accident Year	Volume	Development Year j						
		1	2	3	4	5	6	7
1	4,260/10,000	102	104	209	650	847	847	847
2	5,563/10,000	0	543	1,309	2,443	3,033	3,033	
3	7,777/10,000	412	2,310	3,083	3,358	4,099		
4	8,871/10,000	219	763	1,637	1,423			
5	10,645/10,000	969	4,090	3,801				
6	11,986/10,000	0	3,467					
7	12,873/10,000	932						

Table 12. Estimated parameters of the age-to-age development for the triangle of Brosius (1993)

Model	Parameter	$\hat{f}_j^*$	j: Development from Year j to Year $j+1$					
			1	2	3	4	5	6
generalized linear regression	additive	$\hat{c}_j$	1,920	1,304	463	173	0	0
	multiplicative	$\hat{f}_j$	1.75	0.67	0.99	1.19	1	1
generalized chain ladder	additive	$\hat{c}_j$	not defined	640	972	172	0	0
	multiplicative	$\hat{f}_j$	not defined	0.98	0.85	1.19	1	1
chain ladder	multiplicative	$\hat{f}_j$	6.63	1.29	1.26	1.24	1	1

from a claim cell zero to a positive claim cell. This results in an infinite estimated standard error in the chain-ladder case in Table 13 and indicates that the chain-ladder model is not suitable for such inhomogeneous claims triangles. At least, the reserve itself is calculable in the chain-ladder model, whereas in the generalized chain-ladder model the reserve formulae are no longer applicable, since claim cells of value zero produce zeroes in the denominator. With only one such zero-cell, the reserve could be defined in the generalized chain-ladder model by computing the development parameters for a sequence converging to zero in this zero-cell. The sequence of the development parameters will then be convergent too, meaning that the additive term is fully defined by the development of this zero-cell. However, in the present case in Table 11 with two zero-cells for a given development year, the additive term depends on the two specific sequences chosen for these two particular cells, thus even considering the value of a limit will give no precise amount for the reserve. Table 14 shows the estimated standard errors for the triangle of Brosius (1993).

Table 13. Estimated reserves and their standard errors for the triangle of Brosius (1993)

Accident Year	IBNR Reserves as Projected from the Claims Triangle: Model	
	Generalized Linear Regression	Chain Ladder
1	0	0
2	0	0
3	0	0
4	421	337
5	1,456	2,133
6	1,973	3,491
7	5,207	11,461
Total reserve	9,058	17,422
Estimated standard error = $\widehat{MSEP}^{1/2}$	3,845	∞
Error % = Estimated standard error/ total reserve	**42%**	∞

10. Conclusion

In this paper, we proposed some new affine models for claims reserving and compared them to the well-known multiplicative chain-ladder model, given three sets of real claims data taken from the literature. In

Table 14. Estimated standard errors for the triangle of Brosius (1993)

		$\widehat{MSEP}_j^{1/2} \cdot \hat{f}_{j:n}$					
		j: Development from Year *j* to Year *j* + 1					
Model	$\widehat{MSEP}^{1/2}$	1	2	3	4	5	6
generalized linear regression	3,845	1,079	1,123	3,509	216	18	2
generalized chain ladder	not defined			not defined			
chain ladder	∞	∞	4,903	6,755	426	0	0

our view, the results show that the proposed affine models better correspond to the given claims data than the traditional chain-ladder model.

Does that mean that we have found the ultimate way to compute the exact amount of the predicted reserves and the corresponding prediction error? Not quite. As the numerical examples show, the proposed new models do indeed help to better detect patterns of claims reporting and thus to analyze what is going on in one's business as a whole or in a specific branch. In particular for long-tail branches such as general liability or professional liability insurance, where claims might be reported years after they were incurred, the new models are more appropriate because their additive part takes into account these late reported claims and estimates the average amount to be incorporated in the claims projection, based on the claims data recorded in the past. The prediction error then provides a measure for the stability of the model assumptions in the past years of experience. Also, a fairly high prediction error may be caused by an unrealistic model that is founded on inappropriate parameters or data, and could thus provide us with a stimulus to improve the model—e.g., by using the proposed affine model instead of regular chain-ladder—or the data basis in order to be more in touch with reality.

When applying rigorous and apparently objective mathematical methods in an economic setting, one should always keep in mind how much the corresponding results depend on the choice of models and of the parameters involved. And with the mathematical methods becoming more and more developed and sophisticated, the danger of possible misapplication and implicit trust in such methods rises. For instance, example 4.a mentioned by Mack (1993) in his famous paper establishing the stochastic view in non-life reserving does not seem to be suitable for the purely multiplicative chain-ladder model without additive parameter, as applied by Mack (1993). At least, that is what is suggested by the significant additive parts revealed—and all with positive values!—when applying our affine models to the given data. Hence, not only is in this case the prediction error questionable, but also the much more important reserve estimation itself, in particular for more recent accident years. Thus it comes as no surprise that practicing actuaries are usually rather cautious towards new methods, particularly if they are regarded as highly elaborate and academic. For this reason, we will follow with great interest whether and how our proposed methods will find their way into generally accepted actuarial practice.

Acknowledgments

The author wishes to thank Richard Gorvett for his keen interest in this subject and his valuable advice, Mario Wüthrich for his valuable suggestions, and Andreas Graf for his advice on the English language.

Appendix A. The weighted least squares estimators for the parameters in the different models

Since the weighted least squares estimator $\hat{f}_j^*$ for the age-to-age development is determined, as mentioned in Section 4, by a product of matrices $\hat{f}_j^* = ((X_j^*)^t \cdot W^{-1} \cdot X_j^*)^{-1} \cdot (X_j^*)^t \cdot W^{-1} \cdot X'_{j+1}$, we need to calculate the two matrices involved for all the models under consideration (Table A.1).

Table A.1. Explicit formulae of the fundamental auxiliary terms for the different models

Model	$\left(\left(X_j^*\right)^t \cdot W^{-1} \cdot X_j^*\right)^{-1}$	$\left(X_j^*\right)^t \cdot W^{-1} \cdot X'_{j+1}$
generalized chain ladder	$\frac{1}{n-j} \cdot \frac{1}{\overline{V_j^2/X_j} \cdot \overline{X_j} - \left(\overline{V_j}\right)^2} \begin{pmatrix} \overline{X_j} & -\overline{V_j} \\ -\overline{V_j} & \overline{V_j^2/X_j} \end{pmatrix}$	$(n-j) \cdot \begin{pmatrix} \overline{V_j \cdot X'_{j+1}/X_j} \\ \overline{X'_{j+1}} \end{pmatrix}$
generalized chain ladder with constant volume	$\frac{1}{n-j} \cdot \frac{1}{\overline{1/X_j} \cdot \overline{X_j} - 1} \begin{pmatrix} \overline{X_j} & -1 \\ -1 & \overline{1/X_j} \end{pmatrix}$	$(n-j) \cdot \begin{pmatrix} \overline{X'_{j+1}/X_j} \\ \overline{X'_{j+1}} \end{pmatrix}$
chain ladder	$\left(\left(X_j^*\right)^t \cdot W^{-1} \cdot X_j^*\right)^{-1} = \frac{1}{n-j} \cdot \frac{1}{\overline{X_j}}$	$\left(X_j\right)^t \cdot W^{-1} \cdot X'_{j+1} = (n-j) \cdot \overline{X'_{j+1}}$
generalized linear regression	$\frac{1}{n-j} \cdot \frac{1}{\overline{X_j^2} \cdot \overline{V_j^2} - \left(\overline{V_j \cdot X_j}\right)^2} \begin{pmatrix} \overline{X_j^2} & -\overline{V_j \cdot X_j} \\ -\overline{V_j \cdot X_j} & \overline{V_j^2} \end{pmatrix}$	$(n-j) \cdot \begin{pmatrix} \overline{V_j \cdot X'_{j+1}} \\ \overline{X'_{j+1} \cdot X_j} \end{pmatrix}$
linear regression with intercept	$\frac{1}{n-j} \cdot \frac{1}{\overline{X_j^2} - \left(\overline{X_j}\right)^2} \begin{pmatrix} \overline{X_j^2} & -\overline{X_j} \\ -\overline{X_j} & 1 \end{pmatrix}$	$(n-j) \cdot \begin{pmatrix} \overline{X'_{j+1}} \\ \overline{X_j \cdot X'_{j+1}} \end{pmatrix}$
incremental loss ratio method (1-parameter model)	$\left(\left(V_j\right)^t \cdot W^{-1} \cdot V_j\right)^{-1} = \frac{1}{n-j} \cdot \frac{1}{\left(\overline{V_j}\right)^2}$	$(n-j) \cdot \overline{V_j \cdot X'_{j+1}}$

In this context, the terms with a bar over the letter denote the corresponding average over all accident years from 1 to $n - j$. For these accident years the development from year $j - 1$ to year j has already been experienced. This gives

$$\overline{X_j} = \frac{1}{n-j} \sum_{k=1}^{n-j} X_{k,j},$$

$$\overline{V_j} = \frac{1}{n-j} \sum_{k=1}^{n-j} V_k,$$

$$\overline{V_j \cdot X'_{j+1}/X_j} = \frac{1}{n-j} \sum_{k=1}^{n-j} \frac{V_k X_{k,j+1}}{X_{k,j}}$$

$$\overline{X_j^2} = \frac{1}{n-j} \sum_{k=1}^{n-j} X_{k,j}^2,$$

$$\overline{V_j^2} = \frac{1}{n-j} \sum_{k=1}^{n-j} V_k^2,$$

$$\overline{V_j^2/X_j} = \frac{1}{n-j} \sum_{k=1}^{n-j} \frac{V_k^2}{X_{k,j}}$$

$$\overline{X'_{j+1}} = \frac{1}{n-j} \sum_{k=1}^{n-j} X_{k,j+1},$$

$$\overline{V_j \cdot X'_{j+1}} = \frac{1}{n-j} \sum_{k=1}^{n-j} V_k X_{k,j+1},$$

$$\overline{V_j \cdot X_j} = \frac{1}{n-j} \sum_{k=1}^{n-j} V_k X_{k,j}$$

$$\overline{X_j \cdot X'_{j+1}} = \frac{1}{n-j} \sum_{k=1}^{n-j} X_{k,j} X_{k,j+1}.$$

In each case, the product of the two auxiliary terms in Table A.1 gives the parameter estimators for the corresponding model. For the two-parameter models, as shown in Table A.2, this results in a vector, the first row of which gives the estimated additive parameter and the second row gives the multiplicative one. For the one-parameter models, the estimator describes either the multiplicative or the additive development parameter, depending on the model. For a better understanding of the formulae, we also list in Table A.2 the estimated claims for development year $j + 1$ given the mean of the experienced claims $\overline{X_j}$ in year j and the mean of the volumes of the previous accident years $\overline{V_j}$. For most of the models, this corresponds to the mean of the experienced

Table A.2. Weighted least squares estimators for the year-to-year development parameters, calculated from the auxiliary terms shown in Table A.1

Model	Estimators of Additive and Multiplicative Development Parameters $\hat{f}_j^* = \begin{pmatrix} \hat{c}_j \\ \hat{f}_j \end{pmatrix} = \left(\left(X_j^*\right)^t \cdot W^{-1} \cdot X_j^*\right)^{-1} \cdot \left(X_j^*\right)^t \cdot W^{-1} \cdot X'_{j+1}.$ (formula for the 2-parameter models)	Projected Claims in the Equal Means Case $\left(\hat{f}_j^*\right)^t \cdot \begin{pmatrix} \overline{V_j} \\ \overline{X_j} \end{pmatrix} = \hat{c}_j \overline{V_j} + \hat{f}_j \overline{X_j}$
generalized chain ladder	$\frac{1}{\overline{V_j^2/X_j} \cdot \overline{X_j} - \left(\overline{V_j}\right)^2} \cdot \begin{pmatrix} \overline{V_j \cdot X'_{j+1}/X_j} \cdot \overline{X_j} - \overline{V_j} \cdot \overline{X'_{j+1}} \\ \overline{V_j^2/X_j} \cdot \overline{X'_{j+1}} - \overline{V_j} \cdot \overline{V_j \cdot X'_{j+1}/X_j} \end{pmatrix}$	$\overline{X'_{j+1}}$
generalized chain ladder with constant volume 1	$\frac{1}{\overline{1/X_j} \cdot \overline{X_j} - 1} \cdot \begin{pmatrix} \overline{X'_{j+1}/X_j} \cdot \overline{X_j} - \overline{X'_{j+1}} \\ \overline{1/X_j} \cdot \overline{X'_{j+1}} - \overline{X'_{j+1}/X_j} \end{pmatrix}$	$\overline{X'_{j+1}}$
chain ladder (1-parameter model)	$\hat{f}_j = \frac{\overline{X'_{j+1}}}{\overline{X_j}}$	$\hat{f}_j \cdot \overline{X_j} = \overline{X'_{j+1}}$
generalized linear regression	$\frac{1}{\overline{X_j^2} \cdot \overline{V_j^2} - \left(\overline{V_j \cdot X_j}\right)^2} \begin{pmatrix} \overline{X_j^2} \cdot \overline{V_j \cdot X'_{j+1}} - \overline{X'_{j+1} \cdot X_j} \cdot \overline{V_j} \cdot \overline{X_j} \\ \overline{V_j^2} \cdot \overline{X'_{j+1} \cdot X_j} - \overline{V_j \cdot X'_{j+1}} \cdot \overline{V_j} \cdot \overline{X_j} \end{pmatrix}$	no simple term
linear regression with intercept	$\frac{1}{\overline{X_j^2} - \left(\overline{X_j}\right)^2} \begin{pmatrix} \overline{X_j^2} \cdot \overline{X'_{j+1}} - \overline{X_j} \cdot \overline{X'_{j+1} \cdot X_j} \\ \overline{X_j \cdot X'_{j+1}} - \overline{X_j} \cdot \overline{X'_{j+1}} \end{pmatrix}$	$\left(\hat{f}_j^*\right)^t \cdot \begin{pmatrix} 1 \\ \overline{X_j} \end{pmatrix} = \overline{X'_{j+1}}$
incremental loss ratio method (1-parameter model)	$\hat{c}_j = \frac{\overline{V_j \cdot (X'_{j+1})}}{\left(\overline{V_j}\right)^2}$	$\hat{c}_j \cdot \overline{V_j} = \frac{\overline{V_j \cdot X'_{j+1}}}{\overline{V_j}}$

claims $\overline{X'_{j+1}}$, which is exactly what one would expect in this situation.

Appendix B. The prediction error within the different models

The formulae for the prediction error in the different models are based on the covariance matrix $((X_j^*)^t W^{-1} X_j^*)^{-1}$, depending on the experienced claims of development year j, applied as a quadratic form to the sum of the volumes $j \cdot \hat{\overline{V}}_j = \sum_{k=n-j+1}^{n} V_j$ for the accident years to be developed and to the sum of predicted claims beyond the diagonal in the claims triangle, including the diagonal itself, $j \cdot \hat{\overline{X}}_j = X_{n-j+1,j} + \sum_{k=n-j+2}^{n} \hat{X}_{k,j} = \sum_{k=n-j+1}^{n} \hat{X}_{k,j}$.

For these claims or predicted claims, respectively, no development from year j to year $j + 1$ has been experienced and thus has to be predicted. The predicted entire reserve for all accident years only depends on the sum of the predicted claims. Due to our linear-model assumption, the parameter estimation error thus only depends on the sum of assumed previous-year claims $j \cdot \hat{\overline{X}}_j$, applied in the quadratic form. Table B.1 shows the explicit formulae for the crucial term τ_j in the prediction error

$$MSEP_j = Var_j^{Process} + Err_j^{Param} = \tau_j \cdot \sigma_j^2$$

of the projection $\hat{f}_j^* = \left(j \cdot \begin{pmatrix} \hat{\overline{V}}_j \\ \hat{\overline{X}}_j \end{pmatrix}\right)$ to year $j + 1$, given the sum $j \cdot \hat{\overline{X}}_j$ of predicted claims beyond the diagonal of the claims triangle. This sum $j \cdot \hat{\overline{X}}_j$ is the relevant figure for the computation of the reserve. In order to have a better understanding of the formula, we compute the prediction error in Table B.1 in the case of equal means of experienced and predicted previous-year claims, and of equal means of the volume of both the accident years experienced and of the accident years to be predicted, for the development from year j to $j + 1$. Then, in this particular setting, the prediction error for this

Table B.1. The crucial factor $\hat{\tau}_j$ in the expression $\hat{MSEP}_j = \hat{\tau}_j \cdot \hat{\sigma}_j^2$, the prediction error in year $j+1$ of the sum of claims beyond the diagonal, given the sum $j \cdot \overline{\hat{X}}_j$ in the previous year j

Model	General Case	Equal Means Case: $\overline{\hat{X}}_j = \overline{X}_j, \overline{\hat{V}}_j = \overline{V}_j$
	$\hat{\tau}_j = \frac{\hat{Var}_j^{Process} + \hat{Err}_j^{Param}}{\hat{\sigma}_j^2} = \frac{\hat{Var}_j^{Process}}{\hat{\sigma}_j^2} + j^2 \cdot \left(\overline{\hat{V}}_j \quad \overline{\hat{X}}_j\right)\left(\left(X_j^*\right)^t W^{-1} X_j^*\right)^{-1}\begin{pmatrix}\overline{\hat{V}}_j \\ \overline{\hat{X}}_j\end{pmatrix}$ (formula for the 2-parameter models)	
generalized chain ladder	$j \cdot \overline{\hat{X}}_j + \frac{j^2}{n-j}\frac{\left(\overline{\hat{V}}_j\right)^2 \cdot \overline{X}_j - 2\overline{\hat{V}}_j \cdot \overline{V}_j \cdot \overline{\hat{X}}_j + \left(\overline{\hat{X}}_j\right)^2 \cdot \overline{V_j^2/X_j}}{\overline{V_j^2/X_j} \cdot \overline{X}_j - \left(\overline{V}_j\right)^2}$	$\overline{X}_j \cdot \left(j + \frac{j^2}{n-j}\right)$
generalized chain ladder with constant volume	$j \cdot \overline{\hat{X}}_j + \frac{j^2}{n-j}\frac{\overline{X}_j - 2 \cdot \overline{\hat{X}}_j + \left(\overline{\hat{X}}_j\right)^2 \cdot \overline{1/X_j}}{\overline{1/X_j} \cdot \overline{X}_j - 1}$	$\overline{X}_j \cdot \left(j + \frac{j^2}{n-j}\right)$
chain ladder (1-parameter model)	$j \cdot \overline{\hat{X}}_j + \frac{\left(j \cdot \overline{\hat{X}}_j\right)^2}{(n-j) \cdot \overline{X}_j} = \left(\overline{\hat{X}}_j\right)^2\left(\frac{j}{\overline{\hat{X}}_j} + \frac{j^2}{(n-j) \cdot \overline{X}_j}\right)$	$\overline{X}_j \cdot \left(j + \frac{j^2}{n-j}\right)$
generalized linear regression	$j + \frac{j^2}{n-j}\frac{\left(\overline{\hat{V}}_j\right)^2 \cdot \overline{X_j^2} - 2\overline{\hat{V}}_j \cdot \overline{\hat{X}}_j \cdot \left(\overline{V_j \cdot X_j}\right) + \left(\overline{\hat{X}}_j\right)^2 \cdot \overline{V_j^2}}{\overline{X_j^2} \cdot \overline{V_j^2} - \left(\overline{V_j \cdot X_j}\right)^2}$	$j + \frac{j^2}{n-j}$
linear regression with inception	$j + \frac{j^2}{n-j} \cdot \frac{\overline{X_j^2} - 2\overline{\hat{X}}_j\overline{X}_j + \left(\overline{\hat{X}}_j\right)^2}{\overline{X_j^2} - \left(\overline{X}_j\right)^2}$	$j + \frac{j^2}{n-j}$
incremental claim ratio method (1-parameter model)	$j + \frac{j^2}{n-j} \cdot \left(\frac{\overline{\hat{V}}_j}{\overline{V}_j}\right)^2$	$j + \frac{j^2}{n-j}$

mean of experienced claims only depends on the risk measure.

Appendix C. The dispensability of the correction term $\hat{\sigma}_j$ in the chain-ladder models using the "LINEST" function in Microsoft Excel

The terms returned by the array function "LINEST" for the generalized chain-ladder, the chain-ladder and the generalized linear regression models are (cf. Section 6):

$$\begin{pmatrix} \hat{c}_j & \hat{f}_j & \frac{\sum_{k=n-j+1}^{n} \hat{X}_{k,j+1}}{\left(j \cdot \overline{\hat{X}}_j\right)^{1/2}} \\ \hat{Err}^{Par}\left(\hat{c}_j\right)^{1/2} & \hat{Err}^{Par}\left(\hat{f}_j\right)^{1/2} & \frac{\hat{MSEP}_j^{1/2}}{\left(j \cdot \overline{\hat{X}}_j\right)^{1/2}} \end{pmatrix} \tag{GCL}$$

$$\begin{pmatrix} \hat{f}_j & \frac{\sum_{k=n-j+1}^{n} \hat{X}_{k,j+1}}{\left(j \cdot \overline{\hat{X}}_j\right)^{1/2}} \\ \hat{Err}^{Par}\left(\hat{f}_j\right)^{1/2} & \frac{\hat{MSEP}_j^{1/2}}{\left(j \cdot \overline{\hat{X}}_j\right)^{1/2}} \end{pmatrix} \tag{CL}$$

$$\begin{pmatrix} \hat{c}_j & \hat{f}_j & \sum_{k=n-j+1}^{n} \hat{X}_{k,j+1} \\ Err^{Par}\left(\hat{c}_j\right)^{1/2} & Err^{Par}\left(\hat{f}_j\right)^{1/2} & \left(\hat{MSEP}_j - (j-1)\hat{\sigma}_j^2\right)^{1/2} \\ * & \hat{\sigma}_j & * \end{pmatrix} \tag{GLR}$$

In the chain-ladder models (GCL) and (CL), there is no need for such a correction term involving $\hat{\sigma}_j$, because multiplying the prediction error for the linear regression by $\left(j \cdot \overline{\hat{X}}_j\right)$—where the process variance considers only one additional row, i.e., the term 1 instead of j—and then replacing the transformed coordinates denoted by the tilde (~) with the original ones yields the appropriate formula in the chain-ladder model:

$$\left(j\cdot\overline{\hat{X}_j}\right)\left(\begin{array}{c} 1(\textit{as provided by "LINEST"}) \\ + \dfrac{j^2}{n-j}\dfrac{\left(\begin{array}{c}\left(\overline{\tilde{\hat{V}}_j}\right)^2\cdot\overline{\tilde{X}_j^2} - 2\overline{\tilde{\hat{V}}_j}\cdot\overline{\tilde{\hat{X}}_j} \\ \cdot\left(\overline{\tilde{V}_j\cdot\tilde{X}_j}\right) + \left(\overline{\tilde{\hat{X}}_j}\right)^2\cdot\overline{\tilde{V}_j^2}\end{array}\right)}{\overline{\tilde{X}_j^2}\cdot\overline{\tilde{V}_j^2} - \left(\overline{\tilde{V}_j\cdot\tilde{X}_j}\right)^2}\end{array}\right)$$

$$= j\cdot\overline{\hat{X}_j} + \frac{j^2}{n-j}\frac{\left(\begin{array}{c}\left(\overline{\hat{V}_j}\right)^2\cdot\overline{X_j} - 2\overline{\hat{V}_j}\cdot\overline{V_j}\cdot\overline{\hat{X}_j}\\ +\left(\overline{\hat{X}_j}\right)^2\cdot\overline{V_j^2/X_j}\end{array}\right)}{\overline{V_j^2/X_j}\cdot\overline{X_j} - \left(\overline{V_j}\right)^2}$$

(cf. Table B.1)

This transformation makes use of the identities

$$j\cdot\overline{\hat{X}_j}\cdot\left(\overline{\tilde{\hat{V}}_j}\right)^2 = \left(\overline{\hat{V}_j}\right)^2,\ j\cdot\overline{\hat{X}_j}\cdot\overline{\tilde{\hat{V}}_j}\cdot\overline{\tilde{\hat{X}}_j} = \overline{\hat{V}_j}\cdot\overline{\hat{X}_j},$$

$$j\cdot\overline{\hat{X}_j}\cdot\left(\overline{\tilde{\hat{X}}_j}\right)^2 = \left(\overline{\hat{X}_j}\right)^2 \text{ and}$$

$\overline{X_j} = \overline{\tilde{X}_j^2}$, $\overline{V_j} = \overline{\tilde{V}_j\cdot\tilde{X}_j}$, $\overline{V_j^2/X_j} = \overline{\tilde{V}_j^2}$, where the terms in the coordinates denoted by the tilde correspond to the averages of the first $n-j$ accident years in the transformed coordinates, that is $\overline{\tilde{X}_j^2} = \frac{1}{n-j}\sum_{k=1}^{n-j}\tilde{X}_{k,j}^2$, $\overline{\tilde{V}_j\cdot\tilde{X}_j} = \frac{1}{n-j}\sum_{k=1}^{n-j}\tilde{V}_k\tilde{X}_{k,j}$ and $\overline{\tilde{V}_j^2} = \frac{1}{n-j}\sum_{k=1}^{n-j}\tilde{V}_k^2$.

Appendix D. Estimation of the factor $\hat{\sigma}_j^2$ for the random part of the age-to-age development

If we use the direct calculation method presented in Section 6, we also avoid having to explicitly calculate the well-known estimator $\hat{\sigma}_j^2$ as described below. We therefore distinguish four cases, depending on the number of parameters, *# parameter,* which may be one or two, and on the risk metric. The general case is described by

$$\hat{\sigma}_j^2 = \sum_{k=1}^{n-j}\frac{1}{\left(X_{k,j}\right)^{\rho}}\cdot\frac{\left(\hat{X}_{k,j+1} - X_{k,j+1}\right)^2}{n-j-\#\ parameter},$$

where the *risk measure coefficient* ρ is set to ρ = 0 for the constant risk measure $W_{ii} = 1$, and to ρ = 1 for the proportional risk measure $W_{ii} = X_{i,j}$ as in the chain-ladder models.

$\hat{X}_{k,j+1}$ is given by the estimated development of $X_{k,j}$ in year $j+1$, which in the two-parameter models is determined by

$$X_j^*\cdot\hat{f}_j^* = \hat{X}'_{j+1} = \begin{pmatrix}\hat{X}_{1,j+1}\\ \vdots \\ \hat{X}_{n-j,j+1}\end{pmatrix},$$

and analogously in the one-parameter models. Hence, $\hat{\sigma}_j^2$ estimates the square of the deviations of the predicted claims from the experienced claims, weighted by the inverse of the risk measure in the corresponding model. The number of observations is reduced by the number of parameters, thus only dividing by the number of overdetermined observations.

In the affine models the process variances cannot be estimated for the last two development years $n-2$ and $n-1$, because here the two corresponding development parameters are determined or even underdetermined, respectively, by the observations. Usually, these variances are estimated by resorting to the estimates for previous years, and by assuming certain regularities in their development over time. Here, we use the estimator mentioned in the article by Mack (1993):

$$\hat{\sigma}_{n-t} = \min\left\{\hat{\sigma}_{n-t-1}^2/\hat{\sigma}_{n-t-2},\hat{\sigma}_{n-t-2},\hat{\sigma}_{n-t-1}\right\}$$

$$\text{for } t=1 \text{ and } t=2.$$

Within the usual one-parameter approach such as the chain-ladder method, only the last term, $\hat{\sigma}_{n-1}$, cannot be derived from the observation. Therefore, only the case $t = 1$ is found in the literature.

Moreover, in the affine case, even the two development parameters are undetermined for the last development year, i.e., the development from year n-1 to year n. Therefore, we suggest neglecting the additive parameter for this last year, thus reducing, e.g., the generalized chain-ladder model to the traditional chain-ladder model.

Appendix E. The derivation of *MSEP* as the sum of the scaled-up age-to-age prediction errors

The following correspondence for the prediction error MSEP,

$$MSEP = E\left[\left(\sum_{k=2}^{n}\hat{X}_{k,n} - \sum_{k=2}^{n}X_{k,n}\right)^2\right] = \sum_{j=1}^{n=1} MSEP_j \cdot f_{j:n}^2,$$

will be shown by induction on the development year n.

$$MSEP = E\left[\left(\sum_{k=2}^{n}\hat{X}^*_{k,n-1}\cdot\hat{f}^*_{n-1} - \sum_{k=2}^{n}X^*_{n-1}\cdot f^*_{n-1}\right)^2\right]$$

$$= E\left[\begin{pmatrix}\sum_{k=2}^{n}\hat{f}_{n-1}\cdot\hat{X}_{k,n-1} + \hat{c}_{n-1}\cdot V_k \\ -\sum_{k=2}^{n} f_{n-1}X_{k,n-1} - c_{n-1}\cdot V_k\end{pmatrix}^2\right]$$

$$= E\left[\left(\sum_{k=2}^{n}\hat{f}_{n-1}\cdot\hat{X}_{k,n-1} - \sum_{k=2}^{n} f_{n-1}X_{k,n-1} + C\right)^2\right]$$

with $C = \sum_{k=2}^{n}(\hat{c}_{n-1} - c_{n-1})V_k$

$$= E\left[\begin{pmatrix}\hat{f}_{n-1}\sum_{k=2}^{n}\hat{X}_{k,n-1} - f_{n-1}\sum_{k=2}^{n}\hat{X}_{k,n-1} + C \\ + f_{n-1}\sum_{k=2}^{n}\hat{X}_{k,n-1} - f_{n-1}\sum_{k=2}^{n}X_{k,n-1}\end{pmatrix}^2\right]$$

$$\stackrel{(a)}{=} E\left[\left(\hat{f}_{n-1}\sum_{k=2}^{n}\hat{X}_{k,n-1} - f_{n-1}\sum_{k=2}^{n}\hat{X}_{k,n-1} + C\right)^2\right]$$

$$+ E\left[\left(f_{n-1}\sum_{k=2}^{n}\hat{X}_{k,n-1} - f_{n-1}\sum_{k=2}^{n}X_{k,n-1}\right)^2\right]$$

$$\stackrel{(*)}{=} E\left[\left(\left(\hat{f}_{n-1} - f_{n-1}\right)\cdot\sum_{k=2}^{n}\hat{X}_{k,n-1} + C\right)^2\right]$$

$$+ f_{n-1}^2 E\left[\left(\sum_{k=2}^{n}\hat{X}_{k,n-1} - \sum_{k=2}^{n}X_{k,n-1}\right)^2\right]$$

$$\stackrel{(b)}{=} MSEP_{n-1} + \sum_{j=1}^{n-2} MSEP_j \cdot f_{j:n-1}^2 \cdot f_{n-1}^2$$

$$= \sum_{j=1}^{n-1} MSEP_j \cdot f_{j:n}^2.$$

Dealing with (a), the mixed terms drop out, which is essentially due to the assumption that the age-to-age process is memoryless. This means that the estimators $\hat{f}_{n-1}$ and $\hat{c}_{n-1}$ based on the development from year $n-1$ to year n are independent from the claims development up to year $n-1$, i.e., from $\hat{X}_{k,n-1}$ and $X_{k,n-1}$, and therefore

$$f_{n-1}\cdot E\left[\begin{pmatrix}\left(\hat{f}_{n-1} - f_{n-1}\right)\cdot\left(\sum_{k=2}^{n}\hat{X}_{k,n-1}\right) \\ + \left(\hat{c}_{n-1} - c_{n-1}\right)\sum_{k=2}^{n}V_k\end{pmatrix} \cdot\left(\sum_{k=2}^{n}\hat{X}_{k,n-1} - \sum_{k=2}^{n}X_{k,n-1}\right)\right]$$

$$= f_{n-1}\cdot E\left[\hat{f}_{n-1} - f_{n-1}\right]$$

$$\cdot E\left[\left(\sum_{k=2}^{n}\hat{X}_{k,n-1}\right)\cdot\left(\sum_{k=2}^{n}\hat{X}_{k,n-1} - \sum_{k=2}^{n}X_{k,n-1}\right)\right]$$

$$+ f_{n-1}\cdot E[\hat{c}_{n-1} - c_{n-1}]\cdot\sum_{k=2}^{n}V_k$$

$$\cdot E\left[\sum_{k=2}^{n}\hat{X}_{k,n-1} - \sum_{k=2}^{n}X_{k,n-1}\right]$$

$$= f_{n-1}\cdot\begin{pmatrix}\left(E\left[\hat{f}_{n-1}\right] - E[f_{n-1}]\right)\cdot\ldots \\ + \left(E[\hat{c}_{n-1}] - E[c_{n-1}]\right)\cdot\ldots\end{pmatrix} = 0,$$

because $\hat{c}_{n-1}$ as well as $\hat{f}_{n-1}$ are unbiased estimators, i.e., $E[\hat{c}_{n-1}] = c_{n-1}$ and $E[\hat{f}_{n-1}] = f_{n-1}$.

Dealing with (b), we observe that $MSEP_{n-1}$ corresponds to *MSEP* given the history H_{n-1}, i.e., given the claims development up to age $n-1$. Thus the derivation of *MSEP* outlined above equally holds for $MSEP_{n-1}$, if the history H_{n-1} is assumed., because given the history H_{n-1}, the second term of the development of *MSEP* in the equation noted by (*) drops out, which leads to the relationship $MSEP_{n-1} = E[((\hat{f}_{n-1} - f_{n-1})\sum_{k=2}^{n}\hat{X}_{k,n-1} + \sum_{k=2}^{n}(\hat{c}_{n-1} - c_{n-1})V_k)^2]$ used in the development (b), with the notation $C = \sum_{k=2}^{n}(\hat{c}_{n-1} - c_{n-1})V_k$ as introduced before.

References

Barnett, G., and B. Zehnwirth, "Best Estimate for Reserves," *Proceedings of the Casualty Actuarial Society* 87, Part 2, 2000, pp. 245–321.

Brosius, E., "Loss Development Using Credibility," CAS Study Note, March 1993.

Fahrmeir, L., A. Hammerle, and G. Tutz, *Multivariate statistische Verfahren,* Berlin: de Gruyter, 1996.

Halliwell, L. J., "Chain-Ladder Bias: Its Reason and Meaning," *Variance* 1, 2007, pp. 214–247.

Ludwig, A., and K. D. Schmidt, "Gauss-Markov Loss Prediction in a Linear Model," Casualty Actuarial Society *E-Forum,* Fall 2010.

Mack, T., "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *ASTIN Bulletin* 23:2, 1993, pp. 213–225, http://www.actuaries.org/LIBRARY/ASTIN/vol23no2/213.pdf.

Panning, W. H., "Measuring Loss Reserve Uncertainty," 36th ASTIN Colloquium, Zurich, 2005.

Schnieper, R., "Separating True IBNR and IBNER Claims," *ASTIN Bulletin* 21:1, 1991, pp. 111–127.

Venter, G., and B. Zehnwirth, "Testing the Assumptions of Age-to-Age Factors," *Proceedings of the Casualty Actuarial Society* 85, 1998, pp. 807–847.