AN EXTENSION OF MACK'S MODEL FOR THE CHAIN LADDER METHOD

KLAUS D. SCHMIDT AND ANJA SCHNAUS

ABSTRACT

The chain ladder method is a simple and suggestive tool in claims reserving, and various attempts have been made aiming at its justification in a stochastic model. Remarkable progress has been achieved by Schnieper and Mack who considered models involving assumptions on conditional distributions. The present paper extends the model of Mack and proposes a basic model in a decision theoretic setting. The model allows to characterize optimality of the chain ladder factors as predictors of non-observable development factors and hence optimality of the chain ladder predictors of aggregate claims at the end of the first non-observable calendar year. We also present a model in which the chain ladder predictor of ultimate aggregate claims turns out to be unbiased.

1. INTRODUCTION

The chain ladder method is a simple and suggestive tool in claims reserving, and various attempts have been made aiming at its justification in a stochastic model. Remarkable progress has been achieved by Schnieper [1991] and Mack [1993,1994a, 1994b] who considered models involving assumptions on conditional distributions.

The present paper proposes a basic model in a decision theoretic setting (Section 2) which is analyzed on the background of a general result on conditional prediction (Section 3). The model allows to characterize optimality of the chain ladder factors as predictors of non-observable development factors and hence optimality of the chain ladder predictors of aggregate claims at the end of the first non-observable calendar year (Section 4).

The model considered here is exclusively based on assumptions on the conditional joint distribution (with respect to the past over all occurrence years) of the collection of all development factors from a given development year; by contrast, the model of Mack assumes unconditional independence of the occurrence years and certain properties of the conditional distributions of single development factors. Since our model properly extends the model of Mack (Section 5), we obtain a justification of the chain ladder method under strictly weaker assumptions.

We also present a partial solution to the prediction problem for ultimate aggregate claims: It is shown that in another model which again properly extends the model of

ASTIN BULLETIN, Vol. 26, No. 2, 1996, pp. 247-262

Mack the chain ladder predictor of ultimate aggregate claims is unbiased but shares this property with many other predictors (Section 6). Optimality of the chain ladder predictor of ultimate aggregate claims remains an open problem.

Throughout this paper, let $(\Omega, \mathcal{F}, P)$ be a probability space. We assume that all random variables under consideration have finite second moments.

2. THE PREDICTION PROBLEM AND THE BASIC MODEL

Consider a family of random variables $\{S_{i,k}\}_{i,k \in \{0,1,\ldots,n\}}$. The random variable $S_{i,k}$ is interpreted as the aggregate claim size of all claims which occur in *occurrence year* i and which are settled before the end of *calendar year* $i + k$. We also interpret the subscript k as the *development year.*

We assume that the *aggregate claims* $S_{i,k}$ are strictly positive and that they are observable for $i + k \leq n$ but non-observable for $i + k > n$. The observable aggregate claims can be represented by the *run-off triangle*:

Occurrence year	**0**	**1**	...	***Development year*** $n-i$	$n-i+1$	...	$n-1$	n
0	$S_{0,0}$	$S_{0,1}$	...	$S_{0,n-i}$	$S_{0,n-i+1}$	...	$S_{0,n-1}$	$S_{0,n}$
1	$S_{1,0}$	$S_{1,1}$	...	$S_{1,n-i}$	$S_{1,n-i+1}$	...	$S_{1,n-1}$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$			
i – 1	$S_{i-1,0}$	$S_{i-1,1}$	...	$S_{i-1,n-i}$	$S_{i-1,n-i+1}$			
i	$S_{i,0}$	$S_{i,1}$	...	$S_{i,n-i}$				
$\vdots$	$\vdots$	$\vdots$						
n – 1	$S_{n-1,0}$	$S_{n-1,1}$						
n	$S_{n,0}$							

The problem is to predict the non-observable aggregate claims from the observable ones.

The chain ladder method consists in using the *chain ladder predictors*

$$\hat{S}_{i,m} := S_{i,n-i} \cdot \prod_{l=n-i+1}^{m} \hat{F}_l$$

for all $i \in \{1, \ldots, n\}$ and $m \in \{n-i+1, \ldots, n\}$, where the *chain ladder factors* $\hat{F}_l$ are defined by

$$\hat{F}_l := \frac{\sum_{i=0}^{n-l} S_{i,l}}{\sum_{i=0}^{n-l} S_{i,l-1}}$$

for all $l \in \{1, \ldots, n\}$.

In order to study the properties of the chain ladder factors and of the chain ladder predictors, we introduce the *development factors*

$$F_{i,l} := \frac{S_{i,l}}{S_{i,l-1}}$$

for all $i \in \{0, 1, \ldots, n\}$ and $l \in \{1, \ldots, n\}$. Then the aggregate claims satisfy

$$S_{i,m} = S_{i,n-i} \cdot \prod_{l=n-i+1}^{m} F_{i,l}$$

for all $i \in \{0, 1, \ldots, n\}$ and $m \in \{n-i+1, \ldots, n\}$, and the chain ladder factors can be written as

$$\hat{F}_l = \sum_{i=0}^{n-l} \frac{S_{i,l-1}}{\sum_{j=0}^{n-l} S_{j,l-1}} F_{i,l}$$

for all $l \in \{1, \ldots, n\}$.

Let us now change the point of view by turning from occurrence years to development years.

Occurrence year				*Development year*				
	0	**1**	...	$k-1$	k	...	$n-1$	n
0	$S_{0,0}$	$S_{0,1}$	...	$S_{0,k-1}$	$F_{0,k}$	...	...	...
1	$S_{1,0}$	$S_{1,1}$	...	$S_{1,k-1}$	$F_{1,k}$	...	...	...
⋮	⋮	⋮		⋮	⋮			
$n-k$	$S_{n-k,0}$	$S_{n-k,1}$	...	$S_{n-k,k-1}$	$F_{n-k,k}$	...	...	...
$n-k+1$	$S_{n-k+1,0}$	$S_{n-k+1,1}$	...	$S_{n-k+1,k-1}$	$F_{n-k+1,k}$	...	...	...
⋮	⋮	⋮		⋮	⋮			
$n-1$	$S_{n-1,0}$	$S_{n-1,1}$	...	$S_{n-1,k-1}$	$F_{n-1,k}$	...	...	...
n	$S_{n,0}$	$S_{n,1}$	...	$S_{n,k-1}$	$F_{n,k}$	...	...	...

First of all, it is easy to see that for each $k \in \{1, \ldots, n\}$ the chain ladder factor $\hat{F}_k$ minimizes the expression

$$\sum_{i=0}^{n-k} \frac{S_{i,k-1}}{\sum_{j=0}^{n-k} S_{j,k-1}} (F_{i,k} - \delta)^2$$

over all random variables δ. Thus, for development year k, the chain ladder factor $\hat{F}_k$ is the best *approximation* of the *observable* development factors when the approximation errors are given the weights occurring in the representation of the chain ladder factor as a weighted mean.

In what follows we shall study optimality of the chain ladder factors as *predictors* of *non-observable* development factors. To this end, we first formulate the *prediction problem* and then state the *basic model*:

Prediction Problem: For $k \in \{1, \ldots, n\}$, let $\mathcal{G}_k$ denote the σ-algebra generated by the family of random variables

$$\left\{S_{i,l}\right\}_{i \in \{0,1,\ldots,n\}, l \in \{0,1,\ldots,k-1\}}$$

and let Δ_k denote the collection of all random variables δ which can be written as

$$\delta = \sum_{i=0}^{n-k} W_i F_{i,k}$$

where the weights of the development factors are $\mathcal{G}_k$-measurable random variables satisfying

$$\sum_{i=0}^{n-k} W_i = 1 .$$

For each $j \in \{n-k+1, \ldots, n\}$, the problem is to find some $\delta_j^* \in \Delta_k$ satisfying

$$E\left((F_{j,k} - \delta_j^*)^2 \,\middle|\, \mathcal{G}_k\right) = \inf_{\delta \in \Delta_k} E\left((F_{j,k} - \delta)^2 \,\middle|\, \mathcal{G}_k\right).$$

These quantities can be interpreted as follows:
- The σ-algebra $\mathcal{G}_k$ represents the information provided by the past preceding development year k.
- The non-observable development factors are to be predicted by a weighted mean of observable development factors from the same development year such that the weights are measurable functions of the aggregate claims in the past. (It is not assumed that the weight are positive.)
- The optimality criterion is conditional expected squared error loss, given the information provided by the aggregate claims in the past.

The conditional loss function instead of the usual unconditional one is reasonable since optimality is desired only with regard to the information provided by the past.

Basic Model: For each $k \in \{1, \ldots, n\}$, there exists a random variable F_k such that

$$E(F_{i,k} \mid \mathcal{G}_k) = F_k$$

$$\operatorname{cov}(F_{i,k}, F_{j,k} \mid \mathcal{G}_k) = 0$$

$$\operatorname{var}(F_{i,k} \mid \mathcal{G}_k) > 0$$

holds for all $i, j \in \{0, 1, \ldots, n\}$ such that $i \neq j$.

The following lemma is of interest with regard to the model of Mack which will be studied in Section 5:

2.1. Lemma *Under the assumptions of the basic model and for each $k \in \{1, \ldots, n\}$, the following are equivalent:*

(a) *There exists a real number f_k such that*

$$F_k = f_k.$$

(b) *The identity*

$$\text{cov}[F_{i,k}, F_{j,k}] = 0$$

holds for all $i, j \in \{0, 1, \ldots, n\}$ such that $i \neq j$.

The prediction problem for the basic model will be studied in Section 4 below.

3. CONDITIONAL PREDICTION

In the present section, we study an abstract prediction problem which will later be applied to the prediction of non-observable development factors.

Throughout this section, let $\{X_i\}_{i \in \{1,\ldots,m+1\}}$ be a family of random variables and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. We assume that there exists a random variable X such that

$$E(X_i \mid \mathcal{G}) = X$$

$$\text{cov}(X_i, X_j \mid \mathcal{G}) = 0$$

$$\text{var}(X_i \mid \mathcal{G}) > 0$$

holds for all $i, j \in \{1, \ldots, m, m+1\}$ such that $i \neq j$. We also assume that the random variables $X_1, \ldots, X_m$ are observable whereas X_{m+1} is non-observable.

Let Δ denote the collection of all random variables δ which can be written as

$$\delta = \sum_{i=1}^{m} W_i X_i$$

where the weights are $\mathcal{G}$-measurable random variables satisfying

$$\sum_{i=1}^{m} W_i = 1.$$

The random variables in Δ are called *admissible predictors* of X_{m+1}.

The problem is to find some $\hat{\delta} \in \Delta$ satisfying

$$E((X_{m+1} - \hat{\delta})^2 \mid \mathcal{G}) = \inf_{\delta \in \Delta} E((X_{m+1} - \delta)^2 \mid \mathcal{G}),$$

that is, to predict the non-observable random variable X_{m+1} by a weighted mean of the observable ones such that the weights contain information from outside the sample $\{X_1, \ldots, X_m\}$ and such that conditional expected squared error loss in minimized.

Remark. The *classical case* is the case where $\mathcal{G} = \{\emptyset, \Omega\}$, which means that
- no information from outside the sample is available,
- the random variables $X_1, \ldots, X_m, X_{m+1}$ are uncorrelated with equal expectations and strictly positive variances,
- the admissible predictors have constant weights, and
- the optimality criterion is unconditional expected squared error loss.

The following lemma is immediate:

3.1. Lemma. *The identities*

$$E(\delta \mid \mathcal{G}) = X$$

and

$$E((X_{m+1} - \delta)^2 \mid \mathcal{G}) = \mathrm{var}(X_{m+1} \mid \mathcal{G}) + \mathrm{var}(\delta \mid \mathcal{G})$$

hold for all $\delta \in \Delta$.

The following result establishes existence, uniqueness, and the form of the weights of the optimum predictor of X_{m+1}:

3.2. Theorem. *For*

$$\hat{\delta} = \sum_{i=1}^{m} \hat{W}_i X_i \in \Delta,$$

the following are equivalent:

(a) *There exists a random variable Λ such that*

$$\hat{W}_i = \frac{\Lambda}{\mathrm{var}(X_i \mid \mathcal{G})}$$

holds for all $i \in \{1, \ldots, m\}$.

(b) *The inequality*

$$E((X_{m+1} - \hat{\delta})^2 \mid \mathcal{G}) \leq E((X_{m+1} - \delta)^2 \mid \mathcal{G})$$

holds for all $\delta \in \Delta$.

In this case,

$$\mathrm{var}(\hat{\delta} \mid \mathcal{G}) = \Lambda = \left(\sum_{i=1}^{m} \frac{1}{\mathrm{var}(X_i \mid \mathcal{G})} \right)^{-1},$$

as well as

$$E\left(\frac{1}{m-1} \sum_{i=1}^{m} \hat{W}_i (X_i - \hat{\delta})^2 \,\middle|\, \mathcal{G} \right) = \Lambda$$

when $m \geq 2$.

Proof. Define

$$\Lambda := \left(\sum_{i=1}^{m} \frac{1}{\mathrm{var}(X_i \mid \mathcal{G})} \right)^{-1}$$

and let

$$W_i^* := \frac{\Lambda}{\mathrm{var}(X_i \mid \mathcal{G})}$$

for all $i \in \{1, \ldots, m\}$. For each

$$\delta = \sum_{i=1}^{m} W_i X_i \in \Delta,$$

we have

$$\begin{aligned}
\mathrm{var}(\delta \mid \mathcal{G}) &= \mathrm{var}\left(\sum_{i=1}^{m} W_i X_i \mid \mathcal{G} \right) \\
&= \sum_{i=1}^{m} W_i^2 \,\mathrm{var}(X_i \mid \mathcal{G}) \\
&= \sum_{i=1}^{m} \left(W_i - W_i^* \right)^2 \mathrm{var}(X_i \mid \mathcal{G}) + 2 \sum_{i=1}^{m} W_i W_i^* \,\mathrm{var}(X_i \mid \mathcal{G}) - \sum_{i=1}^{m} \left(W_i^* \right)^2 \mathrm{var}(X_i \mid \mathcal{G}) \\
&= \sum_{i=1}^{m} \left(W_i - W_i^* \right)^2 \mathrm{var}(X_i \mid \mathcal{G}) + 2\Lambda \sum_{i=1}^{m} W_i - \Lambda \sum_{i=1}^{m} W_i^* \\
&= \sum_{i=1}^{m} \left(W_i - W_i^* \right)^2 \mathrm{var}(X_i \mid \mathcal{G}) + \Lambda.
\end{aligned}$$

Because of Lemma 3.1, this proves the equivalence of (a) and (b) as well as the identity for $\mathrm{var}(\hat{\delta} \mid \mathcal{G})$. The final identity follows by straightforward computation.

Remark. In the special case where there exists a random variable V satisfying $\operatorname{var}(X_i|\mathcal{G}) = V$ for all $i \in \{1, \ldots, m\}$, the optimum predictor of X_{m+1} is the sample mean

$$\overline{X} := \frac{1}{m}\sum_{i=1}^{m} X_i$$

and we have

$$\operatorname{var}(\overline{X} \mid \mathcal{G}) = E\left(\frac{1}{m}\frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2 \,\middle|\, \mathcal{G}\right).$$

In the classical case, this reduces to the well-known fact that

$$\frac{1}{m}\frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2$$

is an unbiased estimator of the variance of the sample mean.

4. THE RESULTS

We now turn to the prediction problem for the basic model. Consider $k \in \{1, \ldots, n\}$.

4.1. Lemma. *Under the assumptions of the basic model, the identities*

$$E(\delta \mid \mathcal{G}_k) = F_k$$

and

$$E((F_{j,k} - \delta)^2 \mid \mathcal{G}_k) = \operatorname{var}(F_{j,k} \mid \mathcal{G}_k) + \operatorname{var}(\delta \mid \mathcal{G}_k)$$

hold for all $\delta \in \Delta_k$ and for all $j \in \{n-k+1, \ldots, n\}$.

This is immediate from Lemma 3.1.

The following result characterizes optimality of the chain ladder factor:

4.2. Theorem. *Under the assumptions of th basic model, the following are equivalent:*

(a) *There exists a random variable V_k such that*

$$\operatorname{var}(F_{i,k} \mid \mathcal{G}_k) = \frac{V_k}{S_{i,k-1}}$$

holds for all $i \in \{0, \ldots, n-k\}$.

(b) *The inequality*

$$E((F_{j,k} - \hat{F}_k)^2 \mid \mathcal{G}_k) \leq E((F_{j,k} - \delta)^2 \mid \mathcal{G}_k)$$

holds for all $\delta \in \Delta_k$ and for some $j \in \{n-k+1, \ldots, n\}$.

(c) *The inequality*

$$E((F_{j,k} - \hat{F}_k)^2 \mid \mathcal{G}_k) \leq E((F_{j,k} - \delta)^2 \mid \mathcal{G}_k)$$

holds for all $\delta \in \Delta_k$ *and for all* $j \in \{n-k+1, \ldots, n\}$.

In this case, $\delta^* = \hat{F}_k$ *holds for each* $\delta^* \in \Delta_k$ *such that*

$$E((F_{j,k} - \delta^*)^2 \mid \mathcal{G}_k) \leq E((F_{j,k} - \delta)^2 \mid \mathcal{G}_k)$$

holds for all $\delta \in \Delta_k$ *and for some* $j \in \{n-k+1, \ldots, n\}$; *moreover,*

$$\operatorname{var}(\hat{F}_k \mid \mathcal{G}_k) = \frac{V_k}{\sum_{i=0}^{n-k} S_{i,k-1}} = \left(\sum_{i=0}^{n-k} \frac{1}{\operatorname{var}(F_{i,k} \mid \mathcal{G}_k)}\right)^{-1},$$

as well as

$$E\left(\frac{1}{n-k}\sum_{i=0}^{n-k} \frac{S_{i,k-1}}{\sum_{l=0}^{n-k} S_{l,k-1}}(F_{i,k} - \hat{F}_k)^2 \;\middle|\; \mathcal{G}_k\right) = \operatorname{var}(\hat{F}_k \mid \mathcal{G}_k)$$

when $k \leq n-1$.

Proof. By Theorem 3.2, the chain ladder factor

$$\hat{F}_k := \sum_{i=0}^{n-k} \frac{S_{i,k-1}}{\sum_{l=0}^{n-k} S_{l,k-1}} F_{i,k}$$

minimizes conditional expected squared error loss if and only if the identity

$$\frac{S_{i,k-1}}{\sum\limits_{l=0}^{n-k} S_{l,k-1}} = \frac{\dfrac{1}{\operatorname{var}(F_{i,k} \mid \mathcal{G}_k)}}{\sum\limits_{l=0}^{n-k} \dfrac{1}{\operatorname{var}(F_{l,k} \mid \mathcal{G}_k)}}$$

holds for all $i \in \{0, 1, \ldots, n-k\}$, and this identity is equivalent with

$$\operatorname{var}(F_{i,k} \mid \mathcal{G}_k) = \frac{1}{S_{i,k-1}} \cdot \frac{\sum\limits_{l=0}^{n-k} S_{l,k-1}}{\sum\limits_{l=0}^{n-k} \dfrac{1}{\operatorname{var}(F_{l,k} \mid \mathcal{G}_k)}}.$$

This yields the equivalence of (a) and (b).
The equivalence of (b) and (c) is obvious from Lemma 4.1, and the final assertion follows from Theorem 3.2.

The previous result suggests the definition of the following *general model*:

General Model: For each $k \in \{1, \ldots, n\}$, there exist random variables F_k and $V_k > 0$ such that

$$E(F_{i,k} \mid \mathcal{G}_k) = F_k$$
$$\mathrm{cov}(F_{i,k}, F_{j,k} \mid \mathcal{G}_k) = 0$$
$$\mathrm{var}(F_{i,k} \mid \mathcal{G}_k) = \frac{V_k}{S_{i,k-1}}$$

holds for all $i, j \in \{0, 1, \ldots, n\}$ such that $i \neq j$.

4.3. Corollary. *Under the assumptions of the general model, the chain ladder factors satisfy*

$$E(\hat{F}_k \mid \mathcal{G}_k) = F_k$$

and

$$\mathrm{var}(\hat{F}_k \mid \mathcal{G}_k) = \frac{V_k}{\sum_{i=0}^{n-k} S_{i,k-1}} = \inf_{\delta \in \Delta_k} \mathrm{var}(\delta \mid \mathcal{G}_k)$$

for all $k \in \{1, \ldots, n\}$ as well as

$$E((F_{j,k} - \hat{F}_k)^2 \mid \mathcal{G}_k) = \inf_{\delta \in \Delta_k} E((F_{j,k} - \delta)^2 \mid \mathcal{G}_k)$$

and

$$E((F_{j,k} - \hat{F}_k)^2 \mid \mathcal{G}_k) = \mathrm{var}(F_{j,k} \mid \mathcal{G}_k) + \mathrm{var}(\hat{F}_k \mid \mathcal{G}_k) = \frac{V_k}{S_{j,k-1}} + \frac{V_k}{\sum_{i=0}^{n-k} S_{i,k-1}}$$

for all $k \in \{1, \ldots, n\}$ and for all $j \in \{n-k+1, \ldots, n\}$; moreover, the identity

$$E\left(\frac{1}{n-k} \sum_{i=0}^{n-k} S_{i,k-1} (F_{i,k} - \hat{F}_k)^2 \,\middle|\, \mathcal{G}_k \right) = V_k$$

holds for all $k \in \{1, \ldots, n-1\}$.

Conclusion: Under the assumptions of the general model, we have, for each $\delta \in \Delta_k$,

$$E(\delta \mid \mathcal{G}_k) = E(F_{j,k} \mid \mathcal{G}_k)$$

and hence

$$\begin{aligned}
E(S_{n-k+1,k-1} \cdot \delta \mid \mathcal{G}_k) &= S_{n-k+1,k-1} \cdot E(\delta \mid \mathcal{G}_k) \\
&= S_{n-k+1,k-1} \cdot E(F_{j,k} \mid \mathcal{G}_k) \\
&= E(S_{n-k+1,k-1} \cdot F_{j,k} \mid \mathcal{G}_k) \\
&= E(S_{n-k+1,k} \mid \mathcal{G}_k),
\end{aligned}$$

and this implies that δ and $S_{n-k+1,k-1} \cdot \delta$ are unbiased predictors of $F_{j,k}$ and $S_{n-k+1,k}$, respectively; moreover, we have

$$E((F_{j,k} - \hat{F}_k)^2 \mid \mathcal{G}_k) = \inf_{\delta\in\Delta_k} E((F_{j,k} - \delta)^2 \mid \mathcal{G}_k)$$

and hence

$$E((S_{n-k+1,k} - \hat{S}_{n-k+1,k})^2 \mid \mathcal{G}_k) = \inf_{\delta\in\Delta_k} E((S_{n-k+1,k} - S_{n-k+1,k-1} \cdot \delta)^2 \mid \mathcal{G}_k),$$

which means that the chain ladder factor $\hat{F}_k$ and the chain ladder predictor $\hat{S}_{n-k+1,k} = S_{n-k+1,k-1} \cdot \hat{F}_k$ are the optimum predictors of $F_{j,k}$ and $S_{n-k+1,k}$, respectively. This solves completely the prediction problem for the first non-observable year $n + 1$.

5. THE MODEL OF MACK

For all $i, k \in \{0, 1, ..., n\}$, define

$$\mathcal{S}_{i,k} := \sigma(\{S_{i,l}\}_{l\in\{0,1,...,k\}}).$$

These σ-algebras are needed to formulate the *model of Mack*:

Model of Mack: The family of σ-algebras $\{\mathcal{S}_{i,n}\}_{i\in\{0,1,...,n\}}$ is independent and, for each $k \in \{1, ..., n\}$, there exist real numbers f_k and $v_k > 0$ such that

$$E(F_{i,k} \mid \mathcal{S}_{i,k-1}) = f_k$$

$$\operatorname{var}(F_{i,k} \mid \mathcal{S}_{i,k-1}) = \frac{v_k}{S_{i,k-1}}$$

holds for all $i \in \{0, 1, ..., n\}$.

The main problem when comparing the model of Mack with the general model consists in the fact that (unconditional) independence does not imply conditional independence (and vice versa). Nevertheless, we have the following result:

5.1. Theorem. *The model of Mack is a special case of the general model.*

Proof. Consider $k \in \{1, ..., n\}$. Since the family $\{\mathcal{S}_{i,n}\}_{i\in\{0,1,...,n\}}$ is independent, the family $\{\mathcal{S}_{i,k-1}\}_{i\in\{0,1,...,n\}}$ is independent as well. Also, for all $i \in \{0, 1, ..., n\}$, we have $\mathcal{S}_{i,k-1} \subseteq \mathcal{G}_k$. This yields

$$\begin{aligned} E(F_{i,k} \mid \mathcal{G}_k) &= E(F_{i,k} \mid \mathcal{S}_{i,k-1}) \\ &= f_k \end{aligned}$$

and

$$\begin{aligned} \operatorname{var}(F_{i,k} \mid \mathcal{G}_k) &= \operatorname{var}(F_{i,k} \mid \mathcal{S}_{i,k-1}) \\ &= \frac{v_k}{S_{i,k-1}} \quad . \end{aligned}$$

Furthermore, using independence repeatedly and in a similar manner as before, we obtain, for all $i, j \in \{0, 1, \ldots, n\}$ such that $i \neq j$,

$$\begin{aligned}
E(F_{i,k}F_{j,k} \mid \mathcal{G}_k) &= E\Big(F_{i,k}F_{j,k} \Big| \sigma(\mathcal{S}_{i,k-1} \cup \mathcal{S}_{j,k-1})\Big) \\
&= E\Big(E(F_{i,k}F_{j,k} \big| \sigma(\mathcal{S}_{i,k} \cup \mathcal{S}_{j,k-1}))\Big| \sigma(\mathcal{S}_{i,k-1} \cup \mathcal{S}_{j,k-1})\Big) \\
&= E\Big(F_{i,k} \cdot E(F_{j,k} \big| \sigma(\mathcal{S}_{i,k} \cup \mathcal{S}_{j,k-1}))\Big| \sigma(\mathcal{S}_{i,k-1} \cup \mathcal{S}_{j,k-1})\Big) \\
&= E\Big(F_{i,k} \cdot E(F_{j,k} \big| \mathcal{S}_{j,k-1})\Big| \sigma(\mathcal{S}_{i,k-1} \cup \mathcal{S}_{j,k-1})\Big) \\
&= E\Big(F_{i,k} \Big| \sigma(\mathcal{S}_{i,k-1} \cup \mathcal{S}_{j,k-1})\Big) \cdot E(F_{j,k} \big| \mathcal{S}_{j,k-1}) \\
&= E(F_{i,k} \big| \mathcal{S}_{i,k-1}) \cdot E(F_{j,k} \big| \mathcal{S}_{j,k-1}) \\
&= E(F_{i,k} \mid \mathcal{G}_k) \cdot E(F_{j,k} \mid \mathcal{G}_k)
\end{aligned}$$

and thus

$$\operatorname{cov}(F_{i,k}, F_{j,k} \mid \mathcal{G}_k) = 0$$

The assertion follows.

Because of Lemma 2.1, the model of Mack is even *properly* contained in the general model; this is also true when the random variables F_k and V_k of the general model are assumed to be constant.

6. COMPLEMENT: UNBIASED PREDICTION OF ULTIMATE AGGREGATE CLAIMS IN A MODIFIED MODEL

In the general model, the chain ladder predictor $\hat{S}_{i,n-i+1}$ is the optimum predictor of the aggregate claims $S_{i,n-i+1}$ in the first non-observable calendar year $n + 1$. By contrast, optimum prediction of the ultimate aggregate claims $S_{i,n}$ remains an open problem (except for the case $i = 1$).

Mack proved, in this model, that the chain ladder predictor of ultimate aggregate claims is unbiased. We now formulate a modification of the general model in which *every* predictor of the form

$$S_{i,n-i} \cdot \prod_{l=n-i+1}^{n} \delta_l$$

with $\delta_l \in \Delta_l$ for all $l \in \{n - i + 1, \ldots, n\}$ turns out to be an unbiased predictor of the ultimate aggregate claims.

Modified Model: For each $k \in \{1, \ldots, n\}$, there exists a random variable F_k such that

$$E(F_{i,k} \mid \mathcal{G}_k) = F_k$$

and the identity

$$\operatorname{cov}\left(F_{i,k}, \prod_{l=k+1}^{n} F_l \,\middle|\, \mathcal{G}_k \right) = 0$$

holds for all $k \in \{1, \ldots, n\}$ and $i \in \{0, 1, \ldots, n\}$.

The general model and the modified model can be combined without any problem. Moreover, if in the general model the random variables F_k are assumed to be constant, then the assumptions of the modified model are automatically fulfilled; in particular, the model of Mack is a special case of the modified model.

6.1. Lemma. *Under the assumptions of the modified model, the identities*

$$E(\delta_k \mid \mathcal{G}_k) = F_k$$

and

$$E\left(\prod_{l=k}^{m} \delta_l \cdot \prod_{l=m+1}^{n} F_l \,\middle|\, \mathcal{G}_k \right) = E\left(\prod_{l=k}^{m-1} \delta_l \cdot \prod_{l=m}^{n} F_l \,\middle|\, \mathcal{G}_k \right)$$

hold for all $k \in \{1, \ldots, n\}$ and $m \in \{k, \ldots, n\}$ and for every choice of $\delta_l \in \Delta_l$ for all $l \in \{k, \ldots, m\}$.

Proof. The first identity is obvious. Furthermore, we have

$$E\left(\delta_m \cdot \prod_{l=m+1}^{n} F_l \,\middle|\, \mathcal{G}_m \right) = E\left(\prod_{l=m}^{n} F_l \,\middle|\, \mathcal{G}_m \right)$$

and hence

$$\begin{aligned}
E\left(\prod_{l=k}^{m} \delta_l \cdot \prod_{l=m+1}^{n} F_l \,\middle|\, \mathcal{G}_k \right) &= E\left(E\left(\prod_{l=k}^{m} \delta_l \cdot \prod_{l=m+1}^{n} F_l \,\middle|\, \mathcal{G}_m \right) \,\middle|\, \mathcal{G}_k \right) \\
&= E\left(\prod_{l=k}^{m-1} \delta_l \cdot E\left(\delta_m \cdot \prod_{l=m+1}^{n} F_l \,\middle|\, \mathcal{G}_m \right) \,\middle|\, \mathcal{G}_k \right) \\
&= E\left(\prod_{l=k}^{m-1} \delta_l \cdot E\left(\prod_{l=m}^{n} F_l \,\middle|\, \mathcal{G}_m \right) \,\middle|\, \mathcal{G}_k \right) \\
&= E\left(E\left(\prod_{l=k}^{m-1} \delta_l \cdot \prod_{l=m}^{n} F_l \,\middle|\, \mathcal{G}_m \right) \,\middle|\, \mathcal{G}_k \right) \\
&= E\left(\prod_{l=k}^{m-1} \delta_l \cdot \prod_{l=m}^{n} F_l \,\middle|\, \mathcal{G}_k \right)
\end{aligned}$$

which proves the second identity.

6.2. Theorem. *Under the assumptions of the modified model, the identity*

$$E\left(S_{i,n-i} \cdot \prod_{l=n-i+1}^{n} \delta_l \,\middle|\, \mathcal{G}_{n-i+1} \right) = E(S_{i,n} \mid \mathcal{G}_{n-i+1})$$

holds for all $i \in \{0, 1, \ldots, n\}$ *and for every choice of* $\delta_l \in \Delta_l$ *for all* $l \in \{n-i+1\}$.

Proof. By Lemma 6.1, we have

$$E\left(\prod_{l=n-i+1}^{n} \delta_l \,\middle|\, \mathcal{G}_{n-i+1} \right) = E\left(\prod_{l=n-i+1}^{n} F_l \,\middle|\, \mathcal{G}_{n-i+1} \right)$$

and

$$E\left(\prod_{l=n-i+1}^{n} F_{i,l} \,\middle|\, \mathcal{G}_{n-i+1} \right) = E\left(\prod_{l=n-i+1}^{n} F_l \,\middle|\, \mathcal{G}_{n-i+1} \right),$$

and hence

$$\begin{aligned} E\left(S_{i,n-i} \cdot \prod_{l=n-i+1}^{n} \delta_l \,\middle|\, \mathcal{G}_{n-i+1} \right) &= S_{i,n-i} \cdot E\left(\prod_{l=n-i+1}^{n} \delta_l \,\middle|\, \mathcal{G}_{n-i+1} \right) \\ &= S_{i,n-i} \cdot E\left(\prod_{l=n-i+1}^{n} F_l \,\middle|\, \mathcal{G}_{n-i+1} \right) \\ &= S_{i,n-i} \cdot E\left(\prod_{l=n-i+1}^{n} F_{i,l} \,\middle|\, \mathcal{G}_{n-i+1} \right) \\ &= E\left(S_{i,n-i} \cdot \prod_{l=n-i+1}^{n} F_{i,l} \,\middle|\, \mathcal{G}_{n-i+1} \right) \\ &= E(S_{i,n} \mid \mathcal{G}_{n-i+1}) \end{aligned}$$

as was to be shown.

Conclusion: Under the assumptions of the modified model, the chain ladder predictor is an unbiased predictor of the ultimate aggregate claims, but many other predictors are unbiased as well.

In order to establish optimality, and not only unbiasedness, of the chain ladder predictor, the modified model should be restricted by additional assumptions which are in the spirit of the general model. These additional assumptions should concern products of development factors instead of single ones.

7. REMARKS

At the first glance, it may appear to be somewhat strange that σ-algebras $\mathcal{G}_k$, which are used for conditioning, include (except for the case $k = 1$) non-observable information. However, non-observable information drops out automatically in the formulas for the optimum predictors of non-observable development factors. Moreover, all results remain valid when the σ-algebras $\mathcal{G}_k$ are replaced by the σ-algebras $\mathcal{E}_k := \sigma\left(\left\{S_{i,k-1}\right\}_{i\in\{0,1,\ldots,n-k\}}\right)$ or by any σ-algebras $\mathcal{F}_k$ satisfying $\mathcal{E}_k \subseteq \mathcal{F}_k \subseteq \mathcal{G}_k$; a natural choice would be to take $\mathcal{F}_k := \mathcal{G}_k \cap \mathcal{D}$, where $\mathcal{D}$ denotes the σ-algebra generated by the run-off triangle. The choice of the σ-algebras $\mathcal{G}_k$ considered here allows to capture the model of Mack, which also uses conditioning with respect to σ-algebras including non-observable information.

In the modified model, it is easily seen that the additional assumption

$$E\left[\prod_{l=k}^{n} F_l\right] = \prod_{l=k}^{n} E[F_l]$$

implies

$$E\left[\prod_{l=k}^{n} \hat{F}_l\right] = \prod_{l=k}^{n} E\left[\hat{F}_l\right],$$

which means that successive chain ladder factors are uncorrelated. This assumption is automatically fulfilled if in the general model the random variables F_k are assumed to be constant; in particular, the assumption is fulfilled in the model of Mack. To the present authors, however, uncorrelatedness of chain ladder factors seems to be of minor importance when compared with unbiasedness of the chain ladder predictor, and assumptions on unconditional expectations appear to be a bit strange in the general setting of conditional prediction considered in this paper.

ACKNOWLEDGEMENT

The authors gratefully acknowledge various discussions on the subject with Thomas Mack and Lothar Partzsch.

REFERENCES

MACK, T. (1993) *Distribution-free calculation of the standard error of chain ladder reserve estimates.* ASTIN Bull. **23**, 213-225.

MACK, T. (1994a) *Which stochastic model is underlying the chain ladder method?* Insurance Math. Econom. **15**, 133-138.

MACK, T. (1994b) *Measuring the variability of chain ladder reserve estimates.* Casualty Actuarial Society Forum Spring 1994, vol. 1, 101-182.

SCHNIEPER, R. (1991) *Separating true IBNR and IBNER claims.* ASTIN Bull. **21**, 111-127.

KLAUS D. SCHMIDT
Lehrstuhl für Versicherungsmathematik
Technische Universität Dresden
D-01062 Dresden
Germany

ANJA SCHNAUS
Kölnische Rück
Postfach 10 22 44
D-50642 Köln
Germany