Measurement Claim Reserving Insurance Contract in Accordance with IFRS 17 using Stochastic State Space Model

Adhitya Ronnie Effendie[1*], and *Rafida* 'Alaiyya Hayyin[1]

[1]Department of Mathematic, Universitas Gadjah Mada, Yogyakarta, DI Yogyakarya, Indonesia

Abstract. International Financial Reporting Standard 17 (IFRS 17) is a global accounting standard developed by the International Accounting Standards Board (IASB). It guides the accounting treatment of insurance contracts in financial statements. In IFRS 17, claim reserving is vital for measuring insurance contracts and determining the Contractual Service Margin (CSM). The claim reserving process involves estimating expected claim payouts over the contract's duration. These estimates are crucial for calculating the liability for incurred claims and determining the CSM, representing unearned profit. Claim data is treated as ordered data by sorting the run-off triangle row-wise, resulting univariate time series with missing values. The state space model provides a framework for modeling time series, assuming system development over time. By taking the log of claim data so that the data is normally distributed, the lognormal state space model, along with Kalman recursion algorithms, can be applied for Incurred but Not Reported (IBNR) projections. The lognormal state space model's IBNR reserve estimate is more precise and sufficient to compensate loss than the chain ladder method. Furthermore, the performance of the lognormal state space model outperforms the chain ladder method by means of model validation using RMSE, MAE, RSE, and R^2.

1 Introduction

International Financial Reporting Standard 17 (IFRS 17) which will replace IFRS 4 [1], introduces a new concept to insurance accounting that changes the financial reporting for insurance and reinsurance [2]. IFRS 17 will bring substantial changes to how income from insurance contracts is calculated, especially for contracts involving life and extended coverage, resulting in significant shifts in measurement approaches. In the context of IFRS 17, claim reserving calculation plays a significant role in the measurement of insurance contracts and the determination of the Contractual Service Margin (CSM). CSM represents the unearned profit that an enitity expects to earn as it provides service [3]. While the process of claim reserving requires predicting the amount of insurer needs to pay (best estimate) throughout the duration of the insurance contract [4].

* Corresponding author: adhityaronnie@ugm.ac.id

Various methods have been developed to predict claim reserving. In general, the statistical methods used to estimate claim reserves are divided into two methods, namely deterministic and stochastic. The key strength of stochastic methods is that they yield a wealth of statistical information which would not have been available from deterministic methods [5]. The stochastic claims reserve estimation method makes the observed data, specifically claims, into random variables. In over 40 years, only a few stochastic methods explicitly use time series models. Claim data is usually presented in a run-off triangle using double indexing [6]. There are studies [7] that proposed an approach that constructs claim data into a time series with a single index by stacking the row of run-off triangle, then ordering the values of the run-off triangle row-wisely, resulting a univariate time series with several missing values which can be estimated.

There are several actuarial literatures contains a state space represenation with Kalman algorithms applied to stochastic claim reserving. One that originally develop that method is [8], then [9] establish state space model called linear CL model. The same representation is also used by [10-15].The row-wise stacking method from [7] was most recently extended by [16] using both univariate and multivariate cases of several dependent run-off triangles for correlated business lines using state space representation.

By presenting claim data as a time series, a time series approach can be used, namely the state space model method using Kalman recursion in the form of filtering and smoothing to estimate IBNR claim reserves. The state space model offers a framework for simulating various time series types and assumes that the observed system will change with time, allowing its parameters to evolve as well. The linear Gaussian state space model, commonly referred to as the dynamic linear model, is one of the most basic state space models. This model can be tracked analytically, which makes it popular in many scientific fields, including the projection of IBNR (incurred but not reported) claim reserves.

This study applies what has been developed by [7,16], then compares it to a deterministic method. One of the methods developed by [16], called log-normal state space model will be used in this research because the model seems to be adequate for all datasets from the study of [16]. This research also applies the approach used by [17] to provide time series claim data. The IBNR projection by log-normal state space model with Kalman recursion will be compared to the deterministic method, namely chain ladder that simple and widely used in general insurance [17].

The metrics measurement used in this study to compare the predictive performances are RMSE, MAE, RSE, and R^2 which commonly used as model validation. To measure the predictive performance, this study uses datasets from NAIC(National Association of Insurance Commissioners) that consisting of run-off triangles of six lines of business for all U.S. property casualty insurers, correspond to claims of accident year 1988-1997 with 10 years development lag. The database accommodates run-off triangle along with the future triangle (upper and lower) triangle. The lower triangle of the datasets is used to validate the model discussed in this paper. This kind of model validation that use lower triangle to see the predictive performance also have been used by [18,19,20].

2 Method

In general, claim data is arranged in a run-off triangle using a double index format. The row in the run-off triangle shows the year of occurance and the column in the run-off triangle shows the year of development, as shown in Table 1. For convenience, the accident year is denoted by $i = (0, \ldots, s)$ and the development year denotes by $j = (0, \ldots, s)$.

The value of $X_{i,j}$ denotes the incremental payment that occurred in accident year i and development year j. In this study, Table 1 was rearranged as a time series by stacking the rows of run-off triangle and replaced the double index by a single index t. Table 2 perform a table that shows the log of the incremental claim value. The logarithmic form of the incremental claim Y_{ij} is

$$Y_{ij} = \log(X_{ij}), \tag{1}$$

with $X_{ij} > 0$ for all i and j.

Table 1. Incremental run-off triangle

Accident year i	Development year j							
	0	1	2	3	…	s-2	s-1	s
0	$X_{0,0}$	$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	…	$X_{0,s-2}$	$X_{0,s-1}$	$X_{0,s}$
1	$X_{1,0}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	…	$X_{1,s-2}$	$X_{1,s-1}$	$X_{1,s}$
2	$X_{2,0}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	…	$X_{2,s-2}$	$X_{2,s-1}$	$X_{2,s}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
s-1	$X_{s-1,0}$	$X_{s-1,1}$	$X_{s-1,2}$	$X_{s-1,3}$	…	$X_{s-1,s-2}$	$X_{s-1,s-1}$	$X_{s-1,s}$
s	$X_{s,0}$	$X_{s,1}$	$X_{s,2}$	$X_{s,3}$	…	$X_{s,s-2}$	$X_{s,s-1}$	$X_{s,s}$

The logarithmic incremental claim has been introduced by [21]. If there are negative values in the first row, they will be replaced by the average of the entire value in the first row, if there are negative values in other rows, they will be removed and treated as missing values. The former double index Y_{ij} was replaced by y_t with $t=1,2,\dots,s(s+1)$, except the first column as shown in Table 2. The first column of the run-off triangle in Table 2 has a double index and is used as the initial value (Y_{i0}) in the state space model.

Table 2. Row wise ordered run-off triangle

Accident year i	Development year j							
	0	1	2	3	…	s-2	s-1	s
0	$Y_{0,0}$	y_1	y_2	y_3	…	y_{s-2}	y_{s-1}	y_s
1	$Y_{1,0}$	y_{s+1}	y_{s+2}	y_{s+3}	…	y_{2s-2}	y_{2s-1}	y_{2s}
2	$Y_{2,0}$	y_{2s+1}	y_{2s+2}	y_{2s+3}	…	y_{3s-2}	y_{3s-1}	y_{3s}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
s-1	$Y_{s-1,0}$	$y_{(s-1)s+1}$	$y_{(s-1)s+2}$	$y_{(s-1)s+3}$	…	$y_{s.s-2}$	$y_{s.s-1}$	$y_{s.s}$
s	$Y_{s,0}$	$y_{s.s+1}$	$y_{s.s+2}$	$y_{s.s+3}$	…	$y_{(s+1).s-2}$	$y_{(s+1).s-1}$	$y_{(s+1).s}$

Claim data in time series form can be modeled using the lognormal state space model following this equation below.

$$y_t - Y_{i0} = \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim \text{ independent } N(0, \sigma_\varepsilon^2), \tag{2}$$

$$\alpha_{t+1} = \alpha_{t-s+1} + \eta_t, \quad \eta_t \sim \text{ independent } N(0, \sigma_\alpha^2). \tag{3}$$

The notation α_t becomes the state of the state space model, it represents the column effects of the claim in the run-off triangle, Y_{i0} represents the row effects of the claim in the run-off triangle, while $\boldsymbol{\varepsilon_t}$ and $\boldsymbol{\eta_t}$ denotes the residual of observation equation (2) and state equation (3), respectively. The above lognormal state space models (2) and (3) can be brought into linear Gaussian state space model form with a state vector $\boldsymbol{\vartheta_t} = (\alpha_t, \alpha_{t-1}, \ldots, \alpha_{t-s+1})'$.

$$y_t - Y_{i0} = (1,0,\ldots,0)\boldsymbol{\vartheta_t} + \varepsilon_{t,}\varepsilon_t \sim \text{ independent } N(0,\sigma_\varepsilon^2), \tag{4}$$

$$\boldsymbol{\vartheta_{t+1}} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \boldsymbol{\vartheta_t} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \eta_t,\ \eta_t \sim \text{ independent } N(0,\sigma_\alpha^2). \tag{5}$$

The vector (1, 0, …, 0) in (4) with dimension $(1 \times \mathrm{s})$ and square matrix with dimension $(s \times \mathrm{s})$ in (5) related to matrix Z_t and T_t respectively in observation and state equation of general form of linear gaussian state space model below. This general form can also be found in [22].

$$y_t = Z_t\boldsymbol{\alpha_t} + \varepsilon_{t,} \qquad \varepsilon_t \sim \text{ independent } N(0, H_t), \tag{6}$$

$$\boldsymbol{\alpha_{t+1}} = T_t\boldsymbol{\alpha_t} + R_t\eta_t, \qquad \eta_t \sim \textit{ independent} N(0, Q_t). \tag{7}$$

Table 3. Dimension of linear gaussian state space model

Vector		Matrix	
y_t	$p \times 1$	Z_t	$p \times m$
α_t	$m \times 1$	T_t	$m \times m$
ε_t	$p \times 1$	H_t	$p \times p$
η_t	$k \times 1$	R_t	$m \times k$
a_1	$m \times 1$	Q_t	$k \times k$
		P_1	$m \times m$

Variables y_t and α_t are time varying, while Z_t, T_t, R_t, H_t, Q_t are time invariant. The initial state $\alpha_1 \sim N(a_1,\ P_1)$ and the state α_t can be achieved by Kalman recursion filtering and smoothing, given the observation y_t. The main objectives of recursive algorithm of Kalman filtering are to calculate $a_{t+1} = \mathbb{E}(\alpha_{t+1}|Y_t)$ and $P_{t+1} = Var(\alpha_{t+1}|Y_t)$ with Y_t is set of observation $y_1,\ldots.y_t$. With Kalman filtering algorithm, the one step ahead predictions of state concurently with predictions of error with a_1 and P_1 are known are

$$v_t = y_t - Z_t\boldsymbol{a_t}, \tag{8}$$

$$K_t = T_tP_tZ_t'F_t^{-1}, \tag{9}$$

$$\boldsymbol{a_{t|t}} = \boldsymbol{a_t} + M_tF_t^{-1}v_t, \tag{10}$$

$$\boldsymbol{a_{t+1}} = T_t\boldsymbol{a_{t|t}}, \tag{11}$$

$$F_t = Z_tP_tZ_t' + H_t, \tag{12}$$

$$M_t = P_tZ_t', \tag{13}$$

$$P_{t|t} = P_t - M_tF_t^{-1}M_t' \tag{14}$$

$$P_{t+1} = T_tP_{t|t}T_t' + R_tQ_tR_t'. \tag{15}$$

In this study, the expected initial state $\boldsymbol{a_1} = (0,0,\dots,0)'$ and the covariance matrix $P_1 = \kappa \times I$ is an idenitity matrix multiplied by $\kappa \gg 0$ where κ is set to be $\kappa = 10^6$. When all of the states a_t are already predicted, the role of Kalman smoothing algorithm is to find the smoothed states of each a_t given that $y_1,\dots,y_n$ are known with a backward recursion. The vector of $(y_1', \dots, y_n')$ denoted as y, then the estimate of $\hat{\alpha}_t = \mathbb{E}(\alpha_t|y)$ with error variance denotes as $V_t = Var(\alpha_t - \hat{\alpha}_t) = Var(\alpha_t|y)$ for t=1, …,n. The estimate of the smoothed state vector is

$$\boldsymbol{\hat{\alpha}_t} = \boldsymbol{a_t} + P_t r_{t-1}, \tag{16}$$

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \tag{17}$$

with $r_n = 0$. Then the variance of the smoothed state is

$$V_t = P_t - P_t N_{t-1} P_t, \tag{18}$$

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \tag{19}$$

with $N_n = 0$. In some cases where there are missing values, the vector of v_t and K_t in Kalman filtering and smoothing are set to be zero, $v_t = 0$ and $K_t = 0$. A more complete and detailed formula of Kalman filtering and smoothing can be found at [23].

The unknown parameters of the lognormal state space model are σ_ε^2 and σ_α^2 as the variance of residual ε_t and η_t respectively. The variances σ_ε^2 and σ_α^2 is homoscedastic with constant unknown (positive) variances. They are estimated by maximizing the loglikelihood value of the model. The likelihood and loglikelihood of the state space model are

$$L(y) = p(y_1, \dots, y_n) = p(y_1) \prod_{t=2}^{n} p(y_t|Y_{t-1}), \tag{20}$$

$$log\, L(y) = -\frac{np}{2} log\, 2\pi - \frac{1}{2} \sum_{t-1}^{n} (log|F_t| + v_t' F_t^{-1} v_t). \tag{21}$$

To find the maximum value of loglikelihood, this study simulates 200 initial values for unknown variance parameters σ_ε^2 and σ_α^2 with 50 different seeds using uniform distribution, choosing the maximum value of loglikelihood for each seeds, then choosing the initial values that have the most minimum value of RMSE (root mean squared error). The value of RMSE calculated by the value of estimated incremental claim (upper triangle) and the known upper triangle of the datasets.

After finding σ_ε^2 and σ_α^2 , the recursion algorithm of Kalman filtering and smoothing can be done. Thus, the algorithm can estimate the smoothed state $\hat{\alpha}_t$. Based on the assumption, $\hat{\alpha}_t$ is normally distributed. To estimate the incremental claim $\hat{X}_{ij}$, the smoothed state $\hat{\alpha}_t$ should be transformed into lognormal distribution first, by equating the probability density function (pdf) of the state which is still normally distributed, with the lognormal distribution pdf so that the transformation value can be found.

The density function of normal distribution with parameters μ and σ^2 is given by [24]

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty, \tag{22}$$

while the density function of lognormal distribution if the random variable $Y = \ln(X)$ has a normal distribution with mean μ and standard deviation σ is given by [25]

$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x} e^{-(ln(x)-\mu)^2/2\sigma^2}, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{23}$$

The value of $\hat{\alpha}_t$ put into x in the density function of normal distribution, then the value of the normal distribution density becomes the value of the lognormal distribution density function to find the x as the transformed $\hat{\alpha}_t$. Finally, the transformed state of $\hat{\alpha}_t$ formed into the value of incremental claim by the relation of (2) so the estimate of IBNR claim reserving can be found.

$$y_t - Y_{i0} = \hat{\alpha}_t \tag{24}$$

$$log\ (\hat{X}_{ij}) - log\ (\hat{X}_{i0}) = \hat{\alpha}_t \tag{25}$$

$$log\left(\frac{\hat{X}_{ij}}{\hat{X}_{i0}}\right) = \hat{\alpha}_t \tag{26}$$

$$transformed(\hat{\alpha}_t) = \frac{\hat{X}_{ij}}{\hat{X}_{i0}} \tag{27}$$

$$\hat{X}_{ij} = transformed(\hat{\alpha}_t) \times X_{i0} \tag{28}$$

Thus, the IBNR reserve (R) calculated by

$$R = \sum\nolimits_{ij \in \mathcal{A}} \hat{X}_{ij}, \tag{29}$$

where $\mathcal{A} \equiv \{\ ij : \hat{X}_{ij}$ in the lower triangle $\}$. The estimated incremental claim and IBNR reserve will be compared to the deterministic method, namely chain ladder which formula can be found in [6,26]. The methods compared with several model validation methods, such as RMSE, MAE, RSE, and R^2. Formulas for each method can be found at [27]. The overall algorithm carried out in this study is outlined in Figure 1.

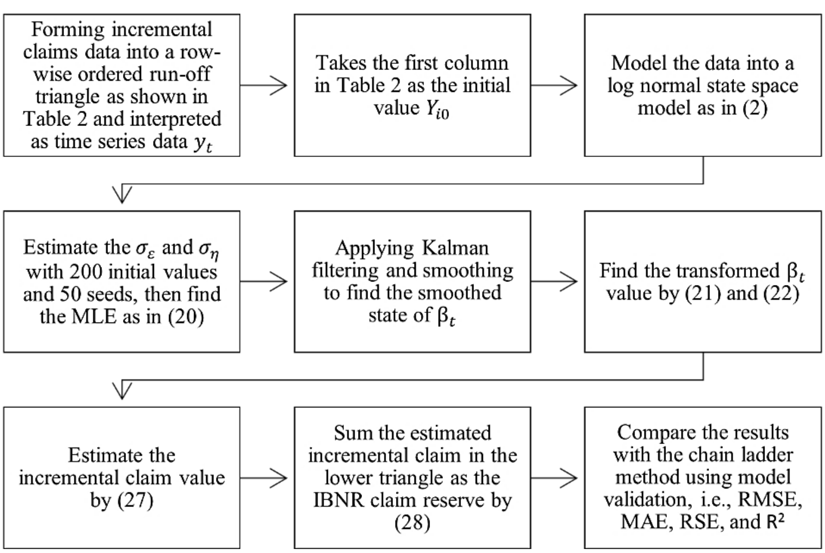

Fig. 1. IBNR claim reserving algorithm

3 Results and Discussion

The datasets used in this study are paid losses in general insurance from commercial auto lines of business of Celina Mut Group with code 353 which also used by [18] as dataset 1 and paid losses from worker's compensation of State Farm Mut Group with code 1767 that also used by [20] as dataset 2. Those datasets consist of 10 accident years and 10 development lags that can be accessed from the CAS (Casualty Actuary Society) loss reserve database. The reason behind selecting those datasets is that the data used can be modeled properly for both models because it does not contain too many negative incremental values of claims as is commonly found in CAS database. The incremental claims of dataset 1 and 2 in the upper triangle with bold value becoming the data observation of the model, while the lower triangle becomes the data validation.

Table 4. Dataset 1

i	j									
	0	1	2	3	4	5	6	7	8	9
0	**952**	**577**	**1284**	**834**	**77**	**108**	**67**	**8**	**4**	**1**
1	**849**	**715**	**638**	**230**	**36**	**19**	**26**	**13**	**5**	-4
2	**983**	**1228**	**619**	**1002**	**207**	**26**	**37**	**53**	113	6
3	**1657**	**1028**	**484**	**431**	**300**	**420**	**12**	6	3	0
4	**932**	**1008**	**686**	**706**	**36**	**123**	40	9	0	43
5	**1162**	**1240**	**397**	**197**	**38**	8	188	8	3	27
6	**1478**	**1502**	**965**	**769**	748	218	2	1	1	0
7	**1240**	**840**	**527**	473	598	-1674	2113	8	3	0
8	**1326**	**1086**	955	476	122	162	6	8	1	2
9	**1413**	1270	490	501	131	200	15	75	37	7

Table 5. Dataset 2

i	j									
	0	1	2	3	4	5	6	7	8	9
0	**22190**	**38644**	**24270**	**15047**	**8661**	**6155**	**3823**	**2768**	**1934**	**1557**
1	**26542**	**51256**	**28609**	**16015**	**10937**	**5240**	**4430**	**2683**	**1646**	1894
2	**32977**	**67517**	**34392**	**22872**	**11233**	**9074**	**4722**	**4973**	4816	3516
3	**38604**	**75824**	**42675**	**24219**	**16089**	**11393**	**4592**	8200	5160	2886
4	**42466**	**83354**	**38956**	**24269**	**15332**	**9527**	7755	6756	3345	4071
5	**46447**	**70317**	**38133**	**24522**	**14257**	12397	6120	4517	4190	3968
6	**41368**	**58976**	**31677**	**19060**	14586	7500	4793	3972	3107	5533
7	**35719**	**47497**	**28052**	15953	9376	5774	3988	2179	2938	1823
8	**28746**	**37287**	21715	12684	8010	7855	3269	2319	2492	2026
9	**25265**	36607	19166	11481	6487	4732	3072	1519	2180	1083

By taking the log of the incremental value in the run-off triangle and picking the first column as the initial value, the lognormal state space model can be applied. As σ_ε^2 and σ_α^2 construct the model, therefore, they play a role in estimating the claims. The difference in the estimate of those parameters could lead to different result of the claim estimate. Therefore, this study also use the role of minimizing RMSE in order to obtain the best model, as discussed in previous section. The results of the estimated claim from dataset 1 and 2 using the lognormal state space model and Kalman recursion presented in Table 6 and Table 7, respectively.

Table 6. Lognormal SSM estimated claim of dataset 1

i	j									
	0	1	2	3	4	5	6	7	8	9
0	952	716.2	703.3	305.9	196.7	96.0	53.2	43.0	66.4	3.6
1	849	638.7	627.2	272.8	175.4	85.6	47.4	38.4	59.2	**3.3**
2	983	739.5	726.2	315.8	203.1	99.1	54.9	44.4	**68.6**	**3.8**
3	1657	1246.6	1224.0	532.4	342.4	167.1	92.6	**74.9**	**115.6**	**6.3**
4	932	701.2	688.5	299.4	192.6	94.0	**52.1**	**42.1**	**65.0**	**3.6**
5	1162	874.2	858.4	373.3	240.1	**117.2**	**64.9**	**52.5**	**81.1**	**4.5**
6	1478	1111.9	1091.8	474.8	**305.4**	**149.1**	**82.6**	**66.8**	**103.1**	**5.7**
7	1240	932.9	916.0	**398.4**	**256.2**	**125.1**	**69.3**	**56.1**	**86.5**	**4.8**
8	1326	997.6	**979.5**	**426.0**	**274.0**	**133.7**	**74.1**	**60.0**	**92.5**	**5.1**
9	1413	**1063.0**	**1043.8**	**454.0**	**291.9**	**142.5**	**79.0**	**63.9**	**98.6**	**5.4**

Table 7. Lognormal SSM estimated claim of dataset 2

i	j									
	0	1	2	3	4	5	6	7	8	9
0	22190	39954	24096	14668	8690	5734	3724	2687	1766	1557
1	26542	50634	28465	16786	10454	6066	4257	3052	1803	**1862**
2	32977	65257	34624	21647	12079	8591	4716	4504	**2240**	**2314**
3	38604	74700	40145	23908	14914	10426	4912	**5272**	**2622**	**2709**
4	42466	78299	39341	24262	15173	10196	**5403**	**5799**	**2885**	**2980**
5	46447	73212	38992	24468	15071	**11152**	**5910**	**6343**	**3155**	**3259**
6	41368	59784	32709	20017	**13423**	**9932**	**5264**	**5650**	**2810**	**2903**
7	35719	48458	28122	**17284**	**11590**	**8576**	**4545**	**4878**	**2426**	**2506**
8	28746	37904	**22632**	**13910**	**9328**	**6902**	**3658**	**3926**	**1953**	**2017**
9	25265	**33314**	**19891**	**12225**	**8198**	**6066**	**3215**	**3450**	**1716**	**1773**

The bold value in Table 6 and Table 7 represents the lower triangle of the incremental run-off triangle, which, when summed up becomes the IBNR reserves. Figure 2 and 3 represents the plot of the run-off trangle from dataset 1 and 2 in black line and a red dashed line that represents the estimated incremental claim obtained from the lognormal state space model. The model seems adequate to model the run-off triangle since it has a similar pattern of the observed datasets. The predictive performance of the methods will be discussed next.

The normal Q-Q plot shown in Figure 4 and 5 represents that the standardized residual of the lognormal state space model meets the assumption of normality because it can be seen on the plot that the quantile points are around the line for both datasets. It is also supported by the *p*-value of Shapiro-Wilk test exceeding the alpha (α=0.05), as shown in in Figure 6 and 7, meaning that the standardized residuals are normally distributed.

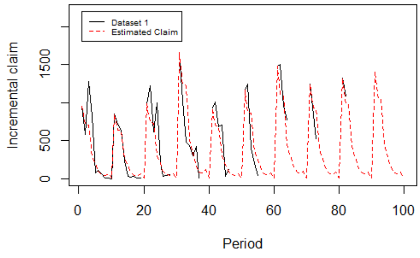

Fig. 2. Plot estimated claim dataset 1

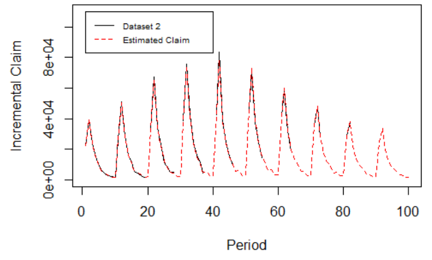

Fig. 3. Plot estimated claim dataset 2

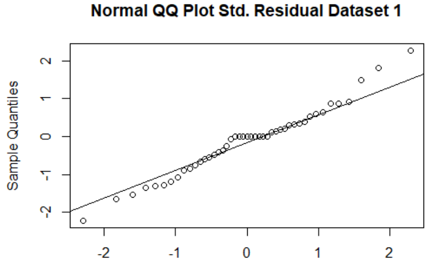

Fig. 4. Normal Q-Q plot residual dataset 1

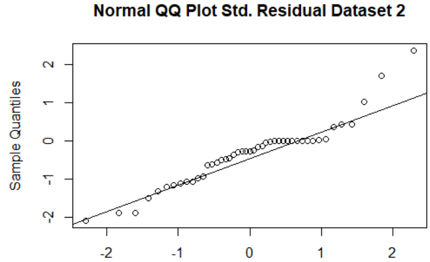

Fig. 5. Normal Q-Q plot residual dataset 2

```
        Shapiro-Wilk normality test

data:  stdres1
W = 0.97742, p-value = 0.5188
```

Fig. 6. Shapiro-Wilk test of residual dataset 1

```
        Shapiro-Wilk normality test

data:  stdres2
W = 0.97096, p-value = 0.3136
```

Fig. 7. Shapiro-Wilk test of residual dataset 2

Since the lognormal state space model seems adequate to model and estimate the run-off triangle, the result will compare to the chain ladder method. The result of the IBNR reserve estimate, along with the model validation presented below.

Table 8. Chain ladder estimate of dataset 1

i	j									
	0	1	2	3	4	5	6	7	8	9
0	952	577	1284	834	77	108	67	8	4	1
1	849	715	638	230	36	19	26	13	5	**0.6**
2	983	1228	619	1002	207	26	37	53	**5.8**	**1.1**

3	1657	1028	484	431	300	420	12	**30.5**	**6.1**	**1.1**
4	932	1008	686	706	36	123	**33.7**	**24.8**	**5.0**	**0.9**
5	1162	1240	397	197	38	**120.7**	**30.5**	**22.4**	**4.5**	**0.8**
6	1478	1502	965	769	**164.9**	**194.1**	**49.0**	**36.0**	**7.2**	**1.3**
7	1240	840	527	**533.2**	**109.8**	**129.3**	**32.6**	**24.0**	**4.8**	**0.9**
8	1326	1086	**776.7**	**652.2**	**134.4**	**158.1**	**39.9**	**29.4**	**5.9**	**1.1**
9	1413	**1232.0**	**851.7**	**715.2**	**147.3**	**173.4**	**43.8**	**32.2**	**6.4**	**1.2**

The sum of the lower triangle, i.e., the IBNR claim reserve estimate of the chain ladder method conceived by the sum of the bold value in the run-off triangle on Table 8 and 9. The comparison of the lognormal state space model and chain ladder is presented below.

Table 9. Chain ladder estimate of dataset 2

i	j									
	0	1	2	3	4	5	6	7	8	9
0	22190	38644	24270	15047	8661	6155	3823	2768	1934	1557
1	26542	51256	28609	16015	10937	5240	4430	2683	1646	**1858**
2	32977	67517	34392	22872	11233	9074	4722	4973	**2515**	**2399**
3	38604	75824	42675	24219	16089	11393	4592	**5003**	**2925**	**2790**
4	42466	83354	38956	24269	15332	9527	**5867**	**5153**	**3013**	**2874**
5	46447	70317	38133	24522	14257	**9860**	**5583**	**4903**	**2867**	**2735**
6	41368	58976	31677	19060	**12428**	**8325**	**4713**	**4139**	**2420**	**2309**
7	35719	47497	28052	**17371**	**10582**	**7088**	**4013**	**3524**	**2061**	**1966**
8	28746	37287	**22592**	**13836**	**8428**	**5646**	**3197**	**2807**	**1641**	**1566**
9	25265	**42555**	**23204**	**14211**	**8656**	**5798**	**3283**	**2883**	**1686**	**1608**

From Table 10, the IBNR reserve estimates of the lognormal state space model of dataset 1 is about 7,751.255171 (in dollar USD) and from the chain ladder method is about 6,576.437291 (in dollar USD), while the actual losses paid by the company is 7,399 (in dollar USD). Afterward, from Table 11, the IBNR reserve estimates of the lognormal state space model of dataset 2 is about 309,865.542 (in dollar USD) and from the chain ladder method is about 304,881.907 (in dollar USD), while the actual losses paid by the company is 377,893 (in dollar USD). Numerically, the difference between the IBNR estimate of lognormal state space model and the actual losses paid is smaller than the chain ladder method for both datasets. It means the IBNR reserve estimate of the lognormal state space model is more precise and more sufficient to compensate the loss even though the IBNR estimate of dataset 2 is underestimated compared to the paid loss.

Table 10. IBNR reserve comparison of dataset 1

Dataset 1			
Accident year	IBNR reserve		Paid Losses
	Lognormal SSM	Chain ladder	
0	0	0	0

1	3.252737	0.6471491	-4
2	72.340525	6.8768695	119
3	196.866232	37.7099125	9
4	162.807981	64.3959225	92
5	320.17002	178.8653526	234
6	712.616037	452.5324012	970
7	996.246848	834.6426125	1521
8	2044.871036	1797.546817	1732
9	3242.083755	3203.220255	2726
Total	**7751.255171**	**6576.437291**	**7399**

Table 11. IBNR reserve comparison of dataset 2

Dataset 2			
Accident year	IBNR reserve		Paid Losses
	Lognormal SSM	Chain ladder	
0	0	0	0
1	1862.366	1857.905	1894
2	4554.089	4913.995	8332
3	10603.237	10719.044	16246
4	17067.385	16906.626	21927
5	29819.145	25947.646	31192
6	39981.507	34333.502	39491
7	51805.509	46604.925	42031
8	64323.807	59713.368	130453
9	89848.497	103884.896	86327
Total	**309865.542**	**304881.907**	**377893**

By validating the model, the value of RMSE, MAE, RSE, and R^2 from datasets 1 and 2 are shown in Table 12 and 13, respectively. To conclude which model is better in this study, it will be chosen which model between the lognormal state space model and the chain ladder has the smallest RMSE, MAE, and RSE values, yet the largest value of R^2, which means having a more accurate model in estimating claims. From dataset 1, the RMSE and RSE values of lognormal state space model are smaller than the chain ladder. But for the MAE value, the chain ladder method has a smaller value than the lognormal state space model. Besides that, the value of R^2 of the lognormal state space model is larger than the chain ladder.

Table 12. Model validation of dataset 1

Dataset 1		
Model validation	IBNR reserve	
	Lognormal SSM	Chain ladder
RMSE	**428.4555**	433.7416
MAE	168.0882	**152.6842**
RSE	**0.7764093**	0.7956854
R^2	**0.2235907**	0.2043146

Table 13. Model validation of dataset 2

Dataset 2		
Model validation	IBNR reserve	
	Lognormal SSM	Chain ladder
RMSE	**1539.958**	1759.806
MAE	**1262.15**	1301.384
RSE	**0.05686552**	0.07426105
R^2	**0.9431345**	0.9257389

Based on dataset 2, the value of RMSE, MAE, and RSE of lognormal state space model are smaller than the chain ladder, then the value of R^2 of the lognormal state space model is larger than chain ladder. So, from both datasets, the lognormal state space model outperforms the chain ladder method. Previous research has estimated IBNR claim reserves using various type of state space model, including the lognormal state space model. This study makes a comparison of IBNR claim reserve estimates using the lognormal state space model and the chain ladder method by looking at the prediction performance in the lower triangle using model validation that was already mentioned. Therefore, by using the datasets chosen in this study, it can be concluded that the lognormal state space model outperforms the chain ladder method.

4 Conclusion

Based on the results of the previous section, we know the performance of the lognormal state space model along with the row-wise ordered run-off triangle by chosen model validation compared to chain ladder. The performance of the lognormal state space model seems adequate to model the run-off triangle and outperform the chain ladder method by means of model validation using RMSE, MAE, RSE, and R^2 using the datasets used in this study. Those are evidenced by model validation with smaller RMSE, RSE, and R^2 values for the first dataset and smaller RMSE, MAE, RSE, and R^2 values for the second dataset . Then, the difference of IBNR estimate between the lognormal state space model and the actual losses paid is smaller than the chain ladder method. It makes the IBNR reserve estimate of the lognormal state space model more precise and more sufficient to compensate the actual losses. For future research, the observed data of lognormal state space model can be obtained from the column-wise ordered run-off triangle that was developed by [8,9], or one can compare the other stochastic and deterministic methods to the lognormal state space model and the other state space model developed by [16].

The authors express their sincere gratitude to Radek Hendrych for his invaluable contributions in shaping the study's method and to Adrian Pizzinga for expertly guiding the model's intricacies. The enhanced quality of this work is attributed to their constructive input. The author also acknowledges the supervisor's instrumental role in providing both opportunity and unwavering support throughout the paper's development.

References

1. O. M. Alhawtmeh, Jordanian Insurance Companies-Field Study, **15**, 11 (2023)
2. P. Longoni, SSRN Electronic Journal (2020)
3. W. Yousuf et al., British Actuarial Journal, **26**, E2 (2020)

4. R. Hartwig, Fundamentals of Risk and Insurance (Insurance Information Institute, 2023)
5. A. Carrato, G. Mcguire, P. Fiaa, and R. Scarth, A Practitioner's Introduction to Stochastic Reserving (2016)
6. M. V. Wuthrich and M. Merz, Stochastic Claims Reserving Methods in Insurance (John Wiley & Sons Ltd, 2012)
7. R. Atherino, A. Pizzinga, and C. Fernandes, Astin Bulletin, **40**, 2 (2010)
8. P. de Jong and B. Zehnwirth, J Inst Actuar, **110**, 01 (1983)
9. R. J. Verral, J Inst Actuar, **116**, 3 (1989)
10. T. S. Wright, J Inst Actuar, **117**, 3 (1990)
11. T. Alpuim and I. Ribeiro, Appl Stoch Models Bus Ind, **19**, 2 (2003)
12. R. J. Verral, ASTIN Bulettin, **24**, 2 (1994)
13. Z. Ben, Insur Math Econ, **20**, 2 (1997)
14. L. Pang and S. He, *The application of state-space model in outstanding claims reserve,* in Proceeding of 2012 International Conference on Information Management, Innovation Management, and Industrial Engineering, ICIII 2012 (2012)
15. N. Chukhrova and A. Johannssen, Risks, **5**, 2 (2017)
16. R. Hendrych and T. Cipra, ASTIN Bulletin, **51**, 1 (2021)
17. F. F. Adam, KnE Social Sciences, **3**, 11 (2018)
18. G. Meyers, Casualty Actuarial Society, **1** (2013)
19. L. Martinek, Risks, **7**, 2 (2019)
20. L. Tee, M. Käärik, and R. Viin, Risks, **5**, 1 (2017)
21. E. Kremer, Scand Actuar J, **1**, 47-55 (1982)
22. J. Durbin and S. J. Koopman, Oxford Statistical Science Series (2012)
23. J. Durbin, S. J. Koopman, and E. Bash, Oxford University Press, **20**, 1 (2001)
24. S. M. Ross, Elservier Inc (2019)
25. J. F. Scott, R. E. Walpole, and R. H. Myers, The Mathematical Gazette, **57**, 400 (1973)
26. T. Mack, ASTIN Bulletin, **23**, 2 (1993)
27. D. Chicco, M. J. Warrens, and G. Jurman, PeerJ Comput Sci, 7 (2021)