



Faculté des Sciences

Département de Mathématiques



Présenté en vue de l'obtention du diplôme de

Doctorat

**Claim provision with development patterns  
via Hachemeister credibility**

(La solvabilité des sociétés d'assurance en assurance non-vie)

Option : **Modélisation Mathématique**

Présentée par :

**GUETTOUCHE Saida.**

Sous la direction de :

**Directeur de thèse:** REMITA Mohamed Riad Prof U.B.M.ANNABA  
**Co-directeur de thèse:** EISELE Karl Theodor Prof Université de Strasbourg. FRANCE

Devant le jury:

**President :** BOUTABIA Hacene Prof U.B.M.ANNABA  
**Examineur :** BENMOSTEFA Fatma Zohra Prof U.B.M.ANNABA  
**Examineur :** BOUKHETALA Kamal Prof U.S.T.H.B  
**Invitée :** ARRAR REMITA Nawel M.C.A U.B.M.ANNABA

**2022/2023**

## **Abstract**

Credibility theory is used in the context that constitutes the subject of this thesis in the aim of establishing a loss reserving model. More specifically, we consider a multivariate model for loss prediction with several contracts for each accident year.

The model includes a Hachemeister's credibility part where the design matrix represents standard development patterns for cumulative quotas. These patterns are determinate by agglomerative hierarchical clustering methods where a numerical example is presented for the design matrix. Following the procedure in regression model of Hachemeister, The credibility estimator yields a mean development pattern. Inverting this mean development pattern allows for an estimation of the final losses of accident years. Volume vectors may be useful in addition, but are not essential parts of the model.

Key words and phrases: Loss reserving, Development patterns, Credibility theory, Regression model of Hachemeister, Cluster analysis, agglomerative hierarchical clustering.

## Résumé

La théorie de la crédibilité est utilisée dans cette thèse dans le but d'établir un modèle de provisionnement. Plus précisément, nous considérons un modèle de provisionnement multivarié avec plusieurs contrats pour chaque année de survenance.

Le modèle comprend une partie basée sur la méthode de régression de Hachemeister où la matrice de dessin représente l'évolution des cadences de paiements pour les quotas cumulatifs. Cette matrice est déterminée par les techniques de la classification hiérarchique agglomérative.

En poursuivant la procédure de la méthode de régression de Hachemeister, l'estimateur de crédibilité donne une estimation moyenne de l'évolution des cadences de paiement et permet d'estimer le montant total des sinistres au cours des années de survenance. Le volume des contrats par année de survenance peut être utile, mais ne représente pas une partie essentielle du modèle. .

Mots clés: Les provisions techniques, IBNR, la théorie de la crédibilité, le modèle de Hachemeister, les cadences de paiement, les techniques de clustering.

## تلخيص

تستخدم نظرية المصدقية في موضوع هذه الأطروحة بهدف انشاء نموذج لتقدير تكلفة التعويضات المستقبلية ، بشكل أكثر تحديداً ، فإننا نعتبر نموذجاً متعدد المتغيرات باعتبار عدد العقود لكل سنة.

يشتمل النموذج على جزء يعتمد على نموذج الانحدار Hachemeister حيث تمثل مصفوفة السحب تطور معدلات السداد التراكمية. يتم تحديد مصفوفة السحب من خلال تقنيات التصنيف الهرمي التكتلي.

باتباع نفس مراحل نموذج الانحدار Hachemeister ، يعطي مقدر المصدقية تقديراً متوسطاً لتطور معدلات السداد ويجعل من الممكن تقدير المبلغ الإجمالي للمطالبات على مدى سنوات حدوثها. يمكن أن يكون حجم العقود حسب سنة حدوثها مفيداً ، لكنه ليس جزءاً أساسياً من النموذج .

### الكلمات المفتاحية والعبارات:

التعويضات المستقبلية ، IBNR ، نظرية المصدقية ، نموذج Hachemeister ، معدلات التطور ، تقنيات التجميع

# Contents

|   |           |
|---|-----------|
| <b>Introduction</b>   | <b>3</b>  |
| <b>1 Preliminaries</b>  | <b>6</b>  |
| 1.1 Experience rating Concept . . . . .   | 6         |
| 1.2 Example . . . . .   | 7         |
| 1.3 The first formula for Credibility premium . . . . .                         | 9         |
| 1.4 Credibility Approaches . . . . .  | 10        |
| 1.4.1 Buhlmann Model . . . . .  | 10        |
| 1.4.2 Buhlmann-Straub Model . . . . .   | 13        |
| 1.4.3 Hachemeister's Regresion Model . . . . .                                  | 16        |
| <b>2 Prediction of claim provisions with Hachemeister credibility</b>           | <b>20</b> |
| 2.1 Model assumptions . . . . .   | 20        |
| 2.2 The credibility estimators for the development quotas . . . . .             | 24        |
| 2.3 Estimators for the structural parameters . . . . .                          | 30        |
| <b>3 Claim development patterns via Agglomerative hierarchical clustering</b>   | <b>35</b> |
| 3.1 Overview of Agglomerative hierarchical clustering . . . . .                 | 35        |
| 3.2 Claim development patterns via agglomerative clustering algorithm . . . . . | 37        |
| 3.3 Implementation and results . . . . .  | 45        |

|  |           |
|--|-----------|
| <b>Conclusion</b>  | <b>49</b> |
| <b>A</b>   | <b>52</b> |
| A.1 Proof of the equivalence between 2.25 and 2.11 . . . . .     | 52        |
| A.1.1 Proof of 2.17 . . . . .                                    | 54        |
| A.1.2 Proof of 2.15 as superposition of 2.35 with 2.11 . . . . . | 54        |
| A.1.3 Calculations of the prediction errors . . . . .            | 55        |
| <b>B Verification of <math>\Delta^{(i)}</math></b>               | <b>59</b> |
| <b>C R code for the design matrix</b>                            | <b>61</b> |
| <b>Bibliography</b>  | <b>63</b> |

# Introduction

Since the onset of insurance contracts, loss reserving has been one of the major challenging tasks actuaries come across. After introducing the new reform of solvency II, more attention has been shifted to run off triangle models. The number of papers on the subject has grown significantly during the past decades. For comprehensive review of loss reserving methods based on run-off triangles, see [33].

Stability of the claim payment or settlement pattern is a widely used assumption. Practically speaking, this ideal is rarely realised. The insurance company is not able to immediately settle a claim because 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. This would lead to a short tail in property damage claims and a medium or a long tail for liability or injury claims. See [5] for farther information.

As long as the claim remains unsettled, developing an accurate forecast for the ultimate claim for different lines of business as well as for the whole insurance portfolio would be a difficult task for actuaries. Moreover, the run-off triangle is heavily fluctuated due to unexpected random perturbation and scarce data base in some business lines. Therefore, taking into account the various claims lines and the fact that some of them have insufficient data, actuaries sometimes use the claims experience from other similar lines instead of the observed individual data. Hence, the main concern now is to what extent actuaries can rely on the portfolio claims experience and to

what extent they can depend on the individual line. These issues can be resolved by means of credibility theory that can also answer the question about how to combine individual and portfolio claims data in order to have a better estimation of the individual ultimate claim amount. Credibility loss reserving was launched by Benktander, in [3], where he replaced the prior estimated ultimate claims amount by a credibility mixture. Later, Esa Hovinen applied the credibility mixture directly to the reserves instead of the ultimates in his paper " Additive and Continuous IBNR". A similar approach was discussed in Vyllder [10] who developed a stochastic multiplicative model for the prediction of IBNR claims where the factor depending on the accident year is credibility adjusted. This latter has been slightly improved upon and generalized by Mack [24]. A similar general aspect was discussed by Schmidt [33] using a typical application of general credibility theory in loss reserving.

The main approach that constitutes the subject of this thesis belongs to credibility loss reserving where a multivariate model for loss prediction with several contracts for each accident year is considered. The model includes a Hachemeister's credibility part where the design matrix represents standard development patterns for cumulative quotas. These patterns are determinate by agglomerative hierarchical clustering methods. For a review of cluster analysis, see [15] [22][16]. Charpentier and Pigeon in [7], Holler in [19], among many other researchers, used clustering techniques in loss reserving. One such significant study was carried out by Kremer [23] 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.

This thesis is organized as follows; the mathematical formulations underlying experience rating including the basic credibility approaches of Buhlmann, Buhlmann-straub and Hachemeister are carefully reviewed in chapter 1. In chapter 2, we start by the necessary assumptions of the model then the explicit solution of the Hachemeister credibility premium estimator is then obtained with the parameter estimators. The design matrix in the Hachemeister part of the main model using agglomerative hierarchical clustering is treated in chapter 4. Numeric experiment is performed at the end of the chapter. Conclusions and future work are given in the final chapter.

renewcommand01

# Chapter 1

## Preliminaries

### 1.1 Experience rating Concept

The basic point in insurance is that those who get exposed to the same degree of risk group together to make what is called "community-at-risk"[6]. Then, they pass their risk to an insurance company in an attempt to pay a premium. The main step in determining premium is to make sure that it covers all its obligations. After that, the insurer will try to allocate fair premiums among its insured individuals.

The first point to get a fair premium allocation is to form a classification system. Therefore, this system puts individuals in lines of business with large number of contracts into homogeneous risk classes in terms of certain observable risk characteristics. Hence, following the Law of Large Numbers, when the volume of similar, independent exposure units goes up, the observed experience will be close to the "true" experience. Moreover, for a sufficiently large number, the observed experience will be equal to the "true" experience. Therefore, the experience of the whole group will be more stable and more predictable by guaranteeing a large number of independent risks. However, to get reasonably homogenous classes of risk, the portfolio is going to be classified into very large number of classes, many of which will include trifle risks and, hence, give few statistical

data to rate future risks. But , when the portfolio is classified into larger classes, the assumption of homogeneous risk profiles within the classes becomes unreal.

Along with the challenges posed by classification, the late claims experience belonging to the insured and which the insurer may have as a result to his previous claims experience will subsequently provide indispensable perception into his unobserved risk characteristics that cannot be disregarded. In fact, in the valuing of Non-life insurance products, the act of rating for the past experience of the insured in claims prediction and in premium calculation is referred to as experience rating. The second point of determining a fair premium allocation, in addition to forming a classification system, is expressed in experience rating systems in general and credibility theory method in particular. In this research, Ragnar Norberg's example, as mentioned in [20], is used in order to introduce a mathematical understanding of the credibility theory concept.

## 1.2 Example

According to the example, 10 drivers were insured by a given insurance company for ten years. Saying that driver "i" had, at large, one accident during year "j", the premium for this portfolio, then, is approximated to be 0.20. This is called the collective premium each insured pays in the first year.

Then, as shown in table 1.1, after a period of one year, the insurer perceived the claim record shown where only the contract number 9 had a claim. The average claim amount, then, is  $1/10 = 0.10$ . The insurer is disposed to maintain its premium unchanged regarding the scarce experience.

Table 1.1: Portfolio claims after one year

| Years | Insureds |   |   |   |   |   |   |   |   |    |
|-------|----------|---|---|---|---|---|---|---|---|----|
|       | 1        | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1     |          |   |   |   |   |   |   |   | 1 |    |

After two years (see table 1.2), the insured number 9 demonstrates the worst record and the average claim amounts costs  $4/20 = 0.20$ . The insurer still keeping its premium unchanged.

Table 1.2: Portfolio claims after two year

| Years | Insureds |   |   |   |   |   |   |   |   |    |
|-------|----------|---|---|---|---|---|---|---|---|----|
|       | 1        | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1     |          |   |   |   |   |   |   |   | 1 |    |
| 2     | 1        | 1 |   |   |   |   |   |   | 1 |    |

After ten years, see Table 1.3, the overall claim average is  $23/100 = 0.23$ . The insured 9 had the worse risk with 7 claims last then years where insureds 7, 8, and 10 incurred no claims. The insurer is better able to infer results about the individual insureds level of risk from the portfolio data. Through the example, The collective premium is globally adequate, but clearly not fair. The insured 9 for example deserves to pay a higher premium, while Insureds 7, 8 should pay less.

Such situations can be modelled by means of Credibility premium which can provide answers to the question of how to create equilibrium between individual and portfolio experiences in order to get better possible forecast of the individual premium. That is to say, the idea that stays at the bottom of credibility theory is that each insured should pay a premium in the next year according to his "behavior" in the previous period of time.

Table 1.3: Portfolio claims after ten year

| Years | Insureds |     |     |     |     |     |   |   |     |    |
|-------|----------|-----|-----|-----|-----|-----|---|---|-----|----|
|       | 1        | 2   | 3   | 4   | 5   | 6   | 7 | 8 | 9   | 10 |
| 1     |          |     |     |     |     |     |   |   | 1   |    |
| 2     | 1        | 1   |     |     |     |     |   |   |     | 1  |
| 3     | 1        |     | 1   |     |     |     |   |   |     | 1  |
| 4     |          |     | 1   |     |     |     |   |   |     | 1  |
| 5     |          |     |     |     |     |     |   |   | 1   |    |
| 6     |          | 1   |     |     |     |     |   |   |     |    |
| 7     | 1        | 1   |     | 1   | 1   |     |   |   |     |    |
| 8     | 1        |     |     | 1   |     | 1   |   |   |     | 1  |
| 9     | 1        |     |     |     | 1   |     |   |   |     |    |
| 10    | 1        |     |     |     |     |     |   |   |     | 1  |
|       | 0.6      | 0.3 | 0.2 | 0.2 | 0.2 | 0.1 | 0 | 0 | 0.7 | 0  |
|       | 0.23     |     |     |     |     |     |   |   |     |    |

### 1.3 The first formula for Credibility premium

According to Whitney's proposition in 1918, the risk premium be a weighted average of both the individual risk experience and the classic risk experience. This can be described using the following formula:

Using both of the individual risk experience  $X$  and the class risk experience  $m$ , Whitney in 1918, proposed that the premium rate be a weighted average described with the help of the following formula:

$$P = zX + (1 - z)m$$

$X$  is the mean from the individual's experience and  $m$  is the collective mean of the insurance portfolio. Therefore, the form of equation (2) is called a credibility premium and  $z$  is called the credibility factor and it takes a value between 0 and 1. As  $z$  grows larger, more consideration is given to the individual experience and vice versa.

## 1.4 Credibility Approaches

### The Classical Credibility Models

Also called the limited fluctuations, attempt to limit the effect that random observations will have on the estimates. It is assumed that risks can be classified into homogeneous groups although in insurance there are no homogeneous risks. This can be explained using the next two models:

#### Buhlmann Credibility

The aim of this model, also known as square credibility, is to reduce the square of the error between the estimate and the true expected value of the quantity being estimated.

#### The Bayesian Analysis

To have a better estimate, current observation and prior information can be combined using this model of credibility. It is proven by research that Buhlmann credibility estimates are the best linear square fits to the Bayesian estimates. Regardless of the fact that Buhlmann credibility is also referred to as Bayesian credibility, it should be noted that this latter will not be discussed in this thesis.

### 1.4.1 Buhlmann Model

We consider a portfolio where each line represents an insured. The portfolio is composed of  $I$  insureds each characterized by an unobservable random risk parameter  $\Lambda_i$ .

The data consist of the available observations  $X_{i,j}$  for  $j = 1, 2, \dots, n$  and  $i = 1, 2, \dots, I$ . The  $X_{i,j}$  consist of insured  $i$ 's claims experience such in year  $j$ .

The mathematical assumptions of the Buhlmann model are the following

**Assumptions 1.1.**

1. *The insureds' vectors  $(X_{ij}, \Lambda_i)$  of the accident years  $i$  are independent and the variables  $\Lambda_i$ ,  $i \leq I$ , are identically distributed .*
2. *For  $i = 1, 2, \dots, I$  and  $j, u = 1, \dots, n$*

$$\begin{aligned}\mathbb{E}(X_{ij}|\Lambda_i) &= \rho(\Lambda_i) \\ \text{cov}(X_{ij}, X_{iu}|\Lambda_i) &= \delta_{ju}\nu(\Lambda_i)\end{aligned}$$

While the first assumption (B1) represents the independence between the insureds, the second assumption (B2) reflects the non correlation within the insured's claims experience across the years and the homogeneity in time. We need to define the next additional notations, namely

$$\begin{aligned}\mathbb{E}(X_{ij}) &= \mathbb{E}(\mathbb{E}(X_{ij}|\Lambda_i)) = \mathbb{E}(\rho(\Lambda_i)) = \rho \\ \text{Var}(X_{ij}|\Lambda_i) &= \nu(\Lambda_i) \\ \mathbb{E}(\text{Var}(X_{ij}|\Lambda_i)) &= \mathbb{E}(\nu(\Lambda_i)) = \nu \\ \text{Var}(\mathbb{E}(X_{ij}|\Lambda_i)) &= \text{Var}(\rho(\Lambda_i)) = a\end{aligned}$$

The Buhlmann model restricts the estimator to one of linear form  $\gamma_0^{(i)} + \sum_{i \leq I} \sum_{j \leq n} \gamma_{ij}^{(i)} X_{ij}$  Because of independence between contracts, we know that the credibility premium of the insured  $i$  is a function of its observations only, we can write it in this form:  $\gamma_0^{(i)} + \sum_{j \leq n} \gamma_j^{(i)} X_{ij}$  where

$\gamma_0^{(i)}, \gamma_j^{(i)}, j = 1, \dots, n$  need to be chosen in the sense of minimizing

$$\mathbb{E} \left[ \left( \rho^{(i)}(\Lambda_i) - \left( \gamma_0^{(i)} + \sum_{j \leq n} \gamma_j^{(i)} X_{ij} \right) \right)^2 \right]. \quad (1.1)$$

To minimise 1.1, we set the derivative for it with respect to  $\gamma_0^{(i)}$  equal to zero, we get

$$\gamma_0^{(i)} = \mathbb{E}[\rho(\Lambda_i)] - \sum_j \gamma_j^{(i)} \cdot \mathbb{E}[X_{ij}] = \left( 1 - \sum_j \gamma_j^{(i)} \right) \cdot \rho^{(i)}. \quad (1.2)$$

so that

$$\sum_j \gamma_j^{(i)} = 1 - \frac{\gamma_0^{(i)}}{\rho^{(i)}} \quad (1.3)$$

than with respect to  $\gamma_u^{(i)}, u = 1, \dots, n$  equal to zero, we get

$$\text{cov}(\rho(\Lambda_i), X_{iu}) = \sum_j \gamma_j^{(i)} \text{cov}(X_{ij}, X_{iu}) \quad (1.4)$$

from which we obtain

$$a = \sum_j \gamma_j^{(i)} (a + \delta_{ju} \nu) = a \sum_j \gamma_j^{(i)} + \gamma_u^{(i)} \nu \quad (1.5)$$

Substituting 1.3 into 1.5 and solving for  $\gamma_u^{(i)}, u = 1, \dots, n$

$$\gamma_1^{(i)} = \gamma_2^{(i)} = \dots = \gamma_n^{(i)} = \gamma^{(i)} = \frac{a}{an + \nu} \quad (1.6)$$

Than from 1.3 we get

$$\gamma_0^{(i)} = (1 - n\gamma_u^{(i)})\rho^{(i)} \quad (1.7)$$

so we have

$$\begin{aligned} \gamma_0^{(i)} + \sum_{j \leq n} \gamma_j^{(i)} X_{ij} &= \frac{an}{an + \nu} \sum_j \frac{X_{ij}}{n} + \left( 1 - \frac{an}{an + \nu} \right) \rho^{(i)} \\ &= z \overline{X_{ij}} + (1 - z) \rho^{(i)} \end{aligned}$$

The credibility premium is a weighted average of the collective premium  $\rho^{(i)}$  and the sample mean  $\overline{X_{ij}}$  where

$$z = \frac{an}{an + \nu} = \frac{n}{n + k}$$

$$k = \frac{\nu}{a} = \frac{\mathbb{E}(\text{Var}(X_{ij}|\Lambda_i))}{\text{Var}(\mathbb{E}(X_{ij}|\Lambda_i))}$$

is the Buhlmann credibility factor.

### 1.4.2 Buhlmann-Straub Model

Buhlmann-Straub model was developed in 1970 in connection with determining claims ratios in reinsurance, has over time found a multitude of applications in insurance practice. It is still by far the most used and the most important credibility model for insurance practice. Buhlmann and Straub notice that the total risk premium grows linearly with the size of the insured unit. For example, the premium in fire insurance depends on the insured value of the building and on the insured value of its contents. In collective accident insurance by which all the employees of a firm are insured against accidents, the size of the firm must be taken into account in order to set the premium. To handle the last variations, we consider the following generalization of the Buhlmann model.

we consider a portfolio of  $I$  risks or risk categories, and use the notation

$X_{ij}$  average claim size of the risk  $i$  in year  $j$

$w_{ij}$  associated known constant measuring exposure(weight).

#### Assumptions 1.2.

1. The pairs  $(X_{ij}, \Lambda_i)$  of the accident years  $i$  are independent and the variables  $\Lambda_i, i \leq I$ , are identically distributed .

2. For  $i = 1, 2, \dots, I$  and  $j, u = 1, \dots, n$

$$\begin{aligned}\mathbb{E}(X_{ij}|\Lambda_i) &= \rho(\Lambda_i) \\ \text{cov}(X_{ij}, X_{iu}|\Lambda_i) &= \delta_{ju} \frac{\nu(\Lambda_i)}{w_{ij}}\end{aligned}$$

As in the Buhlmann model,

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X_{it}|\Lambda_i)) &= \mathbb{E}(\rho(\Lambda_i)) = \rho \\ \text{Var}(X_{it}|\Lambda_i) &= \frac{\nu(\Lambda_i)}{w_{ij}} \\ \mathbb{E}(\text{Var}(X_{it}|\Lambda_i)) &= \mathbb{E}(\nu(\Lambda_i)) = \frac{\nu}{w_{ij}} \\ \text{Var}(\mathbb{E}(X_{it}|\Lambda_i)) &= \text{Var}(\rho(\Lambda_i)) = a\end{aligned}$$

and the credibility estimator is also restricted to the linear form  $\gamma_0^{(i)} + \sum_{i \leq I} \sum_{j \leq n} \gamma_j^{(i)} X_{ij}$  and under the independence assumption, it is reduced to  $\gamma_0^{(i)} + \sum_{j \leq n} \gamma_j^{(i)} X_{ij}$ . The aim will be to find  $\gamma_0^{(i)}, \gamma_1^{(i)}, \dots, \gamma_n^{(i)}$  to minimize the expected squared-error

$$\mathbb{E} \left[ \left( \rho^{(i)}(\Lambda_i) - \left( \gamma_0^{(i)} + \sum_{j \leq n} \gamma_j^{(i)} X_{ij} \right) \right)^2 \right]. \quad (1.8)$$

The derivative for 1.8 with respect to  $\gamma_0^{(i)}$  equal to zero gives

$$\sum_u \gamma_u^{(i)} = 1 - \frac{\gamma_0^{(i)}}{\rho^{(i)}} \quad (1.9)$$

than with respect to  $\gamma_u^{(i)}$ ,  $u = 1, \dots, n$  equal to zero, we get

$$\text{cov}(\rho(\Lambda_i), X_{iu}) = \sum_j \gamma_j^{(i)} \text{cov}(X_{ij}, X_{iu}) \quad (1.10)$$

Which could be written

$$a = \sum_j \gamma_j^{(i)} \left( a + \delta_{ju} \frac{\nu}{w_{iu}} \right) = a \sum_j \gamma_j^{(i)} + \gamma_u^{(i)} \frac{\nu}{w_{iu}} \quad (1.11)$$

Substituting 1.9 into 1.11 and solving for  $\gamma_u^{(i)}$ ,  $u = 1, \dots, n$

$$\frac{\gamma_1^{(i)}}{w_{i1}} = \frac{\gamma_1^{(i)}}{w_{i2}} = \dots = \frac{\gamma_1^{(i)}}{w_{in}} = \frac{a}{a \sum_j w_{ij} + \nu} \quad (1.12)$$

$$\begin{aligned} \gamma_j^{(i)} &= \frac{w_{ij}}{\sum_j w_{ij}} \frac{a \sum_j w_{ij}}{a \sum_j w_{ij} + \nu} \\ &= \frac{w_{ij}}{\sum_j w_{ij}} z_i \end{aligned} \quad (1.13)$$

Than from 1.9 we get

$$\begin{aligned} \gamma_0^{(i)} &= \rho^{(i)} - \sum_j \gamma_j^{(i)} \rho^{(i)} \\ &= \rho^{(i)} - \sum_j \frac{w_{ij}}{\sum_j w_{ij}} z_i \rho^{(i)} \\ &= (1 - z_i) \rho^{(i)} \end{aligned} \quad (1.14)$$

so we have

$$\begin{aligned} \gamma_0^{(i)} + \sum_{j \leq n} \gamma_j^{(i)} X_{it} &= \sum_j \frac{w_{ij}}{\sum_j w_{ij}} z_i X_{ij} + (1 - z_i) \rho^{(i)} \\ &= z_i X_{i\Sigma} + (1 - z_i) \rho^{(i)} \end{aligned}$$

The credibility premium is a weighted average of the collective premium  $\rho^{(i)}$  and the sample mean  $\overline{X_{i\Sigma}}$  Where

$$\begin{aligned} X_{i\Sigma} &= \sum_j \frac{w_{ij}}{\sum_j w_{ij}} X_{it} \\ z_j &= \frac{a \sum_j w_{ij}}{a \sum_j w_{ij} + \nu} = \frac{\sum_j w_{ij}}{\sum_j w_{ij} + \frac{\nu}{a}} \end{aligned} \quad \text{is the Buhlmann-Straub credibility factor.}$$

### 1.4.3 Hachemeister's Regression Model

In 1975, Hachemeister in [14], extends the classical credibility of Buhlmann-Straub by introducing a regression technique. His article deals with private passenger bodily injury insurance, observing the claim amounts for a few U.S.A states for a number of quarters (each quarter is a period of three months). These claims show a tendency to increase in time, due to inflation. He was interested in the state-specific inflation factor, supposing that inflation is not the same in all U.S.A states and then applied credibility theory. For other references on regression credibility see [38]. In practical applications there are cases where the unknown regression coefficients take different values for different observations, i.e. each year has its own line. Hence the regression lines must be considered as random and we need to utilize some prior information about the regression coefficients which are randomly varying. For example, let us consider a portfolio of mortgage insurance claim experience, modelled in a credibility framework. Mortgage insurance indemnifies a mortgage lender against loss of default by the borrower. A lot of different events can lead to a claim, depending not only on the credit worthiness of the borrower, but on some exogenous economic factors as well. Such exogenous factors are the disposable income of the borrower and the real estate prices. However real estate prices can vary from one year to another due to the market demand of a real estate or due to increase of interest or inflation rates. As another example we can consider an automobile insurance portfolio. Differences in claim reporting from one year to another may happen due to the economical instability, which can lead to the fluctuation of the inflation and interest rates.

We consider Hachemeister's regression credibility model with  $j = 1, \dots, K$  contracts and  $i = 1, \dots, n$  years of claims experience. The variables describing the  $j^{\text{th}}$  contract are  $(X_j, \Lambda_j)$  in which  $\Lambda_j$  is an unobservable random risk parameter and  $(X_j) = (X_{1j}, \dots, X_{nj})^{\text{tr}}$  is an observable vector of risks, under the following assumptions

**Assumptions 1.3.**

1. The contracts  $(X_j, \Lambda_j)$  for  $j = 1, \dots, K$  are independent and the variables  $\Lambda_j$  are identically distributed.
2. For  $j = 1, 2, \dots, K$ , the conditional expectations of  $X_j$  have the form

$$\mathbb{E}(X_j | \Lambda_j) = Y_j \beta(\Lambda_j). \quad (1.15)$$

Where  $Y_j$  is a fixed design matrix of full rank  $p \leq n$  and the variables  $\beta(\Lambda_j)$  is a unknown regression vector of length  $p$ .

$$\text{cov}(X_j | \Lambda_j) = \sigma^2(\Lambda_j) W_j^{-1}. \quad (1.16)$$

Where  $\sigma^2(\Lambda_j)$  is the variance term and  $W_j$  is a known positive definite matrix of variance weights.

we need to add the next notation

$$s^2 = \mathbb{E}(\sigma^2(\Lambda_j)) \quad A = \text{Cov}(\beta(\Lambda_j)) \quad b = \mathbb{E}(\beta(\Lambda_j)) \quad (1.17)$$

Based on the above assumptions, a credibility estimator for  $\beta(\Lambda_j)$ , referred to the literature as credibility adjusted estimators,  $B_j^{(a)}$ , is obtained as

$$B_j^{(a)} = z_j + (I - z_j)b$$

with

$$z_j = A [A + s^2(Y_j^{tr} W_j Y_j)^{-1}]^{-1} \quad (1.18)$$

$$B_j = (Y_j^{tr} W_j Y_j)^{-1} Y_j^{tr} W_j X_j \quad (1.19)$$

The collective estimator for  $\mathbb{E}(\beta(\Lambda_j))$  is

$$\hat{b} = \left( \sum_{j \leq K} z_j \right)^{-1} \sum_{j \leq K} z_j B_j \quad (1.20)$$

with  $z_j$  as defined in equation (1.18). The estimators for  $s^2$  and  $A$  are given by

$$\hat{s}^2 = \frac{1}{K(n-p)} \sum_{j \leq K} (X_j - Y_j B_j)^{-1} W_j (X_j - Y_j B_j) \quad (1.21)$$

$$\hat{A} = \frac{1}{(K-1)} \sum_{j \leq K} z_j (B_j - \hat{b})(B_j - \hat{b})^{tr} \quad (1.22)$$



# Chapter 2

## Prediction of claim provisions with Hachemeister credibility

### 2.1 Model assumptions

Let's start with some elementary notations: The well-known Kronecker symbol  $\delta$  is defined as

$$\delta_{i,i'} = \begin{cases} 1 & \text{if } i = i' \\ 0 & \text{otherwise} \end{cases}$$

The  $m \times n$ -matrix whose elements are all 1 is written as  $\mathbb{1}_{m \times n}$ .

For a  $m_1 \times n_1$ -matrix  $X$  and a  $m_2 \times n_2$ -matrix  $Y$  we denote by  $X \otimes Y$  the tensor (or Kronecker) product  $(X_{i_1,j_1} Y_{i_2,j_2})_{(i_1,i_2),(j_1,j_2)}$  as a  $m_1 m_2 \times n_1 n_2$ -matrix. If  $(X^{(\kappa)})_{1 \leq \kappa \leq \ell}$  is a sequence of  $m_\kappa \times n_\kappa$ -matrices  $X^{(\kappa)}$ ,  $\kappa \leq \ell$ , then  $\text{dia}[X^{(\kappa)}]_{\kappa \leq \ell}$  is the  $\sum_{\kappa \leq \ell} m_\kappa \times \sum_{\kappa \leq \ell} n_\kappa$ -matrix whose  $(\kappa, i), (\kappa', j)$ -element is  $\delta_{\kappa,\kappa'} X_{i,j}^{(\kappa)}$  where  $i \leq m_\kappa$  and  $j \leq n_\kappa$ .

We consider a standard insurance run-off situation with the numbers  $I$  of accident years and  $J$  development years,  $J \leq I$ . In addition, for each accident year  $i \in I$ , the portfolio of active contracts is enumerated by  $\kappa \in \{1, \dots, k_i\}$ ,  $k_i \geq 1$ . According to the model settings,  $\kappa$  may also be an enumeration of incurred accidents in year  $i$ . The complete index set of accident years,

contracts, and developing years is

$$\widetilde{\mathcal{I}} := \{(i, \kappa, j) \mid 1 \leq i \leq I, 1 \leq \kappa \leq k_i, j = 1 \leq j \leq J\}. \quad (2.1)$$

The index set of observed contracts and years is

$$\mathcal{I} := \{(i, \kappa, j) \in \widetilde{\mathcal{I}} \mid i + j \leq I + 1\}. \quad (2.2)$$

The set of observed contracts and development years (resp. of development years only) for the accident year  $i \leq I$  is

$$\mathcal{I}_i := \{(\kappa, j) \mid (i, \kappa, j) \in \mathcal{I}\} \quad (2.3)$$

$$\mathcal{J}_i := \{j \mid (i, 1, j) \in \mathcal{I}\}. \quad (2.4)$$

Since for given accident year  $i$  the index set for  $\kappa$  is kept fixed, it is always understood without further notice that  $\kappa$  runs from 1 to  $k_i$ .

By  $\widetilde{D} = (D_{i,\kappa,j})_{(i,\kappa,j) \in \widetilde{\mathcal{I}}}$ , we denote the matrix of the incremental payments of claims from contract  $\kappa$  with accident year  $i$  and development year  $j$ . For fixed accident year  $i$ , the random matrix of incremental payments is  $D_i = (D_{i,\kappa,j})_{(\kappa,j) \in \mathcal{I}_i}$ . If in addition the contract  $\kappa$  is fixed, we get the random vector  $\widetilde{D}_{i,\kappa} = (D_{i,\kappa,j})_{j \leq J}$ . The restriction of  $\widetilde{D}$  to the index set  $\mathcal{I}$  of observed contracts and years is denoted by  $D = (D_{i,\kappa,j})_{(i,\kappa,j) \in \mathcal{I}}$ , while  $D_i = (D_{i,\kappa,j})_{(\kappa,j) \in \mathcal{I}_i}$  and  $D_{i,\kappa} = (D_{i,\kappa,j})_{j \in \mathcal{J}_i}$  are the analogous notations. The observed data of passed years are given by  $\widehat{D} = (\widehat{D}_{i,\kappa,j})_{(i,\kappa,j) \in \mathcal{I}}$ , those of accident year  $i$  by  $\widehat{D}_i = (\widehat{D}_{i,\kappa,j})_{(\kappa,j) \in \mathcal{I}_i}$ , and those of contract  $\kappa$  by  $\widehat{D}_{i,\kappa} = (\widehat{D}_{i,\kappa,j})_{j \in \mathcal{J}_i}$ .

According to the standard setting in hierarchical credibility theory, the active contracts of accident year  $i$  depend on an unobservable structure variable  $\Lambda_i$ , while the claim development of contract  $\kappa$  depends on the unobservable structure variable  $\Lambda_{i,\kappa}$ , too ( $\kappa \leq k_i$  and  $i \leq I$ ). This leads for our model to the following assumptions:

**Assumptions 2.1.**

(H1) The variables  $\Lambda_i$ ,  $i \leq I$ , are identically distributed .

(H2) The families of variables  $\left( \Lambda_i, (\Lambda_{i,\kappa})_{\kappa}, \tilde{D}_i \right)$  of the accident years  $i$  are uncorrelated.

(H3) For  $i \leq I$  and  $\Lambda_i$  fixed, the variables  $\Lambda_{i,\kappa}$ ,  $\kappa \leq k_i$ , are conditionally identically distributed.

(H4) For  $i \leq I$  and  $\Lambda_i$  fixed, the families  $(\Lambda_{i,\kappa}, \tilde{D}_{i,\kappa})$  of the active contracts  $\kappa$  in accident year  $i$  are conditionally uncorrelated.

(H5) For  $i \leq I$ ,  $\kappa \leq k_i$  and  $(\Lambda_i, \Lambda_{i,\kappa})$  fixed, the conditional expectations of the reduced random vectors  $D_{i,\kappa}$  have the form

$$\mathbb{E}(D_{i,\kappa} | \Lambda_i, \Lambda_{i,\kappa}) = S^{(i,\kappa)} \cdot \rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}). \quad (2.5)$$

Here, the reduced design matrices  $S^{(i,\kappa)} := (S_{j,q}^{(i,\kappa)})_{j \in \mathcal{J}_i, q=1, \dots, r_i}$  for  $D_{i,\kappa}$  are supposed to be known (for the moment) and to have full rank  $r_i \leq |\mathcal{J}_i|$ . The random vectors  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) := (\rho_q^{(i)}(\Lambda_i, \Lambda_{i,\kappa}))_{q=1, \dots, r_i}$  is a regression vector of dimension  $r_i$ . Moreover, the covariance matrices of  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa})$  and of  $\rho^{(i)}(\Lambda_i) := \mathbb{E} [\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) | \Lambda_i]$ , given as

$$\begin{aligned} B^{(i)} &:= \mathbb{E} [\text{cov} (\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) \otimes \rho^{(i)\text{tr}}(\Lambda_i, \Lambda_{i,\kappa}) | \Lambda_i)], \quad \text{and,} \\ C^{(i)} &:= \text{cov} (\rho^{(i)}(\Lambda_i) \otimes \rho^{(i)\text{tr}}(\Lambda_i)) \end{aligned} \quad (2.6)$$

are supposed to be positive definite (see remark 2.1 below)).

(H6) The conditional variance matrices of  $D_{i,\kappa}$  have the form

$$\text{cov}(D_{i,\kappa} \otimes D_{i,\kappa}^{\text{tr}} | \Lambda_i, \Lambda_{i,\kappa}) = a(\Lambda_i, \Lambda_{i,\kappa}) T^{(i,\kappa)}, \quad (2.7)$$

where  $a(\Lambda_i, \Lambda_{i,\kappa})$  is the scalar variance term with  $a^{(i)} := \mathbb{E}[a(\Lambda_i, \Lambda_{i,\kappa})] > 0$  and  $T^{(i,\kappa)} := (T_{j,j'}^{(i,\kappa)})_{j,j' \in \mathcal{J}_i}$  is a known positive definite matrix of variance weights.

**Remark 2.1.** *In full generality the dimension of the random regression vector  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa})$  may also depend on  $\kappa$ . In this case also  $B^{(i)}$  and  $a^{(i)}$  depend on  $\kappa$ . However, in most applications the matrices  $S^{(i,\kappa)}$  and  $T^{(i,\kappa)}$  will be independent of  $\kappa$ ; so will be  $B^{(i)}$  and  $a^{(i)}$ .*

We still need some more notations:

$$\begin{aligned}
 R^{(i)} &:= \mathbb{E}[\rho^{(i)}(\Lambda_i)] \\
 U^{(i,\kappa)} &:= a^{(i)} \cdot T^{(i,\kappa)} + S^{(i,\kappa)} \cdot B^{(i)} \cdot S^{(i,\kappa)tr} \\
 S^{(i)} &:= \text{dia}[S^{(i,\kappa)}]_{\kappa} \\
 T^{(i)} &:= \text{dia}[T^{(i,\kappa)}]_{\kappa} \\
 U^{(i)} &:= \text{dia}[U^{(i,\kappa)}]_{\kappa}
 \end{aligned} \tag{2.8}$$

The following result is a first consequence of the assumptions 2.1.

**Lemma 2.1.** *As covariance matrices we find*

$$\text{cov}(D_i \otimes D_i^{tr}) = U^{(i)} + S^{(i)} \cdot \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \cdot S^{(i)tr} =: V^{(i)}. \tag{2.9}$$

Similarly, we have

$$\mathbb{E}[\text{cov}(D_i \otimes D_i^{tr} | \Lambda_i)] = U^{(i)} \tag{2.10}$$

The matrices  $V^{(i)}$  and  $U^{(i)}$  are positive definite

*Proof.* We start with 2.10. By assumptions (H5) and (H6) we have

$$\begin{aligned}
 &\mathbb{E}[\text{cov}(D_i \otimes D_i^{tr} | \Lambda_i)] \\
 &= \mathbb{E}[\text{cov}(D_i \otimes D_i^{tr} | \Lambda_i, (\Lambda_{i,\kappa})_{\kappa})] + \mathbb{E}\left[\text{cov}\left(\mathbb{E}[D_i | \Lambda_i, (\Lambda_{i,\kappa})_{\kappa}] \otimes \mathbb{E}[D_i^{tr} | \Lambda_i, (\Lambda_{i,\kappa})_{\kappa}] \middle| \Lambda_i\right)\right] \\
 &= \text{dia}[\mathbb{E}[a(\Lambda_i, \Lambda_{i,\kappa})] \cdot T^{(i,\kappa)}]_{\kappa} + S^{(i)} \cdot \mathbb{E}\left[\text{dia}[\text{cov}(\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) \otimes \rho^{(i)tr}(\Lambda_i, \Lambda_{i,\kappa}) | \Lambda_i)]_{\kappa}\right] \cdot S^{(i)tr} \\
 &= \text{dia}[U^{(i,\kappa)}]_{\kappa} = U^{(i)}.
 \end{aligned}$$

For the unconditional covariance matrix we find

$$\begin{aligned}
 \text{cov} (D_i \otimes D_i^{\text{tr}}) &= \mathbb{E} [\text{cov} (D_i \otimes D_i^{\text{tr}} | \Lambda_i)] + \text{cov} (\mathbb{E} [D_i | \Lambda_i] \otimes \mathbb{E} [D_i^{\text{tr}} | \Lambda_i]) \\
 &= U^{(i)} + S^{(i)} \cdot \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \cdot S^{(i)\text{tr}},
 \end{aligned}$$

which shows (2.9). □

## 2.2 The credibility estimators for the development quotas

The main result of the credibility estimator is the following:

### Theorem 2.1.

We assume (H1) to (H6).

(i) Based on the observations  $\widehat{D}$ , the credibility estimator for  $\rho^{(i)}(\Lambda_i)$  is given by

$$\bar{\rho}^{(i)} = (Id_{r_i} - \Delta^{(i)}) R^{(i)} + \Delta^{(i)} \cdot \Theta^{(i)} \cdot \widehat{D}_i, \quad (2.11)$$

Here,  $Id_{r_i}$  is the identity matrix of order  $r$ . Further, on the level of accident year  $i$ , the credibility matrix  $\Delta^{(i)}$  and the transition matrix  $\Theta^{(i)}$  are defined by

$$\begin{aligned}
 \Delta^{(i)} &:= [Id_{r_i} + C^{(i)} W^{(i)}]^{-1} C^{(i)} W^{(i)} \quad \text{and} \\
 \Theta^{(i)} &:= W^{(i)-1} \otimes \mathbb{1}_{1 \times k_i} \cdot S^{(i)\text{tr}} \cdot U^{(i)-1}.
 \end{aligned} \quad (2.12)$$

where we set

$$W^{(i)} := \sum_{\lambda} S^{(i,\lambda)\text{tr}} U^{(i,\lambda)-1} S^{(i,\lambda)} \quad (2.13)$$

Using the notation given in lemma 2.1 we have

$$\Delta^{(i)} < Id_{r_i}. \quad (2.14)$$

## 2.2. The credibility estimators for the development quotas

(ii) The credibility estimator for  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa})$ , based on  $\widehat{D}$ , is given by

$$\bar{\rho}^{(i,\kappa)} = (Id_{r_i} - \Delta^{(i,\kappa)})\bar{\rho}^{(i)} + \Delta^{(i,\kappa)} \cdot \Theta^{(i,\kappa)} \cdot \widehat{D}_{i,\kappa}, \quad (2.15)$$

where on the level of accident year  $i$  and contract  $\kappa$ , the credibility matrix  $\Delta^{(i,\kappa)}$  and the transition matrix  $\Theta^{(i,\kappa)}$  are defined by

$$\begin{aligned} \Delta^{(i,\kappa)} &:= B^{(i)} \cdot S^{(i,\kappa)tr} U^{(i,\kappa)-1} S^{(i,\kappa)} && \text{and} \\ \Theta^{(i,\kappa)} &:= \left( S^{(i,\kappa)tr} U^{(i,\kappa)-1} S^{(i,\kappa)} \right)^{-1} S^{(i,\kappa)tr} U^{(i,\kappa)-1}. \end{aligned} \quad (2.16)$$

Again, we have

$$\Delta^{(i,\kappa)} < Id_{r_i}. \quad (2.17)$$

(iii) The prediction errors of  $\bar{\rho}^{(i)}$ , resp. of  $\bar{\rho}^{(i,\kappa)}$ , is

$$\mathbb{E} \left[ \|\rho(\Lambda_i) - \bar{\rho}^{(i)}\|^2 \right] = \text{trace} \left[ C^{(i)-1} + W^{(i)} \right]^{-1} \quad \text{resp.} \quad (2.18)$$

$$\begin{aligned} \mathbb{E} \left[ \|\rho(\Lambda_i, \Lambda_{i,\kappa}) - \bar{\rho}^{(i,\kappa)}\|^2 \right] &= \text{trace} \left( (\mathbf{Id}_{r_i} - \Delta^{(i,\kappa)}) (\mathbf{Id}_{r_i} - \Delta^{(i)}) C^{(i)} (\mathbf{Id}_{r_i} - \Delta^{(i,\kappa)tr}) \right. \\ &\quad \left. + (Id_{r_i} - \Delta^{(i,\kappa)}) B^{(i)} \right). \end{aligned} \quad (2.19)$$

*Proof.*

(i) The credibility estimator 2.11 for  $\rho^{(i)}(\Lambda_i)$  is of the form

$$\bar{\rho}^{(i)} := \gamma_0^{(i)} + \sum_{(m,\ell,n) \in \mathcal{I}} \gamma_{m,\ell,n}^{(i)} D_{m,\ell,n} = \gamma_0^{(i)} + \sum_{m \leq I} \gamma_m^{(i)} \cdot D_m$$

with the vectors  $\gamma_0^{(i)}, \gamma_{m,\ell,n}^{(i)} \in \mathbb{R}^{r_i}$  for  $(m, \ell, n) \in \mathcal{I}$ , or for  $m \leq I$ , the matrices  $\gamma_m^{(i)} = (\gamma_{m,\ell,n}^{(i)})_{(\ell,n) \in \mathcal{I}_m}$ .

The  $\gamma^{(i)}$ 's are determined by minimizing the mean square problem:

$$\begin{aligned} &\mathbb{E} \left[ \left\| \rho^{(i)}(\Lambda_i) - \gamma_0^{(i)} - \sum_m \gamma_m^{(i)} \cdot D_m \right\|^2 \right] \\ &= \text{trace} \mathbb{E} \left[ \left( \rho^{(i)}(\Lambda_i) - \gamma_0^{(i)} - \sum_m \gamma_m^{(i)} \cdot D_m \right) \otimes \left( \rho^{(i)}(\Lambda_i) - \gamma_0^{(i)} - \sum_m \gamma_m^{(i)} \cdot D_m \right)^{tr} \right]. \end{aligned} \quad (2.20)$$

## 2.2. The credibility estimators for the development quotas

Setting the gradient of (2.20) with respect to  $\gamma_0^{(i)}$  equal to 0 and using the notations from (2.6) and (2.8), we get

$$\gamma_0^{(i)} = \mathbb{E}[\rho^{(i)}(\Lambda_i)] - \sum_m \gamma_m^{(i)} \cdot \mathbb{E}[D_m] = \left( Id_{r_i} - \sum_m \gamma_m^{(i)} \cdot S^{(m)} \cdot (\mathbb{1}_{K_m \times 1} \otimes Id_{r_i}) \right) \cdot R^{(i)}. \quad (2.21)$$

We replace  $\gamma_0^{(i)}$  in (2.20) by the right-hand side of (2.21) and then set the gradient of (2.20) with respect to  $\gamma_m^{(i)}$  equal to zero. We obtain

$$0 = \mathbb{E} \left[ \left( \left( \rho^{(i)}(\Lambda_i) - R^{(i)} \right) - \sum_{\mu} \gamma_{\mu}^{(i)} (D_{\mu} - S^{(\mu)} \cdot (\mathbb{1}_{K_{\mu} \times 1} \otimes R^{(i)})) \right) \otimes (D_m - S^{(m)} \cdot (\mathbb{1}_{K_m \times 1} \otimes R^{(i)}))^{\text{tr}} \right]. \quad (2.22)$$

For  $m \neq i$  and using lemma 2.1, this implies

$$\gamma_m^{(i)} \cdot V^{(m)} = 0, \quad (2.23)$$

and therefore  $\gamma_m^{(i)} = 0$ . For  $m = i$ , we get

$$(C^{(i)} \otimes \mathbb{1}_{1 \times k_i}) \cdot S^{(i)tr} - \gamma_i^{(i)} V^{(i)} = 0 \quad \text{or} \quad \gamma_i^{(i)} = (C^{(i)} \otimes \mathbb{1}_{1 \times k_i}) \cdot S^{(i)tr} \cdot V^{(i)-1}. \quad (2.24)$$

Now, the credibility estimator for  $\rho^{(i)}(\Lambda_i)$  is

$$\begin{aligned} \bar{\rho}^{(i)} &= \gamma_0^{(i)} + \gamma_i^{(i)} \cdot \widehat{D}_i \\ &= \left( Id_{r_i} - (C^{(i)} \otimes \mathbb{1}_{1 \times k_i}) \cdot S^{(i)tr} \cdot V^{(i)-1} \cdot S^{(i)} \cdot (\mathbb{1}_{k_i \times 1} \otimes Id_{r_i}) \right) \cdot R^{(i)} \\ &\quad + (C^{(i)} \otimes \mathbb{1}_{1 \times k_i}) \cdot S^{(i)tr} \cdot V^{(i)-1} \cdot \widehat{D}_i. \end{aligned} \quad (2.25)$$

With the identities A.6 and A.7 in the appendix A.1, the representation 2.11 follows. The relation 2.14 is a consequence of the fact that

$$Id_{r_i} - \Delta^{(i)} = [Id_{r_i} + C^{(i)} \cdot W^{(i)}]^{-1} > 0.$$

(ii) Before we directly attack the problem finding the linear non-homogeneous credibility estimator for  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa})$  based on  $(D_i)_{i \in I}$ , we look for the extended estimator of  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa})$  with

## 2.2. The credibility estimators for the development quotas

$(\rho^{(i)}(\Lambda_i), D_i)_{i \in I}$  as its extended basis, i.e. we regard the estimator of the form

$$\bar{\rho}^{(i,\kappa)} = \gamma_0^{(i,\kappa)} + \sum_{m \leq I} \gamma_{1,m}^{(i,\kappa)} \rho^{(i)}(\Lambda_m) + \sum_{m \leq I} \gamma_{2,m}^{(i,\kappa)} D_m$$

with vectors  $\gamma_0^{(i,\kappa)} \in \mathbb{R}^{r_i}$ , and sequences of matrices  $\gamma_1^{(i,\kappa)} := \left( \gamma_{1,m}^{(i,\kappa)} = (\gamma_{1,m,q,p}^{(i,\kappa)})_{q,p \leq r_i} \right)_{m \leq I}$  and

$\gamma_2^{(i,\kappa)} := \left( \gamma_{2,m}^{(i,\kappa)} = (\gamma_{2,m,q,(\ell,n)}^{(i,\kappa)})_{q \leq r_i, (\ell,n) \in \mathcal{A}_m} \right)_{m \leq I}$ . They will be determined by the least square problem:

$$\inf_{\gamma_0^{(i,\kappa)}, \gamma_1^{(i,\kappa)}, \gamma_2^{(i,\kappa)}} \mathbb{E} \left[ \left\| \rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - \bar{\rho}^{(i,\kappa)} \right\|^2 \right]. \quad (2.26)$$

Similarly to 2.21 we get

$$\gamma_0^{(i,\kappa)} = \left[ \left( Id_{r_i} - \sum_m \gamma_{1,m}^{(i,\kappa)} \right) - \sum_m \gamma_{2,m}^{(i,\kappa)} \cdot S^{(m)} \cdot (\mathbf{1}_{K_m \times 1} \otimes Id_{r_i}) \right] \cdot R^{(i)}. \quad (2.27)$$

Setting then the gradient with respect to  $\gamma_{1,m}^{(i,\kappa)}$  equal to zero, we get

$$\begin{aligned} 0 &= \text{cov}(\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) \otimes \rho^{(i)\text{tr}}(\Lambda_m)) - \sum_{\mu} \gamma_{1,\mu}^{(i,\kappa)} \text{cov}(\rho^{(i)}(\Lambda_{\mu}) \otimes \rho^{(i)\text{tr}}(\Lambda_m)) \\ &\quad - \sum_{\mu} \gamma_{2,\mu}^{(i,\kappa)} \text{cov}(D_{\mu} \otimes \rho^{(i)\text{tr}}(\Lambda_m)). \end{aligned} \quad (2.28)$$

For  $m \neq i$ , we get

$$\begin{aligned} 0 &= \gamma_{1,m}^{(i,\kappa)} \cdot C^{(i)} + \gamma_{2,m}^{(i,\kappa)} \cdot S^{(m)} (\mathbf{1}_{k_i \times 1} \otimes C^{(i)}), \quad \text{or equivalently} \\ 0 &= \gamma_{1,m}^{(i,\kappa)} + \gamma_{2,m}^{(i,\kappa)} \cdot S^{(m)} (\mathbf{1}_{k_i \times 1} \otimes Id_{r_i}). \end{aligned} \quad (2.29)$$

while for  $m = i$  it holds that

$$\begin{aligned} 0 &= C^{(i)} - \gamma_{1,i}^{(i,\kappa)} \cdot C^{(i)} - \gamma_{2,i}^{(i,\kappa)} \cdot S^{(i)} (\mathbf{1}_{k_i \times 1} \otimes C^{(i)}), \quad \text{or equivalently} \\ 0 &= (Id_{r_i} - \gamma_{1,i}^{(i,\kappa)}) - \gamma_{2,i}^{(i,\kappa)} \cdot S^{(i)} (\mathbf{1}_{k_i \times 1} \otimes Id_{r_i}). \end{aligned} \quad (2.30)$$

It follows from 2.27, (2.29), and (2.30) that  $\gamma_0^{(i,\kappa)} = 0$ .

Next, we set the derivation of (2.26) w. r. to  $\gamma_{2,m}^{(i,\kappa)}$  equal to zero to find

$$\begin{aligned} 0 &= \text{cov}(\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) \otimes D_m^{\text{tr}}) - \sum_{\mu} \gamma_{1,\mu}^{(i,\kappa)} \text{cov}(\rho^{(i)}(\Lambda_{\mu}) \otimes D_m^{\text{tr}}) \\ &\quad - \sum_{\mu} \gamma_{2,\mu}^{(i,\kappa)} \text{cov}(D_{\mu} \otimes D_m^{\text{tr}}). \end{aligned} \quad (2.31)$$

## 2.2. The credibility estimators for the development quotas

Again for  $m \neq i$ , this gives

$$\begin{aligned} 0 &= \gamma_{1,m}^{(i,\kappa)} (C^{(i)} \otimes \mathbb{I}_{k_i \times 1}) \cdot S^{(m)tr} + \gamma_{2,m}^{(i,\kappa)} \cdot V^{(m)} && \text{or equivalently} \\ \gamma_{2,m}^{(i,\kappa)} &= -\gamma_{1,m}^{(i,\kappa)} (C^{(i)} \otimes \mathbb{I}_{k_i \times 1}) \cdot S^{(m)tr} \cdot V^{(m)^{-1}}. \end{aligned}$$

and together with (2.29), (A.6), and (A.11)

$$\begin{aligned} 0 &= \gamma_{1,m}^{(i,\kappa)} \left[ Id_{r_i} - (C^{(i)} \otimes \mathbb{I}_{1 \times k_i}) \cdot S^{(m)tr} \cdot V^{(m)^{-1}} \cdot S^{(m)} (\mathbb{I}_{k_i \times 1} \otimes Id_{r_i}) \right] \\ &= \gamma_{1,m}^{(i,\kappa)} [Id_{r_i} - \Delta^{(m)}]. \end{aligned}$$

From (2.14), it follows for  $m \neq i$  that  $\gamma_{1,m}^{(i,\kappa)} = 0$  and therefore also  $\gamma_{2,m}^{(i,\kappa)} = 0$ .

In addition to (2.30) we find for  $m = i$  that

$$\begin{aligned} 0 &= \text{cov}(\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) \otimes D_i^{\text{tr}}) - \gamma_{1,i}^{(i,\kappa)} \text{cov}(\rho^{(i)}(\Lambda_i) \otimes D_i^{\text{tr}}) - \gamma_{2,i}^{(i,\kappa)} \cdot \text{cov}(D_i \otimes D_i^{\text{tr}}) \\ &= Id_{r_i} \otimes \mathbb{I}_{1 \times k_i} \cdot \text{dia}[\delta_{\kappa,\ell} B^{(i)} + C^{(i)}]_{\ell} \cdot S^{(i)tr} \\ &\quad - \gamma_{1,i}^{(i,\kappa)} \cdot (C^{(i)} \otimes \mathbb{I}_{1 \times k_i}) \cdot S^{(i)tr} - \gamma_{2,i}^{(i,\kappa)} \cdot V^{(i)}. \end{aligned} \tag{2.32}$$

Combining (2.30) and (2.32) gives

$$\begin{aligned} 0 &= Id_{r_i} \otimes \mathbb{I}_{1 \times k_i} \cdot \text{dia}[\delta_{\kappa,\ell} B^{(i)} + C^{(i)}]_{\ell} \cdot S^{(i)tr} \\ &\quad - (C^{(i)} \otimes \mathbb{I}_{1 \times k_i}) \cdot S^{(i)tr} - \gamma_{2,i}^{(i,\kappa)} \cdot (V^{(i)} - S^{(i)} (C^{(i)} \otimes \mathbb{I}_{k_i \times k_i}) \cdot S^{(i)tr}) \\ &= Id_{r_i} \otimes \mathbb{I}_{1 \times k_i} \cdot \text{dia}[\delta_{\kappa,\ell} B^{(i)}]_{\ell} \cdot S^{(i)tr} - \gamma_{2,i}^{(i,\kappa)} \cdot U^{(i)} \\ &= Id_{r_i} \otimes \mathbb{I}_{1 \times k_i} \cdot \text{dia}[\delta_{\kappa,\ell} B^{(i)} \cdot S^{(i,\ell)tr}]_{\ell} - \gamma_{2,i}^{(i,\kappa)} \cdot U^{(i)}. \end{aligned} \tag{2.33}$$

Hence

$$\begin{aligned} \gamma_{2,i,\ell}^{(i,\kappa)} &= \delta_{\kappa,\ell} B^{(i)} \cdot S^{(i,\kappa)tr} \cdot U^{(i,\kappa)^{-1}} && \text{and} \\ \gamma_{1,i}^{(i,\kappa)} &= Id_{r_i} - B^{(i)} \cdot S^{(i,\kappa)tr} \cdot U^{(i,\kappa)^{-1}} \cdot S^{(i,\kappa)}. \end{aligned} \tag{2.34}$$

The credibility estimator for  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa})$  based on  $(\rho^{(i)}(\Lambda_i), D_i)_{i \in I}$  is

$$\begin{aligned} \bar{\rho}^{(i,\kappa)} &= \left( Id_{r_i} - B^{(i)} S^{(i,\kappa)tr} U^{(i,\kappa)^{-1}} S^{(i,\kappa)} \right) \rho^{(i)}(\Lambda_i) + B^{(i)} S^{(i,\kappa)tr} U^{(i,\kappa)^{-1}} S^{(i,\kappa)} \cdot D_{i,\kappa} \\ &= (Id_{r_i} - \Delta^{(i,\kappa)}) \rho^{(i)}(\Lambda_i) + \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa}. \end{aligned} \tag{2.35}$$

## 2.2. The credibility estimators for the development quotas

---

where we used the definitions in 2.16. The proof of  $\Delta^{(i,\kappa)} < Id_{r_i}$  is given in appendix A.1.1.

Finally, the fact that the composition of the extended estimator  $\bar{\bar{\rho}}^{(i,\kappa)}$  with the estimator  $\bar{\rho}^{(i)}$  yields indeed the estimator of  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa})$  as inhomogeneous linear estimator on  $(D_m)_{m \in I}$  follows from the general lemma 3.3 in [27]. For the convenience of the reader we give a proof adapted to our situation in appendix A.1.2.

(iii) The calculations of the prediction errors are given in appendix A.1.3. □

## 2.3 Estimators for the structural parameters

In order to calculate the credibility estimators from Theorem 2.1, we need to determine the quantities  $R$ ,  $a^{(i)}$ ,  $B^{(i)}$ , and  $C^{(i)}$  for  $i \leq I$ . We recall that and the matrices  $T^{(i,\kappa)}$  of variance weights are supposed to be known for the moment. Estimations of the reduced design matrices  $S^{(i,\kappa)}$  is statistically determined in the next chapter. Only contracts with observation are relevant and so all  $K_i$  are now taken to be greater than 0.

### Theorem 2.2.

(i) For the parameter  $R$  and  $R$  the following estimator is unbiased:

$$\widehat{R} = \left( \sum_{i=1}^I \Delta^{(i)} \right)^{-1} \sum_{i=1}^I \Delta^{(i)} \Theta^{(i)} D_i. \quad (2.36)$$

(ii)

$$\widehat{a}^{(i)} = \frac{1}{I (|\mathcal{J}_i| - r_i)} \sum_{i=1}^I (D_{i,\kappa} - S^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa}) U^{(i,\kappa)^{-1}} (D_{i,\kappa} - S^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa})^{tr}. \quad (2.37)$$

*Proof.*

(i) For the collective mean  $R$  :

$$\begin{aligned} \mathbb{E} [\mathbb{E} [\Theta^{(i,\kappa)} D_{i,\kappa} | \Lambda_i]] &= \mathbb{E} \left[ \mathbb{E} \left[ \left( S^{(i,\kappa)tr} U^{(i,\kappa)^{-1}} S^{(i,\kappa)} \right)^{-1} S^{(i,\kappa)tr} U^{(i,\kappa)^{-1}} D_{i,\kappa} | \Lambda_i \right] \right] \\ &= \left( S^{(i,\kappa)tr} U^{(i,\kappa)^{-1}} S^{(i,\kappa)} \right)^{-1} S^{(i,\kappa)tr} U^{(i,\kappa)^{-1}} \mathbb{E} [\mathbb{E} [D_{i,\kappa} | \Lambda_i]] \\ &= R \\ \mathbb{E} [\widehat{R}] &= \left( \sum_{i=1}^I \Delta^{(i)} \right)^{-1} \sum_{i=1}^I \Delta^{(i)} R = R \end{aligned}$$

(ii) For  $a^{(i)}$

### Lemma 2.2.

For the symmetrical  $(|\mathcal{J}_i| \times |\mathcal{J}_i|)$  matrix  $\zeta$  defined by

$$\zeta = U^{(i,\kappa)^{-1}} - U^{(i,\kappa)^{-1}} S^{(i,\kappa)} \Theta^{(i,\kappa)} \quad (2.38)$$

we have

$$\mathbb{E} [D_{i,\kappa} \zeta D_{i,\kappa}^{\text{tr}} / \Lambda_{i,\kappa}] = (|\mathcal{J}_i| - r_i) a(\Lambda_i, \Lambda_{i,\kappa})$$

*Proof.*

$$\begin{aligned} \mathbb{E} [D_{i,\kappa} \zeta D_{i,\kappa}^{\text{tr}} / \Lambda_i, \Lambda_{i,\kappa}] &= \sum_{j,j'} \zeta_{j,j'} \mathbb{E} [D_{i,\kappa,j} D_{i,\kappa,j'} / \Lambda_i, \Lambda_{i,\kappa}] \\ &= \sum_{j,j'} \zeta_{j,j'} \text{COV} (D_{i,\kappa,j}, D_{i,\kappa,j'} / \Lambda_i, \Lambda_{i,\kappa}) + \sum_{j,j'} \zeta_{j,j'} \mathbb{E} [D_{i,\kappa,j} / \Lambda_i, \Lambda_{i,\kappa}] \mathbb{E} [D_{i,\kappa,j'} / \Lambda_i, \Lambda_{i,\kappa}] \\ &= a(\Lambda_i, \Lambda_{i,\kappa}) \sum_{j,j'} \zeta U_{jj'}^{(i,\kappa)} + (S^{(i,\kappa)} \rho^{(i,\kappa)}(\Lambda_i, \Lambda_{i,\kappa}))^{\text{tr}} \zeta S^{(i,\kappa)} \rho^{(i,\kappa)}(\Lambda_i, \Lambda_{i,\kappa}) \end{aligned}$$

Since

$$\zeta S^{(i)} = 0$$

and

$$\sum_{j,j' \in \mathcal{J}_i} \zeta U_{jj'}^{(i,\kappa)} = |\mathcal{J}_i| - r_i.$$

We find

$$\mathbb{E} [D_{i,\kappa} \zeta D_{i,\kappa}^{\text{tr}} / \Lambda_i, \Lambda_{i,\kappa}] = (|\mathcal{J}_i| - r_i) a(\Lambda_i, \Lambda_{i,\kappa})$$

□

Follow from the fact that

$$E [a^i(\Lambda_i, \Lambda_{i,\kappa})] = a^i \quad (2.39)$$

After obvious simplification, we find

$$\widehat{a}^{(i)} = \frac{1}{I (|\mathcal{J}_i| - r_i)} \sum_{i=1}^I (D_{i,\kappa} - S^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa}) U^{(i,\kappa)-1} (D_{i,\kappa} - S^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa})^{\text{tr}}.$$

(iii) For  $B^{(i)}$  and  $C^{(i)}$

When constructing estimators of  $B^{(i)}$  and  $C^{(i)}$ , noting first for the hierarchy approach  $\beta_1 = C^{(i)}$  for the level  $i$  and  $\beta_2 = B^{(i)}$  for the level  $i, k$ . we even make use of the following result

**Lemma 2.3.** *For the transition matrix defined by (2.12) and (2.16) the following relations hold true*

$$\text{cov} \left( (\Theta^{(i,k)} D_{i,k}) \otimes (\Theta^{(i,k)} D_{i,k})^{\text{tr}} / \Lambda_i, \Lambda_{i,k} \right) = \beta_2 + a \left( S^{(i,k)\text{tr}} U^{(i,k)^{-1}} S^{(i,k)} \right)^{-1} \quad (2.40)$$

$$= \Delta^{(i,k)^{-1}} \beta_2 \quad (2.41)$$

$$\text{cov} \left( (\Theta^{(i)} D_i) \otimes (\Theta^{(i)} D_i)^{\text{tr}} / \Lambda_i \right) = \beta_2 + \sum_{i,\kappa} \Delta^{(i,\kappa)^{-1}} \beta_1 \quad (2.42)$$

*Proof.*

By the assumption (H5) and (H6) we note that

$$\begin{aligned} & \text{cov} \left( D_{i,\kappa} \otimes D_{i,\kappa}^{\text{tr}} \mid \Lambda_i \right) \\ &= \mathbb{E} \left[ \text{cov} \left( D_{i,\kappa} \otimes D_{i,\kappa}^{\text{tr}} \mid \Lambda_i, \Lambda_{i,\kappa} \right) \right] + \mathbb{E} \left[ \text{cov} \left( \mathbb{E} [D_i \mid \Lambda_i, \Lambda_{i,\kappa}] \otimes \mathbb{E} [D_i^{\text{tr}} \mid \Lambda_i, \Lambda_{i,\kappa}] \mid \Lambda_i \right) \right] \\ &= \mathbb{E} \left[ a(\Lambda_i, \Lambda_{i,\kappa}) \cdot T^{(i,\kappa)} + S^{(i)} \cdot \mathbb{E} \left[ \text{cov} \left( \rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) \otimes \rho^{(i)\text{tr}}(\Lambda_i, \Lambda_{i,\kappa}) \mid \Lambda_i \right) \right] \cdot S^{(i)\text{tr}} \right] \\ &= a^{(i)} \cdot T^{(i,\kappa)} + S^{(i,\kappa)} \cdot \beta_2 \cdot S^{(i,\kappa)\text{tr}}. \end{aligned} \quad (2.43)$$

with (H5), (H6) and (2.43) this proves (2.40).  $\square$

Having proved (2.40), using the recursive definition of (2.11) and (2.15) with the conditional independence of the level  $i, k$  given  $\Lambda_i$ , the formulas (2.42) follow by induction. Since  $\beta_\nu$  measures the variation of regression coefficients between level  $\nu$  ( level  $i$  for  $\nu = 1$  and level  $i, k$  for  $\nu = 2$  ) and since  $\widehat{\Theta^{(\nu)} D_\nu}$  estimates  $\rho(\Lambda_\nu, \Lambda_{\nu-1})$ , it seems natural to base the estimation of  $\beta_\nu$  on some (weighted) sum of squared residuals, within  $\nu - 1$  level contracts, of  $\widehat{\Theta^{(\nu)} D_\nu}$ .

### 2.3. Estimators for the structural parameters

The variance of  $\widehat{\Theta^{(\nu)}D_\nu}$  within  $\nu - 1$  level contracts is, by lemma 2.3, proportional to  $\beta_\nu$ , the proportionality factor being  $\Delta^{(\nu)-1}$ . It would therefore be natural to multiply (weight) squared residuals by the credibility matrices. Hence, we consider the quantity

$$M_\nu = \sum_{\nu-1} \sum_{\nu} \Delta^{(\nu)-1} \left( \widehat{\Theta^{(\nu)}D_\nu} - \widehat{\Theta^{(\nu-1)}D_{\nu-1}} \right) \otimes \left( \widehat{\Theta^{(\nu)}D_\nu} - \widehat{\Theta^{(\nu-1)}D_{\nu-1}} \right)^{tr}. \quad (2.44)$$

Where

$$\sum_0 \text{ is an empty sum}$$

$$\widehat{\Theta^{(0)}D_0} = R$$

By lemma 2.3

$$\begin{aligned} \mathbb{E} \left[ \widehat{\Theta^{(\nu)}D_\nu} \otimes \left( \widehat{\Theta^{(\nu)}D_\nu} \right)^{tr} \right] &= RR^{tr} + \text{cov}(\rho^{(i)}(\Lambda_i) \otimes \rho^{(i)tr}(\Lambda_i)) \\ &\quad + \text{cov}((\Theta^{(\nu)}D_\nu) \otimes (\Theta^{(\nu)}D_\nu)^{tr} / \Lambda_\nu, \Lambda_{\nu-1}) \\ &= RR^{tr} + \beta_{\nu-1} + \Delta^{(\nu)-1} \beta_\nu \end{aligned}$$

$$\begin{aligned} \mathbb{E} \left[ \widehat{\Theta^{(\nu-1)}D_{\nu-1}} \otimes \left( \widehat{\Theta^{(\nu)}D_\nu} \right)^{tr} \right] &= RR^{tr} + \sum_{\nu} \Delta^{(\nu)-1} \sum_{\mu} \Delta^{(\mu)} (\text{cov}(\rho^{(i)}(\Lambda_i) \otimes \rho^{(i)tr}(\Lambda_i))) \\ &\quad + \sum_{\nu} \Delta^{(\nu)-1} \sum_{\mu} \Delta^{(\mu)} (\text{cov}((\Theta^{(\nu)}D_\nu) \otimes (\Theta^{(\nu)}D_\nu)^{tr} / \Lambda_\nu, \Lambda_{\nu-1})) \\ &= RR^{tr} + \beta_{\nu-1} + \sum_{\nu} \Delta^{(\nu)-1} \beta_\nu \end{aligned}$$

we obtain finally

$$\mathbb{E} [M_\nu] = \left( \sum_{\nu-1} \right) (\nu - 1) \beta_\nu$$

□



# Chapter 3

## Claim development patterns via Agglomerative hierarchical clustering

### 3.1 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 [15], [21] and [13].

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

### 3.1. Overview of Agglomerative hierarchical clustering

---

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 [15]

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 [22], [4] and [12] 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.

## 3.2 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

$$\mathcal{T} := \{(i, k, j) | 1 \leq i \leq I, 1 \leq k \leq K_i, j \leq I - i + 1\}, \quad (3.1)$$

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

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

$$\begin{aligned} \xi_{i,k} &= (\xi_{i,k,j})_{j \leq J} \\ \text{where } \xi_{i,k,j} &= \frac{X_{i,k,j}}{X_{i,k,J}}, \quad i \leq I, k \leq K_i, j \leq J \quad \text{and} \quad X_{i,k,J} > 0 \end{aligned}$$

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

### Theorem 3.1.

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

$$S := (\xi_{j,\ell})_{1 \leq j \leq J, \ell=1, \dots, L} \in \mathbb{R}^{J \times L} \quad (3.2)$$

*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 clustering method implemented to extract the matrix  $S$  is the following :*

### 3.2. Claim development patterns via agglomerative clustering algorithm

---

(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

$$T^{(j)} := \{(i, k) \mid k \leq K_i, 1 \leq i \leq I + 1 - j\}. \quad (3.3)$$

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  $T^{(j)} \neq \emptyset$

$$T^{(j)} \in \{R \cap \mathcal{T} \neq \emptyset \mid R \in \mathcal{R}^*(T^{(j-1)})\} =: \mathcal{R}^*(T^{(j-1)}) \cap T^{(j)}, \quad (3.4)$$

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

$$\mathcal{R}^{(*j)} := \bigcup_{\emptyset \neq T \in \mathcal{R}^*(T^{(j-1)}) \cap \mathcal{T}} \mathcal{R}^*(T^{(j)}). \quad (3.5)$$

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  $\Xi$  as the set of barycenters of the development patterns in each index set of the repartition  $\mathcal{R}^{(*J)}$ .*

### 3.2. Claim development patterns via agglomerative clustering algorithm

---

*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 (i):**

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

$$\xi_{i,k}^{(j)} = (\xi_{i,k,1}^{(j)}, \dots, \xi_{i,k,j}^{(j)}) \in \mathbb{R}^j \quad \text{with} \quad 0 \leq \xi_{i,k,1}^{(j)} \leq \dots \leq \xi_{i,k,j}^{(j)} = 1. \quad (3.6)$$

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

$$\xi_R^{(j)} := (\xi_{R,1}^{(j)}, \dots, \xi_{R,j}^{(j)}) \quad \text{with} \quad \xi_{R,\nu}^{(j)} = \frac{1}{|R|} \sum_{(i,k) \in R} \xi_{i,k,\nu}^{(j)}, \nu \leq j. \quad (3.7)$$

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

$$\xi_{i,k,j} = \frac{X_{i,k,j}}{X_{i,k,I+1-j}}. \quad (3.8)$$

The initial operation in step (i) is the determination of the reduced set  $T^{(j)}$  given by (3.3) 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 (iii):**

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  $\mathcal{R}^{(j)}$ ,  $T^{(j)} = \dot{\bigcup} R$ , where  $\dot{\bigcup}$  denotes the union of disjoint sets. The

### 3.2. Claim development patterns via agglomerative clustering algorithm

---

finest and the coarsest repartitions are resp.

$$\begin{aligned}\mathcal{R}_{\min}(T^{(j)}) &:= \left\{ \{(i, k)\} \mid (i, k) \in T^{(j)} \right\} && \text{resp.} \\ \mathcal{R}_{\max}(T^{(j)}) &:= \{T^{(j)}\}.\end{aligned}\tag{3.9}$$

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 3.1.**

*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:*

$$d(R_1, R_2) := |R_1| \left\| \xi_{R_1}^{(j)} - \xi_{R_1 \cup R_2}^{(j)} \right\|^2 + |R_2| \left\| \xi_{R_2}^{(j)} - \xi_{R_1 \cup R_2}^{(j)} \right\|^2\tag{3.10}$$

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

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

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

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

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

### 3.2. Claim development patterns via agglomerative clustering algorithm

□

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  $\mathcal{R}$ :

$$\{R_1, R_2\} \in \arg \min_{R, R' \in \mathcal{R}^{(j)}, R \neq R'} d(R, R'). \quad (3.11)$$

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

$$\mathcal{A}(\mathcal{R}^{(j)}) := \left\{ R' \in \mathcal{R}^{(j)} \mid R' \neq R_1, R' \neq R_2 \right\} \cup \{R_1 \cup R_2\}. \quad (3.12)$$

(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

$$\mathcal{I}_{intra}(\mathcal{R}) := \sum_{R \in \mathcal{R}} \sum_{(i,k) \in R} \|\xi_{i,k}^{(j)} - \xi_R^{(j)}\|^2 \geq 0. \quad (3.13)$$

while the inter-inertia of a repartition is

$$\mathcal{I}_{inter}(\mathcal{R}) := \sum_{R \in \mathcal{R}} |R| \|\xi_R^{(j)} - \xi_T^{(j)}\|^2, \quad (3.14)$$

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

#### Lemma 3.2.

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

$$\begin{aligned} \mathcal{I}_{intra}(\mathcal{R}(T^{(j)})) + \mathcal{I}_{inter}(\mathcal{R}(T^{(j)})) &= \mathcal{I}_{intra}(\mathcal{R}_{max}(T^{(j)})) \\ &= \mathcal{I}_{inter}(\mathcal{R}_{min}(T^{(j)})). \end{aligned} \quad (3.15)$$

### 3.2. Claim development patterns via agglomerative clustering algorithm

*Proof.* Indeed, replacing in  $\xi_{i,k}^{(j)}$  for simplicity the double index  $(i, k)$  by  $\ell$  and dropping the superscript  $^{(j)}$ , one easily checks that:

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

□

#### Lemma 3.3.

A consequence of (3.15) is the increase of the intra-partition inertia by aggregation.

$$\mathcal{I}_{intra}(\mathcal{A}(\mathcal{R}^{(j)})) \geq \mathcal{I}_{intra}(\mathcal{R}^{(j)}) \quad (3.16)$$

*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 (3.15) we have

$$\mathcal{I}_{intra}(R_1 \cup R_2) = \mathcal{I}_{intra}(\{R_1, R_2\}) + \mathcal{I}_{inter}(\{R_1, R_2\}). \quad (3.17)$$

### 3.2. Claim development patterns via agglomerative clustering algorithm

---

Using (3.17) with the notation of the aggregated partition  $\mathcal{A}(\mathcal{R}^{(j)})$  in (3.12)

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

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

$$RI(\mathcal{R}(T^{(j)})) := \frac{\mathcal{I}_{intra}(\mathcal{R}(T^{(j)}))}{\mathcal{I}_{intra}(\{T^{(j)}\})} \leq 1 \quad (3.18)$$

is very useful.

In the aim of keeping our final matrix  $S$  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  $\mathcal{A}^{|T^{(j)}|-1}(\mathcal{R}_{min}(T^{(j)})) = \mathcal{R}_{max}(T^{(j)}) = \{T^{(j)}\}$ )

$$\mathcal{R}^{(*j)} := \begin{cases} \{T^{(j)}\} & \text{if } RI(\mathcal{A}^{|T^{(j)}|-2}(\mathcal{R}(T^{(j)}))) \in (0.95, 1], \\ \mathcal{A}^{|T^{(j)}|-2}(\mathcal{R}(T^{(j)})) & \text{if } RI(\mathcal{A}^{|T^{(j)}|-2}(\mathcal{R}(T^{(j)}))) \in (0.9, 0.95], \\ \mathcal{A}^{|T^{(j)}|-3}(\mathcal{R}(T^{(j)})) & \text{else} \end{cases} \quad (3.19)$$

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

### 3.2. Claim development patterns via agglomerative clustering algorithm

#### 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 (iii), 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

$$\xi_{i,k,\nu}^{(j)} := \begin{cases} \xi_{R,\nu}^{(j-1)} \min\left(\frac{X_{i,k,j-1}}{X_{i,k,j}}, 1\right) & \text{for } 1 \leq \nu \leq j-1 \\ 1 & \text{for } \nu = j \end{cases} \quad (3.20)$$

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 (iii) to  $T^{(j)}$  and the associated development patterns, we get first  $\mathcal{R}(T^{(j)})$  by (3.19) and then  $\mathcal{R}^{(*)}$  by (3.5).

#### Step (4):

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

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

we find the matrix  $S := (\xi_{j,\ell}^{(i)})_{j \leq J, \ell \leq L}$  according to (3.20):

$$\xi_{j,\ell}^{(i)} := \xi_{R_\ell,j}^{(J)}. \quad (3.21)$$

This ends the construction of the design matrix. □

### 3.3 Implementation and results

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

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

(i) Initialisation

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

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

$$\mathcal{R}^*(T^{(2)}) = \left[ \begin{array}{c|cc} 2 & 0.33 & 1 \\ 14 & 0.55 & 1 \\ 14 & 0.78 & 1 \end{array} \right] \quad (3.22)$$

(iii) 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.

$$T^{(3)} \in \{R \cap \mathcal{T} \neq \emptyset \mid R \in \mathcal{R}^*(T^{(2)})\} =: \mathcal{R}^*(T^{(2)}) \cap T^{(3)}. \quad (3.23)$$

Table 3.1: 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 | 46   |      |      |      |      |      |
|           | 5     | 1           | 1300 |      |      |      |      |      |      |

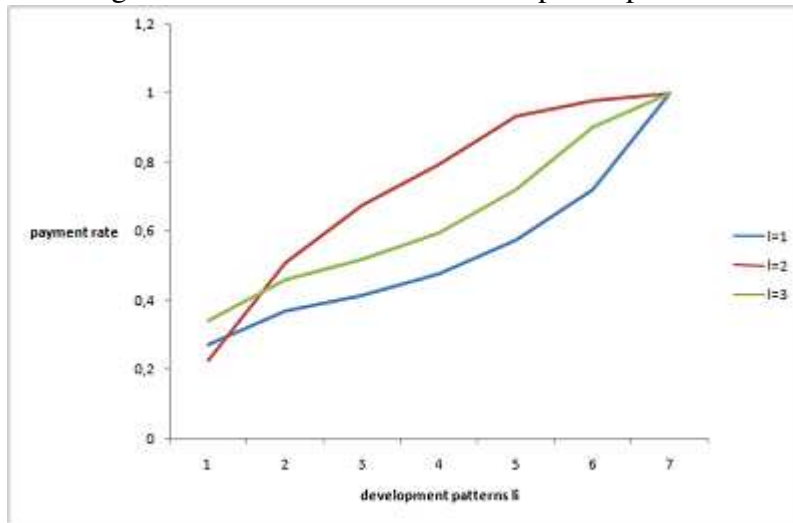
Table 3.2: Cumulated run-off triangle data refined by  $\mathcal{R}^*(T^{(2)})$

| 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 |
|   |       |             | 1         |         |         |         |         |         |         |
| 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 |         |
|   |       |             | 1         |         |         |         |         |         |         |
| 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 |         |         |
|   |       |             | 1         |         |         |         |         |         |         |
| 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 |         |         |         |
|   |       |             | 1         |         |         |         |         |         |         |
| 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  $S$  given by (3.24) and presented by Figure (1) as the set of barycenters of the development patterns.

$$S = \mathcal{R}^{(*7)} = \begin{bmatrix} 0,27 & 0,37 & 0,42 & 0,48 & 0,58 & 0,73 & 1 \\ 0,22 & 0,51 & 0,67 & 0,79 & 0,93 & 0,98 & 1 \\ 0,34 & 0,46 & 0,52 & 0,60 & 0,72 & 0,90 & 1 \end{bmatrix} \quad (3.24)$$

Figure 3.1: Characteristic of development patterns



# 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 thesis, we investigated credibility theory in loss reserving. We started with a review of the classical credibility models of Buhlmann, Buhlmann-Straub and Hachemeister. The main result of this research summarised in the second chapter is the hierarchical Hachemeister credibility approach used to calculate the development quotas of the claim in which we employed a two-step procedure for loss prediction.

Based on the observations  $\widehat{D}$  depending on the unobservable structure variable  $\Lambda_i$ , the credibility estimator for  $\rho^{(i)}(\Lambda_i)$  is derived for the accident year  $i$  level. Similarly, the credibility estimator for  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa})$  is derived based on the observations  $\widehat{D}_{i,\kappa}$  on the active contracts  $\kappa$  level. The hierarchical structure allows more flexibility in the model

A helpful addition would be the adjusting role of the design matrix in the Hachemeister part. Confronted by the shrinkage effect in the prior information in the run-off triangle. In the aim to incorporate as much as possible information, a recursive agglomerative procedure has been

### *3.3. Implementation and results*

---

implemented to identify the potential characteristics of claim development patterns which allows taking into account settlement distribution in loss reserving models without the need of an external volume measure through the design matrix. The latter is particularly useful if one is more interested in the payment pattern in the tails.



# Appendix A

## A.1 Proof of the equivalence between 2.25 and 2.11

With the definitions in 2.12 and 2.13 we want to show that

$$C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot S^{(i)tr} \cdot V^{(i)-1} \cdot S^{(i)} \cdot \mathbb{1}_{k_i \times 1} \otimes Id_{r_i} = \Delta^{(i)}. \quad (\text{A.1})$$

First, it holds that

$$\begin{aligned} & \left( Id_{k_i r_i} + S^{(i)tr} U^{(i)-1} S^{(i)} \cdot \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \right) \cdot S^{(i)tr} \\ &= S^{(i)tr} + S^{(i)tr} \cdot U^{(i)-1} S^{(i)} \cdot \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \cdot S^{(i)tr} \\ &= S^{(i)tr} \cdot U^{(i)-1} [U^{(i)} + S^{(i)} \cdot \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \cdot S^{(i)tr}] = S^{(i)tr} \cdot U^{(i)-1} \cdot V^{(i)}. \end{aligned}$$

Therefore

$$\begin{aligned} & C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot S^{(i)tr} \cdot V^{(i)-1} \\ &= C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot \left( Id_{k_i r_i} + S^{(i)tr} U^{(i)-1} S^{(i)} \cdot \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \right)^{-1} \cdot S^{(i)tr} \cdot U^{(i)-1}. \quad (\text{A.2}) \end{aligned}$$

Let the sequence of matrices  $(Y_\ell)_{\ell \leq k_i}$  be a solution of

$$(\dots, Y_\ell, \dots) = C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot \left( Id_{k_i r_i} + S^{(i)tr} U^{(i)-1} S^{(i)} \cdot \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \right)^{-1}, \quad (\text{A.3})$$

so that with the diagonal structure of  $S^{(i)}$  and  $U^{(i)}$  we get

$$\begin{aligned}
 (\dots, C^{(i)}, \dots) &= (\dots, Y_\ell, \dots) \cdot \left( Id_{k_i r_i} + S^{(i)tr} U^{(i)-1} S^{(i)} \cdot \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \right) \\
 &= \left( \dots, Y_\ell + \sum_\lambda Y_\lambda \cdot S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} \cdot C^{(i)}, \dots \right) \quad \text{or} \\
 Y_\ell &= \left( Id_{r_i} - \sum_\lambda Y_\lambda \cdot S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} \right) \cdot C^{(i)} \tag{A.4}
 \end{aligned}$$

for all  $\ell \leq k_i$ . With the definition 2.13 it follows that

$$\begin{aligned}
 \sum_\lambda Y_\lambda \cdot S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} &= \left( Id_{r_i} - \sum_\lambda Y_\lambda \cdot S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} \right) \cdot C^{(i)} \cdot W^{(i)}, \quad \text{i. e.} \\
 \sum_\lambda Y_\lambda \cdot S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} &= C^{(i)} \cdot W^{(i)} \cdot [Id_{r_i} + C^{(i)} \cdot W^{(i)}]^{-1} \quad \text{and so} \\
 Id_{r_i} - \sum_\lambda Y_\lambda \cdot S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} &= [Id_{r_i} + C^{(i)} \cdot W^{(i)}]^{-1}.
 \end{aligned}$$

From A.4 we get for all  $\ell \leq k_i$  that

$$Y_\ell = [Id_{r_i} + C^{(i)} \cdot W^{(i)}]^{-1} \cdot C^{(i)}. \tag{A.5}$$

Collecting the results from A.2, A.3, and A.5 we get with the definition of  $\Delta^{(i)}$  from 2.12 that

$$\begin{aligned}
 (C^{(i)} \otimes \mathbb{1}_{1 \times k_i}) \cdot S^{(i)tr} \cdot V^{(i)-1} &= (Id_{r_i} - \Delta^{(i)}) \cdot C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot S^{(i)tr} \cdot U^{(i)-1} \\
 &= (Id_{r_i} - \Delta^{(i)}) \cdot C^{(i)} \cdot W^{(i)} \cdot \Theta^{(i)} \\
 &= \Delta^{(i)} \cdot \Theta^{(i)} \tag{A.6}
 \end{aligned}$$

and similarly

$$\begin{aligned}
 (C^{(i)} \otimes \mathbb{1}_{1 \times k_i}) \cdot S^{(i)tr} \cdot V^{(i)-1} \cdot S^{(i)} \cdot (\mathbb{1}_{k_i \times 1} \otimes Id_{r_i}) \\
 &= [Id_{r_i} + C^{(i)} \cdot W^{(i)}]^{-1} \cdot C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot S^{(i)tr} U^{(i)-1} S^{(i)} \cdot \mathbb{1}_{k_i \times 1} \otimes Id_{r_i} \\
 &= [Id_{r_i} + C^{(i)} \cdot W^{(i)}]^{-1} C^{(i)} \cdot W^{(i)} = \Delta^{(i)}. \tag{A.7}
 \end{aligned}$$

This proves A.1 and so 2.11.

### A.1.1 Proof of 2.17

Let  $\xi$  be an eigenvalue of eigenvalue  $\xi$  of  $\Delta^{(i,\kappa)} = B^{(i)} \cdot S^{(i,\kappa)tr} U^{(i,\kappa)-1} S^{(i,\kappa)}$ , so that for some  $\bar{y} \in \mathbb{R}^{r_i} \setminus \{0\}$

$$\Delta^{(i,\kappa)} \bar{y} = B^{(i)} \cdot S^{(i,\kappa)tr} U^{(i,\kappa)-1} S^{(i,\kappa)} \bar{y} = \xi \bar{y}. \quad (\text{A.8})$$

With  $y := C^{(i)-1/2} \bar{y}$ , which we may assume to be normed to it follows that

$$B^{(i)1/2} \cdot S^{(i,\kappa)tr} U^{(i,\kappa)-1} S^{(i,\kappa)} \cdot B^{(i)1/2} y = \xi y \quad (\text{A.9})$$

The fact that matrix on the left-hand side is symmetric and positive semidefinite, implies that  $\xi$  is real and non-negative. Therefore

$$\xi < y^{tr} B^{(i)1/2} \cdot S^{(i,\kappa)tr} (S^{(i,\kappa)} \cdot B^{(i)} \cdot S^{(i,\kappa)tr})^{-1} S^{(i,\kappa)} \cdot B^{(i)1/2} y. \quad (\text{A.10})$$

Setting  $x = (S^{(i,\kappa)} \cdot B^{(i)} \cdot S^{(i,\kappa)tr})^{-1} S^{(i,\kappa)} \cdot B^{(i)1/2} y$ , we find that

$$S^{(i,\kappa)} \cdot B^{(i)1/2} \cdot B^{(i)1/2} \cdot S^{(i,\kappa)tr} x = S^{(i,\kappa)} \cdot B^{(i)1/2} y.$$

Since  $S^{(i,\kappa)}$  has full rank, it is injective, which gives us

$$B^{(i)1/2} \cdot S^{(i,\kappa)tr} x = y.$$

From A.10 we conclude that

$$\xi < y^{tr} B^{(i)1/2} \cdot S^{(i,\kappa)tr} z = y^{tr} y = 1$$

This proves 2.17.

### A.1.2 Proof of 2.15 as superposition of 2.35 with 2.11

It is easy to check that  $\bar{\rho}^{(i,\kappa)}$  has the same mean as  $\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa})$ :

$$\mathbb{E} \left[ \bar{\rho}^{(i,\kappa)} \right] = \mathbb{E} \left[ \mathbb{E} \left[ \bar{\rho}^{(i,\kappa)} \mid \Lambda_i \right] \right] = \mathbb{E} \left[ \rho^{(i)}(\Lambda_i) \right] = R^{(i)} = \mathbb{E} \left[ \rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) \right].$$

Therefore they both have the same inhomogeneous linear estimator on the space generated by the random variables  $(D_m)_{m \in I}$  if and only if their difference is orthogonal to the linear space of the centralized variables  $(D_m - S^{(m)} \cdot (\mathbb{1}_{K_m \times 1} \otimes R_m))_{m \in I}$ .

So, we have with  $\gamma_m \in \mathbb{R}^{r \times K_m \cdot \mathcal{J}|m}$ ,  $m \leq I$

$$\begin{aligned}
 & \mathbb{E} \left[ \left( \rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - \bar{\rho}^{(i,\kappa)} \right) \cdot \left( \sum_m \gamma_m (D_m - S^{(m)} \cdot \mathbb{1}_{K_m \times 1} \otimes R_m) \right) \right] \\
 &= \text{trace} \sum_m \mathbb{E} \left[ \left( \rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) \rho^{(i)}(\Lambda_i) - \Delta^{(i,\kappa)} \cdot \Theta^{(i,\kappa)} \cdot D_{i,\kappa} \right) \right. \\
 & \quad \left. \otimes (D_m - S^{(m)} \cdot \mathbb{1}_{K_m \times 1} \otimes R_m)^{\text{tr}} \right] \gamma_m^{\text{tr}} \\
 &= \text{trace} \sum_m \left[ \text{cov}(\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) \otimes D_m^{\text{tr}}) - (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) \text{cov}(\rho^{(i)}(\Lambda_i) \otimes D_M^{\text{tr}}) \right. \\
 & \quad \left. - \Delta^{(i,\kappa)} \cdot \Theta^{(i,\kappa)} \cdot \text{cov}(D_{i,\kappa} \otimes D_m^{\text{tr}}) \right] \gamma_m^{\text{tr}} \\
 &= \text{trace} \left[ \mathbb{1}_{1 \times k_i} \otimes \text{Id}_{r_i} \cdot \text{dia}[\delta_{\kappa,\ell} B^{(i)} + C^{(i)}]_{\ell} \cdot S^{(i)\text{tr}} - (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) (C^{(i)} \otimes \mathbb{1}_{1 \times k_i}) \cdot S^{(i)\text{tr}} \right. \\
 & \quad \left. - \Delta^{(i,\kappa)} \cdot \Theta^{(i,\kappa)} \cdot (\mathbb{1}_{1 \times k_i} \otimes \text{Id}_{r_i} \cdot \text{dia}[\delta_{\kappa,\ell} U^{(i,\ell)} + S^{(i,\kappa)} \cdot C^{(i)} \cdot S^{(i,\ell)\text{tr}}]_{\ell}) \right] \gamma_i^{\text{tr}} \\
 &= \text{trace} \left[ B^{(i)} \cdot S^{(i,\kappa)\text{tr}} - \Delta^{(i,\kappa)} \cdot \Theta^{(i,\kappa)} \cdot U^{(i,\kappa)} \right] \gamma_{i,\kappa}^{\text{tr}} = 0,
 \end{aligned}$$

since by definition

$$\Delta^{(i,\kappa)} \cdot \Theta^{(i,\kappa)} \cdot U^{(i,\kappa)} = B^{(i)} \cdot S^{(i,\kappa)\text{tr}}.$$

We have shown 2.15.

### A.1.3 Calculations of the prediction errors

We start with the proof of 2.18. By the identity

$$\Theta^{(i)} \cdot S^{(i)} \mathbb{1}_{k_i \times 1} \otimes \text{Id}_{r_i} = \text{Id}_{r_i} \tag{A.11}$$

we rewrite the credibility estimator  $\bar{\rho}^{(i)}$  as

$$\bar{\rho}^{(i)} = R^{(i)} + \Delta^{(i)} (\Theta^{(i)} D_i - R^{(i)}) = R^{(i)} + \Delta^{(i)} \cdot \Theta^{(i)} (D_i - S^{(i)} \mathbb{1}_{k_i \times 1} \otimes R^{(i)}). \tag{A.12}$$

With the help of A.11 and A.6, the prediction error is now calculated as

$$\begin{aligned}
\mathbb{E} \left[ \|\rho^{(i)}(\Lambda_i) - \bar{\rho}^{(i)}\|^2 \right] &= \text{trace} \mathbb{E} \left[ \left( (\rho^{(i)}(\Lambda_i) - R^{(i)}) - \Delta^{(i)} \cdot \Theta^{(i)} (D_i - S^{(i)} \mathbb{1}_{k_i \times 1} \otimes R^{(i)}) \right) \right. \\
&\quad \left. \otimes \left( (\rho^{(i)}(\Lambda_i) - R^{(i)}) - \Delta^{(i)} \cdot \Theta^{(i)} (D_i - S^{(i)} \mathbb{1}_{k_i \times 1} \otimes R^{(i)}) \right)^{\text{tr}} \right] \\
&= \text{trace} \left[ C^{(i)} - \Delta^{(i)} \cdot \Theta^{(i)} \cdot S^{(i)} \mathbb{1}_{k_i \times 1} \otimes C^{(i)} - C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot S^{(i)\text{tr}} \cdot \Theta^{(i)\text{tr}} \cdot \Delta^{(i)\text{tr}} \right. \\
&\quad \left. + \Delta^{(i)} \cdot \Theta^{(i)} \cdot V^{(i)} \cdot \Theta^{(i)\text{tr}} \cdot \Delta^{(i)\text{tr}} \right] \\
&= \text{trace} \left[ C^{(i)} - \Delta^{(i)} \cdot C^{(i)} - C^{(i)} \cdot \Delta^{(i)\text{tr}} + C^{(i)} \otimes \mathbb{1}_{1 \times k_i} S^{(i)\text{tr}} \cdot \Theta^{(i)\text{tr}} \cdot \Delta^{(i)\text{tr}} \right] \\
&= \text{trace} \left[ (\mathbb{I}_{r_i} - \Delta^{(i)}) \cdot C^{(i)} \right] = \text{trace} \left[ C^{(i)-1} + W^{(i)} \right]^{-1}.
\end{aligned}$$

Preparing the proof of 2.19 we use again A.12 to write  $\bar{\rho}^{(i,\kappa)}$  as

$$\bar{\rho}^{(i,\kappa)} = R^{(i)} + \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} (D_{i,\kappa} - S^{(i,\kappa)} R^{(i)}) + (\mathbb{I}_{r_i} - \Delta^{(i,\kappa)}) \Delta^{(i)} \Theta^{(i)} (D_i - S^{(i)} \mathbb{1}_{k_i \times 1} \otimes R^{(i)}). \quad (\text{A.13})$$

In addition, we shall make use of the following two identities:

$$\begin{aligned}
\text{cov} \left( \rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) \otimes (\Delta^{(i)} \Theta^{(i)} D_i)^{\text{tr}} \right) &= \mathbb{I}_{r_i} \otimes \mathbb{1}_{1 \times k_i} \text{dia} [\delta_{\kappa,\ell} B^{(i)} + C^{(i)}]_{\ell} S^{(i)\text{tr}} \Theta^{(i)\text{tr}} \Delta^{(i)\text{tr}} \\
&\stackrel{\text{A.11}}{=} B^{(i)} S^{(i,\kappa)\text{tr}} U^{(i,\kappa)-1} S^{(i,\kappa)} C^{(i)} [\mathbb{I}_{r_i} + W^{(i)\text{tr}} C^{(i)}]^{-1} + C^{(i)} \Delta^{(i)\text{tr}} \\
&= \Delta^{(i,\kappa)} [C^{(i)-1} + W^{(i)}]^{-1} + C^{(i)} \Delta^{(i)\text{tr}} \quad (\text{A.14})
\end{aligned}$$

and

$$\begin{aligned}
\text{cov} \left( (\Delta^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa}) \otimes (\Delta^{(i)} \Theta^{(i)} D_i)^{\text{tr}} \right) &= \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} \otimes \mathbb{1}_{1 \times k_i} \text{dia} [\delta_{\kappa,\ell} U^{(i,\kappa)}]_{\ell} \Theta^{(i)\text{tr}} \Delta^{(i)\text{tr}} + \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} S^{(i,\kappa)} C^{(i)} \Delta^{(i)\text{tr}} \\
&\stackrel{\text{A.11}}{=} \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} \otimes \mathbb{1}_{1 \times k_i} \text{dia} [\delta_{\kappa,\ell} S^{(i,\kappa)}]_{\ell} \mathbb{1}_{k_i \times 1} \otimes C^{(i)} [\mathbb{I}_{r_i} + C^{(i)} W^{(i)}]^{\text{tr},-1} + \Delta^{(i,\kappa)} C^{(i)} \Delta^{(i)\text{tr}} \\
&= \Delta^{(i,\kappa)} [C^{(i)-1} + W^{(i)}]^{-1} + \Delta^{(i,\kappa)} C^{(i)} \Delta^{(i)\text{tr}} \quad (\text{A.15})
\end{aligned}$$

as well as

$$\Delta^{(i,\kappa)} \Theta^{(i,\kappa)} U^{(i,\kappa)} \Theta^{(i,\kappa)\text{tr}} = B^{(i)}. \quad (\text{A.16})$$

With A.13, A.14, and A.15 we get

$$\mathbb{E} \left[ (\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - \bar{\rho}^{(i,\kappa)}) \otimes (\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - \bar{\rho}^{(i,\kappa)})^{\text{tr}} \right]$$

A.1. Proof of the equivalence between 2.25 and 2.11

---

$$\begin{aligned}
&= \text{cov} \left( (\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa}) \otimes (\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa})^{\text{tr}} \right) \\
&\quad + \text{cov} \left( \left( (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) \Delta^{(i)} \Theta^{(i)} D_i \right) \otimes \left( (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) \Delta^{(i)} \Theta^{(i)} D_i \right)^{\text{tr}} \right) \\
&\quad - \text{cov} \left( \left( \rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa} \right) \otimes \left( (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) \Delta^{(i)} \Theta^{(i)} D_i \right)^{\text{tr}} \right) \\
&\quad - \text{cov} \left( \left( (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) \Delta^{(i)} \Theta^{(i)} D_i \right) \otimes \left( \rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} D_{i,\kappa} \right)^{\text{tr}} \right) \\
&= B^{(i)} + C^{(i)} - (B^{(i)} + C^{(i)}) \Delta^{(i,\kappa)tr} - \Delta^{(i,\kappa)} (B^{(i)} + C^{(i)}) \\
&\quad + \Delta^{(i,\kappa)} \Theta^{(i,\kappa)} U^{(i,\kappa)} \Theta^{(i,\kappa)tr} \Delta^{(i,\kappa)tr} + \Delta^{(i,\kappa)} C^{(i)} \Delta^{(i,\kappa)tr} \\
&\quad + (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) \Delta^{(i)} \Theta^{(i)} V^{(i)} \Theta^{(i)tr} \Delta^{(i)tr} (\text{Id}_{r_i} - \Delta^{(i,\kappa)tr}) \\
&\quad - \underset{\text{A.14, A.15}}{(\text{Id}_{r_i} - \Delta^{(i,\kappa)}) C^{(i)} \Delta^{(i)tr} (\text{Id}_{r_i} - \Delta^{(i,\kappa)tr})} \\
&\quad - (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) \Delta^{(i)} C^{(i)} (\text{Id}_{r_i} - \Delta^{(i,\kappa)tr}) \\
&\underset{\text{A.16}}{=} (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) C^{(i)} (\text{Id}_{r_i} - \Delta^{(i,\kappa)tr}) + (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) B^{(i)} \\
&\quad - (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) \Delta^{(i)} C^{(i)} (\text{Id}_{r_i} - \Delta^{(i,\kappa)tr}) \\
&= (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) (\text{Id}_{r_i} - \Delta^{(i)}) C^{(i)} (\text{Id}_{r_i} - \Delta^{(i,\kappa)tr}) + (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) B^{(i)}
\end{aligned}$$

Therefore we have

$$\begin{aligned}
\mathbb{E} \left[ \|\rho^{(i)}(\Lambda_i, \Lambda_{i,\kappa}) - \bar{\rho}^{(i,\kappa)}\|^2 \right] &= \text{trace} \left( (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) (\text{Id}_{r_i} - \Delta^{(i)}) C^{(i)} (\text{Id}_{r_i} - \Delta^{(i,\kappa)tr}) \right. \\
&\quad \left. + (\text{Id}_{r_i} - \Delta^{(i,\kappa)}) B^{(i)} \right)
\end{aligned}$$

which shows 2.19. Let

$$\widetilde{W}^{(i)} := \left( S^{(i)tr} U^{(i)-1} S^{(i)} \right)^{-1} + \mathbb{I}_{k_i \times k_i} \cdot C^{(i)}. \tag{A.17}$$

With the definition of  $\Delta^{(i)}$  from 2.12 we want to show that

$$\Delta^{(i)} = C^{(i)} \otimes \mathbb{I}_{1 \times k_i} \cdot \widetilde{W}^{(i)-1} \cdot \mathbb{I}_{k_i \times 1} \otimes \text{Id}_{r_i}. \tag{A.18}$$

Let the sequence of matrices  $(Z_\ell)_{\ell \leq k_i}$  be a solution of

$$(\dots, Z_\ell, \dots) = C^{(i)} \otimes \mathbb{I}_{1 \times k_i} \cdot \left( (S^{(i)tr} U^{(i)-1} S^{(i)})^{-1} + \mathbb{I}_{k_i \times k_i} \otimes C^{(i)} \right)^{-1}, \tag{A.19}$$

so that with the diagonal structure of  $S^{(i)}$  and  $U^{(i)}$  we get

$$\begin{aligned}
 (\dots, C^{(i)}, \dots) &= (\dots, Z_\ell, \dots) \cdot \left( (S^{(i)tr} U^{(i)-1} S^{(i)})^{-1} + \mathbb{I}_{k_i \times k_i} \otimes C^{(i)} \right) \\
 &= \left( \dots, Z_\ell (S^{(i,\ell)tr} U^{(i,\ell)-1} S^{(i,\ell)})^{-1} + \sum_\lambda Z_\lambda \cdot C^{(i)}, \dots \right) \quad \text{or} \\
 Z_\ell &= \left( Id_{r_i} - \sum_\lambda Z_\lambda \right) \cdot C^{(i)} (S^{(i,\ell)tr} U^{(i,\ell)-1} S^{(i,\ell)}) \quad (\text{A.20})
 \end{aligned}$$

for all  $\ell \leq k_i$ . It follows that

$$\begin{aligned}
 \sum_\lambda Z_\lambda &= \left( Id_{r_i} - \sum_\lambda Z_\lambda \right) \cdot C^{(i)} \sum_\lambda S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} \quad \text{i. e.} \\
 \sum_\lambda Z_\lambda &= C^{(i)} \left( \sum_\lambda S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} \right) \left[ Id_{r_i} + C^{(i)} \sum_\lambda S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} \right]^{-1}
 \end{aligned}$$

From B.4 we get for all  $\ell \leq k_i$  that

$$Z_\ell = \left[ Id_{r_i} + C^{(i)} \sum_\lambda S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} \right]^{-1} \cdot C^{(i)} (S^{(i,\ell)tr} U^{(i,\ell)-1} S^{(i,\ell)}). \quad (\text{A.21})$$

Collecting the results from B.3, and B.5 we get with the definition of  $\Delta^{(i)}$  from 2.12 that

$$\begin{aligned}
 C^{(i)} \otimes \mathbb{I}_{1 \times k_i} \cdot \widetilde{W}^{(i)-1} \cdot (\mathbb{I}_{k_i \times 1} \otimes Id_{r_i}) &= \sum_\lambda Z_\lambda \\
 &= \left[ Id_{r_i} + C^{(i)} \sum_\lambda S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} \right]^{-1} \cdot C^{(i)} \sum_\lambda S^{(i,\lambda)tr} U^{(i,\lambda)-1} S^{(i,\lambda)} = \Delta^{(i)}. \quad (\text{A.22})
 \end{aligned}$$

# Appendix B

## Verification of $\Delta^{(i)}$

Let

$$\widetilde{W}^{(i)} := \left( S^{(i)tr} U^{(i)-1} S^{(i)} \right)^{-1} + \mathbb{1}_{k_i \times k_i} \cdot C^{(i)}. \quad (\text{B.1})$$

With the definition of  $\Delta^{(i)}$  from 2.12 we want to show that

$$\Delta^{(i)} = C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot \widetilde{W}^{(i)-1} \cdot \mathbb{1}_{k_i \times 1} \otimes Id_{r_i}. \quad (\text{B.2})$$

Let the sequence of matrices  $(Z_\ell)_{\ell \leq k_i}$  be a solution of

$$(\dots, Z_\ell, \dots) = C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot \left( \left( S^{(i)tr} U^{(i)-1} S^{(i)} \right)^{-1} + \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \right)^{-1}, \quad (\text{B.3})$$

so that with the diagonal structure of  $S^{(i)}$  and  $U^{(i)}$  we get

$$\begin{aligned} (\dots, C^{(i)}, \dots) &= (\dots, Z_\ell, \dots) \cdot \left( \left( S^{(i)tr} U^{(i)-1} S^{(i)} \right)^{-1} + \mathbb{1}_{k_i \times k_i} \otimes C^{(i)} \right) \\ &= \left( \dots, Z_\ell \left( S^{(i,\ell)tr} U^{(i,\ell)-1} S^{(i,\ell)} \right)^{-1} + \sum_\lambda Z_\lambda \cdot C^{(i)}, \dots \right) \quad \text{or} \\ Z_\ell &= \left( Id_{r_i} - \sum_\lambda Z_\lambda \right) \cdot C^{(i)} \left( S^{(i,\ell)tr} U^{(i,\ell)-1} S^{(i,\ell)} \right)_{\text{fo-Z-ell-1}} \end{aligned} \quad (\text{B.4})$$

for all  $\ell \leq k_i$ . It follows that

$$\begin{aligned} \sum_{\lambda} Z_{\lambda} &= \left( Id_{r_i} - \sum_{\lambda} Z_{\lambda} \right) \cdot C^{(i)} \sum_{\lambda} S^{(i,\lambda)tr} U^{(i,\lambda)^{-1}} S^{(i,\lambda)} && \text{i. e.} \\ \sum_{\lambda} Z_{\lambda} &= C^{(i)} \left( \sum_{\lambda} S^{(i,\lambda)tr} U^{(i,\lambda)^{-1}} S^{(i,\lambda)} \right) \left[ Id_{r_i} + C^{(i)} \sum_{\lambda} S^{(i,\lambda)tr} U^{(i,\lambda)^{-1}} S^{(i,\lambda)} \right]^{-1} \end{aligned}$$

From B.4 we get for all  $\ell \leq k_i$  that

$$Z_{\ell} = \left[ Id_{r_i} + C^{(i)} \sum_{\lambda} S^{(i,\lambda)tr} U^{(i,\lambda)^{-1}} S^{(i,\lambda)} \right]^{-1} \cdot C^{(i)} \left( S^{(i,\ell)tr} U^{(i,\ell)^{-1}} S^{(i,\ell)} \right). \quad (\text{B.5})$$

Collecting the results from B.3, and B.5 we get with the definition of  $\Delta^{(i)}$  from 2.12 that

$$\begin{aligned} C^{(i)} \otimes \mathbb{1}_{1 \times k_i} \cdot \widetilde{W}^{(i)^{-1}} \cdot (\mathbb{1}_{k_i \times 1} \otimes Id_{r_i}) &= \sum_{\lambda} Z_{\lambda} \\ &= \left[ Id_{r_i} + C^{(i)} \sum_{\lambda} S^{(i,\lambda)tr} U^{(i,\lambda)^{-1}} S^{(i,\lambda)} \right]^{-1} \cdot C^{(i)} \sum_{\lambda} S^{(i,\lambda)tr} U^{(i,\lambda)^{-1}} S^{(i,\lambda)} = \Delta^{(i)}. \end{aligned} \quad (\text{B.6})$$

# Appendix C

## R code for the design matrix

The R-code for the design matrix implementation is the following

---

**Algorithm 2** Agglomerative hierarchical clustering algorithm.

---

```
I=35; J=7;
C←read.csv("cummatrix h=0.CSV" ,header=FALSE,sep=";",dec=",")    ▷ the input: the
cumulated run-off triangle matrix for h=0
h = 0    ▷ for h = 0 j = 2 and T(j) = T(2)
R← rep(0, h + 2)
S←function(i,j)
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:
M ← C[1:((I-5)-(h*5)),1:(h+3)]
P ← matrix(data = 0, nrow = ((I-5)-(h*5)) , ncol = (h+3))
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))
for (s in 1:(I-8-(h*5)) )
W ← matrix(data = 0, nrow = I-4-s-(h*5), ncol = I-5-(h*5))    ▷ W is the distance proximity
matrix
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)
```

---

---

*C. R code for the design matrix*

---

```
unique(sort(W))[2]                                ▷ Min of matrix W:
N← which(W == unique(sort(W))[2], arr.ind = TRUE)  ▷ The Position of min distance in the
matrix W
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 :
P←P[-N[1,1],]
P←P[-N[1,2],]                                ▷ delete the two lines of distance min:
P←rbind(P,Q)
print(P)
```

---



# Bibliography

- [1] K.Antonio and R.Plat. Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Jour* 7:649-699, 2014.
- [2] V.Atanasiu. Mathematical models in regression credibility theory. *Buletinul Academiei de Stiinta a Republicii Moldova. Matematica*, 1:18-33, 2008.
- [3] G.Benklander. An approach to credibility in calculating IBNR for casualty excess reinsurance. *The Actuarial Review*, 1976, 7.
- [4] R.K.Blashfield. Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*. 1976; 83(3): 377-388.
- [5] R.Bornhuetter and R.Ferguson. The actuary and ibnr. *Proc. CAS*, 59:181-195, 1972.
- [6] H.Buhlmann and A.Gisler. A Course in Credibility Theory and its Applications. *Springer Verlag. Berlin Heidelberg New York*, 2005.
- [7] A.Charpentier and M. Pigeon. Macro vs. micro methods in non-life claims reserving (an econometric perspective). *Risks*, 4(2):12, 2016.
- [8] F.DeVylder. Credibility for severity revisited. *ASTIN Bull*, 10:99-112, 1978.

## BIBLIOGRAPHY

---

- [9] F.DeVylder. Estimation of IBNR claims by credibility theory. *Insurance: Mathem. a. Econom.*,1:35-40, 1982.
- [10] F.DeVylder. Non-linear regression in credibility theory. *Insurance: Mathem. a. Econom.*, 4:163-172, 1985.
- [11] P.E.Duda, R.O.Hart and D.G.Stork. *Pattern Classification,2nd Edition. John Wiley and Sons, 2000.*
- [12] B.Everitt, S.Hands. A Monte Carlo Study of the Recovery of Cluster Structure in Binary Data by Hierarchical Clustering Techniques. *Multivariate Behavioral Research.* 2010; 22(2): 235-243.
- [13] A.Jain .Data Clustering 50 Years Beyond K Means. *Pattern Recognition Letters.*2010; 31(8):651-666.
- [14] C. Hachemeister. Credibility for regression models with application to trend in credibility, theory and application. *Proceedings of the Berkeley Actuarial Research Conference on credibility;Academic Press, New York, pages 129-163, 1975.*
- [15] J.A.Hartigan. *Clustering Algorithms. New York: John Wiley and Sons; 1975.*
- [16] T.Hastie, R.Tibshirani, and J.Friedman. *The Elements of Statistical Learning.Springer,2009.*
- [17] T.N.Herzog. *Introduction to Credibility Theory. Actex Publ., 3rd ed., 1999.*
- [18] O.Hesselager. Prediction of outstanding claims: a hierarchical credibility approach. *Scandinavian Actuarial Journal, 25-47, 1991.*

## BIBLIOGRAPHY

---

- [19] K. Holler. Something old, something new in classification ratemaking with a novel use of glms for credit insurance. *Casualty Actuarial Forum*, Winter, 1999.
- [20] V.Goulet. Principles and Application of Credibility Theory. *Journal of Actuarial Practice*, 6:5-62, 1998.
- [21] L.Kaufman, P.Rousseeuw. Finding Groups in Data: An Introduction to Cluster Analysis. *New York :John Wiley and Sons*;1990.
- [22] F.Kuiper, L.Fisher. A Monte Carlo Comparison of Six Clustering Procedures. *Biometrics*.1975; 31(3): 777-783.
- [23] E.Kremer. Minimum distance loss reserving. *Blatter DGVM*, 24:629-634, 2000.
- [24] T.Mack. Improved estimation of IBNR claims by credibility theory. *Insurance: Mathem. a.Econom.*, 9:51-57, 1990.
- [25] M.Marvin and A.Hanrik. A survey of matrix theory and matrix inequalities. *Allin and Bacon INC Boston.*, 1964.
- [26] T.Mikosch. Non-Life Insurance Mathematics. *Springer Verlag. Berlin Heidelberg New York*, 2004.
- [27] R.Norberg. Hierarchical credibility: analysis of a random effect linear model with nested classification. *Scand. Acuarial Journ.*, 204-222, 1986.
- [28] R.Norberg. Prediction of outstanding liabilities in non-life insureded. *ASTIN Bull.*, 23:95-115,1993.

## BIBLIOGRAPHY

---

- [29] R.Norberg. Prediction of outstanding liabilities, ii model variation and extensions. *ASTIN Bull.*, 29:5-25, 1999.
- [30] R.Norberg. Credibility theory *Encyclopedia of Actuarial Science*, Wiley , Chichester, UK, 2004.
- [31] D.Pelleg and A.Moore. X-means: Extending k-means with efficient estimation of numbers of clusters. *In Proceedings Int. Conf. Machine Learning ICML*, 727-734, 2000.
- [32] K.D.Schmidt. Versicherungsmathematik. *Springer Verlag. Berlin Heidelberg New York*, 2009.
- [33] K.D.Schmidt. Loss prediction based on run-off triangles. *Adv. Statistical Analysis*, 96:265-310,2012.
- [34] R.Schnieper. Separating true IBNR and IBNER claims. *ASTIN Bull.*, 21:111-127, 1991.
- [35] B.Sundt. A hierarchical regression credibility theory. *Scand. Acuarial Journ.*, 107-114,1979.
- [36] B.Sundt. A multi-level hierarchical regression credibility theory. *Scand. Acuarial Journ.* ,25-32, 1980.
- [37] B.Sundt. Non-Life Insurance Mathematics. *Institut für Versicherungswissenschaft, Universität Mannheim Band 28, VVW Karlsruhe*, 1993.
- [38] B.Sundt. Two credibility regression approaches for the classification of passenger cars in a multiplicative tariff. *ASTIN Bull.*, 17:41-70,1986.
- [39] T.Tarbell. Incurred but not reported claims reserves. *Proc. Casualty Acuarial Soc.*, 20:275-280,1934.

## BIBLIOGRAPHY

---

- [40] G.Taylor. Credibility analysis of a general hierarchical model.*Scand. Acuarial J.*1-12,1979.
- [41] J.H.Ward. Hierarchical grouping to optimize an objective function.*Journal of the American Statistical Association*, 58:236-244, 1968.

