



Thailand Statistician
April 2023; 21(2): 257-267
<http://statassoc.or.th>
Contributed paper

Claim Development Patterns with Cluster Analysis

Guettouche Saida *, Remita Med Riad and Arrar Nawel

Mathematics Department, Badji Moukhtar university, Annaba, Algeria

* Corresponding author; Email: s.guettouche@yahoo.fr

Received: 15 December 2019

Revised: 28 April 2021

Accepted: 5 July 2021

Abstract

Loss reserving models has been in majority very sensitive to long payment delay of incurred claims, the insurance company is not able to, precisely, forecast the amount or the delay which a claim need to be settled definitively. Therefore, this existing contrast between claim development patterns called for the idea of implementing clustering techniques in order to extract patterns of development claims through the set of information aggregated in the run off triangle. The aim of this paper is to shed the light on the way of eliciting the multiple features of development pattern for IBNR claims using Wards agglomerative clustering algorithm.

Keywords: Run-off triangle, claim development patterns, agglomerative hierarchical clustering, Ward's method, clustering

1. Introduction

The actually paid amount in a given development year is a complex result of numerous factors including the uncertain outcomes and costs of medical diagnoses and treatments, and of court proceedings or settlement negotiations. None of these factors can be easily forecast. Consequently, even at an aggregate level, attempts of predicting future claim payments are imprecise.

Stability of the claim payment or settlement pattern is a widely used assumption. Yet, practically speaking, this ideal is rarely realised. Usually, the insurance company is not able to immediately settle a claim. As long as the claim remains unsettled, establishing an accurate reserve would be a difficult task.

The reporting of a claim can take several years; especially in liability insurance where it can take many years to be reported and other several years until the total degree of a claim is known and settled. Usually, injury claims tend to take longer time than property claims to be settled, which would lead to a short tail in the run-off triangle for property damage, and a medium or a long tail for liability or injury claims, see Bornhuetter and Ferguson (1972) for farther information.

Therefore, this existing contrast between claim development patterns called for the idea of implementing data clustering techniques. The aim of clustering is to put the elements in a set so that they are grouped in clusters, each cluster contains very similar elements and is preferably highly heterogeneous from the elements grouped in the other clusters. Using these techniques on aggregated Data with different tails in the run off triangle lead to extract patterns of development claims. For a survey of clustering techniques, see Duda et al. (2001), Hastie et al. (2009) and Ward and Joe (1963). Charpentier and Pigeon (2016), Holler (1999) among many others researchers used clustering techniques in loss reserving. One such significant study was carried out by Kremer (2000) who combined, using

the euclidian distances, link ratio with cluster analysis to calculate the similarity of the development of the i -th accident year with relevant previous ones.

The aim of this paper is to shed the light on the way of eliciting the multiple features of settlement pattern in a design matrix through the aggregated data set in the run-off triangle.

2. Overview of Agglomerative Hierarchical Clustering

As a data mining and an unsupervised learning tool in machine learning, clustering has been widely studied across many disciplines due to its numerous areas of application such as biology, social sciences, business intelligence and marketing.

Clustering is described as the process of creating similar groups or clusters, each cluster is characteristically distinct from one another, and the objects within each one are similar to each other. Methods and algorithms have been developed, and several books and survey articles have been dedicated to clustering methods including Hartigan (1975), Kaufman and Rousseeuw (1990) and Jain (2010).

It can be challenging to identify a relevant categorization of clustering methods since these categories may overlap because of the solid boundary between them. The main ones are partitioning methods, hierarchical methods, density-based methods and grid-based methods.

As the second method of clustering, while partitioning methods meet the essential clustering requirement of organizing a set of objects into several exclusive groups, hierarchical methods seek to partition the data into groups at different levels to build a hierarchy of clusters which can be displayed using a dendrogram. Depending on whether this hierarchical representation is created in top-down, splitting big clusters into smaller ones, or in bottom-up grouping small clusters into larger ones, these hierarchical approaches are called divisive and agglomerative clusterings, respectively. The focus in this paper is agglomerative methods, farther details on divisive methods are available in Hartigan (1975).

The agglomerative approach starts with the single observation clusters and successively merges the closest pairs of clusters. Then, the distances between this new cluster and the other unchanged clusters are updated forming smaller numbers of clusters that contain more observations; this procedure is iterated based on a similarity strategy until the desired cluster structure is obtained or a stopping criterion is met. The basic algorithm is straightforward:

Algorithm 1 Agglomerative hierarchical clustering algorithm.

- 1: Start with each point in a cluster of its own
 - 2: Compute and update the proximity matrix
 - 3: Repeat
 - a: Merge the closest two clusters.
 - b: Update the proximity matrix
- Until the stopping criteria are met
-

The definition of distance proximity is the crucial process of the algorithm; the formed clusters set at each step is updated along with merging nearest clusters. The choice in terms of how the clusters may be merged is determined through the proximity matrix by a linkage criterion, which is a function of the pairwise distances between clusters. A variety of linkage methods are possible such as single linkage, complete linkage, median, average linkage, centroid and Ward's method, which provide different tradeoffs between quality and efficiency.

Comparative researches of Kuiper and Fisher (1975), Blashfield (1976) and Everitt and Hands (2010) among others give rise to Ward's agglomerative clustering mostly when the group proportions were approximately equal, that does not preclude that other distance measures are appropriate when different conditions are considered.

3. Claim Development Patterns via Agglomerative Clustering Algorithm

The basically input is the standard run-off triangle of cumulated payments of claims with $1, \dots, I$ accident years, $1, \dots, J$ development years with $J \leq I$ and $k = 1, \dots, K_i$ contracts with claims in the insurance portfolio for each accident year i given in the set

$$\mathcal{T} := \{(i, k, j) | 1 \leq i \leq I, 1 \leq k \leq K_i, j \leq I - i + 1\}, \quad (1)$$

$S = (S_{i,k,j})_{i \leq I, k \leq K_i, j \leq J}$ is the matrix form of the cumulated payments of claims from accident year i , contract k and development year j .

We set $\mathbb{I}_{K_i}^{\text{tr}} S_i = \sum_{k \leq K_i} S_{i,k} = \left(\sum_{k \leq K_i} S_{i,k,1}, \dots, \sum_{k \leq K_i} S_{i,k,J} \right)$, if we had a complete data set, it would be easy to find characteristic development patterns simply by dividing the cumulative development patterns

$$\xi_{i,k} = (\xi_{i,k,j})_{j \leq J} \text{ where } \xi_{i,k,j} = \frac{S_{i,k,j}}{S_{i,k,J}}, \quad i \leq I, k \leq K_i, j \leq J \quad \text{and} \quad S_{i,k,J} > 0.$$

However we have to deal with the index set $T^{(j)}$.

Theorem 1

- The Claim development patterns in the run-off triangle is characterised by the matrix Ξ defined by:

$$\Xi := (\xi_{j,\ell})_{1 \leq j \leq J, \ell=1, \dots, L} \in \mathbb{R}^{J \times L} \quad (2)$$

whose L columns represent characteristic cumulative development patterns, $L < J$, for all $\ell \leq L$ we have $0 \leq \xi_{1,\ell} \leq \xi_{2,\ell} \leq \dots \leq \xi_{J,\ell} = 1$, and $0 < \xi_{1,\ell}$ for at least one $\ell \leq L$.

- The recursive g implemented to extract the matrix Ξ is the following :

(1) Initialisation

For each subinterval $[1, 2, \dots, j]$ of the whole set of development years $\{1, \dots, J\}$ we regard the reduced index set

$$T^{(j)} := \{(i, k) | k \leq K_i, 1 \leq i \leq I + 1 - j\}. \quad (3)$$

We start with $j = 2$ and $T^{(2)}$, to get the repartition $\mathcal{R}^*(T^{(2)})$.

(2) The aggregation step

For a subset $T^{(j)}$, we construct a repartition $\mathcal{R}^*(T^{(j)})$ representing the clusters of the development patterns within $T^{(j)}$ for this subinterval.

(3) The refinement step

In the refinement step, we construct for each $T^{(j)} \neq \emptyset$

$$T^{(j)} \in \{R \cap \mathcal{T} \neq \emptyset | R \in \mathcal{R}^*(T^{(j-1)})\} =: \mathcal{R}^*(T^{(j-1)}) \cap T^{(j)}, \quad (4)$$

a repartition $\mathcal{R}^*(T^{(j)})$ of the subinterval $[1, 2, \dots, j]$ according to step 1 and set

$$\mathcal{R}^{(*j)} := \bigcup_{\emptyset \neq T \in \mathcal{R}^*(T^{(j-1)}) \cap \mathcal{T}} \mathcal{R}^*(T^{(j)}). \quad (5)$$

So, $\mathcal{R}^{(*j)}$ is a repartition of $T^{(j)}$ which refines the repartition $\mathcal{R}^{(*j-1)}$ on $T^{(j)}$.

- (4) Finally, the repartition $\mathcal{R}^{(*J)}$ for the whole development interval $[1, 2, \dots, J]$ determines our final matrix Ξ as the set of barycenters of the development patterns in each index set of the repartition $\mathcal{R}^{(*J)}$.

Proof:

The main problem lies in the fact that intrinsically the data for future years are missing. If we had complete data, we could apply a standard method of cluster analysis. It is our intention in the following procedure to incorporate as much information available in the given data as possible. The way out chosen of this dilemma is an iterative procedure where the two steps (2) and (3) alternate recursively, until the system converges to a stable situation. We proceed by the following recursive clustering method:

Step 1:

We assume that to each index $(i, k) \in$ from the reduced set $T^{(j)}$ of the subinterval $[1, 2, \dots, j]$ we have associated a cumulative development pattern

$$\xi_{i,k}^{(j)} = (\xi_{i,k,1}^{(j)}, \dots, \xi_{i,k,j}^{(j)}) \in \mathbb{R}^j \quad \text{with} \quad 0 \leq \xi_{i,k,1}^{(j)} \leq \dots \leq \xi_{i,k,j}^{(j)} = 1. \quad (6)$$

For subsets $R \subset T^{(j)}$, the barycenter pattern of R is given by

$$\xi_R^{(j)} := (\xi_{R,1}^{(j)}, \dots, \xi_{R,j}^{(j)}) \quad \text{with} \quad \xi_{R,\nu}^{(j)} = \frac{1}{|R|} \sum_{(i,k) \in R} \xi_{i,k,\nu}^{(j)}, \nu \leq j. \quad (7)$$

where $|R|$ is the cardinality of R . (Here, we assume that every development pattern has the same weight; otherwise one has to use a weighted barycenter.) So, $\xi_T^{(j)}$ is the barycenter pattern of the whole index set $T^{(j)}$. The cumulative development patterns need to be normalized to last available data in the index set $T^{(j)}$ in a preliminary step

$$\xi_{i,k,j} = \frac{S_{i,k,j}}{S_{i,k,I+1-j}}. \quad (8)$$

The initial operation in step 1 is the determination of the reduced set $T^{(j)}$ given by (3) for $j = 2$

$$T^{(2)} := \{(i, k) \mid k \leq K_i, 1 \leq i \leq I - 1\}.$$

To set the partition $\mathcal{R}(T^{(2)})$, we need to move to the next step.

Step 2:

A repartition $\mathcal{R}^{(j)}$ of $T^{(j)}$ is a set of no-empty subsets $R \subset T^{(j)}$, where each index $(i, k) \in T^{(j)}$ appears exactly in one subset of $\mathcal{R}^{(j)}$, $T^{(j)} = \bigcup R$, where $\bigcup$ denotes the union of disjoint sets. The finest and the coarsest repartitions are resp.

$$\begin{aligned} \mathcal{R}_{\min}(T^{(j)}) &:= \left\{ \{(i, k)\} \mid (i, k) \in T^{(j)} \right\} \quad \text{resp.} \\ \mathcal{R}_{\max}(T^{(j)}) &:= \{T^{(j)}\}. \end{aligned} \quad (9)$$

The main operation in step (1) is the recursive aggregation, it consists in merging the two closest subsets R_1 and R_2 of a repartition $T^{(j)}$ based on the distance measure $d(R_1, R_2)$.

Lemma 1 The distance $d(R_1, R_2)$ between two disjoint subsets R_1 and R_2 of $T^{(j)}$ is measured by the inertia of the dipole consisting of their barycenters with the corresponding weights:

$$d(R_1, R_2) := |R_1| \|\xi_{R_1}^{(j)} - \xi_{R_1 \cup R_2}^{(j)}\|^2 + |R_2| \|\xi_{R_2}^{(j)} - \xi_{R_1 \cup R_2}^{(j)}\|^2 \quad (10)$$

with the Euclidian norm $\|\cdot\|$ in $\mathbb{R}^j$. Sometimes, this distance is also called Ward distance and defined as

$$d(R_1, R_2) = \frac{|R_1| |R_2|}{|R_1 \cup R_2|} \|\xi_{R_1}^{(j)} - \xi_{R_2}^{(j)}\|^2.$$

Proof: To see the equivalence between the two formulas it suffices to remember that

$$\xi_{R_1 \cup R_2}^{(j)} = \frac{|R_1|}{|R_1 \cup R_2|} \xi_{R_1}^{(j)} + \frac{|R_2|}{|R_1 \cup R_2|} \xi_{R_2}^{(j)}.$$

$$\begin{aligned} d(R_1, R_2) &= |R_1| \left\| \xi_{R_1}^{(j)} - \xi_{R_1 \cup R_2}^{(j)} \right\|^2 + |R_2| \left\| \xi_{R_2}^{(j)} - \xi_{R_1 \cup R_2}^{(j)} \right\|^2 \\ &= |R_1| \left\| \xi_{R_1}^{(j)} - \frac{|R_1|}{|R_1 \cup R_2|} \xi_{R_1}^{(j)} - \frac{|R_2|}{|R_1 \cup R_2|} \xi_{R_2}^{(j)} \right\|^2 + |R_2| \left\| \xi_{R_2}^{(j)} - \frac{|R_1|}{|R_1 \cup R_2|} \xi_{R_1}^{(j)} - \frac{|R_2|}{|R_1 \cup R_2|} \xi_{R_2}^{(j)} \right\|^2 \\ &= |R_1| \left\| \frac{|R_2|}{|R_1 \cup R_2|} \xi_{R_1}^{(j)} - \frac{|R_2|}{|R_1 \cup R_2|} \xi_{R_2}^{(j)} \right\|^2 + |R_2| \left\| \frac{|R_1|}{|R_1 \cup R_2|} \xi_{R_2}^{(j)} - \frac{|R_1|}{|R_1 \cup R_2|} \xi_{R_1}^{(j)} \right\|^2 \\ &= \frac{|R_1| |R_2|^2}{|R_1 \cup R_2|^2} \left\| \xi_{R_1}^{(j)} - \xi_{R_2}^{(j)} \right\|^2 + \frac{|R_2| |R_1|^2}{|R_1 \cup R_2|^2} \left\| \xi_{R_2}^{(j)} - \xi_{R_1}^{(j)} \right\|^2 \\ &= \frac{|R_1| |R_2|}{|R_1 \cup R_2|} \left\| \xi_{R_1}^{(j)} - \xi_{R_2}^{(j)} \right\|^2 \end{aligned}$$

Now we come to the aggregation procedure, let $\mathcal{R}^{(j)} \neq \mathcal{R}_{\max}(T^{(j)})$ and take $R_1, R_2 \in \mathcal{R}^{(j)}$, $R_1 \neq R_2$, as one pair of closest elements in $\mathcal{R}$:

$$\{R_1, R_2\} \in \arg \min_{R, R' \in \mathcal{R}^{(j)}, R \neq R'} d(R, R'). \quad (11)$$

The aggregated repartition $\mathcal{A}(\mathcal{R}^{(j)})$ of $\mathcal{R}^{(j)}$ is then

$$\mathcal{A}(\mathcal{R}^{(j)}) := \left\{ R' \in \mathcal{R}^{(j)} \mid R' \neq R_1, R' \neq R_2 \right\} \cup \{R_1 \cup R_2\}. \quad (12)$$

(Here, it is not of importance that R_1 and R_2 may not be uniquely determined.) Obviously, one has $\mathcal{A}^{[T^{(j)}]-1}(\mathcal{R}_{\min}(T^{(j)})) = \mathcal{R}_{\max}(T^{(j)})$.

To measure the effectiveness of an aggregation step, one generally uses the increase of the relative intra-partition inertia, defined in the following way: For a repartition $\mathcal{R}^{(j)}$ of $T^{(j)}$, the intra-inertia of $\mathcal{R}^{(j)}$ is given by

$$\mathcal{I}_{\text{intra}}(\mathcal{R}) := \sum_{R \in \mathcal{R}} \sum_{(i,k) \in R} \left\| \xi_{i,k}^{(j)} - \xi_R^{(j)} \right\|^2 \geq 0, \quad (13)$$

while the inter-inertia of a repartition is

$$\mathcal{I}_{\text{inter}}(\mathcal{R}) := \sum_{R \in \mathcal{R}} |R| \left\| \xi_R^{(j)} - \xi_T^{(j)} \right\|^2, \quad (14)$$

as a generalization. Obviously, $\mathcal{I}_{\text{intra}}(\mathcal{R}_{\min}(T^{(j)})) = \mathcal{I}_{\text{inter}}(\mathcal{R}_{\max}(T^{(j)})) = 0$ for all $T^{(j)} \neq \emptyset$.

Lemma 2 *From Huygens Theorem, for any repartition $\mathcal{R}$ of T we have the identity:*

$$\mathcal{I}_{\text{intra}}(\mathcal{R}(T^{(j)})) + \mathcal{I}_{\text{inter}}(\mathcal{R}(T^{(j)})) = \mathcal{I}_{\text{intra}}(\mathcal{R}_{\max}(T^{(j)})) = \mathcal{I}_{\text{inter}}(\mathcal{R}_{\min}(T^{(j)})). \quad (15)$$

Proof: Indeed, replacing in $\xi_{i,k}^{(j)}$ for simplicity the double index (i, k) by ℓ and dropping the super-

script $^{(j)}$, one easily checks that:

$$\begin{aligned}
 & \mathcal{I}_{intra}(\mathcal{R}(T)) + \mathcal{I}_{inter}(\mathcal{R}(T)) \\
 = & \sum_{R \in \mathcal{R}} \left(|R| \|\xi_T - \xi_R\|^2 + \sum_{\ell \in R} \|\xi_\ell - \xi_R\|^2 \right) \\
 = & \sum_{R \in \mathcal{R}} |R| \|\xi_T\|^2 + \sum_{R \in \mathcal{R}} |R| \|\xi_R\|^2 - 2 \sum_{R \in \mathcal{R}} |R| \langle \xi_T, \xi_R \rangle \\
 & + \sum_{R \in \mathcal{R}} \left(\sum_{\ell \in R} \|\xi_\ell\|^2 + \sum_{\ell \in R} \|\xi_R\|^2 - 2 \sum_{\ell \in T} \langle \xi_\ell, \xi_R \rangle \right) \\
 = & \sum_{R \in \mathcal{R}} |R| \|\xi_T\|^2 - 2 \sum_{R \in \mathcal{R}} |R| \langle \xi_T, \xi_R \rangle + \sum_{R \in \mathcal{R}} \sum_{\ell \in R} \|\xi_\ell\|^2 + \sum_{R \in \mathcal{R}} |R| \|\xi_R\|^2 \\
 & - 2 \sum_{R \in \mathcal{R}} \sum_{\ell \in R} \langle \xi_\ell, \xi_R \rangle + \sum_{R \in \mathcal{R}} \sum_{\ell \in R} \|\xi_R\|^2 \\
 = & |T| \|\xi_T\|^2 - 2 \sum_{R \in \mathcal{R}} |R| \left\langle \xi_T, \frac{1}{|R|} \sum_{\ell \in R} \xi_\ell \right\rangle + \sum_{\ell \in T} \|\xi_\ell\|^2 \\
 & + \sum_{R \in \mathcal{R}} \left(|R| \|\xi_R\|^2 - 2 \sum_{\ell \in R} \langle \xi_\ell, \xi_R \rangle + |R| \|\xi_R\|^2 \right) \\
 = & |T| \|\xi_T\|^2 - 2 \sum_{R \in \mathcal{R}} \sum_{\ell \in R} \langle \xi_T, \xi_\ell \rangle + \sum_{\ell \in T} \|\xi_\ell\|^2 \\
 & + \sum_{R \in \mathcal{R}} \left(|R| \|\xi_R\|^2 - 2 |R| \left\langle \frac{1}{|R|} \sum_{\ell \in R} \xi_\ell, \xi_R \right\rangle + |R| \|\xi_R\|^2 \right) \\
 = & \sum_{\ell \in T} \|\xi_T\|^2 - 2 \sum_{\ell \in T} \langle \xi_T, \xi_\ell \rangle + \sum_{\ell \in T} \|\xi_\ell\|^2 + \sum_{R \in \mathcal{R}} \left(|R| \|\xi_R\|^2 - 2 \langle \xi_R, \xi_R \rangle + |R| \|\xi_R\|^2 \right) \\
 = & \sum_{\ell \in T} \|\xi_\ell - \xi_T\|^2 + \sum_{R \in \mathcal{R}} |R| \|\xi_R - \xi_R\|^2 \\
 = & \mathcal{I}_{inter}(\mathcal{R}_{min}(T)).
 \end{aligned}$$

Lemma 3 A consequence of (15) is the increase of the intra-partition inertia by aggregation.

$$\mathcal{I}_{intra}(\mathcal{A}(\mathcal{R}^{(j)})) \geq \mathcal{I}_{intra}(\mathcal{R}^{(j)}) \quad (16)$$

Proof: $\{R_1, R_2\}$ is a partition of the subset $R_1 \cup R_2$, in other terms, taking the repartition $\mathcal{R}(T^{(j)}) = \{R_1, R_2\}$ involves $\mathcal{R}_{max}(\{R_1, R_2\}) = R_1 \cup R_2$. According to (15) we have

$$\mathcal{I}_{intra}(R_1 \cup R_2) = \mathcal{I}_{intra}(\{R_1, R_2\}) + \mathcal{I}_{inter}(\{R_1, R_2\}). \quad (17)$$

Using (17) with the notation of the aggregated partition $\mathcal{A}(\mathcal{R}^{(j)})$ in (12).

$$\begin{aligned}
 \mathcal{I}_{intra}(\mathcal{A}(\mathcal{R}^{(j)})) &= \mathcal{I}_{intra}(\{R' \in \mathcal{R} \mid R' \neq R_1, R' \neq R_2\}) + \mathcal{I}_{intra}(R_1 \cup R_2) \\
 &= \mathcal{I}_{intra}(\{R' \in \mathcal{R} \mid R' \neq R_1, R' \neq R_2\}) + \mathcal{I}_{intra}(\{R_1, R_2\}) \\
 &\quad + \mathcal{I}_{inter}(\{R_1, R_2\}) \\
 &= \mathcal{I}_{intra}(\{R' \in \mathcal{R} \mid R' \neq R_1, R' \neq R_2\}) + \mathcal{I}_{intra}(\{R_1, R_2\}) \\
 &\quad + \mathcal{I}_{inter}(\{R_1, R_2\}) \\
 &= \mathcal{I}_{intra}(\mathcal{R}^{(j)}) + \mathcal{I}_{inter}(\{R_1, R_2\}) \geq \mathcal{I}_{intra}(\mathcal{R}^{(j)})
 \end{aligned}$$

Pointing out that a repartition $\mathcal{R}$ of $T^{(j)}$ is a set of no-empty subsets $R \subset \mathcal{R}^{(j)}$, where each index (i, k) appears exactly in one subset $R_i \in \mathcal{R}^{(j)}$, the third equality stems from

$$\{R' \in \mathcal{R}^{(j)} \mid R' \neq R_1, R' \neq R_2\} \cap \{R_1, R_2\} = \emptyset$$

So, the aggregation step increases the intra-repartition inertia exactly by the inter-repartition inertia of the two aggregated sets.

For the criterion which repartition $\mathcal{R}$ of $T^{(j)}$ should finally be chosen, the relative intra-repartition inertia RI defined as

$$RI(\mathcal{R}(T^{(j)})) := \frac{\mathcal{I}_{intra}(\mathcal{R}(T^{(j)}))}{\mathcal{I}_{intra}(\{T^{(j)}\})} \leq 1 \quad (18)$$

is very useful.

In the aim of keeping our final matrix Ξ very simple, we choose at level of the subinterval $[1, \dots, j]$ as optimal repartition $\mathcal{R}^*(T^{(j)})$ in the following way:

(recall that $\mathcal{A}^{|T^{(j)}|-1}(\mathcal{R}_{min}(T^{(j)})) = \mathcal{R}_{max}(T^{(j)}) = \{T^{(j)}\}$)

$$\mathcal{R}^{(*)j} := \begin{cases} \{T^{(j)}\} & \text{if } RI(\mathcal{A}^{|T^{(j)}|-2}(\mathcal{R}(T^{(j)}))) \in (0.95, 1], \\ \mathcal{A}^{|T^{(j)}|-2}(\mathcal{R}(T^{(j)})) & \text{if } RI(\mathcal{A}^{|T^{(j)}|-2}(\mathcal{R}(T^{(j)}))) \in (0.9, 0.95], \\ \mathcal{A}^{|T^{(j)}|-3}(\mathcal{R}(T^{(j)})) & \text{else} \end{cases} \quad (19)$$

The repartition $\mathcal{R}^{(*)}(T^{(j)})$ represents the clusters of the development patterns within T for the subinterval $[1, 2, \dots, j]$.

Step 3:

The refinement step purpose is to maximise the information included in the recursion because every time j proceed in $[1, 2, \dots, J]$, less partitions R_i are available in the repartition $\mathcal{R}(T^{(j)})$.

At the end of the step 3, a repartition $\mathcal{R}^*(T^{(j-1)})$ is set, To each index $(i, k) \in (T^{(j)})$ we associate from the repartition $\mathcal{R}^{(*)}(T^{(j-1)})$ the cumulative development pattern $\xi_{i,k}^{(*)j-1} = (\xi_{i,k,1}^{(j-1)}, \dots, \xi_{i,k,j-1}^{(j-1)})$ on subinterval $[1, \dots, j-1]$ via

$$\xi_{i,k,\nu}^{(j)} := \begin{cases} \xi_{R,\nu}^{(j-1)} \min\left(\frac{S_{i,k,j-1}}{S_{i,k,j}}, 1\right) & \text{for } 1 \leq \nu \leq j-1 \\ 1 & \text{for } \nu = j \end{cases} \quad (20)$$

according to

$$T^{(j)} \in \{R \cap \mathcal{T} \neq \emptyset \mid R \in \mathcal{R}^*(T^{(j-1)})\} =: \mathcal{R}^*(T^{(j-1)}) \cap T^{(j)}$$

(The min-operator only serves to insure that the cumulative development patterns are increasing, even if the cumulative payments are not).

Applying step 3 to $T^{(j)}$ and the associated development patterns, we get first $\mathcal{R}(T^{(j)})$ by (19) and then $\mathcal{R}^{(*)j}$ by (5).

Step 4:

Having thus constructed the partition $\mathcal{R}^{(*)J}$ with $|\mathcal{R}^{(*)J}| = L$ and an enumeration

$$\mathcal{R}^{(*)J} = \{R_1, \dots, R_L\},$$

we find the matrix $\Xi := \left(\xi_{j,\ell}^{(i)}\right)_{j \leq J, \ell \leq L}$ according to (20):

$$\xi_{j,\ell}^{(i)} := \xi_{R_\ell,j}^{(J)}. \quad (21)$$

This ends the construction of the design matrix.

4. Implementation and Results

To implement the approach steps in practice, a dataset accident claims extracted from a leading automobile insurance company (labeled "CAAR") has been collected in the run-off triangle within a time period of 7 calendar years between 2006 and 2012 with $I = 7$ accident year , $k_i = 5$ contracts for each accident year i and $J = 7$ development years given in the Table 1.

Table 1 Cumulated run-off triangle data between 2006 and 2012

i	k_i	$ R_{i,k} $	j						
			1	2	3	4	5	6	7
			$T^{(2)}$						
1	1	1	1001	2956	3624	4189	4536	4684	4784
	2	1	837	1756	1833	1883	2257	2699	2929
	3	1	1187	2310	3054	3739	4086	6921	4324
	4	1	1011	1275	1528	1850	2392	3145	3445
	5	1	1500	2400	2600	2650	3024	3924	5444
2	1	1	1113	2103	2774	3422	3844	4194	
	2	1	625	789	912	1183	1355	1505	
	3	1	1254	2232	3033	3654	3992	4292	
	4	1	1331	1662	1936	2256	2890	3296	
	5	1	1500	2000	2123	2394	2566	3456	
3	1	1	1265	2433	3334	3878	4328		
	2	1	803	918	1009	1165	1290		
	3	1	1176	2241	3120	3654	4418		
	4	1	1528	1831	2067	2574	2574		
	5	1	1803	2518	2839	2995	3120		
4	1	1	1490	2573	3280	3930			
	2	1	532	739	906	1026			
	3	1	1387	2588	3642	4542			
	4	1	1300	1579	1916	2145			
	5	1	1700	2500	2967	3187			
5	1	1	1725	3161	4361				
	2	1	914	1179	1379				
	3	1	1289	2274	3024				
	4	1	232	701	971				
	5	1	2014	2614	3414				
6	1	1	1889	3489					
	2	1	734	1484					
	3	1	1795	3395					
	4	1	1627	2353					
	5	1	2000	2750					
7	1	1	1300						
	2	1	1100						
	3	1	980						
	4	1	1230						
	5	1	1300						

Once the dataset is summarized in the appropriate run-off triangle format, it is ready to feed into the clustering algorithm.

1. Initialisation

We start with $j = 2$ and $T^{(2)}$ than the data in this set are normalized to last available data.

2. The aggregation step

Each contract (i, k_i) of the set $T^{(2)}$ is considered as a single cluster $R_{i,k}$.Using the algorithm (2) implemented in R code,the distances between all cluster $|R_{i,k}|$ are calculated and the two closest clusters are removed and replaced by their barycenter The operation is than repeated

until the repartition $\mathcal{R}^*(T^{(2)})$ for $T^{(2)}$ is established.

$$\mathcal{R}^*(T^{(2)}) = \left[\begin{array}{c|cc} 2 & 0.33 & 1 \\ 14 & 0.55 & 1 \\ 14 & 0.78 & 1 \end{array} \right] \quad (22)$$

3. The refinement step

The refinement step ensure to incorporate the results of $\mathcal{R}^*(T^{(2)})$ in $T^{(3)}$ because the more data available, the more reliable the results.

$$T^{(3)} \in \{R \cap \mathcal{T} \neq \emptyset \mid R \in \mathcal{R}^*(T^{(2)})\} =: \mathcal{R}^*(T^{(2)}) \cap T^{(3)}. \quad (23)$$

Table 2 Cumulated run-off triangle data refined by $\mathcal{R}^*(T^{(2)})$

i	k_i	$ R_{i,k} $	j						
			1	2	3	4	5	6	7
$T^{(3)}$									
1	1	1	1001	2989,89	3657,89	4222,89	4569,89	4717,89	4817,89
	2	1	837	1514,65	1591,65	1641,65	2015,65	2457,65	2687,65
	3	1	1187	2148,02	2892,02	3577,02	3924,02	4060,02	4162,02
	4	1	1011	1282,53	1535,53	1857,53	2399,53	3152,53	3452,53
	5	1	1500	2714,44	2914,44	2964,44	3338,44	4238,44	5758,44
2	1	1	1113	2014,11	2685,11	3333,11	3755,11	4105,11	
	2	1	625	792,86	915,86	1186,86	1358,86	1508,86	
	3	1	1254	2269,27	3070,27	3691,27	4029,27	4329,27	
	4	1	1331	1688,48	1962,48	2282,48	2916,48	3322,48	
	5	1	1500	1902,87	2025,87	2296,87	2468,87	3358,87	
3	1	1	1265	2289,17	3190,17	3734,17	4184,17		
	2	1	803	1018,67	1109,67	1265,67	1390,67		
	3	1	1176	2128,12	3007,12	3705,12	4305,12		
	4	1	1528	1938,39	2174,39	2441,39	2681,39		
	5	1	1803	2287,25	2608,25	2764,25	2889,25		
4	1	1	1490	2696,34	3403,34	4053,34			
	2	1	532	674,88	841,88	961,88			
	3	1	1387	2509,95	3563,95	4463,95			
	4	1	1300	1649,15	1986,15	2215,15			
	5	1	1700	2156,58	2623,58	2843,58			
5	1	1	1725	3121,60	4321,60				
	2	1	914	1159,48	1359,48				
	3	1	1289	2332,60	3082,60				
	4	1	232	692,96	962,96				
	5	1	2014	2554,92	3354,92				
:	:	:	:	:	:				
:	:	:	:	:	:				

Then, the two steps (2) and (3) alternate recursively forming smaller numbers of clusters until the repartition $\mathcal{R}^{(*)7}$ for the whole development interval $[1, 2, \dots, 7]$ determines our final matrix Ξ given by (24) and presented by Figure (1) as the set of barycenters of the development patterns,

$$\Xi = \mathcal{R}^{(*)7} = \left[\begin{array}{ccccccc} 0,27 & 0,37 & 0,42 & 0,48 & 0,58 & 0,73 & 1 \\ 0,22 & 0,51 & 0,67 & 0,79 & 0,93 & 0,98 & 1 \\ 0,34 & 0,46 & 0,52 & 0,60 & 0,72 & 0,90 & 1 \end{array} \right]. \quad (24)$$

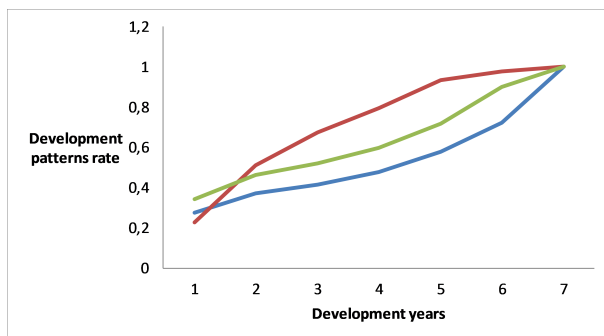


Figure 1 Characteristic of development patterns

Algorithm 2 Agglomerative hierarchical clustering algorithm.

```

1: I=35; J=7;
2: C←read.csv("cummatrix h=0.CSV",header=FALSE,sep=";",dec=",")           ▷ the input: the
   cumulated run-off triangle matrix for h=0
3: h = 0                               ▷ for h = 0 j = 2 and T(j) = T(2)
4: R← rep(0, h + 2)
5: S←function(i,j)
6: for(l in 2:(h+3))R[l-1]←(P[j,l]-P[i,l])2;sum(R)           ▷ choose the level of the matrix T:
7: M ← C[1:((I-5)-(h*5)),1:(h+3)]
8: P ← matrix(data = 0, nrow = ((I-5)-(h*5)) , ncol = (h+3))
9: for (i in 1:((I-5)-(h*5)))
   for (j in 1:(h+3))
   if (j==1)P[i,j]←M[i,j]
   elseP[i,j]←M[i,j]/M[i,h+3]           ▷ The function s for the choice of the repartition R(T(j))
10: for (s in 1:(I-8-(h*5)) )
11: W ← matrix(data = 0, nrow = I-4-s-(h*5), ncol = I-5-(h*5))
   for (j in 1:29)
   if (j←==I-5-s-(h*5))
   for (i in (j+1):(I-4-s-(h*5))) W[i,j]← P[j,1]*P[i,1]/(P[j,1]+P[i,1])*S(i,j)
12: unique(sort(W))[2]           ▷ Min of matrix W:
13: N←which(W == unique(sort(W))[2], arr.ind = TRUE)           ▷ The Position of min distance in the
   matrix W
14: Q ← rep(0, h+3)
   for (r in 1:(h+3))
   if (r==1)Q[r]←P[N[1,1],1]+P[N[1,2],1]
   else
   if (r==h+3)Q[r]←1
   elseQ[r]←(P[N[1,1],r]+P[N[1,2],r])/2           ▷ replace the deleted clusters by their barycenter :
15: P←P[-N[1,1],]
   P←P[-N[1,2],]           ▷ delete the two lines of distance min:
16: P←rbind(P,Q)
17: print(P)

```

5. Conclusion

Insurance companies have to build claim provisions in order to fulfill outstanding claim settlements which present by far the most important line on the passive side of the balance sheet.

Given the high dimensional insurance business data and the changes that which the insurance companies are going through since the introduction of the directive solvency II in Europe. The need for a good understanding of buried patterns is proved before business strategies are put into action because not only a mean estimate of the outstanding claims matters, but also the cash flow of the settlements and their distributions.

In this paper, a recursive agglomerative procedure has been implemented through the dataset summarized in the IBNR run-off triangle to identify the potential characteristics of claim development patterns which permits taking into account settlement distribution in loss reserving models without the need of an external volume measure.

References

- Blashfield RK. Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods. *Psychol Bull.* 1976; 83(3): 377-388.
- Bornhuetter RL, Ferguson RE. The actuary and IBNR. CAS59: Proceedings of the Casualty Actuarial Society; 1972. p.181-195.
- Charpentier A, Pigeon M. Macro vs. micro methods in non-life claims reserving (an econometric perspective). *Risks.* 2016; 4(2):12.
- Duda RO, Hart PE, Stork DG. *Pattern Classification*, 2nd Edition. New York: John Wiley and Sons; 2001.
- Everitt B, Hands S. A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. *Multivar Behav Res.* 2010; 22(2): 235-243.
- Hartigan JA. *Clustering Algorithms*. New York: John Wiley and Sons; 1975.
- Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York: Springer; 2009.
- Holler KD. Something old, Something new in classification ratemaking with a novel use of GLMs for credit insurance. CAF1999: Proceedings of the Casualty Actuarial Society Forum; Winter; 1999. p. 31-84.
- Jain A. Data clustering 50 years beyond K means. *Pattern Recogn Lett.* 2010; 31(8): 651-666.
- Kaufman L, Rousseeuw P. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley and Sons; 1990.
- Kremer E. Minimum distance loss reserving. *Blatter DGVM.* 2000; 24(4): 629-634.
- Kuiper F, Fisher L. A Monte Carlo comparison of six clustering procedures. *Biometrics.* 1975; 31(3): 777-783.
- Schmidt KD. Loss prediction based on run-off triangles. *Adv Stat Anal.* 2012; 96(2): 265-310.
- Schmidt KD. Methods and models of loss reserving based on run-Off trianglesa unifying survey. CAF2006: Proceedings of the Casualty Actuarial Society Forum; Fall; 2006. p.269-317.
- Ward J, Joe H. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc.* 1963; 58(1): 234-244.