risks

MDPI

Article

State Space Models and the KALMAN-Filter in Stochastic Claims Reserving: Forecasting, Filtering and Smoothing

Nataliya Chukhrova and Arne Johannssen *

Faculty of Business Administration, University of Hamburg, 20146 Hamburg, Germany; nataliya.chukhrova@uni-hamburg.de

* Correspondence: arne.johannssen@uni-hamburg.de

Academic Editor: Mogens Steffensen
Received: 1 April 2017; Accepted: 15 May 2017; Published: 27 May 2017

Abstract: This paper gives a detailed overview of the current state of research in relation to the use of state space models and the KALMAN-filter in the field of stochastic claims reserving. Most of these state space representations are matrix-based, which complicates their applications. Therefore, to facilitate the implementation of state space models in practice, we present a scalar state space model for cumulative payments, which is an extension of the well-known chain ladder (CL) method. The presented model is distribution-free, forms a basis for determining the entire unobservable lower and upper run-off triangles and can easily be applied in practice using the KALMAN-filter for prediction, filtering and smoothing of cumulative payments. In addition, the model provides an easy way to find outliers in the data and to determine outlier effects. Finally, an empirical comparison of the scalar state space model, promising prior state space models and some popular stochastic claims reserving methods is performed.

Keywords: state space models; KALMAN-filter; stochastic claims reserving; outstanding loss liabilities; ultimate loss; prediction uncertainty; chain ladder method

1. Introduction

At the end of each fiscal year, non-life insurance companies face the situation that the earned premiums are known, but not the outstanding loss liabilities. Consequently, one of the main tasks for actuaries in non-life insurance is to quantify accurately the outstanding loss liabilities. The outstanding claims reserves are often a large share of the liability side of the balance sheet, so it is very important for every non-life insurer to handle claims reserving adequately. It is not surprising, therefore, that over the past 40 years, numerous reserving methods have been developed, particularly since the early 1990s. These methods are based on various models (see Wüthrich and Merz (2008)), but rarely on time series models. This is surprising, especially in light of the fact that the claims process is a stochastic process, and claims data, as a sequence of discrete time data, represent time series.

In this paper, methods of claims reserving are considered, which are based on time series models, particularly on state space models. A state space model consists of a state equation describing the dynamics of the system and an observation equation establishing a link between the unobservable states of the system and the observations. Compared to other models, state space models have the advantage that the temporal dynamics of a system can often be detected more accurately. In addition, state space representations can be used flexibly to model univariate and multivariate, stationary and non-stationary time series or in cases of structural changes, interventions, missing data or other data irregularities. A consideration of state space models leads directly to the application of the KALMAN-filter algorithms for parameter estimation, forecasting, filtering and smoothing.

The KALMAN-filter is generally distribution-free and provides the best linear predictors in the sense of minimizing the mean squared error (see Kalman (1960)).

There are currently 16 research papers in stochastic claims reserving on the topic of state space models and the KALMAN-filter. Most of these state space representations are based on a calendar year approach, i.e., all available observations of one calendar year are stacked into one observation vector. The matrix-based approach of these models and the resulting KALMAN recursions can complicate their applications. Therefore, to facilitate the implementation of state space models in practice, a new scalar approach is introduced in this paper. The resulting distribution-free scalar state space model for cumulative payments can be considered as an extension of the chain ladder (CL) method under the assumption that the observations in the upper triangle are based on unobservable states.

According to this content, the paper is structured as follows. Section 2 gives a brief overview of claims development triangles and the CL method. In Section 3, we present the previous papers in stochastic claims reserving on the topic of state space models, their contentual similarities and selected modeling approaches. Section 4 introduces the scalar state space model and gives derivations of the formulas for claims reserves and the mean squared error of prediction (MSEP). An empirical comparison of the scalar state space model and promising state space models, as well as popular stochastic claims reserving methods is given in Section 5. Finally, Section 6 provides concluding remarks.

2. Development Triangles and the CL Method

Statistical analysis of claims payments data is usually done by using claims development triangles. In this paper, we denote accident years by $i = 0, \ldots, I$, development years by $j = 0, \ldots, J$ and calendar years by $t = i + j$. We assume further that all claims are settled after the J-th development year. Cumulative payments in accident year i after j development years are denoted by $C_{i,j}$, and incremental payments in the j-th development year are denoted by $X_{i,j}$, respectively. An exemplary claims development triangle (see Table 1) is the dataset of Taylor and Ashe (1983), which includes data from the motor bodily injury class of business in one Australian state (1972–1981, quoted in Australian Dollars). Due to its popularity in claims reserving literature, the dataset of Taylor and Ashe (1983) acts like a kind of benchmark development triangle for claims reserving methods.

Table 1. Cumulative payments in the claims development triangle of Taylor and Ashe (1983).

Accident	Development Year j									
Year i	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
3	310608	1418858	2195047	3757447	4029929	4381982	4588268			
4	443160	1136350	2128333	2897821	3402672	3873311				
5	396132	1333217	2180715	2985752	3691712					
6	440832	1288463	2419861	3483130						
7	359480	1421128	2864498							
8	376686	1363294								
9	344014									

The CL method is probably the most popular loss reserving technique, and there are several stochastic models that could be used as a basis for the application of the CL method. In this paper, we introduce briefly the distribution-free CL model of Mack (1993).

Model Assumptions 1 (Distribution-free CL model of Mack (1993))**.**

- ▷ Cumulative payments $C_{i,j}$ of different accident years i are stochastically independent.
- ▷ There exist factors $f_0, \ldots, f_{J-1} > 0$ and variance parameters $\sigma_0^2, \ldots, \sigma_{J-1}^2 > 0$, such that for all $i = 0, \ldots, I$ and all $j = 0, \ldots, J-1$, we have:

$$\begin{aligned} \mathbb{E}[C_{i,j+1}|C_{i,0}, \ldots, C_{i,j}] &= \mathbb{E}[C_{i,j+1}|C_{i,j}] = f_j C_{i,j} \\ \mathrm{Var}(C_{i,j+1}|C_{i,0}, \ldots, C_{i,j}) &= \mathrm{Var}(C_{i,j+1}|C_{i,j}) = \sigma_j^2 C_{i,j} \end{aligned} \tag{1}$$

◁

The development factors and variance parameters are then estimated using the (unconditionally) unbiased estimators:

$$\widehat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} \tag{2}$$

and:

$$\widehat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f}_j \right)^2 \tag{3}$$

for f_j and σ_j^2, respectively (for more details, see Mack (1993)).

3. Prior Applications in Stochastic Claims Reserving

This section presents the previous papers in claims reserving literature on the topic of state space models and the KALMAN-filter, their relationships and selected modeling approaches of claims development data, as well as their state space representations.

3.1. Chronology and Categorization of the Papers

There are 16 papers in the claims reserving literature that are based on state space models and the KALMAN-filter. Figure 1 displays these papers in chronological order and classifies them into four categories considering their substantial contentual similarities. The formulated categories "parametric development", "log-normal model", "different approaches" and "comparisons" need not be taken as mutually exclusive. Instead, the choice of the appropriate category is made considering the main approach used in the respective paper. In the papers of the first category a parametric development of claims data over the development years is assumed. The papers of the second category are based on the log-normal model for incremental payments, while in the third category, different approaches are aggregated. The fourth category unites comparisons of various reserving methods. The solid arrows in Figure 1 represent the contentual similarities among the papers in modeling approaches. The dashed arrows indicate, however, that the respective models are included in the papers of the fourth category that compares various models.

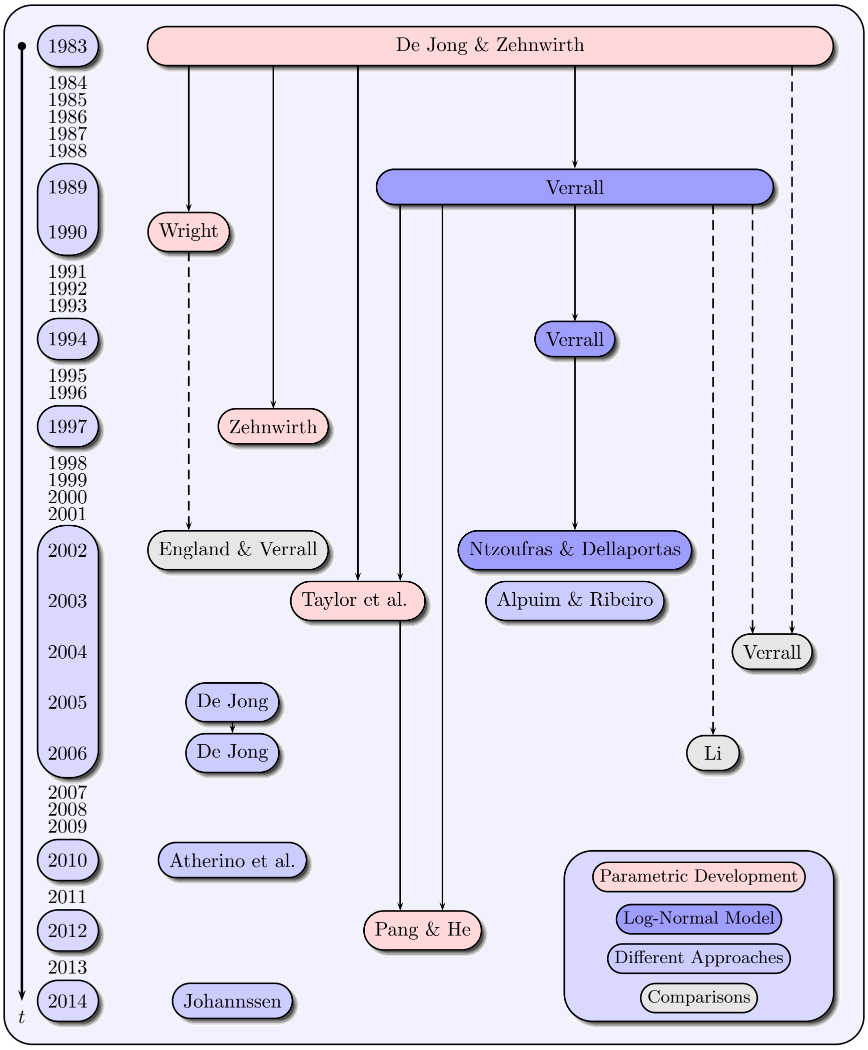

Figure 1. Chronology and categorization of the papers based on state space models.

3.2. Modeling of Claims Development Data

In the first category, the authors De Jong and Zehnwirth (1983), Wright (1990), Zehnwirth (1997), Taylor et al. (2003) and Pang and He (2012) mainly use variations on the Hoerl curve to model the claims data over the development years. The general exponential-logarithmic Hoerl curve is given as $\beta_j = \exp(\kappa j + \delta \log j)$ with development year parameter β_j for all $j = 0, \ldots, J$ and $\kappa, \delta \in \mathbb{R}$. An advantage of treating development time j as a continuous covariate is that extrapolation is possible beyond the range of development times observed (see Frees (2010) and Kaas et al. (2009)). As a variation, De Jong and Zehnwirth (1983), for example, define incremental payments $X_{i,j}$ as:

$$X_{i,j} = \alpha_i (j+1) e^{-j} + w_{i,j} \tag{4}$$

with accident year parameter α_i and $w_{i,j} \sim WN\left(0; \sigma_w^2\right)$ for all $i = 0, \ldots, I$ and $j = 0, \ldots, J$. To allow a dynamic recursive estimation of the accident year parameters and to avoid over-parametrization, they assume the accident year parameters evolve in the way $\alpha_{i+1} = \alpha_i + v_i$ with $\alpha_0 = 0$ and $v_i \sim WN\left(0; \sigma_v^2\right)$. The modeling (4) leads to the general shape of incremental payments curve over the development years, i.e., the incremental payments are supposed to rise very fast in early development years and to decrease exponentially over the following development years.

The models in the papers of the first category are mostly distribution-free in contrast to the papers of the second category (Verrall (1989,1994), Ntzoufras and Dellaportas (2002)), in which incremental payments are assumed to follow a log-normal distribution. The logarithmic incremental payments are specified by the log-normal model:

$$Y_{i,j} = \log\left(X_{i,j}\right) = \mu + \alpha_i + \beta_j + w_{i,j} \tag{5}$$

with $X_{i,j} > 0$, mean μ, accident year parameter α_i, development year parameter β_j and the GAUSSian white noise process $w_{i,j} \sim WN\left(0;\sigma_w^2\right)$ for all $i = 0,\dots,I$ and $j = 0,\dots,J$. The accident and development year parameters are further assumed to evolve as follows for the same reasons as in De Jong and Zehnwirth (1983):

$$\begin{aligned} \alpha_{i+1} &= \alpha_i + v_i^{(\alpha)} \\ \beta_{j+1} &= \beta_j + v_j^{(\beta)} \end{aligned} \tag{6}$$

with $\alpha_0 = \beta_0 = 0$ and white noise processes $v_i^{(\alpha)}$ and $v_j^{(\beta)}$ for all $i = 0,\dots,I$ and $j = 0,\dots,J$. Model (5) was also named as the "linear CL model" by Verrall (1989,1994), since it is very similar to an additive representation of the CL method (see also Kremer (1982)). In addition to the basic model (5), which is used by Verrall (1989) and Li (2006), Verrall (1994) and Ntzoufras and Dellaportas (2002) extend the basic model by integrating varying run-off evolutions.

As a first example of the third category, Atherino et al. (2010) consider a different data ordering of non-cumulative run-off triangles, in which the stacked rows of a triangle form a univariate time series y_t with several runs of missing data. They choose a structural model for the incremental payments with a local level component ν_t, a stochastic periodic component η_t and a regression term $\mathbf{h}_t^T\mathbf{u}$,

$$y_t = \nu_t + \eta_t + \mathbf{h}_t^T\mathbf{u} + w_t \tag{7}$$

$$\nu_{t+1} = \nu_t + v_t^{(\nu)} \tag{8}$$

$$\eta_{t+1} = -\sum_{d=1}^{J-1} \eta_{t+1-d} + v_t^{(\eta)} \qquad (t = J-1, J, \dots) \tag{9}$$

with the GAUSSian white noise processes w_t, $v_t^{(\nu)}$, $v_t^{(\eta)}$. This model and its components are motivated by the claims process behavior: The level component shall respond for the average value of claims along each accident year, while the periodic component is supposed to capture the development year effect. The regression term is mainly motivated by the need for intervention effects due to the presence of outliers.

Another example of the third category to model the claims data can be found in Alpuim and Ribeiro (2003). They assume that the incremental payments $X_{i,j}$ for the i-th accident year ($i = 0,\dots,I$) in the j-th development year ($j = 1,\dots,J$) depend on the payments $X_{i,0}$ of the respective accident year,

$$X_{i,j} = \lambda_{i,j} X_{i,0} + w_{i,j} \tag{10}$$

with $w_{i,j} \sim WN\left(0;\sigma_w^2\right)$. That is, the total amount of claims incurred in year i and paid j years later is proportional to the claims incurred and paid in accident year i. This proportion varies randomly with development and accident year, so that they assume for the proportion parameter $\lambda_{i,j}$ a first order autoregressive process:

$$\lambda_{i,j} = \mu_j + \phi_j(\lambda_{i-1,j} - \mu_j) + v_{i,j} \tag{11}$$

with mean μ_j and $v_{i,j} \sim WN\left(0; \sigma_v^2\right)$. Moreover, the $X_{i,j}$ of different development years $j = 1, \ldots, J$ are stochastically independent.

Since the fourth category (England and Verrall (2002), Verrall (2004), Li (2006)) provides comparisons of various methods, there is only a recapitulation of existing approaches in the remaining papers.

All of these modeling approaches can be converted into a state space representation, and then, the KALMAN-filter can be used for prediction, filtering and smoothing of the claims development data.

3.3. Modeling Approaches of State Space Representations

Most of state space representations in the introduced papers are based on the calendar year approach, which provides the claims data of each calendar year to be stacked into separate observation vectors of the respective calendar years. In addition to this approach, there are further differing approaches considering an accident year- or a development year-based modeling of the observation vectors (see Figure 2) as also detached approaches, which model run-off triangle data, for example as a univariate time series.

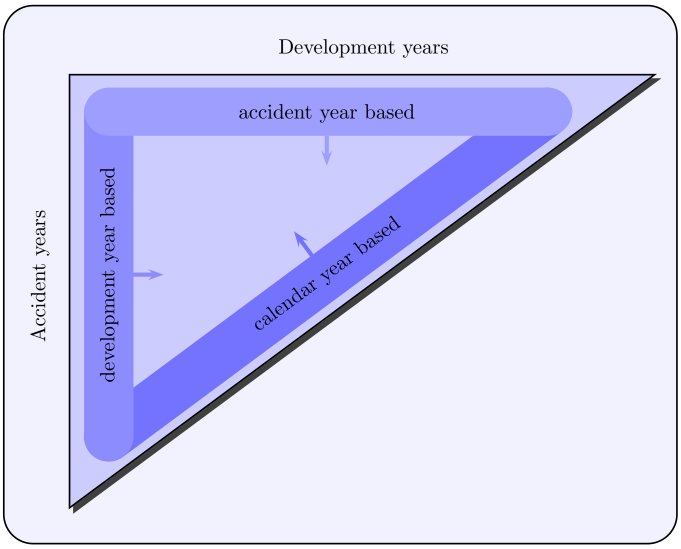

Figure 2. Modeling approaches of claims development data.

The popularity of the calendar year-based approach can be justified as follows:

- Annually-added observations build a new diagonal in the run-off triangle. Therefore, the calendar year approach corresponds to natural modeling of the claims data.
- The observations of the same calendar year are subjected to calendar year effects of the same level, such as the inflation factor or changes in legislation.
- As for estimating and forecasting, the recent observations should be weighted higher compared to past observations. This proposition is also consistent with the view of many authors such as Verrall (1994), Taylor (2000) or De Jong (2005,2006). Therefore, the use of the KALMAN-filter is justified here. Its recursive and dynamic nature complies with this requirement especially in relation to the calendar year approach.

Four representative state space models, which are also used for the empirical comparison in Section 5, are presented below. Two of these models (Verrall (1989), Li (2006)) are based on the calendar year approach, and the other two models (Atherino et al. (2010), Alpuim and Ribeiro (2003)) consider detached approaches. As for accident year and development year approaches, they can only be found in Taylor et al. (2003) and De Jong and Zehnwirth (1983). However, these approaches have no significant advantages compared to the calendar year approach, so they are not introduced in this paper.

The state space representation of the log-normal model for incremental payments in Verrall (1989) has the observation equation:

$$\underbrace{\begin{pmatrix} Y_{0,t} \\ Y_{1,t-1} \\ Y_{2,t-2} \\ \vdots \\ Y_{t-1,1} \\ Y_{t,0} \end{pmatrix}}_{\substack{\text{observation} \\ \text{vector}}} = \underbrace{\begin{pmatrix} 1 & 0 & \dots & & & & & & & \dots & 0 & 1 \\ 1 & 1 & 0 & \dots & & & & \dots & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & & & & & \ddots & & & & & & \vdots \\ \vdots & & & & & & \ddots & & & & & \vdots \\ 1 & 0 & 1 & 0 & \dots & & \dots & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & \dots & & & & & & \dots & 0 & 1 & 0 \end{pmatrix}}_{\text{system matrix}} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_t \\ \beta_t \end{pmatrix}}_{\substack{\text{state} \\ \text{vector} \\ \text{in } t}} + \underbrace{\begin{pmatrix} w_{0,t} \\ w_{1,t-1} \\ w_{2,t-2} \\ \vdots \\ w_{t-1,1} \\ w_{t,0} \end{pmatrix}}_{\substack{\text{measurement} \\ \text{noise} \\ \text{vector}}}$$

for calendar year $t = 0, \dots, I$, which implies (5) for each $Y_{i,j}$ of calendar year $t = i + j$, and the state equation:

$$\underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_{t+1} \\ \beta_{t+1} \end{pmatrix}}_{\substack{\text{state} \\ \text{vector} \\ \text{in } t+1}} = \underbrace{\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}}_{\text{transition matrix}} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_t \\ \beta_t \end{pmatrix}}_{\substack{\text{state} \\ \text{vector} \\ \text{in } t}} + \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ v_t^{(\alpha)} \\ v_t^{(\beta)} \end{pmatrix}}_{\substack{\text{process} \\ \text{noise} \\ \text{vector}}} \tag{12}$$

for $t = 0, \dots, I$, where (12) allows a dynamic recursive estimation of the accident and development year parameters via (6). The state space model of Verrall (1989) is also used in Li (2006), but slightly modified.

The state space representation of the structural model in Atherino et al. (2010) has the observation equation:

$$y_t = \underbrace{\begin{pmatrix} 1 & 1 & 0 & \dots & 0 \end{pmatrix}}_{\text{system matrix}} \underbrace{\begin{pmatrix} \nu_t \\ \eta_t \\ \eta_{t-1} \\ \vdots \\ \eta_{t-J+2} \end{pmatrix}}_{\substack{\text{state} \\ \text{vector} \\ \text{in } t}} + \mathbf{h}_t^T \mathbf{u} + w_t,$$

that stands for (7), as well as the state equation:

$$\underbrace{\begin{pmatrix} \nu_{t+1} \\ \eta_{t+1} \\ \eta_t \\ \vdots \\ \eta_{t-J+3} \end{pmatrix}}_{\substack{\text{state} \\ \text{vector} \\ \text{in } t+1}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & -1 & \dots & -1 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}}_{\text{transition matrix}} \underbrace{\begin{pmatrix} \nu_t \\ \eta_t \\ \eta_{t-1} \\ \vdots \\ \eta_{t-J+2} \end{pmatrix}}_{\substack{\text{state} \\ \text{vector} \\ \text{in } t}} + \underbrace{\begin{pmatrix} v_t^{(\nu)} \\ v_t^{(\eta)} \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\substack{\text{process} \\ \text{noise} \\ \text{vector}}},$$

which includes (8) and (9). In both state space representations Verrall (1989) and Atherino et al. (2010), the GAUSSian white noise processes are assumed to be stochastically independent as also independent of the initial state.

Both examples and the most state space representations in the other papers have a matrix-based approach in common, so the KALMAN recursions also contain numerous matrices of high dimensions (see, for example, Brockwell and Davis (2006)). This complicates parameter estimation and practical application considerably. Therefore, a scalar structure could likewise be a preferable approach. One alternative is the state space model in Alpuim and Ribeiro (2003), which consists of the observation Equation (10), the state Equation (11) and the additional assumption $\mathbb{E}[v_{i,j}w_{k,l}] = 0$ for all $i, k = 0, \dots, I$ and $j, l = 1, \dots, J$. Another promising alternative is the newly-developed scalar state space model introduced in Section 4.

4. Scalar State Space Model for Cumulative Payments

The idea behind the scalar state space model is a modification of the CL method to get a meaningful state space representation and to use the KALMAN-filter for calculating the claims reserves, as well as for measuring their precision. Since the CL method operates under the assumption that the cumulative payments are a linear function of the cumulative payments of the previous development year, we consider a linear state space model.

4.1. Model Assumptions and KALMAN *Recursions*

It is assumed that a run-off triangle of observed cumulative payments $C_{i,j}^{\text{obs}}$ is based on a run-off triangle of unobservable states $C_{i,j}$ with $i + j \leq I$ for all $i = 0, \dots, I$ and $j = 0, \dots, J$ with $I = J$. Therefore, there is a probable observation error in the claims data, and we can not definitely observe "real cumulative payments", i.e., the payments made do not necessarily correspond with the payments actually incurred. One possible reason for this difference is that a claim is not reported correctly, for example if the claim is not reported, only partially reported or reported too late. Therefore, we model these unobservable "real cumulative payments" as latent variables. The scalar state space model for cumulative payments presented below provides a basis for determining the entire unobservable upper and lower run-off triangles using KALMAN-filter, that is prediction, filtering and smoothing of all states $C_{i,j}$ with $i = 0, \dots, I$ and $j = 0, \dots, J$ (see Figure 3).

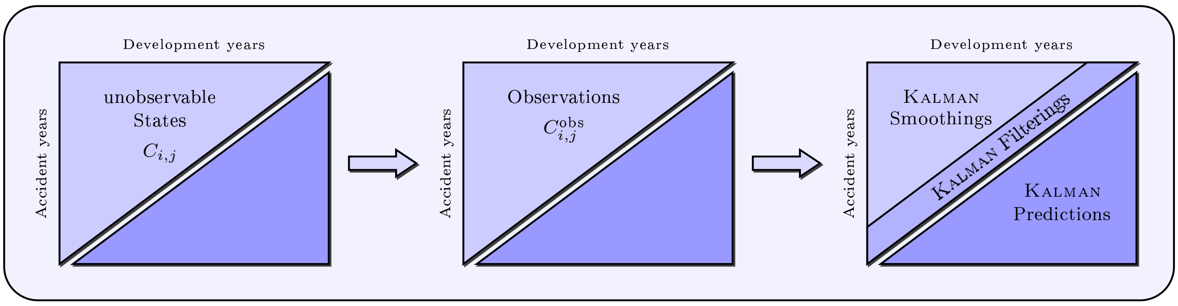

Figure 3. Unobservable states, observations and KALMAN smoothings ($i + j < I$), KALMAN filterings ($i + j = I$) and KALMAN predictions ($i + j > I$).

Model Assumptions 2 (Scalar state space model for cumulative payments)**.**

▷ There exist parameters $g_j > 0$ and $\sigma_w^2 > 0$, such that:

$$C_{i,j}^{\text{obs}} = g_j C_{i,j} + w_{i,j} \qquad \text{(observation equation)} \tag{13}$$

with $w_{i,j} \sim WN\left(0; \sigma_w^2\right)$ for $i = 0, \dots, I$ and $j = 0, \dots, J$.

▷ There exist parameters $f_j > 0$ and $\sigma_v^2 > 0$, such that:

$$C_{i,j+1} = f_j C_{i,j} + v_{i,j} \qquad \text{(state equation)} \tag{14}$$

with $v_{i,j} \sim WN\left(0;\sigma_v^2\right)$ for $i = 0,\ldots,I$ and $j = 0,\ldots,J-1$.

▷ The white noise processes $(w_{i,j})_{j=0,\ldots,J}^{i=0,\ldots,I}$ and $(v_{i,j})_{j=0,\ldots,J-1}^{i=0,\ldots,I}$ are uncorrelated and therefore satisfy $\mathbb{E}[v_{i,j}w_{k,l}] = 0$ for all $i,k = 0,\ldots,I$, $j = 0,\ldots,J-1$ and $l = 0,\ldots,J$.

▷ Cumulative payments $C_{i,j}$ of different accident years i are stochastically independent.

◁

The assumption of uncorrelated white noise processes is motivated by the fact that there is no reason to assume a systematic relationship between the measurement noise $(w_{i,j})_{j=0,\ldots,J}^{i=0,\ldots,I}$ and the process noise $(v_{i,j})_{j=0,\ldots,J-1}^{i=0,\ldots,I}$. Nevertheless, this assumption is not necessary, but simplifies the KALMAN recursions. The further assumption of stochastically independent accident years is frequently applied to claims reserving and especially used in the CL method (see Model Assumptions 1).

The state equation and the observation equation can also be stated as follows:

$$C_{i,j} = f_{j-1}C_{i,j-1} + v_{i,j-1} = \ldots = a_{i,j}(C_{i,0}, v_{i,0}, \ldots, v_{i,j-2}, v_{i,j-1}) \tag{15}$$

$$C_{i,j}^{\text{obs}} = g_j C_{i,j} + w_{i,j} = \ldots = b_{i,j}(C_{i,0}, v_{i,0}, \ldots, v_{i,j-2}, v_{i,j-1}, w_{i,j}) \tag{16}$$

Here, $a_{i,j}$ and $b_{i,j}$ for $i = 0,\ldots,I$ and $j = 0,\ldots,J$ are appropriate linear functions. Considering (15), (16) and the assumptions regarding the measurement and process noise, it is clear that:

$$\mathbb{E}[C_{i,j}v_{i,l}] = 0 \qquad \text{and} \qquad \mathbb{E}[C_{i,j}w_{i,k}] = 0$$

for all $j,k = 0,\ldots,J$, $l = 0,\ldots,J-1$ with $j \leq k$, $j \leq l$. Consequently, the initial state $C_{i,0}$ of an accident year $i = 0,\ldots,I$ is uncorrelated with $v_{i,j}$ and $w_{i,j}$ for all j.

Remark 1.

- To forecast future cumulative payments $C_{i,j}$ with $i + j > I$ for $i = 1,\ldots,I$, $j = 1,\ldots,J$ (lower triangle) the corresponding KALMAN predictions are required. The one-step predictor provides forecasts for the next calendar year $t = I + 1$, while the h-step predictor shall be used for forecasting cumulative payments in calendar years $t = I + h$ with $h > 1$.
- As for the underlying states $C_{i,j}$ of the observations $C_{i,j}^{\text{obs}}$ in the upper triangle, the KALMAN filterings (for $i + j = I$) and smoothings (for $i + j < I$) are useful to identify outliers in the observations and to replace them with smoothed or filtered observations, as well as to obtain an adjusted presentation of the observed quantities and to determine outlier effects. Another key application of smoothing and filtering is the determination of missing values in the upper run-off triangle (for example, resulting from a merger) to interpolate gaps in the data.

Remark 2.

- We denote the one-step predictor by $\widehat{C}_{i,j+1}^{(\text{P})}$ and its error variance by $\left(\widehat{\sigma}_{i,j+1}^{(\text{P})}\right)^2 = \mathbb{E}\left[\left(C_{i,j+1} - \widehat{C}_{i,j+1}^{(\text{P})}\right)^2\right]$ for $j = 0,\ldots,I-i$, as well as the h-step predictor by $\widehat{C}_{i,j+h}^{(\text{P})}$ and its error variance by $\left(\widehat{\sigma}_{i,j+h}^{(\text{P})}\right)^2 = \mathbb{E}\left[\left(C_{i,j+h} - \widehat{C}_{i,j+h}^{(\text{P})}\right)^2\right]$ for $j = 0,\ldots,J-h$. The superscript (P) stands for "prediction" and the subscript indicates the cell $(i,j+1)$ or $(i,j+h)$ in the lower triangle, for which we predict cumulative payments.
- We denote the filtering by $\widehat{C}_{i,j}^{(\text{F})}$ and its error variance by $\left(\widehat{\sigma}_{i,j}^{(\text{F})}\right)^2 = \mathbb{E}\left[\left(C_{i,j} - \widehat{C}_{i,j}^{(\text{F})}\right)^2\right]$ for $j = 0,\ldots,I-i$, as well as the smoothings by $\widehat{C}_{i,j}^{(\text{S})}$ and their error variances by

$\left(\widehat{\sigma}_{i,j}^{(S)}\right)^2 = \mathbb{E}\left[\left(C_{i,j} - \widehat{C}_{i,j}^{(S)}\right)^2\right]$ for $j = I-i, \ldots, 0$. The superscript (F) or (S) stands for "filtering" or "smoothing", and the subscript indicates the (i,j)-cell in the upper triangle, for which we filter or smooth cumulative payments.

The KALMAN recursions for prediction, filtering and smoothing are given below. There are two different smoothing approaches: fixed-point and fixed-interval smoothing. While the fixed-point approach calculates smoothed values for a few fixed, predetermined points of time, the fixed-interval approach provides an ex post reconstruction of the behavior of a system in order to understand the phenomenon underlying observations. Since we are going to smooth all observations in the upper run-off triangle for identifying outliers in the data, the fixed-interval algorithm is presented (the fixed-point algorithm can be found for example in Brockwell and Davis (2006)).

Theorem 1 (KALMAN-filter algorithms for the scalar state space model)**.**
Given the Model Assumptions 2, the one-step, h-step, filtering and fixed-interval smoothing predictors, as well as their error variances are uniquely determined by the initial conditions $\widehat{C}_{i,0}^{(P)}$ *and* $\left(\widehat{\sigma}_{i,0}^{(P)}\right)^2$ *and the recursions* $(i = 0, \ldots, I,\ h \geq 2)$*:*

$$\widehat{C}_{i,j+1}^{(P)} = f_j \widehat{C}_{i,j}^{(P)} + \frac{f_j \gamma_{i,j}}{\Delta_{i,j}} \left(C_{i,j}^{\text{obs}} - g_j \widehat{C}_{i,j}^{(P)}\right) \qquad (j = 0, \ldots, I-i) \tag{17}$$

$$\widehat{C}_{i,j+h}^{(P)} = (f_{j+h-1} \cdots f_{j+1}) \widehat{C}_{i,j+1}^{(P)} \qquad (j = 0, \ldots, J-h) \tag{18}$$

$$\widehat{C}_{i,j}^{(F)} = \widehat{C}_{i,j}^{(P)} + \frac{\gamma_{i,j}}{\Delta_{i,j}} \left(C_{i,j}^{\text{obs}} - g_j \widehat{C}_{i,j}^{(P)}\right) \qquad (j = 0, \ldots, I-i) \tag{19}$$

$$\widehat{C}_{i,j}^{(S)} = \widehat{C}_{i,j}^{(F)} + \psi_{i,j} \left(\widehat{C}_{i,j+1}^{(S)} - \widehat{C}_{i,j+1}^{(P)}\right) \qquad (j = I-i, \ldots, 0) \tag{20}$$

and:

$$\left(\widehat{\sigma}_{i,j+1}^{(P)}\right)^2 = f_j^2 \left(\widehat{\sigma}_{i,j}^{(P)}\right)^2 + \sigma_v^2 - \frac{(f_j \gamma_{i,j})^2}{\Delta_{i,j}} \qquad (j = 0, \ldots, I-i) \tag{21}$$

$$\left(\widehat{\sigma}_{i,j+h}^{(P)}\right)^2 = f_{j+h-1}^2 \left(\widehat{\sigma}_{i,j+h-1}^{(P)}\right)^2 + \sigma_v^2 \qquad (j = 0, \ldots, J-h) \tag{22}$$

$$\left(\widehat{\sigma}_{i,j}^{(F)}\right)^2 = \left(\widehat{\sigma}_{i,j}^{(P)}\right)^2 - \frac{\gamma_{i,j}^2}{\Delta_{i,j}} \qquad (j = 0, \ldots, I-i) \tag{23}$$

$$\left(\widehat{\sigma}_{i,j}^{(S)}\right)^2 = \left(\widehat{\sigma}_{i,j}^{(F)}\right)^2 + \psi_{i,j}^2 \left(\left(\widehat{\sigma}_{i,j+1}^{(S)}\right)^2 - \left(\widehat{\sigma}_{i,j+1}^{(P)}\right)^2\right) \qquad (j = I-i, \ldots, 0) \tag{24}$$

where $\gamma_{i,j} = \left(\widehat{\sigma}_{i,j}^{(P)}\right)^2 g_j$, $\Delta_{i,j} = g_j^2 \left(\widehat{\sigma}_{i,j}^{(P)}\right)^2 + \sigma_w^2$ *and* $\psi_{i,j} = f_{j+1} \frac{\left(\widehat{\sigma}_{i,j}^{(F)}\right)^2}{\left(\widehat{\sigma}_{i,j+1}^{(P)}\right)^2}$.

◁

Proof: See Appendix A.

□

Remark 3.

- The KALMAN gain $0 \leq \frac{\gamma_{i,j}}{\Delta_{i,j}} \leq 1$ represents the relative importance of the innovation $\varepsilon_{i,j} = C_{i,j}^{\text{obs}} - g_j \widehat{C}_{i,j}^{(P)}$ with respect to the prior predictor $\widehat{C}_{i,j}^{(P)}$. The higher the covariance $\gamma_{i,j}$ between the innovation and the state to be predicted and/or the lower the variance $\Delta_{i,j}$ of the innovation, the higher the trust in the new observation $C_{i,j}^{\text{obs}}$ and therefore the higher the KALMAN gain.

- Due to the fact that there is no observation after the recent calendar year, and therefore no innovation $\varepsilon_{i,j}$, the covariance $\gamma_{i,j}$ is equal to zero for $i+j>I$. This implies that the KALMAN gain is equal to zero in the h-step recursions.
- Since the KALMAN smoother is a backwards recursive algorithm and its initializations $\widehat{C}^{(\mathrm{F})}_{i,I-i}$ and $\left(\widehat{\sigma}^{(\mathrm{F})}_{i,I-i}\right)^2$ are values of the last filtering recursion, the smoothings and filterings are identical for the current calendar year $t=I$.

4.2. Determination of KALMAN Reserves and MSEP

Using Theorem 1, the ultimate claims $C_{i,J}$ of accident years $i=1,\dots,I$ can be forecasted. The ultimate claim of the first accident year is predicted by the one-step predictor:

$$\widehat{C}^{(\mathrm{P})}_{1,J} = f_{J-1}\widehat{C}^{(\mathrm{P})}_{1,J-1} + \frac{f_{J-1}\gamma_{1,J-1}}{\Delta_{1,J-1}}\left(C^{\mathrm{obs}}_{1,J-1} - g_{J-1}\widehat{C}^{(\mathrm{P})}_{1,J-1}\right), \tag{25}$$

and the ultimate claims of the accident years $i=2,\dots,I$ are predicted by the h-step predictors:

$$\widehat{C}^{(\mathrm{P})}_{i,J} = (f_{J-1}\cdots f_{J-i+1})\widehat{C}^{(\mathrm{P})}_{i,J-i+1}. \tag{26}$$

Since we predict the ultimate claims in (26), we have $h=i$. The KALMAN reserve $\widehat{R}_i$ for a single accident year $i=1,\dots,I$, as well as the total KALMAN reserve $\widehat{R}$ across all accident years can therefore be determined by:

$$\widehat{R}_i = \widehat{C}^{(\mathrm{P})}_{i,J} - C^{\mathrm{obs}}_{i,I-i} \quad \text{and} \quad \widehat{R} = \sum_{i=1}^{I}\widehat{R}_i,$$

respectively. The error variance of the predictor $\widehat{C}^{(\mathrm{P})}_{1,J}$ according to (25) results in:

$$\left(\widehat{\sigma}^{(\mathrm{P})}_{1,J}\right)^2 = f^2_{J-1}\left(\widehat{\sigma}^{(\mathrm{P})}_{1,J-1}\right)^2 + \sigma^2_v - \frac{\left(f_{J-1}\gamma_{1,J-1}\right)^2}{\Delta_{1,J-1}}, \tag{27}$$

while the error variances of $\widehat{C}^{(\mathrm{P})}_{i,J}$ for $i=2,\dots,I$ according to (26) are given by:

$$\left(\widehat{\sigma}^{(\mathrm{P})}_{i,J}\right)^2 = f^2_{J-1}\left(\widehat{\sigma}^{(\mathrm{P})}_{i,J-1}\right)^2 + \sigma^2_v. \tag{28}$$

The error variances (27) and (28) provide a basis for estimating the (unconditional) MSEP. The MSEP for a single accident year $i=1,\dots,I$ using the KALMAN predictor $\widehat{C}^{(\mathrm{P})}_{i,J}$ is defined as follows:

$$\mathrm{MSEP}\left(\widehat{C}^{(\mathrm{P})}_{i,J}\right) = \mathbb{E}\left[\left(C_{i,J} - \widehat{C}^{(\mathrm{P})}_{i,J}\right)^2\right]$$

Therefore, an appropriate estimator of the MSEP for a single accident year results in:

Definition 1 (KALMAN estimator of MSEP for single accident years)**.**
Using Model Assumptions 2 and the KALMAN predictor (25) for $i=1$ or (26) for $i=2,\dots,I$ at time $t=I$, an estimator of the MSEP for a single accident year $i=1,\dots,I$ is given by:

$$\widehat{\mathrm{MSEP}}\left(\widehat{C}^{(P)}_{i,J}\right) = \left(\widehat{\sigma}^{(P)}_{i,J}\right)^2$$

$\triangleleft$

To determine an estimator of the MSEP for aggregated accident years, we consider two different accident years $i, k = 1, \ldots, I$ with $i < k$ in the first step.

$$\begin{aligned}
\text{MSEP}\left(\widehat{C}_{i,J}^{(\mathrm{P})} + \widehat{C}_{k,J}^{(\mathrm{P})}\right) &= \mathbb{E}\left[\left(C_{i,J} - \widehat{C}_{i,J}^{(\mathrm{P})} + C_{k,J} - \widehat{C}_{k,J}^{(\mathrm{P})}\right)^2\right] \\
&= \mathbb{E}\left[\left(C_{i,J} - \widehat{C}_{i,J}^{(\mathrm{P})}\right)^2\right] + \mathbb{E}\left[\left(C_{k,J} - \widehat{C}_{k,J}^{(\mathrm{P})}\right)^2\right] \\
&\quad + 2\mathbb{E}\left[\left(C_{i,J} - \widehat{C}_{i,J}^{(\mathrm{P})}\right)\left(C_{k,J} - \widehat{C}_{k,J}^{(\mathrm{P})}\right)\right] \\
&= \text{MSEP}\left(\widehat{C}_{i,J}^{(\mathrm{P})}\right) + \text{MSEP}\left(\widehat{C}_{k,J}^{(\mathrm{P})}\right) \\
&\quad + 2\mathbb{E}\left[\left(C_{i,J} - \widehat{C}_{i,J}^{(\mathrm{P})}\right)\left(C_{k,J} - \widehat{C}_{k,J}^{(\mathrm{P})}\right)\right]
\end{aligned} \tag{29}$$

Therefore, the MSEP for two accident years is the sum of two single accident year MSEPs and a mixed term based on both accident years. Using independence of the different accident years (see Model Assumptions 2), we obtain for the expectation in (29):

$$\begin{aligned}
\mathbb{E}\left[\left(C_{i,J} - \widehat{C}_{i,J}^{(\mathrm{P})}\right)\left(C_{k,J} - \widehat{C}_{k,J}^{(\mathrm{P})}\right)\right] &= \mathbb{E}\left[C_{i,J} - \widehat{C}_{i,J}^{(\mathrm{P})}\right]\mathbb{E}\left[C_{k,J} - \widehat{C}_{k,J}^{(\mathrm{P})}\right] \\
&= \left(\mathbb{E}\left[C_{i,J}\right] - \mathbb{E}\left[\widehat{C}_{i,J}^{(\mathrm{P})}\right]\right)\left(\mathbb{E}\left[C_{k,J}\right] - \mathbb{E}\left[\widehat{C}_{k,J}^{(\mathrm{P})}\right]\right)
\end{aligned} \tag{30}$$

Since $\widehat{C}_{i,J}^{(\mathrm{P})}$ is an unbiased predictor for the ultimate claim $C_{i,J}$, the term (30) is equal to zero. Consequently, the MSEP for two accident years results from the sum of two single accident year MSEPs:

$$\text{MSEP}\left(\widehat{C}_{i,J}^{(\mathrm{P})} + \widehat{C}_{k,J}^{(\mathrm{P})}\right) = \text{MSEP}\left(\widehat{C}_{i,J}^{(\mathrm{P})}\right) + \text{MSEP}\left(\widehat{C}_{k,J}^{(\mathrm{P})}\right)$$

The MSEP of all considered accident years is therefore, by independence, determined as:

$$\text{MSEP}\left(\sum_{i=1}^{I} \widehat{C}_{i,J}^{(\mathrm{P})}\right) = \sum_{i=1}^{I} \text{MSEP}\left(\widehat{C}_{i,J}^{(\mathrm{P})}\right)$$

Thus, an estimator of the MSEP for aggregated accident years is given by:

Definition 2 (**KALMAN estimator of MSEP for aggregated accident years**).
Using Model Assumptions 2, we have the following estimator for the MSEP of the ultimate claim for aggregated accident years:

$$\widehat{\text{MSEP}}\left(\sum_{i=1}^{I} \widehat{C}_{i,J}^{(P)}\right) = \sum_{i=1}^{I} \widehat{\text{MSEP}}\left(\widehat{C}_{i,J}^{(P)}\right)$$

◁

5. Empirical Applications

In this section, various empirical applications are considered. The applications are based on the claims development triangle of Taylor and Ashe (1983) (see Table 1). Firstly, the scalar state space model is used to calculate predicted, filtered and smoothed values for cumulative payments, as well as outlier effects in the data. Secondly, an empirical comparison of representative state space models and other popular methods in claims reserving is made.

5.1. Applications of Scalar State Space Model

Using Theorem 1, we can calculate forecasts, as well as filtered and smoothed cumulative payments, but in a prior step, we need to estimate the model parameters g_j, f_j, σ_w^2, σ_v^2 and the initializations $\widehat{C}_{i,0}^{(\mathrm{P})}$ and $\left(\widehat{\sigma}_{i,0}^{(\mathrm{P})}\right)^2$ for all i, j. Alternatively, they can also be determined by consulting expert opinion, market statistics or similar portfolios. In this section, we estimate the parameters as follows. Taking conditional expectations of both sides of (14) with respect to $C_{i,j}$ leads directly to $\mathbb{E}[C_{i,j+1}|C_{i,j}] = f_j C_{i,j}$. If we compare this result to (1), we find that f_j plays the role of the usual CL factor. This fact motivates the estimation of the factors $f_0, \ldots, f_{J-1}$ using the CL estimator (2). To avoid over-parametrization of the model, the parameter $g_j = g$ is assumed to be time-invariant. The parameters $g, \sigma_w^2, \sigma_v^2$ are estimated using maximum likelihood (ML) method in conjunction with the expectation-maximization (EM) algorithm. Since a direct maximization of the GAUSSian log likelihood can cause local maxima or convergence problems, we use the EM algorithm, which is an iterative method to find ML estimates based on the expected conditional GAUSSian log likelihood (for more details on the EM algorithm, see Shumway and Stoffer (1982 2010) or Johannssen (2016)).

Taking the observations $C_{i,0}^{\text{obs}}$ in the respective accident years $i = 0, \ldots, I$ and the CL variance parameter $\widehat{\sigma}_0^2$ (see (3)) as initializations, we get the following parameter estimates for the run-off triangle of Taylor and Ashe (1983) in Table 2:

Table 2. Estimated parameter values for the dataset of Taylor and Ashe (1983).

$\widehat{f}_0$	$\widehat{f}_1$	$\widehat{f}_2$	$\widehat{f}_3$	$\widehat{f}_4$	$\widehat{f}_5$	$\widehat{f}_6$	$\widehat{f}_7$	$\widehat{f}_8$	$\widehat{g}$
3.4906	1.7473	1.4574	1.1739	1.1038	1.0863	1.0539	1.0766	1.0177	1.0014

Since individual estimates of g for different years slightly deviate between 0.98 and 1.02, the assumption of time invariance has no significant impact on the calculated results. On the contrary: Due to the small number of observations in more recent years, the quality of the respective estimates would be questionable. Moreover, the estimated value $\widehat{g} \approx 1$ suggests the presumption for the dataset of Taylor and Ashe (1983) that the observations do not differ systematically upwards or downwards from the underlying states.

The predicted, filtered and smoothed cumulative payments are given in Table 3.

Table 3. Predictions (lower triangle), Filterings (last diagonal of upper triangle), Smoothings (upper triangle) for the data set of Taylor and Ashe (1983).

Accident	**Development Year j**									
Year i	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	357846	1088646	1724703	2344677	2801926	3216482	3460558	3608732	3849709	3907933
1	352118	1245805	2202664	3283457	3811445	4186179	4624394	4910422	5318601	5412740
2	290510	1218431	2204841	3256998	3888722	4213313	4619217	4892887	5267683	5360921
3	310611	1301186	2304955	3558299	4046805	4374444	4652591	4903365	5278963	5372401
4	443156	1255658	2103481	2949026	3444357	3845121	4176955	4402093	4739293	4823179
5	396131	1303545	2175261	3073819	3659435	4039284	4387874	4624381	4978608	5066730
6	440830	1384198	2411162	3494015	4101625	4527373	4918086	5183170	5580201	5678971
7	359483	1455238	2750622	4008757	4705880	5194350	5642622	5946760	6402282	6515602
8	376686	1344075	2348503	3422708	4017917	4434977	4817715	5077390	5466318	5563072
9	344014	1200815	2098185	3057894	3589662	3962269	4304213	4536210	4883683	4970125

Using the smoothed values $\widehat{C}_{i,j}^{(S)}$ for $i + j \leq 9$ (upper triangle in Table 3), outliers in the data can be identified in the first place. Subsequently, outlier effects can be isolated, and observations can be robustified. Table 4 shows the outlier effects $\widehat{Z}_{i,j} = C_{i,j}^{\text{obs}} - \widehat{C}_{i,j}^{(S)}$ for all $i = 0, \ldots, 9$ and $j = 0, \ldots, 9 - i$.

Table 4. KALMAN outlier effects.

Accident	Development Year j									
Year i	0	1	2	3	4	5	6	7	8	9
0	2	36142	10627	−126407	−56330	103512	5778	−2446	−16194	−6470
1	0	−9666	−32631	69865	−12378	−66116	23473	3617	20484	
2	−3	73875	13684	−21819	97273	−80395	9693	16428		
3	−3	117672	−109908	199148	−16876	7538	−64323			
4	4	−119308	24852	−51205	−41685	28190				
5	1	29672	5454	−88067	32277					
6	2	−95735	8699	−10885						
7	−3	−34110	113876							
8	0	19219								
9	0									

The outlier effects at $j = 0$ are negligible or equal to zero for all $i = 0, \dots, 9$ due to the chosen initializations. As for the magnitude of the observed cumulative payments, the outlier effects are relatively minor with a few exceptions. The largest outlier effect $\widehat{Z}_{3,3}$ has an (absolute) deviation of 199148. For example, the outlier effect $\widehat{Z}_{0,9} = -6470$ indicates that we have an observation error in the data, which implies that the cumulative payments made are way too low compared to the unobservable "real cumulative payments", i.e., there are still outstanding payments. It is for some claims reserving methods (such as the CL method), which are not robust against outliers, of high relevance, if the observations $C^{\text{obs}}_{i,I-i}$ for $i = 1, \dots, I$ of the current calendar year represent outliers. Therefore, two cumulative payments $C^{\text{obs}}_{3,6}$ and $C^{\text{obs}}_{7,2}$ in the run-off triangle of Taylor and Ashe (1983) should be treated with caution. In particular, such observations should be robustified, for example, they could be replaced by the KALMAN smoothed values. As for the KALMAN recursion algorithms, outliers entail less problems because of the lack of credibility of these observations and, consequently, the minor KALMAN gain, i.e., outliers do not change the forecast decisively.

The robustness of the CL method and the impact of outliers were also considered by Verdonck et al. (2009), who primarily surveyed how (simulated) outliers affect the CL total reserve. The results of their study show that the most problematic areas in a run-off triangle are the lower left corner and the upper right corner, since there are too few observations. In these areas, the CL method is particularly sensitive to outliers, because the CL reserves directly and indirectly (via the CL estimators (2)) depend on the observations of the recent calendar year. Further papers on the subject of robustness of the CL method are, for example, Van Wouwe et al. (2009), Verdonck and Debruyne (2011) and Verdonck and Van Wouwe (2011).

5.2. Empirical Comparison of Selected Models

The models considered in the empirical comparison are the scalar state space model, but the models in Verrall (1989), Alpuim and Ribeiro (2003), Li (2006), Atherino et al. (2010) and the well-known methods CL and BORNHUETTER–FERGUSON (BF), as well as the (overdispersed) POISSON (ODP) model. The chosen BF method is rather conservative, i.e., we use a priori estimates for the expected ultimate claims, which generally exceed estimates based on development triangle data. It should be pointed out that the BF results depend largely on the quality of the a priori estimates.

The most results in Table 5 suggest total claims reserves of approximately 18500000 ± 200000. The model in Atherino et al. (2010) and the conservative BF method present differing results with more optimistic and more conservative total reserves, respectively. Higher reserves tend to lead to higher

MSEP, just like in the ODP model, Li (2006), Verrall (1989) and the CL method, with the exception of the BF method and the scalar state space model. The model of Atherino et al. (2010), which provides by far the lowest total claims reserves, also leads to the smallest MSEP for aggregated accident years, but not to the smallest variational coefficient (VCO $= \sqrt{\text{MSEP}}/\text{reserve}$) for aggregated accident years.

Table 5. Reserves, standard errors and VCOs for selected models.

i	Scalar State Space Model			Verrall (1989)		
	Reserve	$\sqrt{\text{MSEP}}$	VCO	Reserve	$\sqrt{\text{MSEP}}$	VCO
1	73655	167499	227.4%	143834	72675	50.5%
2	451606	221667	49.1%	465847	166438	35.7%
3	784133	270524	34.5%	673175	194229	28.9%
4	949868	317331	33.4%	1060794	266228	25.1%
5	1375018	366006	26.6%	1479407	339755	23.0%
6	2195841	422159	19.2%	2218738	487975	22.0%
7	3651104	507337	13.9%	3287633	735669	22.4%
8	4199778	662654	15.8%	4517179	1040596	23.0%
9	4626111	797161	17.2%	4570683	1167068	25.5%
aggr.	18307113	1376670	7.5%	18417290	2627190	14.3%
i	**Alpuim and Ribeiro (2003)**			**Atherino et al. (2010)**		
	Reserve	$\sqrt{\text{MSEP}}$	VCO	Reserve	$\sqrt{\text{MSEP}}$	VCO
1	66860	161177	241.1%	78904	18385	23.3%
2	321421	227246	70.7%	433790	75046	17.3%
3	551625	278017	51.0%	663312	90874	13.7%
4	1243900	322745	25.9%	891774	107013	12.0%
5	1535502	422709	27.5%	1336361	144327	10.8%
6	2356440	625125	26.5%	2009913	207021	10.3%
7	2817779	1667381	59.2%	2919587	303637	10.4%
8	4472888	1448251	32.4%	3810769	411563	10.8%
9	4942889	763241	15.4%	4726935	571959	12.1%
aggr.	18309304	1637284	8.9%	16871345	1197865	7.1%
i	**Li (2006)**			**BF Method**		
	Reserve	$\sqrt{\text{MSEP}}$	VCO	Reserve	$\sqrt{\text{MSEP}}$	VCO
1	101374	54755	54.0%	104097	117241	112.6%
2	457788	178242	38.9%	516462	218187	42.2%
3	651123	198744	30.5%	780602	255401	32.7%
4	1035739	271135	26.2%	1083378	284276	26.2%
5	1473338	360715	24.5%	1561405	334286	21.4%
6	2190410	522967	23.9%	2395405	409247	17.1%
7	3442432	808061	23.5%	4312331	550065	12.8%
8	4269816	1054731	24.7%	4706870	560833	11.9%
9	5027791	1425522	28.4%	5088393	578565	11.4%
aggr.	18649811	2809220	15.1%	20548942	1220525	5.9%

Table 5. *Cont.*

i	CL Method			ODP Model		
	Reserve	$\sqrt{\text{MSEP}}$	VCO	Reserve	$\sqrt{\text{MSEP}}$	VCO
1	94634	75535	79.8%	94634	110100	116.3%
2	469511	121700	25.9%	469511	216043	46.0%
3	709638	133551	18.8%	709638	260872	36.8%
4	984889	261412	26.5%	984889	303550	30.8%
5	1419459	411028	29.0%	1419459	375014	26.4%
6	2177641	558356	25.6%	2177641	495378	22.7%
7	3920301	875430	22.3%	3920301	789961	20.2%
8	4278972	971385	22.7%	4278972	1046514	24.5%
9	4625811	1363385	29.5%	4625811	1980101	42.8%
aggr.	18680856	2447618	13.1%	18680856	2945661	15.8%

The conservative BF method produces the smallest VCO for aggregated accident years, closely followed by Atherino et al. (2010), the scalar state space model and Alpuim and Ribeiro (2003). The model of Alpuim and Ribeiro (2003) and the scalar state space model provide a large VCO in the first accident year, resulting from the interaction of relatively low reserves and a larger MSEP for this year. In particular, for the scalar state space model the VCO decreases noticeably in subsequent accident years. Other models, like the log-normal models in Verrall (1989) and Li (2006), lead to relatively low VCO in the first accident years, but without a remarkable reduction in later accident years, so they remain mostly at the level of 25% $\pm$ 3%.

Compared with the results of the other seven models, the scalar state space model produces quite precise results with relatively low MSEPs and VCOs. Thus, the scalar structure of this model leads not only to facilitated practical application, but it is also in no way inferior to more complex models for the considered dataset.

6. Conclusions

The previously written papers in the stochastic claims reserving literature based on state space models and the KALMAN-filter have numerous contentual similarities, and we can classify them into four categories related to the methodology or modeling used in the respective papers. Models for incremental payments dominate in this field of research, in particular variations on the Hoerl curve, the log-normal model, as well as calendar year-based state space representations. However, most state space representations in these papers entail a matrix-based approach, which complicates their direct application in practice.

In contrast, the newly-developed scalar state space model for cumulative payments is quite elegant due to its simple, yet powerful structure. Since the CL method is the most commonly-used reserving method in practice and the scalar state space model is an extension of the CL method, the scalar state space model can readily be applied to claims reserving practice. In particular, this model provides facilitated calculation of forecasts for the cumulative payments in the lower run-off triangle and of smoothed values for the cumulative payments in the upper triangle. Moreover, the determination of claims reserves and the estimation of their MSEP for single and aggregated accident years are straightforward calculations using KALMAN-filter. The scalar state space model can also be used to identify and smooth outliers. This subject is of particular importance for less robust claims reserving methods such as the CL method. Summarizing, the scalar state space model is a very promising, robustified extension of the CL method, which renounces a complex matrix structure compared to most state space models in claims reserving and therefore simplifies its practical applications.

Because of its recursive and dynamic nature, the KALMAN-filter is predestined for use in stochastic claims reserving. The authors McGuire (2007), Taylor and McGuire (2008) and Chima-Okereke (2013) even recommend the use of state space models as a part of a nearly completely automated script, a so-called reserving robot in stochastic claims reserving. Due to the high flexibility of state space models and the KALMAN-filter, it is also possible to perform multivariate analyses in a simple manner and to take several run-off triangles simultaneously into account. In this way, dependencies, in particular correlations, between run-off triangles can be surveyed and included into the modeling.

Acknowledgments: The authors would like to thank three anonymous reviewers for their valuable feedback and suggestions.

Author Contributions: The authors have equally contributed to the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

This Appendix provides a brief proof of Theorem 1; for a more detailed proof, see Johannssen (2016) and Shumway and Stoffer (2010). For the derivation of the KALMAN recursions, we do not need any distributional assumptions. However, to simplify the notation, we can think of $\mathbb{E}$ for conditional expectations as a projection operator instead of an expectation. Then, the predictors obtained are the minimum MSEP predictors within the class of linear predictors.

Appendix A.1. One-Step Predictors and Error Variances

To derive recursion (17), we use the decomposition:

$$\underbrace{\mathbb{E}\left[C_{i,j+1}|C_{i,0}^{\text{obs}},\ldots,C_{i,j}^{\text{obs}}\right]}_{\text{1st component}} = \underbrace{\mathbb{E}\left[C_{i,j+1}|C_{i,0}^{\text{obs}},\ldots,C_{i,j-1}^{\text{obs}}\right]}_{\text{2nd component}} + \underbrace{\mathbb{E}\left[C_{i,j+1}|\varepsilon_{i,j}\right]}_{\text{3rd component}}. \tag{A1}$$

For the first component, we have:

$$\mathbb{E}\left[C_{i,j+1}|C_{i,0}^{\text{obs}},\ldots,C_{i,j}^{\text{obs}}\right] = \widehat{C}_{i,j+1}^{(\mathrm{P})}. \tag{A2}$$

For the second component, the following expression can be specified with (14):

$$\mathbb{E}\left[C_{i,j+1}|C_{i,0}^{\text{obs}},\ldots,C_{i,j-1}^{\text{obs}}\right] = \mathbb{E}\left[f_j C_{i,j} + v_{i,j}|C_{i,0}^{\text{obs}},\ldots,C_{i,j-1}^{\text{obs}}\right] = f_j \widehat{C}_{i,j}^{(\mathrm{P})} \tag{A3}$$

The third component:

$$\mathbb{E}\left[C_{i,j+1}|\varepsilon_{i,j}\right] = \underbrace{\mathbb{E}\left[C_{i,j+1}\varepsilon_{i,j}\right]}_{\text{1st term}} \Big(\underbrace{\mathbb{E}\left[\varepsilon_{i,j}^2\right]}_{\text{2nd term}}\Big)^{-1} \underbrace{\varepsilon_{i,j}}_{\text{3rd term}} \tag{A4}$$

consists of three terms, which we determine separately below. Since we need the innovation $\varepsilon_{i,j}$ for each of these three terms, we start with the third term using (13):

$$\varepsilon_{i,j} = C_{i,j}^{\text{obs}} - g_j \widehat{C}_{i,j}^{(\mathrm{P})} = g_j \left(C_{i,j} - \widehat{C}_{i,j}^{(\mathrm{P})}\right) + w_{i,j} \tag{A5}$$

Then, the first term in (A4) can be specified with the help of (14) and (A5) as follows:

$$\mathbb{E}\left[C_{i,j+1}\varepsilon_{i,j}\right] = \mathbb{E}\left[\left(f_j C_{i,j} + v_{i,j}\right)\left(g_j\left(C_{i,j} - \widehat{C}_{i,j}^{(\mathrm{P})}\right) + w_{i,j}\right)\right] = f_j \left(\widehat{\sigma}_{i,j}^{(\mathrm{P})}\right)^2 g_j \tag{A6}$$

The second term in (A4) can be determined using (A5):

$$\mathbb{E}\left[\varepsilon_{i,j}^2\right] = \mathbb{E}\left[\left(g_j\left(C_{i,j} - \widehat{C}_{i,j}^{(\mathrm{P})}\right) + w_{i,j}\right)^2\right] = g_j^2\left(\widehat{\sigma}_{i,j}^{(\mathrm{P})}\right)^2 + \sigma_w^2 \tag{A7}$$

Substituting three terms (A5), (A6) and (A7) into (A4) leads to:

$$\mathbb{E}\left[C_{i,j+1}|\varepsilon_{i,j}\right] = f_j\left(\widehat{\sigma}_{i,j}^{(\mathrm{P})}\right)^2 g_j\left(g_j^2\left(\widehat{\sigma}_{i,j}^{(\mathrm{P})}\right)^2 + \sigma_w^2\right)^{-1}\left(C_{i,j}^{\mathrm{obs}} - g_j\widehat{C}_{i,j}^{(\mathrm{P})}\right) \tag{A8}$$

Inserting components (A2), (A3) and (A8) into (A1) directly implies:

$$\widehat{C}_{i,j+1}^{(\mathrm{P})} = f_j\widehat{C}_{i,j}^{(\mathrm{P})} + f_j\left(\widehat{\sigma}_{i,j}^{(\mathrm{P})}\right)^2 g_j\left(g_j^2\left(\widehat{\sigma}_{i,j}^{(\mathrm{P})}\right)^2 + \sigma_w^2\right)^{-1}\left(C_{i,j}^{\mathrm{obs}} - g_j\widehat{C}_{i,j}^{(\mathrm{P})}\right),$$

which is equal to (17) with $\gamma_{i,j} = \left(\widehat{\sigma}_{i,j}^{(\mathrm{P})}\right)^2 g_j$ and $\Delta_{i,j} = g_j^2\left(\widehat{\sigma}_{i,j}^{(\mathrm{P})}\right)^2 + \sigma_w^2$.

□

The verification of recursion (21) can be carried out using the decomposition:

$$\left(\widehat{\sigma}_{i,j+1}^{(\mathrm{P})}\right)^2 = \underbrace{\mathbb{E}[C_{i,j+1}^2]}_{\text{1st component}} - \underbrace{\mathbb{E}\left[\left(\widehat{C}_{i,j+1}^{(\mathrm{P})}\right)^2\right]}_{\text{2nd component}}. \tag{A9}$$

For the first component, we have with (14):

$$\mathbb{E}[C_{i,j+1}^2] = \mathbb{E}[\,(f_jC_{i,j} + v_{i,j})^2\,] = f_j^2\mathbb{E}[C_{i,j}^2] + \sigma_v^2, \tag{A10}$$

while we get for the second component with (17), (A5) and (A7):

$$\mathbb{E}\left[\left(\widehat{C}_{i,j+1}^{(\mathrm{P})}\right)^2\right] = \mathbb{E}\left[\left(f_j\widehat{C}_{i,j}^{(\mathrm{P})} + \frac{f_j\gamma_{i,j}}{\Delta_{i,j}}\left(C_{i,j}^{\mathrm{obs}} - g_j\widehat{C}_{i,j}^{(\mathrm{P})}\right)\right)^2\right] = f_j^2\mathbb{E}\left[\left(\widehat{C}_{i,j}^{(\mathrm{P})}\right)^2\right] + \frac{(f_j\gamma_{i,j})^2}{\Delta_{i,j}} \tag{A11}$$

Inserting (A10) and (A11) into (A9) directly leads to:

$$\left(\widehat{\sigma}_{i,j+1}^{(\mathrm{P})}\right)^2 = f_j^2\mathbb{E}[C_{i,j}^2] + \sigma_v^2 - \left(f_j^2\mathbb{E}\left[\left(\widehat{C}_{i,j}^{(\mathrm{P})}\right)^2\right] + \frac{(f_j\gamma_{i,j})^2}{\Delta_{i,j}}\right) = f_j^2\left(\widehat{\sigma}_{i,j}^{(\mathrm{P})}\right)^2 + \sigma_v^2 - \frac{(f_j\gamma_{i,j})^2}{\Delta_{i,j}}$$

□

Appendix A.2. h-Step Predictors and Error Variances

Recursion (18) can be easily determined by updating the one-step predictions (17) (see Remark 3, second bullet point):

$$\widehat{C}_{i,j+h}^{(\mathrm{P})} = f_{j+h-1}\widehat{C}_{i,j+h-1}^{(\mathrm{P})} = \ldots = (f_{j+h-1}\cdots f_{j+1})\widehat{C}_{i,j+1}^{(\mathrm{P})} \tag{A12}$$

□

Recursion (22) for the error variances $\left(\widehat{\sigma}_{i,j+h}^{(\mathrm{P})}\right)^2$ with $h \geq 2$ can be obtained using (14) and (A12):

$$\left(\widehat{\sigma}_{i,j+h}^{(\mathrm{P})}\right)^2 = \mathbb{E}\left[\left(f_{j+h-1}C_{i,j+h-1} + v_{i,j+h-1} - f_{j+h-1}\widehat{C}_{i,j+h-1}^{(\mathrm{P})}\right)^2\right] = f_{j+h-1}^2\left(\widehat{\sigma}_{i,j+h-1}^{(\mathrm{P})}\right)^2 + \sigma_v^2$$

□

Appendix A.3. Filtering Predictors and Error Variances

To derive recursion (19), we use the decomposition:

$$\underbrace{\mathbb{E}\left[C_{i,j}|C_{i,0}^{\text{obs}},\ldots,C_{i,j}^{\text{obs}}\right]}_{\text{1st component}} = \underbrace{\mathbb{E}\left[C_{i,j}|C_{i,0}^{\text{obs}},\ldots,C_{i,j-1}^{\text{obs}}\right]}_{\text{2nd component}} + \underbrace{\mathbb{E}\left[C_{i,j}|\varepsilon_{i,j}\right]}_{\text{3rd component}}. \tag{A13}$$

The specification of three components in (A13) is analogous to the derivation of the recursion for one-step prediction. Therefore, we have for the first and second component:

$$\mathbb{E}\left[C_{i,j}|C_{i,0}^{\text{obs}},\ldots,C_{i,j}^{\text{obs}}\right] = \widehat{C}_{i,j}^{(\text{F})} \tag{A14}$$

$$\mathbb{E}\left[C_{i,j}|C_{i,0}^{\text{obs}},\ldots,C_{i,j-1}^{\text{obs}}\right] = \widehat{C}_{i,j}^{(\text{P})} \tag{A15}$$

The third component consists of three terms, whose second and third terms are identical to those in (A4), but the first term differs slightly:

$$\mathbb{E}\left[C_{i,j}|\varepsilon_{i,j}\right] = \underbrace{\mathbb{E}\left[C_{i,j}\varepsilon_{i,j}\right]}_{\text{1st term}} \Big(\underbrace{\mathbb{E}\left[\varepsilon_{i,j}^2\right]}_{\text{2nd term}} \Big)^{-1} \underbrace{\varepsilon_{i,j}}_{\text{3rd term}} \tag{A16}$$

Whereas the last both terms in (A16) are already given by (A5) and (A7), the first term can be specified using (A5):

$$\mathbb{E}\left[C_{i,j}\varepsilon_{i,j}\right] = \mathbb{E}\left[C_{i,j}\left(g_j\left(C_{i,j}-\widehat{C}_{i,j}^{(\text{P})}\right)+w_{i,j}\right)\right] = \left(\widehat{\sigma}_{i,j}^{(\text{P})}\right)^2 g_j \tag{A17}$$

Substituting three terms (A5), (A7) and (A17) into (A16) leads to:

$$\mathbb{E}\left[C_{i,j}|\varepsilon_{i,j}\right] = \left(\widehat{\sigma}_{i,j}^{(\text{P})}\right)^2 g_j \left(g_j^2\left(\widehat{\sigma}_{i,j}^{(\text{P})}\right)^2+\sigma_w^2\right)^{-1}\left(C_{i,j}^{\text{obs}}-g_j\widehat{C}_{i,j}^{(\text{P})}\right) \tag{A18}$$

Inserting components (A14), (A15) and (A18) into (A13) directly implies:

$$\widehat{C}_{i,j}^{(\text{F})} = \widehat{C}_{i,j}^{(\text{P})} + \left(\widehat{\sigma}_{i,j}^{(\text{P})}\right)^2 g_j \left(g_j^2\left(\widehat{\sigma}_{i,j}^{(\text{P})}\right)^2+\sigma_w^2\right)^{-1}\left(C_{i,j}^{\text{obs}}-g_j\widehat{C}_{i,j}^{(\text{P})}\right),$$

which is equal to (19) with $\gamma_{i,j} = \left(\widehat{\sigma}_{i,j}^{(\text{P})}\right)^2 g_j$ and $\Delta_{i,j} = g_j^2\left(\widehat{\sigma}_{i,j}^{(\text{P})}\right)^2+\sigma_w^2$.

□

To establish recursion (23), we use (19) as follows:

$$C_{i,j}-\widehat{C}_{i,j}^{(\text{P})} = C_{i,j}-\widehat{C}_{i,j}^{(\text{F})}+\widehat{C}_{i,j}^{(\text{F})}-\widehat{C}_{i,j}^{(\text{P})} = C_{i,j}-\widehat{C}_{i,j}^{(\text{F})}+\frac{\gamma_{i,j}\varepsilon_{i,j}}{\Delta_{i,j}} \tag{A19}$$

Hence, from (A19), we obtain:

$$\widehat{C}_{i,j}^{(\text{P})} = \widehat{C}_{i,j}^{(\text{F})} - \frac{\gamma_{i,j}\varepsilon_{i,j}}{\Delta_{i,j}} \tag{A20}$$

Using (A7) and (A20), we get the following expression:

$$\left(\widehat{\sigma}_{i,j}^{(\text{P})}\right)^2 = \mathbb{E}\left[\left(C_{i,j}-\widehat{C}_{i,j}^{(\text{F})}+\frac{\gamma_{i,j}\varepsilon_{i,j}}{\Delta_{i,j}}\right)^2\right] = \left(\widehat{\sigma}_{i,j}^{(\text{F})}\right)^2+\frac{\gamma_{i,j}^2}{\Delta_{i,j}},$$

which directly leads to (23).

□

Appendix A.4. Fixed-Interval Smoothing Predictors and Error Variances

In order to obtain fixed-interval recursion (20), first, we need to derive the KALMAN fixed-point recursion:

$$\widehat{C}^{(S)}_{i,j|s} = \widehat{C}^{(S)}_{i,j|s-1} + K_{i,j,s}\left(C^{\text{obs}}_{i,s} - g_s\widehat{C}^{(P)}_{i,s}\right) \tag{A21}$$

with $K_{i,j,s} = \frac{\gamma_{i,j,s}}{\Delta_{i,s}}$, $\gamma_{i,j,s} = g_s\mathbb{E}\left[\left(C_{i,j} - \widehat{C}^{(P)}_{i,j}\right)\left(C_{i,s} - \widehat{C}^{(P)}_{i,s}\right)\right]$, $\Delta_{i,s} = g_s^2\left(\widehat{\sigma}^{(P)}_{i,s}\right)^2 + \sigma_w^2$ for $s = j+1, j+2, \ldots, I-i$, a fixed $i = 0, \ldots, I$ and a fixed $j = 0, \ldots, I-i$. For derivation of fixed-point recursion (A21), we use the decomposition:

$$\underbrace{\mathbb{E}\left[C_{i,j}|C^{\text{obs}}_{i,0}, \ldots, C^{\text{obs}}_{i,s}\right]}_{\text{1st component}} = \underbrace{\mathbb{E}\left[C_{i,j}|C^{\text{obs}}_{i,0}, \ldots, C^{\text{obs}}_{i,s-1}\right]}_{\text{2nd component}} + \underbrace{\mathbb{E}\left[C_{i,j}|\varepsilon_{i,s}\right]}_{\text{3rd component}}. \tag{A22}$$

The specification of three components in (A22) is analogous to the derivation of the recursion for one-step prediction. Therefore, we have for the first and second component:

$$\mathbb{E}\left[C_{i,j}|C^{\text{obs}}_{i,0}, \ldots, C^{\text{obs}}_{i,s}\right] = \widehat{C}^{(S)}_{i,j|s} \tag{A23}$$
$$\mathbb{E}\left[C_{i,j}|C^{\text{obs}}_{i,0}, \ldots, C^{\text{obs}}_{i,s-1}\right] = \widehat{C}^{(S)}_{i,j|s-1} \tag{A24}$$

The third component consists of three terms, whose second and third terms are identical to those in (A4), but the first term differs slightly:

$$\mathbb{E}\left[C_{i,j}|\varepsilon_{i,s}\right] = \underbrace{\mathbb{E}\left[C_{i,j}\varepsilon_{i,s}\right]}_{\text{1st term}}\Big(\underbrace{\mathbb{E}\left[\varepsilon^2_{i,s}\right]}_{\text{2nd term}}\Big)^{-1}\underbrace{\varepsilon_{i,s}}_{\text{3rd term}} \tag{A25}$$

Whereas the last both terms in (A25) are already given by (A5) and (A7), the first term can be specified using (A5):

$$\begin{aligned}\mathbb{E}\left[C_{i,j}\varepsilon_{i,s}\right] &= \mathbb{E}\left[C_{i,j}\left(g_s\left(C_{i,s} - \widehat{C}^{(P)}_{i,s}\right) + w_{i,s}\right)\right]\\ &= \mathbb{E}\left[\left(\left(C_{i,j} - \widehat{C}^{(P)}_{i,j}\right) + \widehat{C}^{(P)}_{i,j}\right)\left(g_s\left(C_{i,s} - \widehat{C}^{(P)}_{i,s}\right) + w_{i,s}\right)\right]\\ &= g_s\mathbb{E}\left[\left(C_{i,j} - \widehat{C}^{(P)}_{i,j}\right)\left(C_{i,s} - \widehat{C}^{(P)}_{i,s}\right)\right]\end{aligned} \tag{A26}$$

Substituting three terms (A5), (A7) and (A26) into (A25) leads to:

$$\mathbb{E}\left[C_{i,j}|\varepsilon_{i,s}\right] = g_s\mathbb{E}\left[\left(C_{i,j} - \widehat{C}^{(P)}_{i,j}\right)\left(C_{i,s} - \widehat{C}^{(P)}_{i,s}\right)\right]\left(g_s^2\left(\widehat{\sigma}^{(P)}_{i,s}\right)^2 + \sigma_w^2\right)^{-1}\left(C^{\text{obs}}_{i,s} - g_s\widehat{C}^{(P)}_{i,s}\right) \tag{A27}$$

Inserting components (A23), (A24) and (A27) into (A22) directly implies:

$$\widehat{C}^{(S)}_{i,j|s} = \widehat{C}^{(S)}_{i,j|s-1} + g_s\mathbb{E}\left[\left(C_{i,j} - \widehat{C}^{(P)}_{i,j}\right)\left(C_{i,s} - \widehat{C}^{(P)}_{i,s}\right)\right]\left(g_s^2\left(\widehat{\sigma}^{(P)}_{i,s}\right)^2 + \sigma_w^2\right)^{-1}\left(C^{\text{obs}}_{i,s} - g_s\widehat{C}^{(P)}_{i,s}\right),$$

which is equal to (A21) with $\gamma_{i,j,s} = g_s\mathbb{E}\left[\left(C_{i,j} - \widehat{C}^{(P)}_{i,j}\right)\left(C_{i,s} - \widehat{C}^{(P)}_{i,s}\right)\right]$, $\Delta_{i,s} = g_s^2\left(\widehat{\sigma}^{(P)}_{i,s}\right)^2 + \sigma_w^2$ and $K_{i,j,s} = \frac{\gamma_{i,j,s}}{\Delta_{i,s}}$.

Now, we can derive fixed-interval recursion (20) by considering (A21), i.e.,

$$\widehat{C}^{(S)}_{i,j|s} = \widehat{C}^{(S)}_{i,j|s-1} + K_{i,j,s}\varepsilon_{i,s} = \ldots = \widehat{C}^{(F)}_{i,j} + \sum_{l=j+1}^{s} K_{i,j,l}\varepsilon_{i,l}. \tag{A28}$$

In a similar way, we get:

$$\widehat{C}^{(S)}_{i,j+1|s} = \widehat{C}^{(P)}_{i,j+1} + \sum_{l=j+1}^{s} K_{i,j+1,l}\varepsilon_{i,l} \tag{A29}$$

Comparing (A28) and (A29), we find that the relation $K_{i,j,l} = \psi_{i,j} K_{i,j+1,l}$ with $\psi_{i,j} = f_{j+1} \frac{\left(\widehat{\sigma}^{(F)}_{i,j}\right)^2}{\left(\widehat{\sigma}^{(P)}_{i,j+1}\right)^2}$ holds. Therefore, using these findings, we get the following equation:

$$\widehat{C}^{(S)}_{i,j|s} = \widehat{C}^{(F)}_{i,j} + \psi_{i,j} \sum_{l=j+1}^{s} K_{i,j+1,l}\varepsilon_{i,l} = \widehat{C}^{(F)}_{i,j} + \psi_{i,j}\left(\widehat{C}^{(S)}_{i,j+1|s} - \widehat{C}^{(P)}_{i,j+1}\right) \tag{A30}$$

Since we smooth observations in the upper triangle within the fixed-interval recursion using all available observations, we have $s = I - i$ (s is therefore no longer required) and can simplify the notation $\widehat{C}^{(S)}_{i,j|s}$ to $\widehat{C}^{(S)}_{i,j}$ for all $j = I - i, \ldots, 0$. Then, (A30) is equivalent to (20).

□

The verification of (24) follows from (20) with some straightforward calculations. Using (20), we obtain:

$$\left(C_{i,j} - \widehat{C}^{(S)}_{i,j}\right) + \psi_{i,j}\widehat{C}^{(S)}_{i,j+1} = \left(C_{i,j} - \widehat{C}^{(F)}_{i,j}\right) + \psi_{i,j}\widehat{C}^{(P)}_{i,j+1}. \tag{A31}$$

Squaring and taking expectations of both sides of (A31), as also using the fact the cross products are zero, imply:

$$\left(\widehat{\sigma}^{(S)}_{i,j}\right)^2 + \psi^2_{i,j}\mathbb{E}\left[\left(\widehat{C}^{(S)}_{i,j+1}\right)^2\right] = \left(\widehat{\sigma}^{(F)}_{i,j}\right)^2 + \psi^2_{i,j}\mathbb{E}\left[\left(\widehat{C}^{(P)}_{i,j+1}\right)^2\right]. \tag{A32}$$

Now, we use decompositions for the remaining expectations (see (A9)):

$$\mathbb{E}\left[\left(\widehat{C}^{(S)}_{i,j+1}\right)^2\right] = \mathbb{E}[C^2_{i,j+1}] - \left(\widehat{\sigma}^{(S)}_{i,j+1}\right)^2 \tag{A33}$$

$$\mathbb{E}\left[\left(\widehat{C}^{(P)}_{i,j+1}\right)^2\right] = \mathbb{E}[C^2_{i,j+1}] - \left(\widehat{\sigma}^{(P)}_{i,j+1}\right)^2 \tag{A34}$$

Inserting the right-hand sides of (A33) and (A34) into (A32) directly leads to (24).

□

References

Alpuim, Teresa, and Isabel Ribeiro. 2003. A State Space Model for Run-Off Triangles. *Applied Stochastic Models in Business and Industry* 19: 105–20.

Atherino, Rodrigo, Adrian Pizzinga, and Cristiano Fernandes. 2010. A row-wise Stacking of the Runoff Triangle: State Space Alternatives for IBNR Reserve Prediction. *ASTIN Bulletin* 40: 917–46.

Brockwell, Peter J., and Richard A. Davis. 2006. *Time Series: Theory and Methods*, 2nd ed. New York: Springer.

Chima-Okereke, Chibisi. 2013. A practical Approach to Claims Reserving using State Space Models with Growth Curves. Paper presented at the R in Insurance Conference 2013, London, UK, July 15.

De Jong, Piet. 2005. State Space Models in Actuarial Science. Paper presented at the Second Brazilian Conference on Statistical Modelling in Insurance, Institute of Mathematics and Statistics, University of São Paulo, Maresias, Brazil, August 28–September 3.

De Jong, Piet. 2006. Forecasting Runoff Triangles. *North American Actuarial Journal* 10: 28–38.

De Jong, Piet, and Ben Zehnwirth. 1983. Claims Reserving, State-Space Models and the Kalman Filter. *Journal of the Institute of Actuaries* 110: 157–81.

England, Peter D., and Richard J. Verrall. 2002. Stochastic Claims Reserving in General Insurance. *British Actuarial Journal* 8: 443–518.

Frees, Edward W. 2010. *Regression Modeling with Actuarial and Financial Applications*. Cambridge: Cambridge University Press.

Johannssen, Arne. 2016. *Stochastische Schadenreservierung unter Verwendung von Zustandsraummodellen und des Kalman-Filters*. Hamburg: Dr. Kovac.

Kaas, Rob, Marc Goovaerts, Jan Dhaene, and Michel Denuit. 2009. *Modern Actuarial Risk Theory—Using R*, 2nd ed. Berlin: Springer.

Kalman, Rudolf E. 1960. A New Approach to Linear Filtering and Prediction Problems. *Trans. of the ASME—Journal of Basic Engineering (Series D)* 82: 35–45.

Kremer, Erhard. 1982. IBNR-Claims and the Two-Way Model of ANOVA. *Scandinavian Actuarial Journal* 1982: 47–55.

Li, Jackie. 2006. Comparison of Stochastic Reserving Methods. *Australian Actuarial Journal* 12: 489–569.

Mack, Thomas. 1993. Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin* 23: 213–25.

McGuire, Gráinne. 2007. Building a Reserving Robot. Paper presented at the Biennial Convention 2007, Institute of Actuaries of Australia, Christchurch, New Zealand, September 23–26.

Ntzoufras, Ioannis, and Petros Dellaportas. 2002. Bayesian Modelling of Outstanding Liabilities incorporating Claim Count Uncertainty. *North American Actuarial Journal* 6: 113–28.

Pang, Liyan, and Siqi He. 2012. The Application of State-Space Model in Outstanding Claims Reserve. Paper presented at the 2012 International Conference on Information Management, Innovation Management and Industrial Engineering (ICIII), Sanya, China, October 20–21, pp. 271–74.

Shumway, Robert H., and David S. Stoffer. 1982. An Approach to Time Series Smoothing and Forecasting using the EM Algorithm. *Journal of Time Series Analysis* 3: 253–64.

Shumway, Robert H., and David S. Stoffer. 2010. *Time Series Analysis and Its Applications (With R Examples)*, 3rd ed. New York: Springer.

Taylor, Greg C. 2000. *Loss Reserving: An Actuarial Perspective*. Boston: Kluwer Academic Publishers.

Taylor, Greg C., and Frank R. Ashe. 1983. Second Moments of Estimates of Outstanding Claims. *Journal of Econometrics* 23: 37–61.

Taylor, Greg C., and Gráinne McGuire. 2008. Robotic Reserving. *GIRO Convention 2013*, Italy.

Taylor, Greg C., Gráinne McGuire, and Alan Greenfield. 2003. *Loss Reserving: Past, Present and Future*. Research Paper No. 109. Melbourne: University of Melbourne.

Van Wouwe, Martine, Tim Verdonck, and Kristel Van Rompay. 2009. Application of Classical and Robust Chain-Ladder Methods: Results for the Belgian Non-Life Business. *Global Business and Economics Review* 11: 99–115.

Verdonck, Tim, and Michiel Debruyne. 2011. The Influence of Individual Claims on the Chain-Ladder Estimates: Analysis and Diagnostic Tool. *Insurance: Mathematics and Economics* 48: 85–98.

Verdonck, Tim, and Martine Van Wouwe. 2011. Detection and Correction of Outliers in the Bivariate Chain-Ladder Method. *Insurance: Mathematics and Economics* 49: 188–93.

Verdonck, Tim, Martine Van Wouwe, and Jan Dhaene. 2009. A Robustification of the Chain-Ladder Method. *North American Actuarial Journal* 13: 280–98.

Verrall, Richard J. 1989. A State Space Representation of the Chain Ladder Linear Model. *Journal of the Institute of Actuaries* 116: 589–610.

Verrall, Richard J. 1994. A Method for Modelling Varying Run-off Evolutions in Claims Reserving. *ASTIN Bulletin* 24: 325–32.

Verrall, Richard J. 2004. Kalman Filter, Reserving Methods. In *Encyclopedia of Actuarial Science*. Edited by J. L. Teugels and B. Sundt. Chichester: John Wiley & Sons, vol. 1, pp. 952–55.

Wright, Thomas S. 1990. A Stochastic Method for Claims Reserving in General Insurance. *Journal of the Institute of Actuaries* 117: 677–731.

Wüthrich, Mario V., and Michael Merz. 2008. *Stochastic Claims Reserving Methods in Insurance*. Chichester: John Wiley & Sons.

Zehnwirth, Ben. 1997. Kalman Filters with Applications to Loss Reserving. *Insurance: Mathematics and Economics* 20: 149–218.