Advances in Economics and Business 12(1): 1-7, 2024
DOI: 10.13189/aeb.2024.120101 http://www.hrpub.org

Estimation of Individual Claims Reserves Using K Modes Cluster and Reserving by Detailed Conditioning

Yuciana Wilandari[1,2], **Gunardi**[1,*], **Adhitya Ronnie Effendie**[1]

[1]Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Indonesia

[2]Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro, Indonesia

Received September 21, 2023; Revised January 15, 2024; Accepted February 17, 2024

Cite This Paper in the following Citation Styles

***(a)**: [1] Yuciana Wilandari, Gunardi, Adhitya Ronnie Effendie, "OEstimation of Individual Claims Reserves Using K Modes Cluster and Reserving by Detailed Conditioning," Advances in Economics and Business, Vol.12, No.1, pp. 1-7, 2024. DOI: 10.13189/aeb.2024.120101*

***(b)**: Yuciana Wilandari, Gunardi, Adhitya Ronnie Effendie (2024). Estimation of Individual Claims Reserves Using K Modes Cluster and Reserving by Detailed Conditioning. Advances in Economics and Business, 12(1), 1-7. DOI: 10.13189/aeb.2024.120101*

Abstract The estimation of claims reserves is of great importance to insurance companies. In liability insurance, payments for claims often take a long time to be made after the incident. To complete these payments, claims reserves are necessary. The individual claims reserve estimates provided by the individual run-off triangle provide information related to each policyholder's claims data, such as the ID claim, the occurrence period, the reporting period, the period of claims closed, and the payment data. To calculate individual claims reserve estimates, the Reserving by Detailed Conditioning (RDC) method is used. However, this method does not take into account the policyholder's personal information, such as age and gender. Therefore, in this article, the Cluster method is used. The aim of this research is to calculate individual claims reserve estimates using the RDC method and the K-Mode Clustering method. The K-mode clustering is used because the claims data used is health insurance data, which is in the form of categories. The data information from patients on health insurance includes gender, clinic, severity level, number of days of treatment, treatment class, type of return, and type of patient. The data was first grouped into 4 clusters, and then each cluster calculated its claims reserve estimate using the RDC method. The claims reserve estimates obtained were compared with individual claims reserve estimates without grouping using Mean Square Error Prediction (MSEP). The results of this article show that the estimated individual claims reserves using grouping have MSEP values that are smaller than the estimated individual claims reserves without grouping.

Keywords Individual Claims Reserve, Reserve by Detailed Conditioning, K Modes Cluster

1 Introduction

Humans are exposed to a variety of risks in their daily lives that can have a destabilizing effect on the economy. One way to address and manage these risks is through insurance. Insurance is divided into two categories: life insurance and non-life insurance. In some cases of non-life insurance, payments are made more than once and can take a long time to be settled, sometimes more than a year, and may not even be resolved at the time of evaluation [1]. This extended period between the time the claim is reported and the payment is made leads to the emergence of outstanding claims liability for the insurance company. To counter this problem, companies can set aside funds specifically for paying claims, known as claims reserves. The amount of reserves for claims is usually based on an estimate.

Generally, the calculation of estimated claims reserves is based on the aggregate data presented in the runoff triangle. Mack [2] introduced the Chain Ladder method for estimating reserves for collective claims. Godecharle and Antonio [3] noted that when aggregated data is used, much important claim information is omitted when estimating claims reserves, such as the length of claims for each claim that occurs, which is calculated from claims reported until claims are paid. This has led to the development of claims reserves estimation methods based on individual data. The data to calculate individual claim reserve estimates are presented in individual run-off triangles. Various studies on claims reserves based on individual data have been proposed by Godecharle and Antonio [3], Drienkens [4], Rosenlund [5], and Kroon [6]. Rosenlund [5] proposed the Detailed Conditioning (RDC) Reserving method for estimating claims reserves using an individual approach.

The Reserve Development Method (RDC) requires the in-

clusion of information about each policyholder's claim, such as the period of occurrence, the delay period, the settlement period, and the payment amount, in order to estimate the claim reserves. Rosenlund [5] identifies three key characteristics of the claim that must be taken into account when calculating the reserves: the duration of the claim, the delay period, and the group number of quantile intervals of cumulative payments from each individual claim. The following steps are then taken to calculate the reserves: determining the characteristics of the claim, calculating the estimated probability of the claim's length, calculating the estimated average payment, and calculating the estimated reserves for the claim [5].

In the RDC approach, some of the available claim data is taken into account when calculating claims reserves, while other information is not required. Each non-life insurance claim contains details such as the policyholder's characteristics (age or gender if the policyholder is an individual, type of business if it is a company, etc.), the characteristics of the insured object (age or car model, type of building, etc.), and the characteristics of the geographical area (per capita income or population density of the policyholder's residential area, etc.).

When calculating premiums, claim information is a key factor in the assessment process and is referred to as a background variable [7-10]. The addition of claim information to the RDC method can lead to an unstable estimate due to the large number of calculation combinations. Yulita and Effendie used RDC and Gamma GLM to find estimates of claims reserves and compared the results with the Chain Ladder, RDC and RDC and Poisson GLM methods using MSEP. The results showed that the RDC and Gamma GLM method had the smallest values of MSEP [10]. Wu trich, Lopez and Milhaud used machine learning to estimate individual claims reserves [11-12]. This article adds claim information to the RDC method by first grouping this claim information. The data used in this article are health insurance data, with claim information from policyholders in the form of category data, so the K-mode cluster method was used [13].

2 Materials and Methods

In this article, we will explore the estimation of individual claims reserves through the Reserving by Detailed Conditioning (RDC) approach, with data initially grouped using K mode clustering. This section will focus on the estimation of claims reserves using the RDC method and the K mode clustering technique.

2.1 Claim Reserving

In liability insurance, claims are not usually settled right away. This is due to a variety of factors, such as a lag in reporting the claim, a time gap between the incident and the report, or a delay in resolving the claim. In some cases, a claim that has been closed may need to be reopened and further payments made. This results in a debt to the insurer, which must be prepared to cover the cost of settling the claim. These funds are known as claims reserves.

2.2 Reserving by Detailed Conditioning

Rosenlund [5] and Kroon [6] both suggest that the Detailed Conditioning (RDC) reserve method is a way to calculate individual claims reserves that take into account specific claim characteristics.

2.2.1 Claim Information

Suppose that there are N claims, with each k ($k = 1, 2, ..., N$) claim having the following claim information,

1. The period of occurrence of the claim k, denoted by $i(k)$, $i(k) \in \{1, 2, ..., n\}$.
2. The period of delay in reporting claims k, indicated by $W(k)$, $W(k) \in \{1, 2, ..., n\}$. If the claim is reported in the same period as the period in which it occurred, the value of $W(k) = 1$.
3. The period of settlement of the claim k, denoted by $S(k)$, $S(k) \in \{1, 2, ..., n\}$. If the claim is settled in the same period as the period in which it occurred, the value of $S(k) = 1$.
4. Payments for claims k in the development period j, denoted by $Y(k, j)$, $j \in \{1, 2, ..., S(k)\}$.

Using the notation above, the information for each claim k ($k = 1, 2, ..., n$) can be expressed as,

$$i(k), W(k), S(k), Y(k, 1), ..., Y(k, S(k)).$$

2.2.2 Claim Characteristics

For each claim k ($k = 1, 2, ..., N$) three characteristics of claims are defined as conditions in the estimation of claims reserves, claim duration k, delay period in reporting claims k and quantile interval of the group number of cumulative claims payments k.

1. Claim length
   Claim length is the period of time from the claim being reported until the claim is settled. The length of the claim k ($k = 1, 2, ..., N$) is denoted by $L(k)$ which can be calculated by:

   $$L(k) = S(k) - W(k) + 1$$

   From the values of $L(k)$ and $W(k)$, two claim statuses can be obtained as follows,

   (a) If $W(k) \leq n - i(k) + 1$, then the claim status is reported
   (b) If $L(k) \leq n - i(k) - W(k) + 2$, then the claim status is settled.
2. Claim Reporting Delay Period
   The delay period in reporting claims k ($k = 1, 2, ..., N$) that have occurred is indicated by $W(k)$. For example, set the value w_0, with $w_0 \in \{1, 2, ..., n\}$. This value w_0 is used as the maximum limit of $W(k)$, $k = 1, 2, ..., N$.

Thus, it can be defined that a claim is said to be reported late to the insurance company if $W(k) \geq w_0$. The value used as a condition for the estimation of the claims reserve is $min(W(k), w_0)$.

In general, the estimation of claims reserves using the RDC method will go through several calculation stages, namely as follows [5],

1. Determine the characteristics of the claim.
2. Calculate the estimated probability of the length of the claim.
3. Calculate the estimated average claim payment.
4. Calculate the estimated reserve of claims.

2.2.3 Estimation of Claim Length Probability

We give $0 \leq t \leq n-1$ and $t+1 \leq \lambda \leq n$, the probability of claim length can be described by,

$$p_\lambda(t,q,w) = P(L = \lambda | L > t, Q_t = q, min(W, w_0) = w) \quad (1)$$

for each $q \in \{1,2,...,q_0\}$ and $w \in \{1,2,...,w_0\}$ with $0 \leq t \leq n-1$ and $t+1 \leq \lambda \leq n$ [4].

Estimation of the probability of length of the claim can be obtained by estimating the probability that a claim can be settled in a certain period, knowing that the claim has not been resolved in the previous period. Given $0 \leq t \leq n-1$ and $t+1 \leq \lambda \leq n$, the probability that a claim can be settled in a certain period, knowing that the claim has not been resolved in the previous period, according to Roselund [5], we can be expressed in the equation,

$$r_\lambda(t,q,w) = P(L = \lambda | L \geq t, Q_t = q, min(W, w_0) = w) \quad (2)$$

for each $q \in \{1,2,...,q_0\}$ and $w \in \{1,2,...,w_0\}$.

The relationship with $p_\lambda(t,q,w)$ and $r_\lambda(t,q,w)$ is,

$$p_\lambda(t,q,w) = r_\lambda(t,q,w) \prod_{m=t+1}^{\lambda-1} (1 - r_m(t,q,w)) \quad (3)$$

Estimation of $r_\lambda(t,q,w)$ can be obtained through observation of claim data, namely setting$I_\lambda^F(t,q,w)$ is the number of claims that have been settled knowing $L = \lambda$, $Q_t = q$, $min(W, w_0) = w$ and$J_\lambda(t,q,w)$ is the known number of reported claims $L \geq \lambda$, $Q_t = q$, $min(W, w_0) = w$.

The results of the observation data are used to estimate $r_\lambda(t,q,w)$, so $\hat{r}_\lambda(t,q,w)$, we defined

$$\begin{cases} \hat{r}_n(t,q,w) = 1 \\ \hat{r}_\lambda(t,q,w) = \frac{I_\lambda^F(t,q,w)}{J_\lambda(t,q,w)}, \quad \text{if } \lambda < n. \end{cases} \quad (4)$$

and $\hat{p}_\lambda(t,q,w)$, we defined

$$\hat{p}_\lambda(t,q,w) = \hat{r}_\lambda(t,q,w) \prod_{m=t+1}^{\lambda-1} (1 - \hat{r}_m(t,q,w)) \quad (5)$$

2.2.4 Estimation of Claim Payment Mean

The estimated average claim payment is obtained through a combination of calculations between settled and unsettled claim payments. According to Roselund [5], the estimated average claim payment denoted by $\hat{\mu}_{\lambda h}(t,q,w)$ is obtained through a combination of calculations between settled and unsettled claim payments, by making several observations as follows,

1. For claims that have been settled, observation of claim data is carried out in the period $h \leq n-i-W+2$. Sum of $Y(h+W-1)$ is the total amount of claim payments that have been settled in the entire event period $i \in \{1,2,...,n\}$.
   Given $0 \leq t \leq n-1$, $t+1 \leq \lambda \leq n$, and $t+1 \leq h \leq \lambda$, t we can be stated that $I_{\lambda h}^F(t,q,w)$ is the number of claims that have been settled knowing $L \leq n-i-W+2$, $L = \lambda$, $Q_t = q$, $min(W, w_0) = w$ and $Y_{\lambda h}^F(t,q,w) = \sum_{h=t+1}^{\lambda} Y(h+W-1)$, with knowing $L \leq n-i-W+2$, $L = \lambda$, $Q_t = q$, $min(W, w_0) = w$, for each $q \in \{1,2,...,q_0\}$ and $w \in \{1,2,...,w_0\}$.

2. For unresolved claims, observation of claim data is carried out in the period $t \leq n-2$. We given $0 \leq t \leq n-2$, $t+1 \leq r \leq n-1$, and $t+1 \leq h \leq r$, it can be stated that $I_r^o(t,q,w)$ is the number of known claims $n-i-W+2 = r$, $L > r$, $Q_t = q$, $min(W, w_0) = w$ and $Y_{rh}^o(t,q,w) = \sum_{h=t+1}^{r} Y(h+W-1)$, with knowing that $n-i-W+2 = r$, $L > r$, $Q_t = q$, $min(W, w_0) = w$, for each $q \in \{1,2,...,q_0\}$ and $w \in \{1,2,...,w_0\}$.

3. For claims that have not been settled in the period r can be settled in the period $L = \lambda$. Given $\lambda = r+1,...,n$, the number of outstanding claims at time r can be settled at time $L = \lambda$ define,

$$I_{r\lambda}^o(t,q,w) = \frac{\hat{p}_\lambda(t,q,w)}{\hat{p}_{r+1}(t,q,w) + ... + \hat{p}_n(t,q,w)} I_r^o(t,q,w) \quad (6)$$

for each $q \in \{1,2,...,q_0\}$ and $w \in \{1,2,...,w_0\}$.
The recursive calculation from the period $r = \lambda$ to $r = h$ is carried out to produce the value $\hat{\mu}_{\lambda h}(t,q,w)$. To start a recursive calculation, an initial value is required. We given $0 \leq t \leq n-1$, $t+1 \leq \lambda \leq n$, and $t+1 \leq h \leq \lambda$, the initial value of the recursive calculation is determined by the claims that have been settled, that is,

$$\begin{cases} I_\lambda^{(\lambda)}(t,q,w) = I_\lambda^F(t,q,w) \\ Y_{\lambda h}^{(\lambda)}(t,q,w) = Y_{\lambda h}^F(t,q,w). \end{cases} \quad (7)$$

and recursive equation can describe,

$$\begin{cases} I_\lambda^{(r)}(t,q,w) = I_\lambda^{(r+1)}(t,q,w) + I_{r\lambda}^o(t,q,w) \\ Y_{\lambda h}^{(r)}(t,q,w) = Y_{\lambda h}^{(r+1)}(t,q,w) + Y_{r\lambda h}^o(t,q,w). \end{cases} \quad (8)$$

for $r = \lambda-1, \lambda-2;, ..., h$. Value $Y_{r\lambda h}^o(t,q,w)$ can get

$$Y_{r\lambda h}^o(t,q,w) = \beta_{rh}(t,q,w) \frac{Y_{\lambda h}^{r+1}(t,q,w)}{I_\lambda^{r+1}(t,q,w)} I_{r\lambda}^o(t,q,w). \quad (9)$$

with

$$\beta_{rh}(t,q,w) = Y^{o}_{rh}(t,q,w)\left(\sum_{v=r+1}^{n} Y^{r+1}_{vh}(t,q,w)\frac{I^{o}_{rv}(t,q,w)}{I^{r+1}_{v}(t,q,w)}\right)^{-1}. \quad (10)$$

From (10), it is possible that the value $\beta_{rh}(t,q,w)$ is undefined if $\left(\sum_{v=r+1}^{n} Y^{r+1}_{vh}(t,q,w)\frac{I^{o}_{rv}(t,q,w)}{I^{r+1}_{v}(t,q,w)}\right)^{-1} = 0$. If this happens, it will result in $Y^{o}_{r\lambda h}(t,q,w)$ undefined. To overcome this, another alternative is used to find the value $Y^{o}_{r\lambda h}(t,q,w)$, with

$$Y^{o}_{r\lambda h}(t,q,w) = \frac{\hat{p}_{\lambda}(t,q,w)}{\hat{p}_{r+1}(t,q,w) + ... + \hat{p}_{n}(t,q,w)} Y^{o}_{rh}(t,q,w) \quad (11)$$

After the recursive calculation is completed up to period h, an estimate of the average claim payment can be obtained which is expressed in,

$$\hat{\mu}_{\lambda h}(t,q,w) = \frac{Y^{(h)}_{\lambda h}(t,q,w)}{I^{(h)}_{\lambda}(t,q,w)} \quad (12)$$

with $h = t+1, ..., n$ dan $\lambda = h, ..., n$.

2.2.5 Estimation of Claim Reserves

Given $0 \leq t \leq n-1$, $t+1 \leq \lambda \leq n$, and $t+1 \leq h \leq \lambda$, the estimated reserve per claim can be expressed as,

$$\hat{R}(t,q,w) = \sum_{\lambda=t+1}^{n}\sum_{h=t+1}^{\lambda} \hat{p}_{\lambda}(t,q,w)\hat{\mu}_{\lambda,h}(t,q,w) \quad (13)$$

for each $q \in \{1,2,...,q_0\}$ and $w \in \{1,2,...,w_0\}$.

2.3 K Modes Clustering

According to Hair et al. [14], cluster analysis is a collection of several multivariate data processing techniques which have the main objective of grouping objects based on their characteristics. The results of the clusters formed must show high internal homogeneity within one cluster and high heterogeneity between clusters.

The k-Modes clustering algorithm was first introduced by Huang in 1997 [15]. This algorithm is a development of the k-Means clustering algorithm for grouping categorical data. The standard k-Means clustering algorithm cannot be applied to categorical data. This is due to the Euclidean distance function and the use of averages to represent cluster centers. The steps in the k-Modes clustering algorithm are as follows:

1. Choose k initial modes as center points, one for each cluster.
2. Calculate the distance of each data to all cluster center points. Allocate each object to the closest cluster using a simple dissimilarity measure.

$$d(X,Y) = \sum_{i=1}^{n} \delta(x_i, y_i) \quad (14)$$

with,

$$\delta(x_i, y_i) = \begin{cases} 0, x_i = y_i \\ 1, x_i \neq y_i \end{cases} \quad (15)$$

3. After all objects are allocated to the cluster, the differences between objects are retested against mode. If the object is closer to another cluster than the current cluster, then reallocate the object to that cluster and update the mode of both clusters.
4. Repeat step 3 until there are no objects that change clusters after one full iteration of all data.

2.4 Case Study

The data used is liability insurance data from health insurance company (BPJS Health) in Indonesia. The research data were taken from the period January 2014 to December 2014 with research time units in months. The claim information that will be used in data analysis is the identity of the claim (claim ID), the month the loss experienced by the policy holder occurred (occurrence period), the month the loss was reported to the insurance company (reporting period), the month the insurance claim was settled (settling period) and the amount paid for each claim (in IDR). As for other information from the data, the variables used are information from each policy holder, i.e., gender, clinic, severity, number of days of care, care class, return type, and patient type. The number of liability insurance claims taken as samples is 664 claims.

In the first step, we group the data according to gender, clinic, severity, number of days of care, care class, return type, patient type, using the cluster k modes method. The best clusters obtained are four clusters, with details of each cluster member presented in Table 1.

Next, we calculate the number of delays in each cluster. This information is presented in Table 2.

After grouping, the estimated reserve claims for each cluster will be calculated using the RDC method. Before beginning the calculation of reserve estimation for liability insurance claims, it is necessary to determine the values of w_0 and q_0. According to Rosenlund [5], if the majority of claims are reported at the same time as the claim occurrence period, then the value of $w_0 = 1$ is set, but if most claims are reported one period after the claim occurrence period, then the value of $w_0 = 2$ is set, and so on. Table 2 shows that the majority of claims reported to insurance companies in each cluster 1, cluster 2, cluster 3 and cluster 4 experienced a delay of one to two periods after the claim incident period. The percentage of claims that have been reported up to $W = 2$ is more than 50%. Based on this percentage, it has been decided that the maximum limit for a claim reported to an insurance company is a reporting delay of two periods after the claim event. This is because the percentage of claims that have been reported up to $W = 3$ is more than 50%, meaning that up to the W period the majority of claims (more than half of the total claims) have been reported to the company. Therefore, the w_0 value set for liability insurance claim data is $w_0 = 3$. For q_0, it has been determined that cumulative claim payments are grouped into three categories, namely small, medium and large, so $q_0 = 3$ is set. The following shows the result of reserve estimation of liability insurance claims with $w_0 = 3$ and $q_0 = 3$ for occurrence period $i \in \{2,3,,12\}$ in IDR. The estimation of claims

Table 1. Cluster Member Based Variable

Cluster	Gender	Clinic	Severity	Days of care	class of care	Return type	Patient type	Cluster member
1	1	12	0	1	3	1	21	411
2	2	12	0	3	2	1	3	72
3	1	12	0	4	2	1	13	45
4	2	6	0	1	3	1	21	136

Table 2. Number of Claim from Reporting Delay of Liability Insurance Claim

W	1	2	3	4	5	6	7	8	9	10	11	12
Cluster 1	14	242	92	12	12	3	2	7	6	9	9	3
	3.41%	58.88%	22.38%	2.92%	2.92%	0.73%	0.49%	1.70%	1.46%	2.19%	2.19%	0.73%
Cluster 2	5	32	24	5	2	1	1	0	0	2	0	0
	6.94%	44.44%	33.33%	6.94%	2.78%	1.39%	1.39%	0.00%	0.00%	2.78%	0.00%	0.00%
Cluster 3	3	27	13	1	0	0	1	0	0	0	0	0
	6.67%	60.00%	28.89%	2.22%	0.00%	0.00%	2.22%	0.00%	0.00%	0.00%	0.00%	0.00%
Cluster 4	5	82	35	4	4	1	3	0	0	2	0	0
	3.67%	60.29%	25.74%	2.94%	2.94%	0.74%	2.21%	0.00%	0.00%	0.00%	1.47%	0.00%

Table 3. Claims Reserves Estimation Per Accident Period for Cluster 1

Accident Period	Development Period											
	1	2	3	4	5	6	7	8	9	10	11	12
201401												
201402												7853339
201403											12876932	10492315
201404										6840005	12240008	9973340
201405									3538317	7533192	13480449	10984069
201406								1665059	1891597	4027272	7206697	5872124
201407							587229	2325058	2641393	5623610	10063302	8199728
201408						1087385	842369	3335256	3789032	8066972	14435634	11762368
201409					2416412	724924	561579	2223504	2526021	5377981	9623756	7841579
201410				4209234	4676926	1403078	1086928	4303556	4889074	10408996	18626624	15177249
201411			31058000	4835116	5372352	1611706	1248547	4943464	5616044	11956739	21396269	17433997
201412		0	0	0	0	0	0	0	0	0	0	0

Table 4. Claims Reserves Estimation Per Accident Period for Cluster 2

Accident Period	Development Period											
	1	2	3	4	5	6	7	8	9	10	11	12
201401												
201402												0
201403											0	0
201404										666477	0	0
201405									0	2332669	0	0
201406								0	0	3665622	0	0
201407							378680	0	0	1363248	0	0
201408						443136	483421	0	0	1740317	0	0
201409					499858	276517	301655	0	0	1085958	0	0
201410				1301712	551926	305321	333077	0	0	1199078	0	0
201411			18513244	4097984	1737545	961195	1048577	0	0	3774876	0	0
201412		0	0	0	0	0	0	0	0	0	0	0

reserve for cluster 1, cluster 2, cluster 3 and cluster 4 are presented in Table 3, Table 4, Table 5 and Table 6 respectively. These calculations were carried out using R software.

The results of the estimated claim reserves for each cluster are presented in Table 7. It can be seen that the total estimated claim reserves for cluster 1 is 390,783,703 IDR, for cluster 2 is 47,062,092 IDR, for cluster 3 is 11,062,080 IDR and for cluster 4 is 5,159,891 IDR. We compared the results of estimating individual claim reserves when they were grouped into 4 clusters with the results of estimating individual claim reserves without grouping. The Mean Square Error Prediction (MSEP) was then calculated. The results are also presented in Table 7. The total estimated claim reserves when they were grouped using K mode into 4 clusters was 454,067,766 IDR. This means that the insurance company must provide funds amounting to 454,067,766 IDR in the coming year, 2015, to pay claims that occurred in 2014. The total estimated claim reserves without grouping was 546,064,535 IDR. This means that by using the RDC method, the insurance company must provide funds amounting to 454,067,766 IDR in the coming year to pay

Table 5. Claims Reserves Estimation Per Accident Period for Cluster 3

Accident Period	Development Period											
	1	2	3	4	5	6	7	8	9	10	11	12
201401												
201402												0
201403											0	0
201404										0	0	0
201405									0	0	0	0
201406								0	0	0	0	0
201407							709481	0	0	0	0	0
201408						0	709481	0	0	0	0	0
201409					0	0	709481	0	0	0	0	0
201410				581996	0	0	914565	0	0	0	0	0
201411			6052758	538346	0	0	845973	0	0	0	0	0
201412		0	0	0	0	0	0	0	0	0	0	0

Table 6. Claims Reserves Estimation Per Accident Period for Cluster 4

Accident Period	Development Period											
	1	2	3	4	5	6	7	8	9	10	11	12
201401												
201402												0
201403											191591	0
201404										0	229909	0
201405									0	0	134113	0
201406								0	0	0	153272	0
201407							99480	0	0	0	202644	0
201408						53708	192110	0	0	0	391335	0
201409					145206	47608	170291	0	0	0	346890	0
201410				141080	170781	55994	200286	0	0	0	407990	0
201411			935016	128717	155815	51087	182733	0	0	0	372235	0
201412		0	0	0	0	0	0	0	0	0	0	0

Table 7. Total of Claims Reserves Estimation and MSEP

	Claim Reserves Estimation using Cluster	Claim Reserves Estimation without Cluster
Cluster	Claim Reserves Estimation	
1	390783703	
2	47062092	
3	11062080	
4	5159891	
Total	454067766	546064535
MSEP	56819.164	57563.117

claims that occurred in 2014. The estimated MSEP of claim reserves with grouping and without grouping was 56,819,164 and 57,563,117 respectively. It can be seen that the total estimated claim reserves with grouping are smaller than without grouping, and the MSEP value with grouping is also smaller than without grouping.

3 Discussion

This study has demonstrated that the Mean Square Error Prediction (MSEP) value is smaller when individual claim reserves are estimated using the Reserving by Detailed Conditioning (RDC) and K modes cluster methods, which involve grouping data based on policyholder claim information, than when the RDC method is used without grouping. Kartikasari et al., Effendie and Pebriawan, and Yulita and Effendie have all previously conducted research that showed the MSEP value was smaller when individual claims reserves were estimated using the RDC and GLM methods than when aggregate claims reserves were estimated using the Chain Ladder method [8-10]. Additionally, Wütrich's research using Machine Learning, specifically Regression Trees, predicted a smaller total number of claim payments than the Chain Ladder method [11]. Therefore, it is recommended that insurance companies use multiple individual claim reserve estimation methods to calculate their claim reserves.

4 Conclusions

The Reserving by Detailed Conditioning (RDC) method combined with K modes cluster can be used to estimate individual claims reserves. This approach involves grouping data based on policyholder claim information background variables using K mode clustering. By looking at the estimated claim re-

serves in each cluster, it is possible to estimate individual claim reserves. A case study using BPJS data showed that the MSEP claim reserve estimate was smaller than the claim reserve estimate using the RDC method without grouping. Consequently, insurance companies can benefit from applying the RDC and K modes clustering methods to calculate estimated claims reserves.

Acknowledgements

We are grateful to the editors and reviewers for their thorough review of our manuscript and the helpful feedback that enabled us to make significant improvements to the original version. We are also thankful to LPDP, Gadjah Mada University and Diponegoro University for their support of our research.

REFERENCES

[1] Wütrich, M. and Merz, M., "Stochastic Claims Reserving Methods in Insurance", John Wiley and Sons Ltd., 2008

[2] Mack, T., "Distribution-Free Calculation of the Standard Error of Chain Ladder Method Reserves Estimates", ASTIN Bulletin, Vol. 23, pp. 213-225, 1993. https://www.actuaries.org/LIBRARY/ASTIN/vol23no2/213.pdf

[3] Godecharle, E. and Antonio, K., "Reserving by Conditioning on Markers of Individual Claims: A Case Study Using Historical Simulation", Faculty of Economics and Business, Belgie, 2014. DOI: https://doi.org/10.1080/10920277.2015.1046607

[4] Drieskens, "Stochastic Projection for Large Individual Losses", Scandinavian Actuarial Journal, Vol. 1, pp. 1-39, 2012. DOI: https://doi.org/10.1080/03461231003759708

[5] Rosenlund, S., "Bootstrapping Individual Claim Histories", ASTIN Bulletin, Vol. 42, No. 1, pp. 291-324, 2012. DOI: https://doi.org/10.2143/AST.42.1.2160744

[6] Kroon, R., "Individual Reserving by Detailed Conditioning-A Parametric Approach", Faculty of Economics and Business: Belgie, 2014.

[7] Ohlsson, E. and Johansson, B., "Non-Life Insurance Pricing with Generalized Models", Springer-Verlag, Berlin, 2010.

[8] Kartikasari, M.D., Effendie, A.R. and Wilandari, Y., "Reserving by Detailed Conditioning on Individual Claim", AIP Conference Proceedings, Vol. 1827, No. 1, 2017, 020012. DOI: https://doi.org/10.1063/1.4979428

[9] Effendie, A., and Pebriawan, R., "Estimation of IBNR and RBNS Reserve by Detailed Conditioning Method", Far East Journal of Mathematical Sciences (FJMS), Vol. 101, pp. 2785–2801, 2017. https://doi.org/10.17654/MS101122785.

[10] Yulita, T and Effendie, A.R., "Estimation of IBNR and RBNS Reserves Using RDC Method and Gamma Generalized Linear Model", Media Statistika, Vol. 15, No. 1, pp. 24-35, 2022. DOI: 10.14710/medstat.15.1.24-35.

[11] Wütrich, M., "Machine Learning in Individual Claims Reserving", Scandinavian Actuarial Journal, Vol. 6, pp. 465-480, 2018. DOI: 10.1080/03461238.2018.1428681

[12] Lopez, O. and Milhaud, X., "Individual Reserving and Nonparametric Estimation of Claim Amounts Subject to Large Reporting Delays", Scandinavian Actuarial Journal, Vol. 2021, No. 1, pp. 34-53, 2021. DOI: 10.1080/03461238.2020.1793218

[13] Saida, G., Riad, R.M. and Nawel, R., "Claim Development Patterns with Cluster Analysis", Thailand Statistician, Vol. 21, No. 2, pp. 257-267, 2023. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/248999/168678

[14] Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E., "Multivariate Data Analysis", Springer-Verlag, Berlin, 2010.

[15] Huang, Z. and Ng, M. K., 2003. "A Note on K-modes Clustering", Journal of Classification, Vol. 20, No. 2, pp. 257-261, 2003. DOI: 10.1007/s00357-003-0014-4