INP Grenoble – ENSIMAG
École Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble

Rapport de projet de fin d'études
Mémoire de Recherche

Effectué à l'Université Laval (Québec)

Modèles de crédibilité avec régression au barycentre du temps

Xavier Milhaud
3e année – Option Ingénierie Financière (3A à l'ISFA Lyon)

28 avril 2008 – 29 août 2008

Université Laval
1045 Avenue de la Médecine
Local 1620, Ecole d'Actuariat
99401 Québec Canada

Tuteur du laboratoire
Professeur Vincent Goulet
Tuteur de l'école
Yves Denneulin

Topics in Credibility Theory: Hachemeister model in the general framework of regressions, adjustment of structural parameters, heterogeneity of the portfolio.

Xavier Milhaud

August 29, 2008

Résumé

La théorie de la crédibilité est une des branches de recherche des sciences actuarielles qui a été développée à l'origine pour être capable de calculer la prime du risque associé à un assuré en combinant le fruit de l'expérience individuelle de ce risque avec l'expérience collective du groupe auquel il appartient.
Quand une compagnie d'assurance calcule la prime versée par les assurés, elle sectorise ses assurés en groupes ou "classes" de mêmes caractéristiques. Ainsi, en regroupant ses risques semblables, elle est en mesure de calculer la prime du groupe après une analyse statistique. Le problème du calcul de la prime pour un groupe de contrats se résoud ensuite grâce à la théorie de la crédibilité.
Un actuaire veut déterminer le montant de la prime pour l'année à venir, il va donc prendre en compte non seulement l'expérience individuelle des contrats mais aussi l'expérience collective du portefeuille. On distingue ainsi deux visions opposées :

- nous chargons la même prime à absolument tous les assurés, notée $\bar{X}$, qui est la moyenne des données de montants de sinistres sur tout le portefeuille : ceci n'est logique que si le portefeuille est parfaitement homogène (impossible en réalité),
- ou nous chargons à chaque groupe j sa moyenne de montant de sinistre, notée $\bar{X}_j$.

Dans le cas d'un portefeuille hétérogène et avec un minimum d'expérience, l'idée est donc de pondérer ces deux positions extrêmes avec des poids choisis de manière à tenir compte de la différence d'expérience entre les contrats. Cette combinaison permet à la fois d'éviter de faire payer une prime plus élevée qu'elle ne devrait l'être à des assurés peu risqués, ou de faire payer une prime "insuffisante" à des assurés risqués. La théorie de la crédibilité joue ainsi le rôle de vase communiquant entre une expérience collective et une expérience individuelle.

Le premier chapitre traite de différents modèles de crédibilité, dans le cas de la régression ou non. Nous abordons les modèles de `Bühlmann`, premier modèle de cette théorie, de `Bühlmann-Straub` (fortement inspiré du modèle de `Bühlmann`), et enfin le modèle de `Hachemeister` considéré comme différent mais dans lequel l'idée sous-jacente est identique. Le but du second chapitre est de représenter l'hétérogénéité d'un portefeuille graphiquement pour une visualisation pratique, et donc de choisir quelles données représenter et comment. Nous verrons que les paramètres de structure y jouent un rôle déterminant.
Tous les modèles et leurs calculs sont implémentés dans le package **actuar** (développé en majorité par le professeur Vincent Goulet) du logiciel libre **R**.

Mots-clés : **crédibilité** ; **collective** ; **individuelle** ; **prime** ; **moyenne pondérée** ; **experience** ; **modèle de régression** ; **orthogonalisation** ; **ajustement** ; **estimateurs des paramètres de structure** ; **homogénéité** ; **hétérogénéité**.

Abstract

Credibility theory is a branch of actuarial science, developed originally as a method to calculate the risk premium by combining the individual risk experience with the class risk experience.
When an insurance company calculates the premium it will charge, it divides the policy holders into groups. The division is made balancing the two requirements that the risks in each group are sufficiently similar, and the group sufficiently large that a meaningful statistical analysis of the claims experience can be done to calculate the premium. This compromise means that none of the groups contains only identical risks. Then, the insurance company wants to combine the experience of the group with the experience of the individual risk the better in order to calculate the premium. For actuaries, credibility theory provides a solution to the problem of calculating the premium for a group of contracts.
The goal is to set up an experience rating system to determine next year's premium, taking into account not only the individual experience with the group, but also the collective experience. There are two extreme positions :

- we charge the same premium to everyone, estimated by the overall mean $\bar{X}$ of the data : this makes sense only if the portfolio is homogeneous, which means that all risks cells have identical mean claims,
- if the portfolio is not homogeneous, the other solution is to charge to group j its own average claims $\bar{X}_j$.

This methods are used if the portfolio is heterogeneous, provided a fairly large claim experience. To compromise these two extreme positions, we take the weighted average of these two extremes.
Charging a premium based on collective as well as individual experience is justified because portfolio in general is neither completely homogeneous, nor completely heterogeneous. Besides, there is every chance that we do not overcharge "good" people and undercharge "bad" risk people.

The first chapter focuses on different credibility models in the regression case or not. It deals with the well-known model of `Bühlmann` which has been the first model developed, the `Bühlmann-Straub` model strongly based on it and the `Hachemeister` model considered as different but in which we can find the same idea. The aim of the second chapter is to represent graphically the heterogeneity of the portfolio, to see for each contract the main difference from the others.
All the computations have been inserted into the software **R**, more precisely into the package **actuar** (implemented first by Pr. Vincent Goulet).

Keywords : **credibility**; **collective**; **individual**; **premium**; **weighted average**; **experience**; **regression model**; **orthogonalization**; **adjustment**; **structural estimators**; **homogeneity**; **heterogeneity**.

Acknowledgements

Despite the relative short time (18 weeks) I spent in Quebec city, working in "`act.ulaval`*" was a real pleasure. So I would like to thank everyone I have worked with during this internship, and without who this experience would have never happened.

First of all, I would like to express my gratitude to Pr. Vincent Goulet who spent time for me and suggested me new ways to understand a problem when I needed it. Besides he also gave me a lot of attention, dealt with me very well when I arrived, and has everytime done the best for me to feel great during all my stay here. So thanks to him and his parents who welcomed me and offered me a place to live at the beginning of the internship.
Moreover the work I had to do made me deeply improve my knowlegde of the statistical software **R**. It has been an honor to contribute to the package **actuar**.

Finally, I would also like to thank the Ph.D. student Florent Toureille, with whom I talked about many topics and permited me to spread my general knowledge on the actuarial theory.

∗. the actuarial lab of Université Laval

Contents

Introduction

The topic of credibility theory has been for many years one of the major interest in insurance. Credibility theory can be seen as a method to price insurance products, and the basic understanding of insurance risk is built upon two fundamental concepts : the individual and the collective risks. The insurer has typically little knowledge about the individual risk, but quite extensive statistical information about the collective. The theorical concept of a *homogeneous* collective is a fiction and credibility theory relies on the more realistic concept of a *heterogeneous* collective. Hence, credibility theory solves rigorously the problem of determining “insurance premiums” from different sources.
“Credibility” designates the weight given to the experience of the “individual risk”. That theory is a mathematical discipline which borrows its methods from many fields of mathematics; e.g. bayesian statistics, L^2 Hilbert-space, least squares method, and state space modelling for example.
But credibility theory remains interesting when it is closely linked with its applications, and it is only through its applications that credibility won its status in insurance thinking. It provides us with techniques to adjust insurance premiums for contracts that belong to an heterogeneous portfolio, in case there is limited or irregular claims experience for each contract.

The basic idea underlying insurance is that individuals, who have the “same” exposure to a particular risk, join together to form a “community-at-risk” in order to bear this perceived risk. In modern society, this idea is most often realized in the form of insurance. On payment of a premium, the individuals in the “community-at-risk” transfer their risk to an insurance company.
We will refer to the premium payable by the insured to the insurer as the gross premium. The premium volume is the sum, over the whole portfolio, of all gross premiums in the insurance period. The basic task underlying the rating of a risk is the determination of the so-called pure risk premium $P_i = E[X_i]$. The classical point of view assumes that, on the basis of some objectively quantifiable characteristics, the risks can be classified into homogeneous groups (risk classes), and that statistical data and theory then allow one to determine the risk premium to a high degree of accuracy.
Of course in reality, it is clear that a huge number of factors contribute to the size of the risk premium, and so to have reasonably homogeneous risk classes, one would have to subdivide the portfolio into a large number of classes. On the other hand, subdividing the portfolio into larger classes, we loose the assumption of homogeneous risk profiles within the classes.

This research memoir is divided into two independent chapters with common topics : credibility premiums and homogeneity of the portfolio.
The first chapter extends the work of Hachemeister (1975) by assuming that it is possible to get the credibility premium always between the individual and the collective ones. One uses mathematical concepts and an adjustment of the credibility coefficients so as to do that. First, It provides some generalities on credibility models like `Bühlmann-Straub` model for the reader to understand better the underlying concepts on easier models.

In the second chapter, we discuss about the representation of the heterogeneity of the portfolio. This section deals in majority with the problem of representing the variance and covariance components which are computed in the previous models. The aim is to show the between variance of the contracts and the within variance within a given portfolio, very important structural parameters to price a contract and check the consistency of results from applied models. We implement methods to do it for some models : the `Bühlmann` model, the `Bühlmann-Straub` model, the `Hachemeister` model and the `hierarchical` model.

Chapter 1

Modelling in credibility theory

1.1 Introduction

The well-known credibility formula

$$\pi = (1 - Z).B + Z.A$$

born in United States just before the World War I has been suggested for the first time in the field of workmen's compensation insurance. We can see in this formula the weights given to "individual risk" and "collective", respectively Z called the *credibility factor* and (1-Z).
The idea is to balance C between the two extremes A and B, suggested by *Whitney* who wrote :
"the problem of experience rating arises out of the necessity, from the standpoint of equity to the individual risk, of striking a balance between class-experience on the one hand and risk experience on the other". Of course the more experience you have for an individual risk, the greater Z will be.
So the question in the theory of credibility is : how much weight should be given to this actual claims experience?

1.2 Credibility theory, History

1.2.1 The american credibility theory

North-american actuaries developed the "american credibility theory" also called "limited fluctuation credibility theory", a new branch which was born with a paper by Mowbray [1914] "How extensive a payroll exposure is necessary to give a dependable pure premium ?".
The issue is to know how much individual claims experience is needed ? Here is his solution :
"A dependable pure premium is one for which the probability is high, that it does not differ from the true pure premium by more than an arbitrary limit".

Hence the main idea is to control the precision of the premium. Mathematical modelling enables us to answer this question using some tools to compute the number of insureds required for a reliable estimate of the "true" premium. But what can we do when the number of insureds is too small ? Later, this problem known as partial credibility was solved thanks to heuristic formulas for Z.

1.2.2 The european credibility theory

Bühlmann published in 1967 his "distribution free" credibility formula based on a least squares criterion and gave birth to the "european credibility theory". The Bühlmann model explicits the pure premium as a convex combination of the individual experience and the experience on the portfolio. Since it, researches concentrated in this scope and other famous papers were written by *Bailey and Robbins* for example. There are many fields of application with this theory; the models we are going to study in the following and the most famous papers in credibility theory are based on it.

1.2.3 In practice...

Most of the actuaries agree with the top-down approach in tarification as proposed by *Bühlmann*. The matching of premiums and liabilities must be the insurer's main goal.
Except exceptionnal cases, the larger part of the premium payable consists of the risk premium ot the expected future claim amount.
The actuary's first task is to determine the characteristics of insureds so as to distinguish between them. His opinion is really important because he has to choose the tariff variables. Unfortunately, only some of the tariff variables are observable and have data available. The past claims experience is the most used tariff variable, but is representative of both observable and non-observable variables. The aim is to calculate optimal tariff classes, it implies a structure that consists of so-called cells. Within every cell, insureds have identical risk characteristics.
Next question is how to determine insurance premiums for each of the cells?
First, we can estimate by the method of maximum likelihood the parameters of a pre-specified distribution function for the tariff variables. The other possibility consists of the credibility approach. The Bühlmann model and its generalizations allow for a distribution free estimation of the insurance premiums as a weighted average of the cell-experience and the portfolio experience. But this model is restrictive because it only suits to deal with the claims experience variable and exactly one other tariff variable. By consequence, the model should be modified when more tariff variables are involved. We can also combinate both methods (credibility and maximum likelihood) in theory.

Let us take an example so as to illustrate the problem : in automobile insurance, given the tariff classes (age of the driver, weight of the car, sex...), the heterogeneous portfolio is split into groups of insureds which are less heterogeneous. Then a premium for each group is calculated, it reflects the average claim amount within that group. But all drivers within a group do not have the same behavior and individual claims experience can tell us more about these hidden risk characteristics. Notice that there is a second selection, and the final premium will result from a weighted average of the individual driver's credibility adjusted premium on his own and the group's claims experience. Another part of the heterogeneity will be eliminated and the new premium is closer to the true premium. Of course, in practice this procedure is too laborious to handle ; insurers use a bonus-scale with fixed discounts and surcharges in percentages of the group-premium to incorporate the individual claims experience.

Credibility techniques are able to handle with different types of data, so we are going to study which model fits with which set of data.

1.2.4 The mathematical model

Some introductory information is necessary to have a better understanding of the techniques in use today. Our aim is to present the basic ideas of credibility without going into pure mathematics, so as to state basic results and introduce notations. As everybody knows, a portfolio is compound by more or less similar contracts ; that is why actuars consider a risk parameter, say θ_{js}, that describes the risk characteristics of contract j in period s.
θ_{js} is supposed to be time-homogeneous and does not change over time for a fixed contract, so we will drop the subscript s in the following. Of course and in most practical situations we do not know the risk parameters which are unobservable and we have to estimate these random variables as accurately as possible. $U(\theta)$, called the structure function, is the distribution function of all these variables. In fact, the contracts are similar in the way their risk parameter have the same structure function, but differ because of the different realizations θ_j.

1.2.5 The exact credibility

First let us present the best experience premium (where X represents the amount of the claims) :

Definition 1. *The Bayes premium is defined by*

$$P^{Bayes} = E[\mu(\Theta)|X]$$

The exact credibility formula uses the Bayes premium as it is stated as follows

$$E[\mu(\theta)|X_1, X_2, ..., X_t] = \frac{\int \mu(\theta).f(x_1|\theta).f(x_2|\theta)...f(x_t|\theta).dU(\theta)}{\int f(x_1|\theta).f(x_2|\theta)...f(x_t|\theta).dU(\theta)}$$

where $\mu(\theta) = E[X_{t+1}|\Theta]$.

It is clear that we must know the functions $U(\theta)$ and $F(x|\theta)$ to evaluate this integration.
The classical solution is to consider a linear estimate for $\mu(\Theta)$ in which only the first and second order moments of the unknown distribution are involved. We could evaluate the formula above for different pairs $(F(x|\theta), U(\theta))$ but it is a waste of time in most cases, although some pairs of distributions turn into atonishing linear credibility premiums. In general, the estimator of the Bayes premium cannot be expressed in a closed analytical form and can only be calculated by numerical procedures. Therefore it does not fulfil the requirement of simplicity.
Besides to calculate this premium one has to specify the conditional distribution as well as the *a-priori* distribution, which, in practice, can often neither be inferred from data nor guessed by intuition.
The basic idea underlying credibility is to force the required simplicity of the estimator by restricting the class of the allowable estimator functions to those which are linear in the vector of observations.

In other words we look for the best estimator in the class of all linear estimator functions. But "best" is to be understood in the bayesian sense : the optimality criterion is again the *quadratic loss*. Thus we will get the *credibility estimators* as *linear Bayes estimators*.

1.3 Introduction to credibility estimators

The Bayes premium is the best possible estimator in the class of all estimator functions. But it is often difficult to have a simple analytical form, and that is why we restrict the class of allowable estimators to those which are linear in the observations.
In a simple credibility model, we are used to assuming the results presented below.

1.3.1 Some basic clues

Model Assumptions.
The risk i is characterized by an individual risk profile which is itself the realization of a random variable Θ_i, and we have that :

i) Conditionnally, given Θ_i, the random variables X_j $(j = 1, ..., n)$ are independant with the same distribution function F_θ with the conditional moments

$$\begin{aligned} \mu(\Theta_i) &= E[X_{ij}|\Theta_i], \\ \sigma^2(\Theta_i) &= Var[X_{ij}|\Theta_i] \end{aligned}$$

ii) Θ is a random variable with distribution U(θ).

Under these assumptions, we have the following result :

Theorem 1. *The credibility estimator is given by*

$$\hat{\mu}(\Theta) = \alpha\bar{X} + (1-\alpha)\mu_0$$

where

$$\begin{aligned} \mu_0 &= E[\mu(\Theta)], \\ \alpha &= \tfrac{n}{n+\sigma^2/\tau^2} \end{aligned}$$

Proof : cf appendix, A.1

1.3.2 Intuitive principles

We can study certain important quantities to see their characteristics and their importance on a general case, and we get

- $\mu_0 =$ collective premium, has the quadratic loss $E\big[(\mu_0 - \mu(\Theta))^2\big] = Var(\mu(\Theta)) = \tau^2$,
- $\bar{X}$ is the best linear individually unbiased estimator of the individual mean, and has the quadratic loss $E\big[(\bar{X} - \mu(\Theta))^2\big] = E[\sigma^2(\Theta)/n] = \sigma^2/n$,
- P^{cred} is a weighted mean of this two, where the weights are proportional to the inverse quadratic loss.

We need two sources of information : the a priori knowledge and the individual observations, respectively for the collective and the individual risk. It is also intuitively reasonable to weight the estimators according to their precision, the inverse value of the quadratic loss.

1.3.3 The quadratic loss of the credibility premium

We can easily compute the quadratic loss of the credibility premium thanks to the previous results and we have,

Theorem 2. *The quadratic loss of the credibility estimator $\mu(\Theta)$ given by the previous theorem is*

$$E\big[(\hat{\mu}(\Theta) - \mu(\Theta))^2\big] = (1-\alpha)\tau^2 = \alpha\frac{\sigma^2}{n}$$

Proof : cf appendix, A.2

1.4 The Bülhmann model

The most famous models in the credibility theory are Bühlmann and Bühlmann-Straub's ones. They can be seen as an empiric bayesian credibility.
Consider the risk i during the period j. Actually you do not know the distribution of the risk parameter Θ_i, or the conditionnal distribution of $(X_{ij}|\Theta_i)$. Θ_i is not directly observable, so it is difficult to determine it and to solve this problem; that is why our aim is to find the best linear estimator of

$$\mu(\Theta_i) = E[X_{ij}|\Theta_i]$$

The Bühlmann model is restrictive because all the data must have the same weight (not specified).

Remark 3. *We need to use ratios as claim amounts in all the following models. The crucial assumption is to consider ratio for claim amount, because it has a real impact on the figures of the structure parameters (it is a kind of "weight").*

1.4.1 Motivation

An actuary has to rate a risk in motor third-party liability insurance for example. On the basis of certain risk characteristics, the risks have been grouped into various risk classes and now a risk premium has to be calculated for each of these classes. Suppose statistical information is available. For the ith risk, we have typically these useful information to use the model :

- S_{ij} is the aggregate claim amount in year j,
- V_{ij} is the number of years at risk in year j,
- $X_{ij} = S_{ij}/V_{ij}$ is the average claim costs per year at risk in year j (claim ratio),
- N_{ij} is the number of claims in year j,
- $F_{ij} = N_{ij}/V_{ij}$ is the claim frequency in year j,
- $Y_{ij} = S_{ij}/N_{ij}$ is the average claim size in year j.

The actuary's task is to calculate for the ith risk class the "true" individual pure risk premium for a future period. He can do it directly from the observations X_{ij}, or indirectly by the analysis of the components "claim frequency" and "average claim size".

1.4.2 Notations and assumptions

Let us introduce the notations used in the present chapter :
- Θ_i is the non-observable risk parameter of the contract i ;
- $((X_{i1}, X_{i2}, ..., X_{im}, \Theta_i) : 1 \leq i \leq n)$ is the vector of observations ;
- $\mu(\Theta_i) = E[X_{ij}|\Theta_i]$: conditionnal expectation of amount for risk i ;
- $\mu = E[\mu(\Theta_i)]$: expectation of global claim amounts for individual risk i ;
- $s^2(\Theta_i) = Var[X_{ij}|\Theta_i]$: conditionnal variance of claim amounts, variance within ind. risk ;
- $\sigma^2 = E[s^2(\Theta_i)]$: expectation of the conditionnal variance, average variance within ind. risk ;
- $\tau^2 = Var[\mu(\Theta_i)]$: variance of the conditionnal expectation, variance between ind. risk premiums.

Model Assumptions. (*Bühlmann model*) :
The risk i is characterized by an individual risk profile which is itself the realization of a random variable Θ_i, and we have that :

B1 : Conditionnally, given Θ_i, the $((X_{i1}, X_{i2}, ..., X_{im}, \Theta_i) : 1 \leq i \leq n)$ are independant with

$$\begin{aligned} E[X_{ij}|\Theta_i] &= \mu(\Theta_i), \\ Var[X_{ij}|\Theta_i] &= s^2(\Theta_i) \end{aligned}$$

B2 : the pairs (Θ_1, X_1), (Θ_2, X_2), ..., are independant ; and Θ_1, Θ_2, ... are independant and identically distributed (i.i.d.).

It is also interesting to see the following result.

Lemma 1. *We have (decomposition of the variance)*

$$\begin{aligned} Var[X_{ij}] &= E[Var[X_{ij}|\Theta_i = \theta]] + Var[E[X_{ij}|\Theta_i = \theta]] \\ &= E[s^2(\Theta_i)] + \left(E[(\mu(\Theta_i))^2] - E[\mu(\Theta_i)]^2\right) \\ &= \sigma^2 + E[(\mu(\Theta_i))^2] - \mu^2 \\ &= \sigma^2 + \tau^2 \end{aligned}$$

1.4.3 Determining the credibility premium

We want to estimate $\mu(\Theta_i)$ by $\mu(\hat{\Theta}_i)$, often called the credibility premium. We suggest it is linear (cf proof in appendix, A.1) because we are in the same framework as the previous section, therefore

$$\hat{\mu(\Theta_i)} = a_{i0} + \sum_{j=1}^{m} a_{ij} X_{ij}$$

Proposition 1. *The estimator of the credibility premium is given by the expression*

$$\begin{aligned}\hat{\mu(\Theta_i)} &= \frac{m\tau^2}{\sigma^2 + m\tau^2}\bar{X}_i + (1 - \frac{m\tau^2}{\sigma^2 + m\tau^2})\mu \\ &= z\bar{X}_i + (1-z)\mu\end{aligned}$$

where $z = \dfrac{1}{1 + \frac{\sigma^2}{m\tau^2}}$

Proof. cf. appendix, A.3

Remark 4. *Looking at this formula, we can interpret some quantities :*

- *the coefficient z is the same one for all the contracts ;*
- *the more you have experience (m increasing), the closer to 1 the coefficients z will be. It means that if you have many years of experience for a contract, you need not consider the collective premium because the actuary has enough information ;*
- $\tau^2 = Var[\mu(\Theta_i)]$ *corresponds to the variance of the conditionnal expectation, called also the between variance. The bigger* τ^2 *is, the more the conditionnal expectation is different from one risk to another. It means that you have an heterogeneous portfolio and z is close to 1; you had better use the average of claim amounts over one risk instead of considering the portfolio one ;*
- σ^2 *represents the expectation of the conditional variance : z is close to 0 when* σ^2 *increases. In fact if you have a small average of within variance, it is because your portfolio is homogeneous and you will consider more the collective experience ;*
- *assume that we have* $Var[X_{ij}] = \tau^2 + \sigma^2 = \delta^2$*, it means that*
  $\frac{\tau^2}{\tau^2+\sigma^2} = \frac{\tau^2}{\delta^2} \rightarrow 1 \Rightarrow$ *variability of amounts for a given risk is predominantly due to the heterogeneity of the portfolio.*
  $\frac{\sigma^2}{\tau^2+\sigma^2} = \frac{\sigma^2}{\delta^2} \rightarrow 1 \Rightarrow$ *variability of amounts for a given risk is predominantly due to the average of the variability of amounts in the whole portfolio.*

1.4.4 Estimation of the parameters

The expression of the credibility premium depends only on three parameters : μ, σ^2, and τ^2. Consequently, you need not know or suppose a distribution for the risk parameter Θ_i. We are going to give the estimators for these quantities, and see how to correct the bias when it is the case.
Given m as the number of periods and n as the number of contracts, we get :

Estimator of μ

The natural estimator of μ is

$$\hat{\mu} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij} = \frac{1}{n} \sum_{i=1}^{n} \bar{X}_i$$

This estimator is unbiased.
Proof. cf. appendix, A.4.

Estimator of σ^2

To estimate $s^2(\Theta_i)$, we use the unbiased estimator

$$\frac{1}{m-1}\sum_{j=1}^{m}(X_{ij}-\bar{X}_i)^2$$

That is why we get for σ^2 (simply the mean of the $s^2(\Theta_i)$)

$$\frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij}-\bar{X}_i)^2$$

Proof. cf. appendix, A.5

Estimator of τ^2

The unbiased estimator of the between variance is

$$\hat{\tau}^2=\frac{1}{n-1}\sum_{i=1}^{n}(\bar{X}_i-\hat{\mu})^2-\frac{1}{m}\hat{\sigma}^2$$

This quantity is less natural and we can not find it directly because in fact there is a bias. We correct it in order to have the best etismator.

Proof : cf appendix, A.6.

You can have a numerical example of this model applied to Hachemeister data (defined below) in the appendix section, B.1.
The next section presents the *Bühlmann-Straub* model, and we are going to see it is very similar to this one.

1.5 The Bühlmann-Straub model

The *Bühlmann-Straub* model is a generalization of the Bühlmann model. By the same token we can price credibility premiums except that if we want to give more weight to specific data it is now possible. This is the most used model in the field of the credibility theory as it is easy to apply on real problems and is more general than the Bühlmann model.
Generally, ratios (X_{ij}) are given by the global claim amount of the risk i during the year j over a quantity which represents the collective contribution (typically the number of insured or the salarial mass). We have $X_{ij}=\frac{S_{ij}}{m_{ij}}$. The risk parameter is still represented by the ramdom variable Θ_i.

1.5.1 Notations and assumptions

We need to make some assumptions to be able to calculate the credibility premium. Notice it is nearly the same ones than in Bühlmann model :

- w_{ij} is the weight associated with the ratio X_{ij}, equal to 0 if no data ;
- Θ_i is the non-observable risk parameter of the contract i ;
- $((X_{i1}, X_{i2}, ..., X_{im}, \Theta_i) : 1 \leq i \leq n)$ is the vector of observations ;
- $\mu(\Theta_i) = E[X_{ij}|\Theta_i]$: conditionnal expectation of amount for risk i ;
- $\mu = E[\mu(\Theta_i)]$: expectation of global claim amounts for individual risk i ;
- $s^2(\Theta_i) = w_{ij}Var[X_{ij}|\Theta_i]$: weighted conditionnal variance of amounts, variance within ind. risk ;
- $\sigma^2 = E[s^2(\Theta_i)]$: expectation of the conditionnal variance, average variance within ind. risk ;
- $\tau^2 = Var[\mu(\Theta_i)]$: variance of the conditionnal expectation, variance between ind. risk premiums.

Model Assumptions. (*Bühlmann-Straub model*) :
The risk i is characterized by an individual risk profile which is itself the realization of a random variable Θ_i, and we have :

BS1 : Conditionnally, given Θ_i, the $((X_{i1}, X_{i2}, ..., X_{im}, \Theta_i) : 1 \leq i \leq n)$ are independant with

$$\begin{aligned} E[X_{ij}|\Theta_i] &= \mu(\Theta_i), \\ Var[X_{ij}|\Theta_i] &= \frac{s^2(\Theta_i)}{w_{ij}} \end{aligned}$$

BS2 : the pairs $(\Theta_1,\ X_1)$, $(\Theta_2,\ X_2)$, ..., are independant ; and Θ_1, Θ_2, ... are independant and identically distributed (i.i.d.).

Remark 5. *In the Bühlmann model, the assumption* $Var[X_{ij}|\Theta_i] = s^2(\Theta_i)$ *was not reasonable. Indeed, the variance should depend on the volume measure* V_{ij}*. Here the variance is inversely proportional to a certain volume measure, which seems to be a good idea.*

1.5.2 How to determine the credibility premium?

Thanks to these assumptions and notations, it is easy to see these two results :

$$\begin{aligned} E[X_{ij}] &= E\big[E[X_{ij}|\Theta_i]\big] = E[\mu(\theta_i)] = \mu \\ Var[X_{ij}] &= E[Var[X_{ij}|\Theta_i]] + Var[E[X_{ij}|\Theta_i]] = E[\tfrac{s^2(\theta_i)}{w_{ij}}] + Var[\mu(\theta_i)] = \tfrac{s^2(\theta_i)}{w_{ij}} + \tau^2 \end{aligned}$$

Like in the Buhlmann model, assume that the estimator has a linear form : $\hat{\mu}(\Theta_i) = c_0 + \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} X_{ij}$. We look for c_0 and c_{ij} which minimize the average of the quadratic error

$$E\big[(\hat{\mu}(\Theta_i) - X_{i,j+1})^2\big], \forall i = 1, ..., n$$

Of course, the results of this minimization are the same as in the Bühlmann model except that you have to consider the weights here, included in the credibility factor z_i.

The result of the minimization (with the same proof adapted to take account of weights) is

Theorem 6. *The credibility estimator is* $\hat{\mu}(\Theta_i) = E[X_{i,j+1}|X_{i,1}, ..., X_{i,m}] = z_i\bar{X}_i + (1 - z_i)\mu$

$$\text{where}\quad \begin{aligned} z_i &= \frac{w_i}{w_i + \frac{\sigma^2}{\tau^2}} \\ \bar{X}_i &= \sum_{j=1}^{m} \frac{w_{i,j} X_{i,j}}{w_i} \\ w_i &= \sum_{j=1}^{m} w_{i,j} \end{aligned}$$

Consequently, the credibility premium depends only on $\tau^2 = Var[\mu(\theta_i)]$ and $\sigma^2 = E[s^2(\theta_i)]$. The next section concerns the estimation of these two parameters.

1.5.3 Estimation of the structural parameters

Bühlmann and Straub suggested to use unbiased estimators (the expectation of the estimators is equal to the parameter itself) which had the minimal variance. Indeed, the smaller the variance of the estimator is, the closer to the theoric value and the more reliable the estimator will be.

Estimation of the mean of the within variances, σ^2

The best linear and unbiased estimator for this quantity is

$$E[\hat{s}^2(\theta)] = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m-1} \sum_{j=1}^{m} w_{i,j} \left(X_{i,j} - \bar{X}_i\right)^2$$

Proof : same idea of the proof in the Bühlmann model.

Estimation of the between variance τ^2

We get for this estimator

$$\hat{\tau}^2 = c\left(\frac{n}{n-1} \sum_{i=1}^{n} \frac{w_{i.}}{w_{..}} (X_i - \bar{X})^2 - \frac{n\hat{\sigma}^2}{w_{i.}}\right) \qquad \text{where} \qquad c = \frac{n-1}{n}\left(\sum_{i=1}^{n} \frac{w_{i.}}{w_{..}}(1 - \frac{w_{i.}}{w_{..}})\right)^{-1}$$

Proof : cf appendix, A.7.

Remark 7. *$\hat{\tau}^2$ can possibly be negative, which means that there is no detectable difference between the risks. Naturally in this case we put $\hat{\tau}^2 = 0$. Hence we finally use the estimator*

$$\hat{\tau}^2 = max(0, \hat{\tau}^2)$$

See also Dannenberg & collab. (1996), and Klugman & collab. (1998) for more precision and other estimators. A numerical example of applying this model with Hachemeister data (defined below) is available in appendix, B.2.

1.5.4 Quadratic losses

Analogously as in the introduction of the credibility estimators, we find here that
- the quadratic loss of the collective premium μ_0 is $E[(\mu_0 - \mu(\Theta_i))^2] = Var(\mu(\Theta_i)) = \tau^2$,
- X_i is the best linear and individually unbiased estimator and has the quadratic loss $\sigma^2/w_{i.}$,
- the credibility estimator is again a weighted mean between the collective and the individual estimators,
- the quadratic loss of the credibility estimator is $E\Big[\big(\hat{\mu}(\Theta_i) - \mu(\Theta_i)\big)^2\Big] = (1-z_i)\tau^2 = z_i\frac{\sigma^2}{w_{i.}}$.

The intuitive principle is still valid.

1.6 Hachemeister modelling, the first regression model in credibility

This section is the heart of the job I made during the internship.
When data are affected by time trends, the homogeneity (in time) assumption made by Bühlmann-Straub model, $E[X_{ij}|\Theta_i]$ being independant of j, is not appropriated. That is why Charles Hachemeister "gave birth" to his model (see Hachemeister (1975)), with the idea to fit regression lines to the individual data "observed claim amounts versus time", taking also account of the country-wide data. Then the same question as in the Bühlmann-Straub model arised : how much weight one should assign to the individual regression line and the collective one ?
This model was first presented in 1975 and deals with private passenger bodily injured insurance. Claim amounts for a few U.S.A. states are observed for a number of quarters. They show a tendency to increase in time, due to inflation. We are interested in the state-specific inflation factor, supposed different in all states, hence it forms an heterogeneous collective. The main difference between this model and others is that it is adapted to data with a trend.
In general, time is responsible for an increase of premiums because of exogen parameters such as for example the inflation rate or the interest rate. Using this model, the premium would increase in the future, so it seems to be a well-adapted model to the reality. Predictions should make sense now.

1.6.1 Some generalities on regression models

The general linear model in classical statistics is given by

$$\begin{aligned} \text{X} &= \text{Y}\,\beta + \epsilon, \\ \text{where X} &= \text{obs. vector of dimension n} \times 1, \\ \text{Y} &= \text{known design matrix of dimension n} \times \text{p (p} \leq \text{n)}, \\ \beta &= \text{unknown parameter vector of dimension n} \times 1, \\ \epsilon &= \text{vector of the random deviations of dimension n} \times 1, \\ &\quad \text{with } E[\epsilon] = 0 \text{ and } Cov(\epsilon, \epsilon') = \Sigma. \end{aligned}$$

Remark 8. .
In the case of ordinary least squares, $\Sigma = \sigma^2 \ . \ I$

And if weighted least squares, $\Sigma = \sigma^2 \ . \ W^{-1}$ *where* $W = \begin{pmatrix} w_1 & 0 & \ldots \\ 0 & \ddots & 0 \\ \ldots & 0 & w_n \end{pmatrix}$

Theorem 9. *The best linear unbiased estimator of* β *is given by* $\hat{\beta} = (Y'\Sigma^{-1}Y)^{-1}Y'\Sigma^{-1}X$

And the covariance matrix of $\hat{\beta}$ *is* $\Sigma_{\hat{\beta}} = (Y' \Sigma^{-1} Y)^{-1}$

We get therefore in the case of weighted least squares $\hat{\beta} = (Y'WY)^{-1} Y'WX$ and $\Sigma_{\hat{\beta}} = \sigma^2 \, . \, (Y'WY)^{-1}$
"Best" estimator means that $E\left[\left(\hat{\beta}_i - \beta_i\right)^2\right]$ = minimum for i=1,...,p.

This best estimator can be derivated from the data as the least squares estimator, i.e. in the case of a weighted least squares $\hat{\beta} = argmin_{\beta_1,\ldots,\beta_p} \sum_{j=1}^{n} w_j (X_j - (Y\beta)_j)^2$

1.6.2 The Hachemeister data

The famous data used by Hachemeister are extracted from a practical problem. The data have five different states which represent the contracts and twelve quarters of claim experience, representing the periods. The experience consists of average claim amounts for total private passenger bodily injury insurance from July 1970 until June 1973. Here the subscript j represents the number of the contract and s is the period considered. Obviously, some contracts might have more influence on the overall figures than others, and it's why we also have data which represent the weight of contract j in the period s, denoted by w_{js}.

X_{js}	j = 1	j = 2	j = 3	j = 4	j = 5
s = 1	1738	1364	1759	1223	1456
s = 2	1642	1408	1685	1146	1499
s = 3	1794	1597	1479	1010	1609
s = 4	2051	1444	1763	1257	1741
s = 5	2079	1342	1674	1426	1482
s = 6	2234	1675	2103	1532	1572
s = 7	2032	1470	1502	1953	1606
s = 8	2035	1448	1622	1123	1735
s = 9	2115	1464	1828	1343	1607
s = 10	2262	1831	2155	1243	1573
s = 11	2267	1612	2233	1762	1613
s = 12	2517	1471	2059	1306	1690

w_{js}	j = 1	j = 2	j = 3	j = 4	j = 5
s = 1	7861	1622	1147	407	2902
s = 2	9251	1742	1357	396	3172
s = 3	8706	1523	1329	348	3046
s = 4	8575	1515	1204	341	3068
s = 5	7917	1622	998	315	2693
s = 6	8263	1602	1077	328	2910
s = 7	9456	1964	1277	352	3275
s = 8	8003	1515	1218	331	2697
s = 9	7365	1527	896	287	2663
s = 10	7832	1748	1003	384	3017
s = 11	7849	1654	1108	321	3242
s = 12	9077	1861	1121	342	3425

To be persued of a time-dependance, here is the graph of the Hachemeister data for each contract :

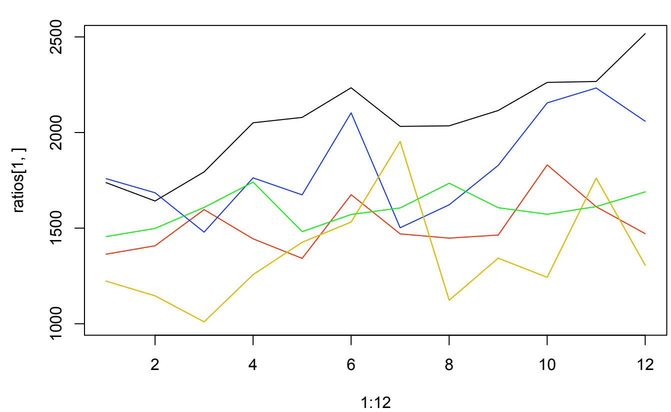

Figure 1.1: Trend-time of the data

This graph clearly shows a trend (it seems to increase with the time), so we can not consider that the risk is time-independant and it is enough to justify the use of this model with these data.

1.6.3 Model assumptions

Two assumptions are necessary to use efficiently the model.

Model Assumptions (*Hachemeister model*) :
The individual risk profile θ_i is the realization of a random variable Θ_i, caracterizing the risk i.

H1 : Given Θ_i, the entries X_{ij}, $j = 1, 2, ..., n$ are independant and $E[X_i|\Theta_i] = Y_i\beta(\Theta_i)$, where

$$\begin{array}{rcl} \beta(\Theta_i) & = & \text{regression vector of length } p \leq n, \text{ with linearly independant components,} \\ Y_i & = & \text{known "design" matrix of rank p,} \\ Cov(X_i, X_i'|\Theta_i) & = & \Sigma_i(\Theta_i). \end{array}$$

H2 : the pairs (Θ_1, X_1), (Θ_2, X_2), ... are independant, and Θ_1, Θ_2, ... are independant and identically distributed.

Remark 10. *First, here are very useful remarks to compute some quantities we are going to use :*
- *the Hachemeister model includes the standard regression as a special case,*
- *it implies that the matrices below are the structural parameters :*

$$S_i = E[\Sigma_i(\Theta_i)] \ \ (i=1,2,...,I)$$

instead of $\sigma^2 W_i^{-1}$ *with only one parameter* σ^2 *in the standard regression case.*

We usually assume $\Sigma_i(\Theta_i) = W_i^{-1/2}\Sigma(\Theta_i)W_i^{-1/2}$, thus

$$S_i = W_i^{-1/2} S W_i^{-1/2} \quad \text{where} \quad W_i^{-1/2} = \begin{pmatrix} w_{i1}^{-1/2} & 0 & ... & 0 \\ 0 & w_{i2}^{-1/2} & 0 & 0 \\ .. & 0 & ... & 0 \\ .. & .. & 0 & w_{in}^{-1/2} \end{pmatrix}$$

This assumption simplifies the problem but we still have more structural parameters than in the regression case because the matrix

$$S = E[\Sigma(\Theta_i)]$$

contains $n(n+1)/2$ structural parameters instead of only one parameter σ^2 in the standard regression model. So as to find the credibility estimator, we first focus on the data compression.

Conditionnaly, given Θ_i, the optimal unbiased estimator resulting from classical statistics (as seen before!) is

$$\beta(\hat{\Theta}_i) = \left(Y_i^{'}\Sigma_i(\Theta_i)^{-1}Y_i\right)^{-1}Y_i^{'}\Sigma_i(\Theta_i)^{-1}X_i$$

and the covariance matrix is $\Sigma_{\beta(\hat{\Theta}_i)} = \left(Y_i^{'}\Sigma_i(\Theta_i)^{-1}Y_i\right)^{-1}$.

But the covariance matrix $\Sigma_i(\Theta_i)$ depends on the unknown Θ_i and is therefore itself unknown.

We suggest to replace the covariance matrix by the structural parameter matrix $S_i = E[\Sigma_i(\Theta_i)]$ because it is the analog of the Buhlmann-Sraub model where the structural parameter in the credibility weight is $\sigma^2 = E[\sigma^2(\Theta_i)]$. Indeed we obtain the optimal data compression as shown in the next theorem.

Theorem 11. *Data compression theorem.*
Under Hachemeister model assumptions, we have :

i) The best linear and individually unbiased estimator of $\beta(\Theta_i)$ based on X_i is

$$\hat{B}_i = \left(Y_i^{'}S_i^{-1}Y_i\right)^{-1}Y_i^{'}S_i^{-1}X_i$$

ii) The quadratic loss matrix of B_i is

$$E[(B_i - \beta(\Theta_i)).(B_i - \beta(\Theta_i))^{'}] = \left(Y_i^{'}S_i^{-1}Y_i\right)^{-1}$$

Proof : see appendix, A.8.

Remark 12. *$\hat{B}_i$ is individually unbiased.*

Proof : see appendix, A.9.

Theorem 13. *Hachemeister formula.*
Under Hachemeister model assumptions, we get that the credibility estimator for $\beta(\Theta_i)$ satisfies

$$\begin{aligned} \beta(\hat{\Theta}_i) &= A_iB_i + (I - A_i)\beta, \qquad \textit{where} \\ A_i &= T\left(T + (Y_i^{'}S_i^{-1}Y_i)^{-1}\right)^{-1}, \\ B_i &= \left(Y_i^{'}S_i^{-1}Y_i\right)^{-1}Y_i^{'}S_i^{-1}X_i, \\ S_i &= E[\Sigma_i(\Theta_i)] = E[Cov(X_i, X_i^{'}|\Theta_i)], \\ T &= Cov(\beta(\Theta_i), \beta(\Theta_i)^{'}), \\ \beta &= E[\beta(\Theta_i)] \end{aligned}$$

The quadratic loss is given by $E[(\beta(\hat{\Theta}_i) - \beta(\Theta_i)).(\beta(\hat{\Theta}_i) - \beta(\Theta_i))^{'}] = (I - A_i)T$

Proof : The quadratic loss can be found thanks to the abstract multidimensional credibility model and the point of view of projections. See Goovaerts & Hoogstad (1987) or Bühlmann & Gisler (2005) for further explanations.

The homogeneous credibility estimator as well as the quadratic loss are given by inserting the structural parameters presented before and the credibility matrix found above. See also Concordia University & Statistics (2007) for a summary on Hachemeister formulae.

1.6.4 The simple linear regression case, a linear trend model

The crucial problem of Hachemeister modelling

Using his model, Hachemeister was faced with a big problem when he discovered that the slope of the credibility regression line was not between the individual and the collective ones.
First, remind us what the simple linear regression model is because it is the most important application of Hachemeister model in actuarial theory.
The simple linear regression model with time as covariable satisfies the following equation :

$$E[X_{ij}|\Theta_i] = \beta_0(\Theta_i) + j\ \beta_1(\Theta_i)$$

where $\beta_0(\Theta_i)$ is the intercept and $\beta_1(\Theta_i)$ the slope.
Assume also that we are in the standard regression case, so

$$S_i = E[Cov(X_i, X_i^{'}|\Theta_i)] = \sigma^2.W_i^{-1}\ ,$$

$$\text{where } W_i = \begin{pmatrix} w_1 & 0 & .. & 0 \\ 0 & w_2 & 0 & .. \\ 0 & 0 & \ddots & 0 \\ 0 & .. & 0 & w_n \end{pmatrix}$$

Although the Hachemeister model in the standard regression case is well-known to predict premiums, we could also imagine a multiple linear regression model, but it is not relevant to use it because time seems to be the only interesting variable.
This model fits exactly into the framework of the credibility regression model. But when Hachemeister used this model to his bodily injured data, he was surprised to see that in some states, results were very stange and opposed to commonsense judgement, see fig. 1.2.
Indeed, the slope of the credibility adjusted regression line was smaller than that in the collective as well as that for the individual risk. It should have been somewhere between the two!
The predictions based on this regression line did not make sense because the trend was not the correct one. They were less than what one would predict using either the individual regression line as well as being less than the predictions one would make using all the states together, i.e. the portfolio. Noone would trust such a prediction!

The solution was found only twenty years later by Bühlmann & Gisler (1997). The trick consists of modelling the intercept in the centre of the time range instead of in the time origin. Results working

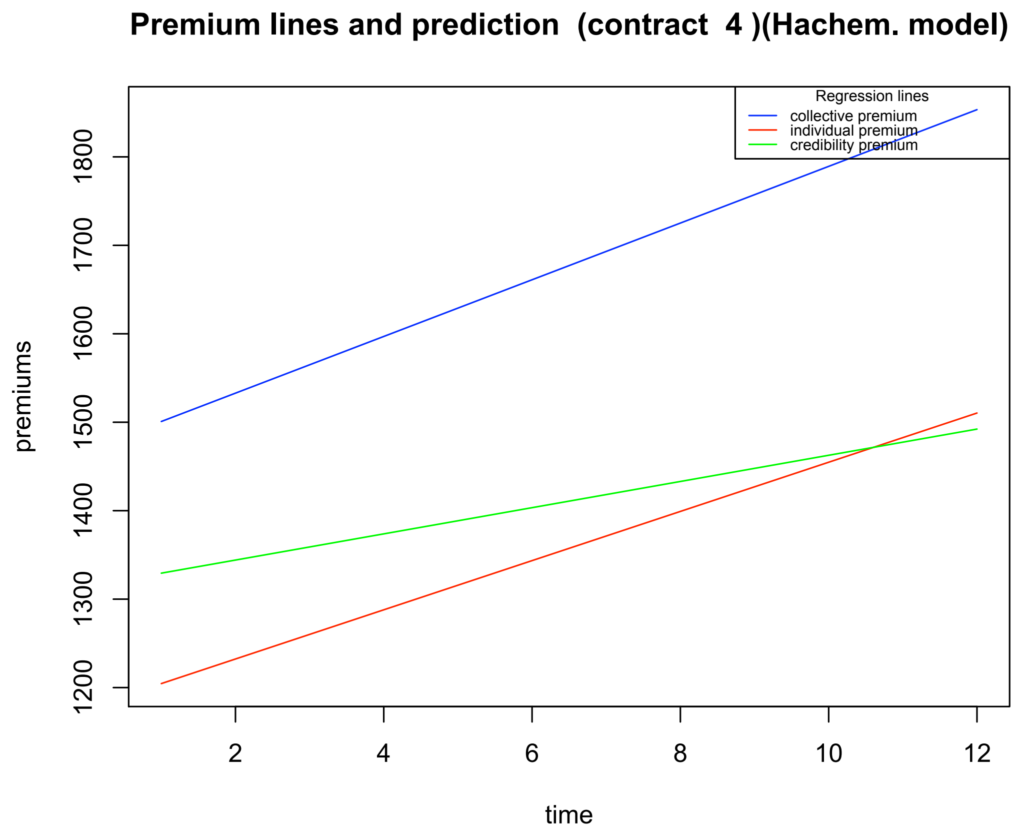

Figure 1.2: Evolution of the premium with intercept at the origin.

in the general case with original model (intercept in the time origin) are very complicated, that is why we will directly focus on the trick that simplifies the calculations.

Modelling with the intercept at the time origin

Consider that the slope and the intercept are independant of each other, which is reasonable, it gives the following form for the matrix T :

$$T = Cov(\beta(\Theta_i), \beta(\Theta_i)') = \begin{pmatrix} \tau_0^2 & 0 \\ 0 & \tau_1^2 \end{pmatrix}$$

We will omit the subscript i in order to simplify the notations. First, we introduce this quantity where Y is called the design matrix

$$V = Y'WY$$

Let us define some notations and other useful quantities : we regard the relative weights $w_j/w_.$ as "sample weights" and denote with $E^{(s)}$ and $Var^{(s)}$ the moments with respect to this distribution.

Then, we have :
- $E^{(s)}[j] = \sum_{j=1}^{n} j \frac{w_j}{w_.}$
- $E^{(s)}[X_j] = \sum_{j=1}^{n} X_j \frac{w_j}{w_.}$
- $Var^{(s)}[j] = \sum_{j=1}^{n} j^2 \frac{w_j}{w_.} - \left(\sum_{j=1}^{n} j \frac{w_j}{w_.}\right) = E^{(s)}[j^2] - \left(E^{(s)}[j]\right)^2$

We get

$$\begin{aligned} V &= \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & n \end{pmatrix} \begin{pmatrix} w_1 & 0 & .. & 0 \\ 0 & w_2 & 0 & .. \\ 0 & 0 & .. & 0 \\ 0 & .. & 0 & w_n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ . & . \\ . & . \\ 1 & n \end{pmatrix} \\ &= w_. \begin{pmatrix} \sum_{j=1}^{n} \frac{w_j}{w_.} & \sum_{j=1}^{n} j \frac{w_j}{w_.} \\ \sum_{j=1}^{n} j \frac{w_j}{w_.} & \sum_{j=1}^{n} j^2 \frac{w_j}{w_.} \end{pmatrix} \\ &= w_. \begin{pmatrix} 1 & E^{(s)}[j] \\ E^{(s)}[j] & E^{(s)}[j^2] \end{pmatrix} \end{aligned}$$

Now, we can calculate the best linear and individually unbiased estimator of β, applying the theorem of Hachemeister formula :

$$\begin{aligned} B &= V^{-1} Y' W X \\ &= \begin{pmatrix} 1 & E^{(s)}[j] \\ E^{(s)}[j] & E^{(s)}[j^2] \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{n} \frac{w_j}{w_.} X_j \\ \sum_{j=1}^{n} \frac{w_j}{w_.} j X_j \end{pmatrix} \\ &= \frac{1}{Var^{(s)}[j]} \begin{pmatrix} E^{(s)}[j^2] & -E^{(s)}[j] \\ -E^{(s)}[j] & 1 \end{pmatrix} \begin{pmatrix} E^{(s)}[X_j] \\ E^{(s)}[jX_j] \end{pmatrix} \\ &= \frac{1}{Var^{(s)}[j]} \begin{pmatrix} E^{(s)}[j^2] E^{(s)}[X_j] - E^{(s)}[j] E^{(s)}[jX_j] \\ E^{(s)}[jX_j] - E^{(s)}[j] E^{(s)}[X_j] \end{pmatrix} \end{aligned}$$

What about the credibility matrix A ?
The general expression of the credibility matrix is $A = T(T + \sigma^2 V^{-1})^{-1}$, which can be rewritten differently to reduce the calculations :

$$\begin{aligned} A &= T(T + \sigma^2 V^{-1})^{-1} \\ &= (I + \sigma^2 V^{-1} T^{-1})^{-1} \\ &= (V + \sigma^2 T^{-1})^{-1} V \end{aligned}$$

For the inverse of T, it is very simple as T is diagonal and we get $T^{-1} = \frac{1}{\tau_0^2 \tau_1^2} \begin{pmatrix} \tau_1^2 & 0 \\ 0 & \tau_0^2 \end{pmatrix}$

So, $T^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} \kappa_0 & 0 \\ 0 & \kappa_1 \end{pmatrix}$ where $\kappa_0 = \frac{\sigma^2}{\tau_0^2}$, $\kappa_1 = \frac{\sigma^2}{\tau_1^2}$ and $\kappa_2 = \frac{\sigma^2}{\tau_2^2}$

We can now write $V + \sigma^2 T^{-1} = \begin{pmatrix} w_. + \kappa_0 & w_. E^{(s)}[j] \\ w_. E^{(s)}[j] & w_. E^{(s)}[j^2] + \kappa_1 \end{pmatrix}$

So,

$$(V+\sigma^2T^{-1})^{-1} = \frac{1}{N}\begin{pmatrix} w_.E^{(s)}[j^2]+\kappa_1 & -w_.E^{(s)}[j] \\ -w_.E^{(s)}[j] & w_.+\kappa_0 \end{pmatrix}$$

where $N = (w_.+\kappa_0)\big(w_.E^{(s)}[j^2]+\kappa_1\big) - (w_.E^{(s)}[j])^2 = w_.^2Var^{(s)}[j] + \kappa_0 w_.E^{(s)}[j^2] + w_.\kappa_1 + \kappa_0\kappa_1$

Finally, we get for the credibility matrix A

$$A = \frac{w_.}{N}\begin{pmatrix} w_.Var^{(s)}[j]+\kappa_1 & \kappa_1 E^{(s)}[j] \\ \kappa_0 E^{(s)}[j] & w_.Var^{(s)}[j]+\kappa_0 E^{(s)}[j^2] \end{pmatrix}$$

Conclusion.
The credibility matrix is not diagonal and this is the source of all the unpleasantness : in fact that is why this regression can lead to implausible results as we have seen before. The interpretation of the parameters is also difficult (with respect to Bühlmann-Straub results) and their effect is hard to predict. Now we understand why this model can not be applied in some situations and we are going to apply the "trick" in the next section in order to see the change.

Modelling with the intercept at the barycenter of time

What is really amazing using the trick is that you have explicit formulas to compute the credibility estimators in the case of simple linear regression, so it is very simple to apply. There may also exist a recurrence which permit us to find a generic formula, valid in other cases of regression too. For a simple linear regression, the regression equation becomes

$$\mu_j(\Theta) = \beta_0(\Theta) + (j - E^{(s)}[j]).\beta_1(\Theta)$$

and the design matrix is now

$$Y = \begin{pmatrix} 1 & 1-E^{(s)}[j] \\ 1 & 2-E^{(s)}[j] \\ .. & .. \\ 1 & n-E^{(s)}[j] \end{pmatrix}$$

Everybody knows that a linear transformation of the time axis has no effect on the credibility estimator because it is a simple translation. Here $\beta_0(\Theta_i)$ has a slightly different meaning as it is the intercept at the "centre of gravity" of the time variable. We model therefore a "new" vector and assume that the covariance matrix T is still diagonal with respect to it.
Repeating the same calculations, we get for the individual estimators of vector B

$$B = \begin{pmatrix} E^{(s)}[X_j] \\ \dfrac{Cov^{(s)}(j,X_j)}{Var^{(s)}(j)} \end{pmatrix}$$

and for the credibility matrix A

$$A = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}, \text{ with } a_{11} \text{ and } a_{22} \text{ defined below}$$

Thanks to the fact that A is diagonal, the interesting result is that you can split the credibility formula in the simple linear regression model into the two following one-dimensionnal credibility formulae :

$$\hat{\beta}_0(\Theta) = a_{11}B_0 + (1 - a_{11})\beta_0$$
$$\hat{\beta}_0(\Theta) = a_{22}B_1 + (1 - a_{22})\beta_1$$

where β_0 and β_1 are the collective coefficients from the regression and

$$B_0 = E^{(s)}[j] \qquad B_1 = \frac{Cov^{(s)}(j, X_j)}{Var^{(s)}[j]},$$
$$a_{11} = \frac{w_{\cdot}}{w_{\cdot} + \frac{\sigma^2}{\tau_0^2}} \qquad a_{22} = \frac{w_{\cdot} Var^{(s)}[j]}{w_{\cdot} Var^{(s)}[j] + \frac{\sigma^2}{\tau_1^2}}$$

and τ_0^2, τ_1^2 are defined below.

Notice then that these formulae are equivalent to those of Bühlmann-Straub model. Indeed, the credibility weights a_{11} and a_{22} have the same simple form. What is really "beautiful" is that we have splitted the two-dimensionnal problem into two separated one-dimensionnal problems : one to estimate the intercept ($\hat{\hat{\beta}}_0(\Theta)$) and the other to estimate the slope ($\hat{\hat{\beta}}_1(\Theta)$). This technique can also be used with n-dimensionnal problems, but calculations are extremly complicated because of the products of matrix. To simplify it, the solution would be to change the design matrix as above in order to make the credibility matrix become diagonal. But there does not exist one simple and explicit multidimensionnal formula for credibility factors a_{ii}.

We can use directly in practice these formulae, furthermore these results are intuitive and make sense. We are sure that the intercept and the slope of the credibility regression line lie between the corresponding values for the individual and collective ones, which was the big problem.

Remark 14. *: it is now possible to understand the impact of each structural parameter :*

- τ_1^2 *small compared to* $\sigma^2/\big(w_{\cdot}Var^{(s)}[j]\big) \Rightarrow$ *the slope of the credibility regression line lies close to the slope of the collective line because* $a_{22} \to 0$,
- *if we have the opposite phenomenom, of course the slope of the credibility regression line will lie close to the slope of the individual line, as* $a_{22} \to 1$,
- *we can observe the same results with the intercept.*
- *we took the centre of gravity of the collective (and not for each contract), whose form is* $j_0 = \sum_{j=1}^{n} \frac{w_{\cdot j}}{w_{\cdot\cdot}} j$
  The difference between the result taking the centre of gravity of the collective instead of taking the individual centres of gravity (which gives the exact credibility) is negligible in most cases.

Estimators

Of course we need the estimators of quantities such as the collective intercept and slope, the individual ones, and structural parameters appearing in a_{11} and a_{22}.
For the collective intercept and slope, we use a classic and natural result : the mean of the individual ones given by the regression coefficients. Structural parameters are estimated analogously to those in the Bühlmann-Straub model.

Estimation of the within variance σ^2

Consider

$$\hat{\sigma}_i^2 = \frac{1}{n-2}\sum_{j=1}^{n} w_{ij}(X_{ij} - \hat{\mu}_{ij})^2$$

where $\hat{\mu}_{ij}$ = fitted values from the individual regression lines, given by the result in **R**.

These estimators for the $\sigma^2(\Theta_i)$ are conditionally unbiased, so the natural estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{I}\sum_{i=1}^{I}\hat{\sigma}_i^2$$

Estimation of the between variance τ^2

From the expression of the design matrix and the definition of the covariance matrix, we get that the conditional covariance matrix of the individual regression parameters with time origin at the individual centre of gravity is given by

$$Cov(B_i, B_i^{'}|\Theta_i) = \sigma^2(\Theta_i)\begin{pmatrix} w_{i.} & 0 \\ 0 & w_{i.}d_i \end{pmatrix}^{-1}$$

where $\quad d_i = Var^{(s_i)}(j) = \sum_{j=1}^{n}\frac{w_{ij}}{w_{i.}}(j - j_0^{(i)})^2 \quad$ and $\quad j_0^{(i)} = \sum_{j=1}^{n}\frac{w_{ij}}{w_{i.}}j.$

The individual centres of gravity $j_0^{(i)}$ vary little from one risk to another, so it is possible to use the collective centre of gravity without loosing all these results.

We notice that the conditional variance of the elements of B_i have the same structure as the conditional variance of X_i in the Bühlmann-Straub model, that is why we take the same technique of estimation here, hence

$$\hat{\tau}_0^2 = c_0\left(\frac{I}{I-1}\sum_{i=1}^{I}\frac{w_{i.}}{w_{..}}(B_{0i} - \bar{B}_0)^2 - \frac{I\hat{\sigma}^2}{w_{..}}\right)$$
$$\hat{\tau}_1^2 = c_1\left(\frac{I}{I-1}\sum_{i=1}^{I}\frac{w_{i.}^*}{w_{..}^*}(B_{1i} - \bar{B}_1)^2 - \frac{I\hat{\sigma}^2}{w_{..}^*}\right)$$

using $c_0 = \frac{I-1}{I}\left(\sum_{i=1}^{I}\frac{w_{i.}}{w_{..}}(1 - \frac{w_{i.}}{w_{..}})\right)^{-1}$, $c_1 = \frac{I-1}{I}\left(\sum_{i=1}^{I}\frac{w_{i.}^*}{w_{..}^*}(1 - \frac{w_{i.}^*}{w_{..}^*})\right)^{-1}$ and $\bar{B}_1 = \sum_{i=1}^{I}\frac{w_{i.}^*}{w_{..}^*}B_{1i}$

Notation.
$w_{i.}^* = d_i w_{i.}$

A lot of well-known articles and books are available on this topic : see Cossette & Luong (2003) and Ho Lo & Zhu (2006).

Results

Let us visualize the graphical result and notice that the credibility regression line now lies between the collective and the individual ones. It seems to work very well.
A numerical example is available in appendix, B.3.

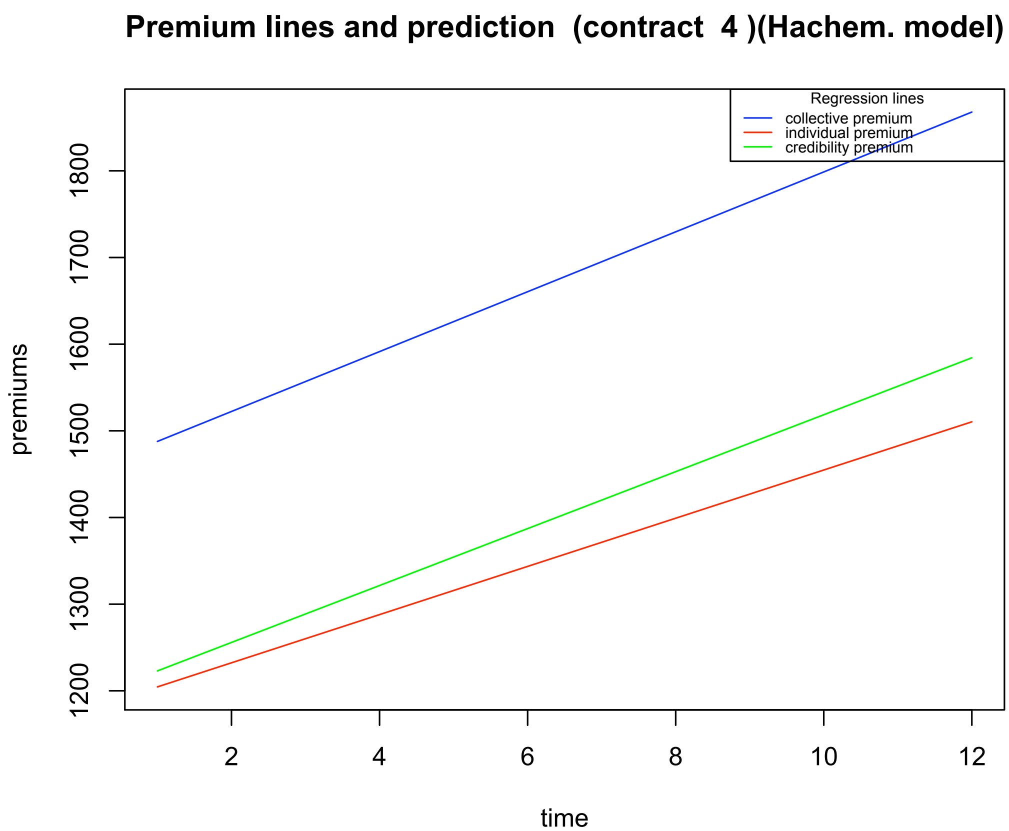

Figure 1.3: Evolution of the premium with intercept at the barycentre of time.

1.6.5 What about non-linear regressions ?

Here the topic is to wonder if we could generalize the Hachemeister model in the simple linear regression case to the case of a quadratic regression for example. Indeed, one can have data with a special trend, like a parabol, and so the model studied above should not correspond perfectly to the aspect of data. A classic regression could be applied but we can expect that predictions will not be the best ones, but this is the main issue for actuars. For a better understanding of linear models, see McCulloch & Searle (2008) and Radhakrishna Rao & Toutenburg (1999)
If we consider a quadratic regression, a lot of questions remain :

- can we apply a quadratic regression model with Hachemeister hypothesis ?
- can we apply a non-simple quadratic regression model ?

- can we find explicit formulae for credibility factors (or credibility weights) ?
- do we have the same problem as in the simple linear model for predictions if we do not adjust the regression ?
- what is the “new” centre of gravity if we adjust our regression ?
- what is the factor d_i to adjust weights in the credibility weight formula ?

We are going to calculate once more the essential quantities to answer all these questions, and see if we can generalize our results to every type of regression. Then we will illustrate the concrete benefit of using such a model for predictions.

1.6.6 The example of a quadratic regression model

It is crucial to notice that the time evolution of the data is not negligible in choosing our model : if data show clearly a linear trend, apply a quadratic model would not be efficient as results can look like very stange (for example we saw that without adjustment curve is the opposite than with adjustment).

Without adjustment of the regression

Consider the quadratic regression model $E[X_{ij}|\Theta_i] = \beta_0(\Theta_i) + j\ \beta_1(\Theta_i) + j^2\ \beta_2(\Theta_i)$

The design matrix is now $Y = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ .. & 3 & 9 \\ .. & .. & .. \\ 1 & n & n^2 \end{pmatrix}$

Remember $V = Y^{'}WY$ (where W represents the matrix of weights), we get

$$\begin{aligned} V &= \begin{pmatrix} 1 & 1 & .. & 1 \\ 1 & 2 & .. & n \\ 1 & 4 & .. & n^2 \end{pmatrix} \begin{pmatrix} w_1 & 0 & .. & 0 \\ 0 & w_2 & 0 & .. \\ 0 & 0 & .. & 0 \\ 0 & .. & 0 & w_n \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ .. & 3 & 9 \\ .. & .. & .. \\ 1 & n & n^2 \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^n w_j & \sum_{j=1}^n jw_j & \sum_{j=1}^n j^2w_j \\ \sum_{j=1}^n jw_j & \sum_{j=1}^n j^2w_j & \sum_{j=1}^n j^3w_j \\ \sum_{j=1}^n j^2w_j & \sum_{j=1}^n j^3w_j & \sum_{j=1}^n j^4w_j \end{pmatrix} \\ &= w_. \begin{pmatrix} 1 & E^{(s)}[j] & E^{(s)}[j^2] \\ E^{(s)}[j] & E^{(s)}[j^2] & E^{(s)}[j^3] \\ E^{(s)}[j^2] & E^{(s)}[j^3] & E^{(s)}[j^4] \end{pmatrix} \end{aligned}$$

Consider also that we have the same hypothesis,

- $S_i = \sigma^2 W_i^{-1}$
- $T = Cov(\beta(\Theta_i), \beta(\Theta_i)^{'}) = \begin{pmatrix} \tau_0^2 & 0 & 0 \\ 0 & \tau_1^2 & 0 \\ 0 & 0 & \tau_2^2 \end{pmatrix}$

We are now able to compute the best unbiased estimators of β_p in the regression model.

We know that $B = V^{-1}Y'WX$ and that $Y'WX = \begin{pmatrix} \sum_{j=1}^n w_j X_j \\ \sum_{j=1}^n w_j j X_j \\ \sum_{j=1}^n w_j j^2 X_j \end{pmatrix}$

Thus $B = \begin{pmatrix} 1 & E^{(s)}[j] & E^{(s)}[j^2] \\ E^{(s)}[j] & E^{(s)}[j^2] & E^{(s)}[j^3] \\ E^{(s)}[j^2] & E^{(s)}[j^3] & E^{(s)}[j^4] \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^n \frac{w_j}{w_.} X_j \\ \sum_{j=1}^n \frac{w_j}{w_.} j X_j \\ \sum_{j=1}^n \frac{w_j}{w_.} j^2 X_j \end{pmatrix}$

Let us calculate the inverse of this matrix, first we begin by determining the determinant thanks to the Sarrus rule :

$$\begin{aligned} |V| &= E^{(s)}[j^2]E^{(s)}[j^4] + E^{(s)}[j]E^{(s)}[j^2]E^{(s)}[j^3] + E^{(s)}[j]E^{(s)}[j^2]E^{(s)}[j^3] - (E^{(s)}[j^2])^3 \\ &- (E^{(s)}[j^3])^2 - (E^{(s)}[j])^2E^{(s)}[j^4] \\ &= E^{(s)}[j^2]E^{(s)}[j^4] + 2E^{(s)}[j]E^{(s)}[j^2]E^{(s)}[j^3] - (E^{(s)}[j^2])^3 - (E^{(s)}[j^3])^2 - (E^{(s)}[j])^2E^{(s)}[j^4] \end{aligned}$$

We tried to simplify this expression but it was impossible to have a great form for it. Then we calculate the cofactor-matrix to get to the final right form of V^{-1} :

$$\frac{1}{|V|}\begin{pmatrix} E^{(s)}[j^2]E^{(s)}[j^4] - (E^{(s)}[j^3])^2 & E^{(s)}[j^2]E^{(s)}[j^3] - E^{(s)}[j]E^{(s)}[j^4] & E^{(s)}[j]E^{(s)}[j^3] - (E^{(s)}[j^2])^2 \\ E^{(s)}[j^2]E^{(s)}[j^3] - E^{(s)}[j]E^{(s)}[j^4] & E^{(s)}[j^4] - (E^{(s)}[j^2])^2 = Var^{(s)}[j^2] & E^{(s)}[j]E^{(s)}[j^2] - E^{(s)}[j^3] \\ E^{(s)}[j]E^{(s)}[j^3] - (E^{(s)}[j^2])^2 & E^{(s)}[j]E^{(s)}[j^2] - E^{(s)}[j^3] & E^{(s)}[j^2] - (E^{(s)}[j])^2 = Var^{(s)}[j] \end{pmatrix}$$

Note that this matrix is symetric.
We could now get the expression of B to see if we could nullify some terms of the matrix and finally get the adjustment to do. But it is not the most important because we can have an accurate numerical estimation of B by extracting the results of the regression in **R**. It is better to have the expression of credibility matrices in order to get the credibility weights and see weither it is diagonal or not, and what are the conditions to get it diagonal ?
Indeed, to decompose this three-dimensionnal problem into three one-dimensionnal problems, the credibility matrix A has to be diagonal if we want to adjust the regression later.
The general expression of the credibility matrix is $A = T(T + \sigma^2 V^{-1})^{-1}$, which can be rewritten differently to reduce the calculations :

$$\begin{aligned} A &= T(T + \sigma^2 V^{-1})^{-1} \\ &= (I + \sigma^2 V^{-1} T^{-1})^{-1} \\ &= (V + \sigma^2 T^{-1})^{-1} V \end{aligned}$$

For the inverse of T, it is very simple as T is diagonal and we get $T^{-1} = \frac{1}{\tau_0^2\tau_1^2\tau_2^2}\begin{pmatrix} \tau_1^2\tau_2^2 & 0 & 0 \\ 0 & \tau_0^2\tau_2^2 & 0 \\ 0 & 0 & \tau_0^2\tau_1^2 \end{pmatrix}$

which yields to $T^{-1} = \frac{1}{\sigma^2}\begin{pmatrix} \kappa_0 & 0 & 0 \\ 0 & \kappa_1 & 0 \\ 0 & 0 & \kappa_2 \end{pmatrix}$, where $\kappa_0 = \frac{\sigma^2}{\tau_0^2}$, $\kappa_1 = \frac{\sigma^2}{\tau_1^2}$ and $\kappa_2 = \frac{\sigma^2}{\tau_2^2}$

Let us denote $(V + \sigma^2 T^{-1})^{-1}$ by Q^{-1}, then $Q^{-1} = \begin{pmatrix} w_. + \kappa_0 & w_.E^{(s)}[j] & w_.E^{(s)}[j^2] \\ w_.E^{(s)}[j] & w_.E^{(s)}[j^2] + \kappa_1 & w_.E^{(s)}[j^3] \\ w_.E^{(s)}[j^2] & w_.E^{(s)}[j^3] & w_.E^{(s)}[j^4] + \kappa_2 \end{pmatrix}^{-1}$

After a lot of calculations,

$$Q^{-1} = \frac{1}{det(Q)} \times$$

$$\begin{pmatrix} (w_. E^{(s)}[j^2] + \kappa_1)(w_. E^{(s)}[j^4] + \kappa_2) - w_.^2 (E^{(s)}[j^3])^2 & w_.^2 E^{(s)}[j^2] E^{(s)}[j^3] - w_. E^{(s)}[j](w_. E^{(s)}[j^4] + \kappa_2) & w_.^2 E^{(s)}[j] E^{(s)}[j^3] - w_. E^{(s)}[j^2](w_. E^{(s)}[j^2] + \kappa_1) \\ w_.^2 E^{(s)}[j^2] E^{(s)}[j^3] - w_. E^{(s)}[j](w_. E^{(s)}[j^4] + \kappa_2) & (w_. + \kappa_0)(w_. E^{(s)}[j^4] + \kappa_2) - w_.^2 (E^{(s)}[j^2])^2 & w_.^2 E^{(s)}[j] E^{(s)}[j^2] - w_. E^{(s)}[j^3](w_. + \kappa_0) \\ w_.^2 E^{(s)}[j] E^{(s)}[j^3] - w_. E^{(s)}[j^2](w_. E^{(s)}[j^2] + \kappa_1) & w_.^2 E^{(s)}[j] E^{(s)}[j^2] - w_. E^{(s)}[j^3](w_. + \kappa_0) & (w_. + \kappa_0)(w_. E^{(s)}[j^2] + \kappa_1) - w_.^2 (E^{(s)}[j])^2 \end{pmatrix}$$

Because $A = Q^{-1}V$, we finally get

$$A = \frac{w_.}{det(Q)} \times$$

$$\begin{pmatrix} a_{11} & w_. \kappa_1 (E^{(s)}[j] E^{(s)}[j^4] - E^{(s)}[j^2] E^{(s)}[j^3]) + \kappa_1 \kappa_2 E^{(s)}[j] & w_. \kappa_2 ((E^{(s)}[j^2])^2 - E^{(s)}[j] E^{(s)}[j^3]) + \kappa_1 \kappa_2 E^{(s)}[j^2] \\ w_. \kappa_0 (E^{(s)}[j] E^{(s)}[j^4] - E^{(s)}[j^2] E^{(s)}[j^3]) + \kappa_0 \kappa_2 E^{(s)}[j] & a_{22} & w_. \kappa_2 (E^{(s)}[j^3] - E^{(s)}[j] E^{(s)}[j^2]) + \kappa_0 \kappa_2 E^{(s)}[j^3] \\ w_. \kappa_0 ((E^{(s)}[j^2])^2 - E^{(s)}[j] E^{(s)}[j^3]) + \kappa_0 \kappa_1 E^{(s)}[j^2] & w_. \kappa_1 (E^{(s)}[j^3] - E^{(s)}[j] E^{(s)}[j^2]) + \kappa_0 \kappa_1 E^{(s)}[j^3] & a_{33} \end{pmatrix}$$

Looking at this matrix, we felt there could exist a generic formula for a_{ij} and you also felt that this formula could be extend to other dimensions. But the problem remains : it looks impossible to nullify non-diagonal terms with only one condition over moments $E^{(s)}[j^k]$. This condition could have given us the idea about the adjustment to do in the regression. Another way to find would be to change the design matrix and calculate it again.
Anyway it is not an automatic proceeding so it does not seem to be he right way. Moreover, results are not so accurate as we could expect, certainly because some assumptions are not respected. It is the case for the covariance matrix of B which is supposed to be diagonal but which is not diagonal in practice.

On a graphical point of view, let us see the evolution of the credibility premium compared to the individual and the collective ones and notice that we have the same problem than with the simple linear regression, predicitions are obviously insignificant, see fig. 1.4 on the next page.

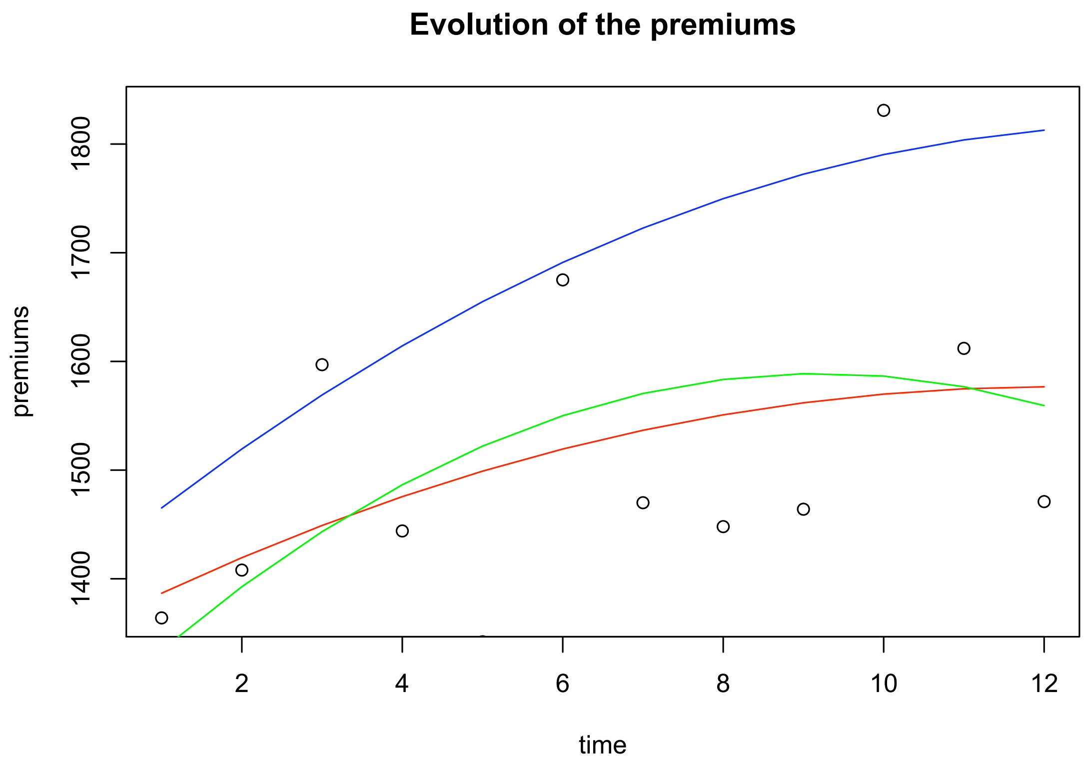

Figure 1.4: Evolution of the premium without adjustment.

With adjustment of the regression

In fact, we discovered that the mathematical key was the orthogonalization of the design matrix. First, we thought that the trick was to adjust the regression with the barycentre of the time, so that the results would be easy to calculate and all the matrices (covariance, credibility...) would be diagonal.
But we realized that the problem was difficult to solve when we took the most general case where you have the following regression equation

$$E[X_{ij}|\Theta_i] = \beta_0(\Theta_i) + j.\beta_1(\Theta_i) + ... + j^p.\beta_p(\Theta_i)$$

How could we find the centre of gravity in such a situation ?

Mathematically, the real trick is to make the columns of the design matrix Y be mutually orthogonal. It is what we did (but did not realize first) by taking

$$Y = \begin{pmatrix} 1 & 1 - E^{(s)}[j] \\ 1 & 2 - E^{(s)}[j] \\ .. & .. \\ 1 & n - E^{(s)}[j] \end{pmatrix}$$

Indeed, the inner product of the two columns is zero. By consequence the credibility matrix A is diagonal and splitting the n-dimensionnal problem into n one-dimensionnal problems is possible. Besides statistics tell us, when the columns of the design matrix are mutually independant, that the estimators for the fixed parameters β_p are stochastically uncorrelated (conditionnal on Θ).
One method could be to orthogonalize the design matrix by the Gram-Schmidt algorithm and then to make all the calculations. We just have to change the basis (to come back at the one of origin) before predicting. It is the algorithm now implemented in the package **actuar**. Have a look to Christensen (n.d.) for further explanations.
But how is the QR decomposition done with weighted matrix ?

In a general framework, we look for a QR decomposition of the matrix X such as to get Q orthogonal and R a squared-upper-triangular matrix, e.g. $Q^{'}WQ = I$, where $Q^{'}$ is the transpose matrix of Q and W the matrix of weights.
We have to get X = QR. The idea is to take a weighted Q, denoted by Qw, and a weighted matrix R, denoted by Rw, which satisfy the condition of orthogonality for Qw.
Let us define $Xw = X/\sqrt{W}$, $Qw = Q/\sqrt{W}$ and $Rw = R$, then we have

$$\begin{aligned} Qw^{'}WQw &= Q^{'}/\sqrt{W} \times W \times Q/\sqrt{W} = Q^{'}Q = I \\ Xw &= QwRw = Q/\sqrt{W} \times R \Rightarrow Xw \times \sqrt{W} = QR \Rightarrow X = QR \end{aligned}$$

(the R code is available in appendix, B.5).

The graphical result in the case of an orthogonalization of the matrix is the following :

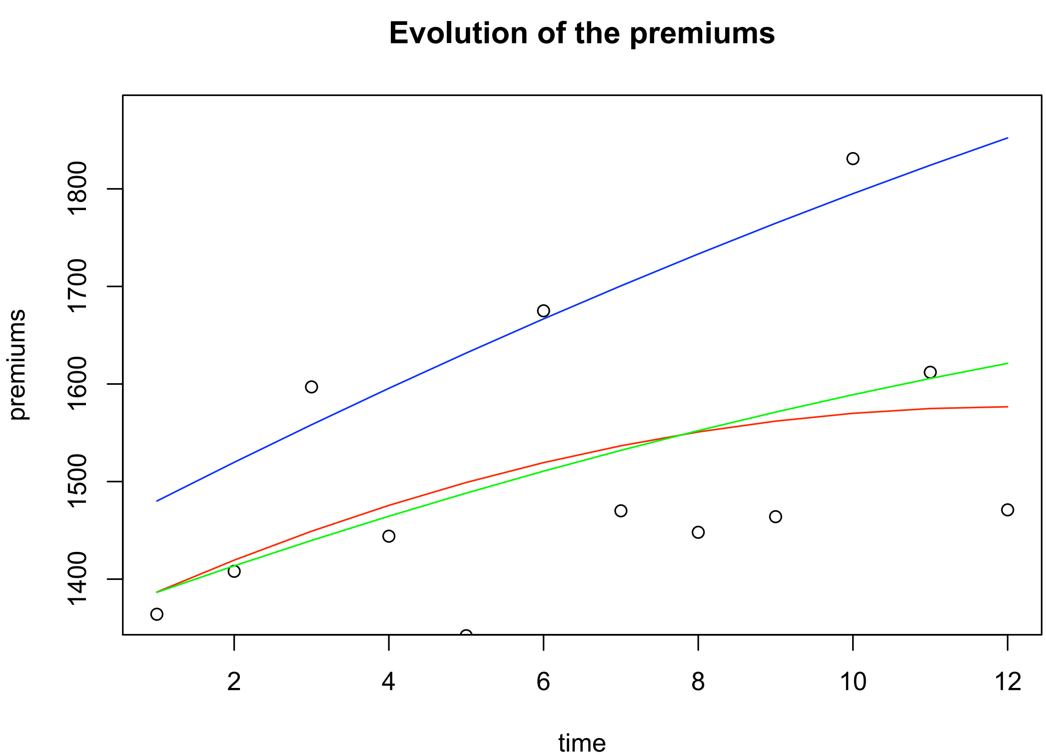

Figure 1.5: Evolution of the premium with orthogonalization of the design matrix.

What a beautiful result : it lies between the two lines. Now, a user of **actuar** is able to fit a Hachemeister model to his data with any regression, and see the prediction of the premiums for each contract. But we must not forget that a study of the trend of the data at the beginning is really important to choose the regression model to apply.

1.7 Conclusion

This general study on credibility models has been the opportunity to discover a new way of predicting insurance premiums, and furthermore the most applied models used by actuaries, even if these are not used in practice yet, because of their difficulties. I really enjoyed it because it helped me to understand the basics of the credibility theory, which I had heard of but which I did not understand before. I had not understood that it was a different thoughout comparing to classical statistics which necessit choosing a model at the beginning so as to do estimations and predictions. After a period of theory, it was almost programming in the package **actuar** to incorpore all the possibilities (features) of regression models and also to fix some bugs in the Bühlmann and Bühlmann-Straub models.

Chapter 2

The heterogeneity of the portfolio

2.1 Introduction

The heterogeneity of the portfolio is a very important information for actuaries to manage their company because it can be the origin of various decisions. Let us begin with a short definition of the word *heterogeneity* (ref. *www.wordreference.com*):

Definition. *heterogeneity*
-noun
the quality or state of being heterogeneous ; composition from dissimilar parts; disparateness.

Everybody knows that one can reduce his risk by diversification. Therefore we could think that the more diversificated the portfolio is, the less risky it is. Obviously this principle remains true in this context, but it also implies that one applies the credibility models with many “small” classes, and the question of the effectiveness of these techniques in this case remains. An actuary can decide weither he prefers to favor the homogeneity of his portfolio and an easier management or to have an heterogeneous portfolio which represents better the real data but which is more complicated to manage.
Sometimes, one prefers to keep a certain homogeneity because of the specialization of the company, the heterogeneity could be the source of a bad choice of the risk parameters. In other words, when the actuary knows the composition of his portfolio and its homogeneity, he is able to realize for example the perspicacity of his choice of the risk parameters.

2.2 The approach

In fact, this part of the internship was more a discussion on how to represent the heterogeneity of the portfolio than a choice of the parameters to represent it.
Indeed, it is clear that the parameters we are going to use are the *between variance* and the *within variance* because these are the best description of the difference between contracts within the portfolio.

It is also these parameters which are used to calculate the credibility factors. If we look at the expression of the credibility factor in the Bühlmann-Straub model (generalization of the Bühlmann model),

$$z_i = \frac{w_i}{w_i + \frac{\sigma^2}{\tau^2}} \ ,$$

and which of the Hachemeister model with intercept at the barycenter of time,

$$\begin{aligned} a_{11} &= \frac{w_.}{w_. + \frac{\sigma^2}{\tau_0^2}}, \\ a_{22} &= \frac{w_. Var^{(s)}[j]}{w_. Var^{(s)}[j] + \frac{\sigma^2}{\tau_1^2}} \end{aligned}$$

we realize that the factor σ^2/τ^2 (where σ^2 is the average of the within variances and τ^2 is the between variance), which represents a kind of percentage of homogeneity, is very important on the interpretation of the credibility factor z.
As we have said before, it is obvious that :

- if $\sigma^2/\tau^2 \to 1$, the weight will be perfectly shared between the individual experience and the collective one,
- if $\sigma^2 << \tau^2$ alors $z_i \to 1$, which means that the individual experience is more relevant and we need not consider a lot the collective experience,
- if $\sigma^2 >> \tau^2$ alors $z_i \to 0$, then the collective experience is the major factor to consider.

By consequence, we have to show graphically these two quantities, where actually the within variance is a measure of dispersion between the observations in each contract and the between variance is the dispersion between the averages of the observations of each contract. In case of regression, it is the the dispersion between the regression coefficients of each contract (intercept and slope).

The problem we faced with was that there did not exist a function in `R` with which we could directly plot the information we wanted to. Bar plots, pie charts, ..., did not correspond to our criterions since we wanted to be able to show all the contracts together and furthermore with a measure of dispersion between a certain figure.
In fact, there exists a method implemented to do this job, except that it computes his measure of dispersion differently as us and uses it in the plot. So we had to adapt this function to our data and set parameters specifically to obtain what we wanted, see Murell (2005) and Cleveland (1993) for further information on R graphics. The work was almost programming the interface and the appearance of the homogeneity, taking account that sometimes you can have several "levels" with a different heterogeneity. It took time but there is nothing else about theory to add. We decided to show the two following main information : the individual mean will be represented by a red point for each contract and the within-variance of the contracts will be represented by the length of the "whiskers". Hence, the vertical space between red points is naturally the between-variance of the contracts.
Let us see in the next section the final graphical result on an example.

2.3 An example with Hachemeister data

We work with the Hachemeister data because we already know all the quantities with it.
We could first compare the estimators of the homogeneity in two different models like the Bühlmann-Straub one and the Hachemeister one. Of course, it is not exactly the same information which is represented because it is the homogeneity of regression coefficients in the Hachemeister model, but it gives a comparison and permit us to realize that the way of computing them is almost the same one. See fig. 2.1 and fig. 2.2 for a better visualization.

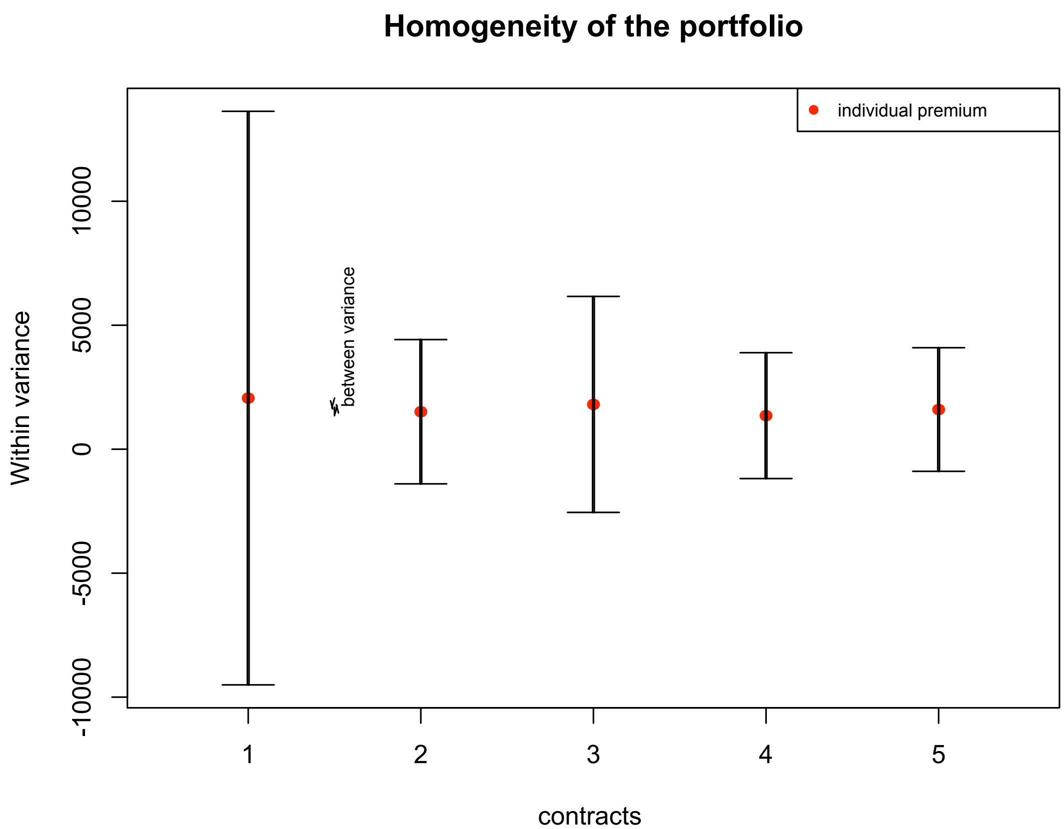

Figure 2.1: Homogeneity of the portfolio in the Bühlmann-Straub model.

And,

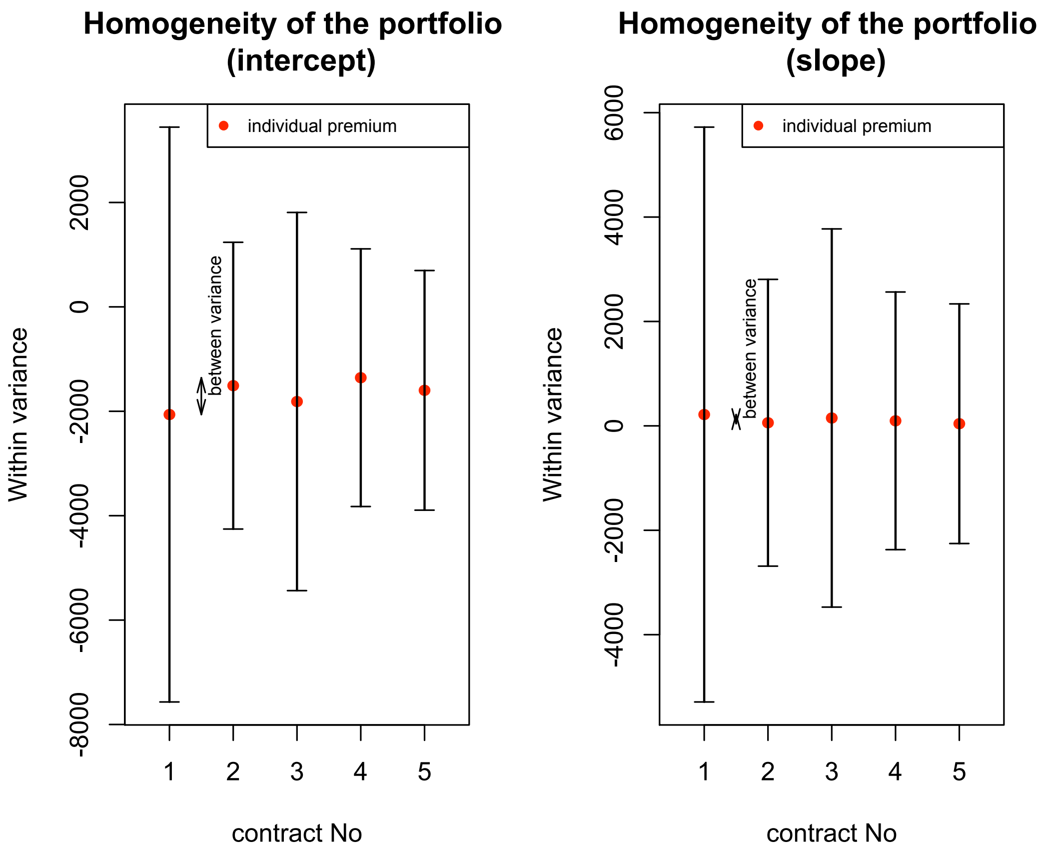

Figure 2.2: Homogeneity of the portfolio in the Hachemeister model.

Then, it could be interesting to show the differences between the homogeneity in each level in the case of a hierarchical portfolio. For example let us simulate a three-levels portfolio, then we fit this portfolio with the hierarchical model. The portfolio is represented in fig. 2.3.

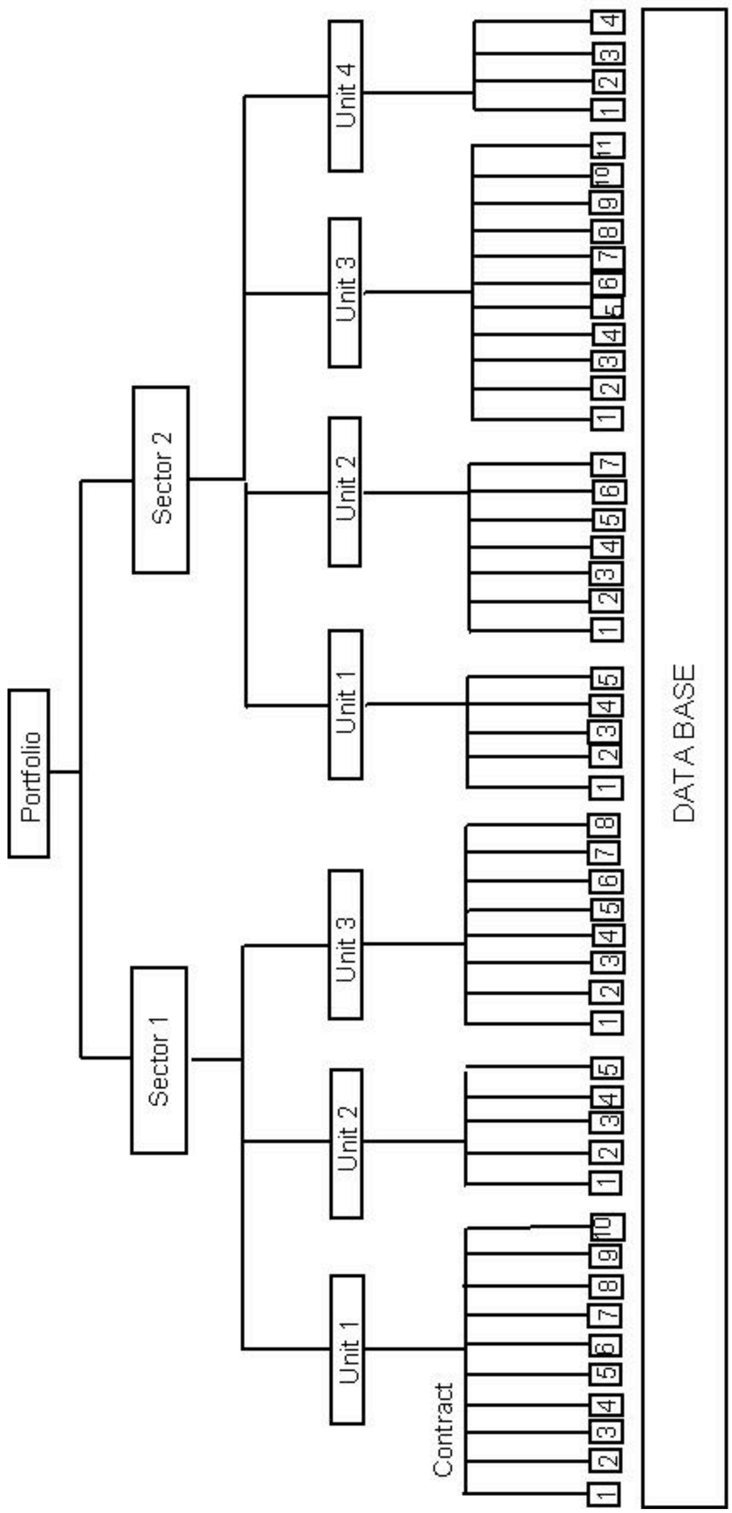

Figure 2.3: Three levels portfolio simulated.

With :

- at the level "sector" : severity (amount of claims) $\sim$ N(2, $\sqrt{0.1}$) and frequency $\sim$ Exp(2),
- at the level "unit" : severity $\sim$ N(nb sectors, $\sqrt{0.1}$) and frequency $\sim$ Gamma(nb sectors, 0.1),
- at the level "contract" : frequency $\sim$ Gamma(3, 4).

We first calculate all the structural parameters at each level of the portfolio, then we want to represent the homogeneity at each level of the portfolio.
We did not speak about the hierarchical model previously because it follows the same laws and formulae as the Bühlmann-Straub model except that you have different levels. In fact, the collective premium is given by the level above, and the individual one is computed with the credibility formulae seen before. Mathematically, we can write this in the following form, not exhaustive but simpler (if g is the third level (top), h is the second level and i is the first level)

$$\begin{aligned} \hat{\mu}_3 &= \alpha_3 B_3 + (1-\alpha_3)\mu_0 \\ \hat{\mu}_2 &= \alpha_2 B_2 + (1-\alpha_2)\hat{\mu}_3 \\ \hat{\mu}_1 &= \alpha_1 B_1 + (1-\alpha_1)\hat{\mu}_2 \end{aligned}$$

Where μ_0 is the collective premium and B_i the individual ones.

Finally we get for the sectors,

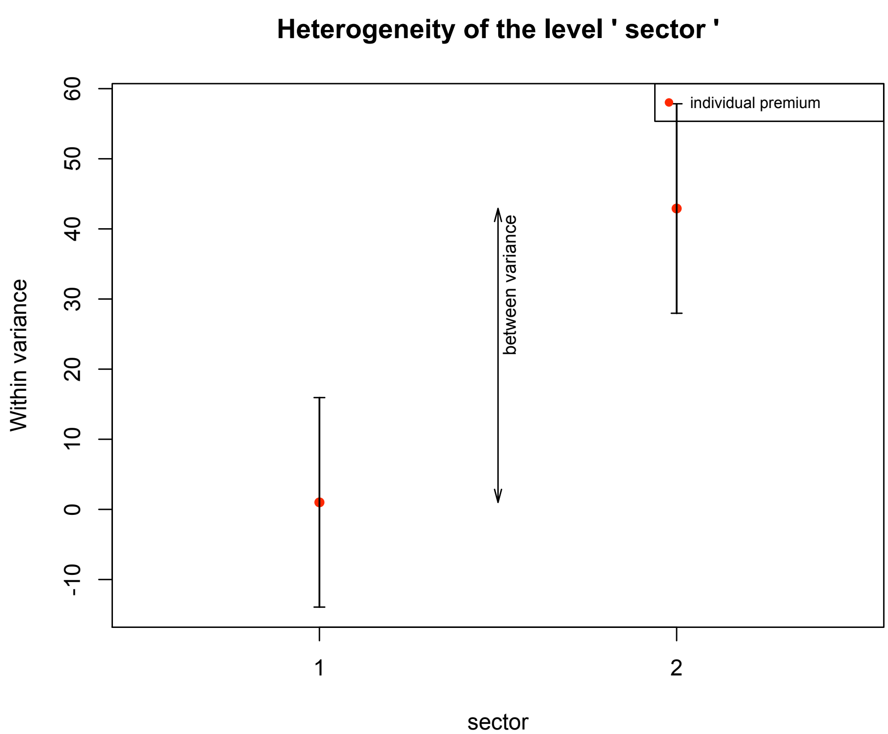

Figure 2.4: Homogeneity of the level "sector" in the hierarchical model.

For the units

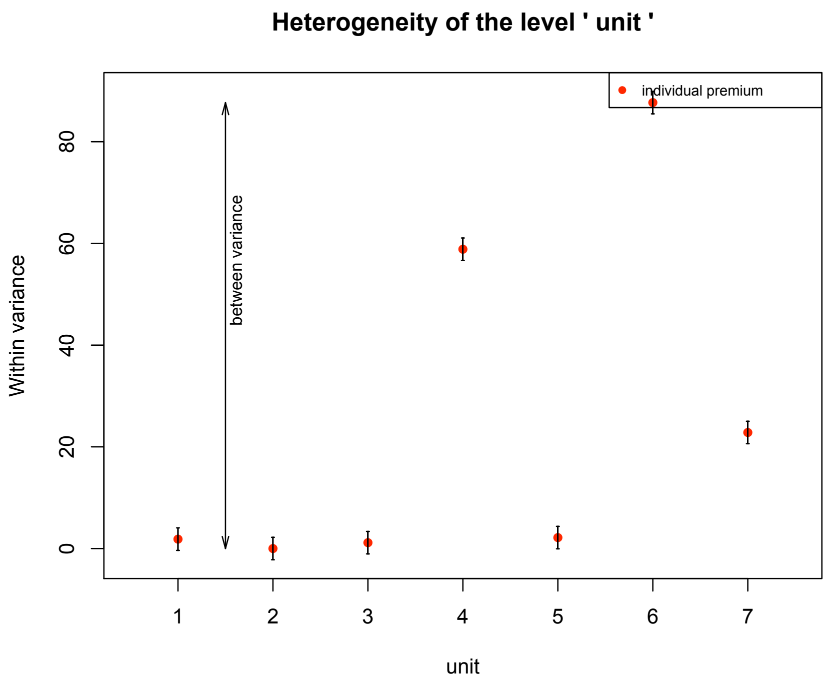

Figure 2.5: Homogeneity of the level “unit” in the hierarchical model.

And for the contracts

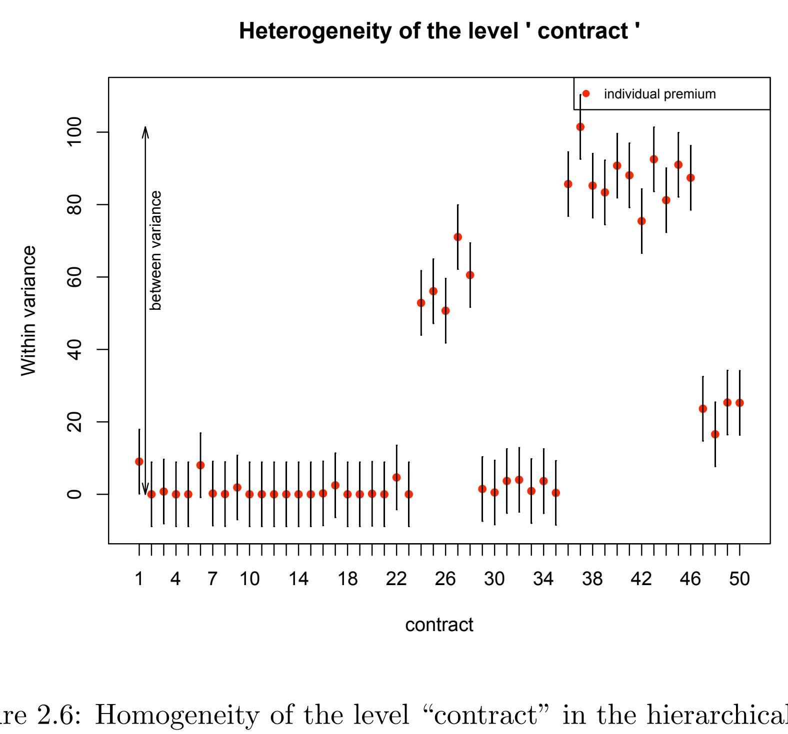

Figure 2.6: Homogeneity of the level “contract” in the hierarchical model.

Conclusion

It is clear that this way of representing the heterogeneity is very useful so as to make groups because we can easily distinguish different classes in the data. Let us notice that the more bottom you are the easier it is to see classes.

In our example, there are clearly four different risky classes : the group which contains the contracts number 1 to 23 and 29 to 35, the groups with the contracts from 24 to 28, the one from 36 to 46 and the last from contracts 47 to 50. The within-variance seems more or less equal in all the contracts, comparing to the between-variance which looks very big. These conclusions could have been deduced with the results at the level *unit* even if it is less precise.

Conclusion

In this memoir, we dealt with different credibility models to price the credibility premiums in various cases. It was a good way to understand the difference between this kind of models and the classical reasoning we are used to doing when we want to price a financial contract ; e.g. beginning with descriptive statistics, choose a model, apply it to data, estimate parameters of the model and finally make predictions. Here, one just takes account of the experience (idea of the bayesian statistics) without making any assumption on the model of the severity or on the frequency of the claims.
It looks more robust and more adapted to the insurance context, since we can not be mistaken about the choice of the models.

First, we studied the models of `Bühlmann`, `Bühlmann-Straub` and `Hachemeister` in a context of independant risk classes, and tried to understand the influence of various structural parameters on the credibility premium. This study was the occasion to realize the advantages and the drawbacks of each model : the `Bühlmann` model does not take account of the weight of the contracts although it is obvious that different contracts do not a similar influence in the portfolio, the `Bühlmann-Straub` model is not adapted with data which present a trend (always the case in reality !), and the `Hachemeister` model takes account of such a trend but predictions are very bad in some cases. Then we implemented a method to adjust the `Hachemeister` model because of absurd predictions. This leaded to a very interesting conclusion : we can modify the slope and the intercept of a regression model so as to get the credibility regression line between the individual and the collective ones. Furthermore it is possible to generalize this result in every dimension thanks to a mathematical trick.
With this results, the `Hachemeister` model seems to be the more relevant because it provides the possibility to compute the premiums in case of weighted data with a trend. An improvement of the package could be the development of others models like *De Vylder* model for example.

Second, we looked for a way to represent the heterogeneity of the portfolio in each of the models. We had to make choices on what to represent and how, the solution found was to customize special graphs provided by the software `R` to adapt them to the data we wanted to show : the between variance and the within variance estimators of the contracts within the portfolio.
This work was particularly difficult to implement for the `hierarchical` model, in its understanding but also in its adaptability to multiple levels (≥ 3).

Finally, the last part of the internship was a good opportunity to deal with something totally different but very important because it is used in several algorithms of actuarial theory : the matrix exponential, and its computation. It appears in applications involving the phase-type distributions (see Marceau (n.d.)). This work does not appear in this memoir because there is no link with other parts.
In fact, we are used to computing the matrix exponential with algorithms already implemented but

we have no idea of the errors we can have. I improved my general knowledge of this calculation and will give heed in the future about the algorithm I use to do such a computation. This work was to study nineteen different ways to do such a calculation, and to check if the algorithm implemented in **R** used the "right" one. See Matthews (1991), Moler & Van Loan (1978), Moler & Van Loan (2003) and Bates & Maechler (2008*a*) for further applications and understandings.

To speak a bit about the time I spent on each topic, the first one has been longer than the two other ones because it was a total discovery of this technique, and because we searched for a solution during a long time to find the trick (how to adjust the model) and implement it. Besides, we had no numerical results to check our figures. Then the second topic took time because of the decisions we had to take and its implementation; and we spent about two weeks on the last one to have an exhaustive list of various techniques.

In conclusion, this internship has really been an excellent job to discover new features in statistical theory (the credibility theory), understand them and be able to apply them in a real situation. I think that what I learnt in my training course will be very useful for further applications in my future job. I also improved my knowledge of the more and more famous (and used) statistical software **R**, which was really pleasant.

Appendices

Appendix A

Proofs of some theorems

A.1 Proof: $\mu(\hat{\Theta}_i)$ has a linear expression of μ_0 and $\bar{X}$

We want to find a linear estimator for the credibility premium $\mu(\Theta)$. We will denote it by P^{cred} or $\hat{\mu}(\Theta)$. By definition, we have

$$\hat{\mu}(\Theta) = \hat{a}_0 + \sum_{j=1}^{n} \hat{a}_j X_j$$

where the coefficients $\hat{a}_0$, $\hat{a}_1$,..., $\hat{a}_n$ need to solve

$$E\Big[\Big(\mu(\Theta) - \hat{a}_0 - \sum_{j=1}^{n} \hat{a}_j X_j\Big)^2\Big] = min_{a_0,\dots,a_n} E\Big[\Big(\mu(\Theta) - a_0 - \sum_{j=1}^{n} a_j X_j\Big)^2\Big]$$

We said the probability distribution of X_1,..., X_n was invariant (by assumptions) under permutations of X_j and $\hat{\mu}(\Theta)$ is unique so we have

$$\hat{a}_1 = \hat{a}_2 = ... = \hat{a}_n$$

which means that the estimator has the form $\quad \hat{\mu}(\Theta) = \hat{a} + \hat{b}\bar{X} \quad$ where $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$.

The partial derivatives with respect to a, resp. b hold that

$$\begin{aligned} E[\mu(\Theta) - a - b\bar{X}] &= 0, \\ Cov(\bar{X}, \mu(\Theta)) - bVar(\bar{X}) &= 0 \end{aligned}$$

From the dependency structure imposed by assumptions we have

$$\begin{aligned} Cov(\bar{X}, \mu(\Theta)) &= Var(\mu(\Theta)) = \tau^2, \\ Var(\bar{X}) &= \frac{E[s^2(\Theta)]}{n} + Var(\mu(\Theta)) = \frac{\sigma^2}{n} + \tau^2 \end{aligned}$$ From which we get

$$\begin{aligned} \text{b} &= \frac{\tau^2}{\tau^2 + \sigma^2/n}, \\ \text{a} &= (1-b)\mu_0 \end{aligned}$$

We have proved the theorem.

A.2 Proof: expression of the quadratic loss of the credibility estimator in a general context

$$\begin{aligned}
E\big[(\mu(\Theta)-\hat{\mu}(\Theta))^2\big] &= E[(\alpha(\mu(\Theta)-\bar{X})+(1-\alpha)(\mu(\Theta)-\mu_0))^2] \\
&= \alpha^2\frac{\sigma^2}{n}+(1-\alpha)^2\tau^2 \\
&= (\frac{n}{n+\sigma^2/\tau^2})\frac{\sigma^2}{n}+(\frac{\sigma^2/\tau^2}{n+\sigma^2/\tau^2})\tau^2 \\
&= (\frac{\sigma^2}{n+\sigma^2/\tau^2})=\alpha\frac{\sigma^2}{n} \\
&= (\frac{\sigma^2/\tau^2}{n+\sigma^2/\tau^2})\tau^2=(1-\alpha)\tau^2
\end{aligned}$$

A.3 Proof: $\mu(\hat{\Theta}_i) = z\bar{X}_i + (1-z)\mu$

Assume that the estimator is linear and can be written $a_{i0}+\sum_{j=1}^m a_{ij}X_{ij}$.
We want to minimize the quadratic loss (associated with the least squares criterion) with respect to the parameters a_{ij} :

$$E\Big[\big(a_{i0}+\textstyle\sum_{j=1}^m a_{ij}X_{ij}-\mu(\Theta_i)\big)^2\Big]$$

So,

$$\begin{aligned}
\frac{1}{2}\frac{d}{da_{i0}}E\Big[\big(a_{i0}+\sum_{j=1}^m a_{ij}X_{ij}-\mu(\Theta_i)\big)^2\Big] &= E\Big[a_{i0}+\sum_{j=1}^m a_{ij}X_{ij}-\mu(\Theta_i)\Big] \\
&= a_{i0}+\big(\sum_{j=1}^m a_{ij}-1\big)\mu=0
\end{aligned}$$

It holds that

$$a_{i0}=(1-\textstyle\sum_{j=1}^m a_{ij})\mu$$

By choosing k as $1\le k\le m$ (k is a given period), we get

$$\begin{aligned}
\frac{1}{2}\frac{d}{da_{ik}}E\Big[\big(a_{i0}+\sum_{j=1}^m a_{ij}X_{ij}-\mu(\Theta_i)\big)^2\Big] &= E\Big[X_{ik}\big(a_{i0}+\sum_{j=1}^m a_{ij}X_{ij}-\mu(\Theta_i)\Big] \\
&= a_{i0}\mu+\sum_{j=1,j\neq k}^m a_{ij}E[X_{ik}Y_{ij}]+a_{ik}E[X_{ik}^2]-E[X_{ik}\mu(\Theta_i)] \\
&= 0
\end{aligned}$$

The assumptions permit us to have $E[X_{ik}^2]=\sigma^2+\tau^2+\mu^2$, and for $j\neq k$ we get

$$E[X_{ik}X_{ij}]=E\big[E[X_{ik}X_{ij}|\Theta_i]\big]=E[(\mu(\Theta_i))^2]=\tau^2+\mu^2$$

Besides,

$$E[X_{ik}\ \mu(\Theta_i)] = E\big[E[X_{ik}\mu(\Theta_i)|\Theta_i]\big] = E[(\mu(\Theta_i))^2] = \tau^2 + \mu^2$$

Hence, we just have to solve

$$\begin{aligned} a_{ik}\sigma^2 - (1 - \sum_{j=1}^{m} a_{ij})\tau^2 &= 0 \\ a_{ik}\sigma^2 &= (1 - \sum_{j=1}^{m} a_{ij})\tau^2 \end{aligned}$$

And notice that the term on the right is independant from k, so we have the expression

$$a_{i1} = a_{i2} = ... = a_{im} = \tfrac{\tau^2}{\sigma^2 + m\tau^2} = z/m$$

Finally the estimator of the credibility premium is

$$\begin{aligned} a_{i0} + \sum_{j=1}^{m} a_{ij}X_{ij} &= (1 - \sum_{j=1}^{m} a_{ij})\mu + \sum_{j=1}^{m} \frac{\tau^2}{\sigma^2 + m\tau^2} X_{ij} \\ &= \frac{m\tau^2}{\sigma^2 + m\tau^2}\bar{X}_i + (1 - \frac{m\tau^2}{\sigma^2 + m\tau^2})\mu \end{aligned}$$

A.4 Proof: $\hat{\mu}$ is unbiased

We just have to show that $E[\hat{\mu}] = \mu$. So,

$$\begin{aligned} E[\hat{\mu}] &= E[\frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} = \frac{1}{n}\sum_{i=1}^{n}\bar{X}_i] \\ &= \frac{1}{mn}E[\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} = \frac{1}{n}\sum_{i=1}^{n}\bar{X}_i] \\ &= \frac{1}{mn} mnE[X_{ij}] \\ &= E[X_{ij}] = E[E[X_{ij}]|\Theta_i] \\ &= E[\mu(\Theta_i)] \\ &= \mu \end{aligned}$$

A.5 Proof: $\hat{s}^2(\Theta_i)$ is unbiased

We just have to show that $E[\hat{s}^2(\Theta_i)] = s^2(\Theta_i)$. So,

$$\begin{aligned}
E[\hat{s}^2(\Theta_i)] &= E[\frac{1}{m-1}\sum_{j=1}^{m}(X_{ij}-X_i)^2] \\
&= \frac{1}{m-1}E[\sum_{j=1}^{m}(X_{ij}-\mu(\Theta_i)+\mu(\Theta_i)-X_i)^2] \\
&= \frac{1}{m-1}E[\sum_{j=1}^{m}(X_{ij}-\mu(\Theta_i))^2-(\mu(\Theta_i)-X_i)^2] \\
&= s^2(\Theta_i)
\end{aligned}$$

A.6 Proof: $\hat{\tau}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{X}_i-\hat{\mu})^2 - \frac{1}{m}\hat{\sigma}^2$

First, we could suggest that $\bar{X}_i$ is an unbiased estimator of $\mu(\Theta_i)$. Then, we have a natural estimator of τ^2 which is

$$\tfrac{1}{n-1}\sum_{i=1}^{n}(\bar{X}_i-\hat{\mu})^2$$

But this estimator is biased and we must correct it. We have :

$$\begin{aligned}
\frac{1}{n-1}\sum_{i=1}^{n}E[(\bar{X}_i-\hat{\mu})^2] &= \frac{1}{n-1}\sum_{i=1}^{n}E\Big[(\frac{n-1}{n}\bar{X}_i-\frac{1}{n}\sum_{i\neq j}\bar{X}_j)^2\Big] \\
&= \frac{n}{n-1}E\Big[\frac{n-1}{n}\bar{X}_1-\frac{1}{n}\sum_{j=2}^{n}\bar{X}_j)^2\Big] \\
&= \frac{n}{n-1}E\Big[\frac{n-1}{n}(\bar{X}_1-\mu)-\frac{1}{n}\sum_{j=2}^{n}(\bar{X}_j-\mu))^2\Big] \\
&= \frac{n}{n-1}\Big((\frac{n-1}{n})^2E[(\bar{X}_1-\mu)^2]+\frac{n-1}{n^2}E[(\bar{X}_1-\mu)^2]\Big) \\
&= E[(\bar{X}_1-\mu)^2] \\
&= E\Big[\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}X_{1i}X_{1j}-\frac{2}{m}\mu\sum_{j=1}^{m}X_{1j}+\mu^2\Big] \\
&= \frac{1}{m}(\sigma^2+\tau^2+\mu^2)+\frac{m-1}{m}(\tau^2+\mu^2)-\mu^2 \\
&= \tau^2+\frac{\sigma^2}{m}
\end{aligned}$$

We just have to adjust the bias in order to have the unbiased estimator of τ^2

$$\hat{\tau}^2 = \tfrac{1}{n-1}\sum_{i=1}^{n}(\bar{X}_i-\hat{\mu})^2 - \tfrac{1}{m}\hat{\sigma}^2$$

A.7 Proof: $\hat{\tau^2} = c\Big(\frac{n}{n-1}\sum_{i=1}^n \frac{w_{i.}}{w_{..}}(X_i - \bar{X})^2 - \frac{n\hat{\sigma}^2}{w_{i.}}\Big)$

Consider a particular case where $w_{1.} = w_{2.} = ... = w_{n.}$;

$T = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

Since $E[T] = Var[X_i] = \frac{\sigma^2}{w_{i.}} + \tau^2$, we get that

$$\hat{\tau^2} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 - \frac{\hat{\sigma}^2}{w_{i.}}$$

is an unbiased estimator of τ^2.

Consider now the general case where sums of weights over a contract are different, then we can find analogously the best linear estimator :

we define $T = \frac{1}{n-1}\sum_{i=1}^n \frac{w_{i.}}{w_{..}}(X_i - \bar{X})^2$, and after some calculation we get the following unbiased estimator,

$$\hat{\tau^2} = c\Big(\frac{n}{n-1}\sum_{i=1}^n \frac{w_{i.}}{w_{..}}(X_i - \bar{X})^2 - \frac{n\hat{\sigma}^2}{w_{i.}}\Big) \qquad \text{where} \qquad c = \frac{n-1}{n}\Big(\sum_{i=1}^n \frac{w_{i.}}{w_{..}}(1 - \frac{w_{i.}}{w_{..}})\Big)^{-1}$$

A.8 Proof : expression of the quadratic loss

Given Θ_i, we know (by theorem) that the best unbiased linear estimator of $\beta(\Theta_i)$ is

$$B_i = \big(Y_i^{'}W_iY_i\big)^{-1}Y_i^{'}W_iX_i$$

This expression does not depend on Θ_i, therefore it must be the optimal data compression.
We saw that the expression of the covariance matrix of β is $Cov(B_i, B_i^{'}|\Theta_i) = \sigma^2(\Theta_i)(Y_i^{'}W_iY_i)^{-1}$

But by definition of the covariance, we have $Cov(B_i, B_i^{'}|\Theta_i) = E[(\beta_i - \beta(\Theta_i))(\beta_i - \beta(\Theta_i))^{'}|\Theta_i]$, and we take the expected value of the first covariance formula above.

A.9 Proof : $\hat{B}_i$ is individually unbiased

$\hat{B}_i$ is individually unbiased since

$$\begin{aligned} E[\hat{B}_i|\Theta_i] &= \big(Y_i^{'}S_i^{-1}Y_i\big)^{-1}Y_i^{'}S_i^{-1}E[X_i|\Theta_i] \\ &= \big(Y_i^{'}S_i^{-1}Y_i\big)^{-1}Y_i^{'}S_i^{-1}Y_i\beta(\Theta_i) \\ &= \beta(\Theta_i) \end{aligned}$$

Appendix B

R commands and outputs

For an excellent knowledge on the free software **R** and more particularly on the package **actuar**, see the following references : Goulet (n.d.), Goulet (2007) and Bates & Maechler (2008*b*).

B.1 The Bühlmann model

```
## Fitting of a Buhlmann model to the Hachemeister data set using
## function 'cm'. The interface of the function is similar to 'lm'.
> fit <- cm(~state, hachemeister, ratios = ratio.1:ratio.12)
> fit                          # print method
Call:
cm(formula = ~state, data = hachemeister, ratios = ratio.1:ratio.12)

Structure Parameters Estimators

  Collective premium: 1671.017

  Between state variance: 72310.02
  Within state variance: 46040.47

> predict(fit)                 # credibility premiums
[1] 2044.041 1518.588 1814.234 1375.987 1602.233
```

B.2 The Bühlmann-Straub model

```
## Fitting of a Buhlmann-Straub model require weights. Here, iterative
## estimators of the variance components are used.
> fit <- cm(~state, hachemeister, ratios = ratio.1:ratio.12,
+           weights = weight.1:weight.12, method = "iterative")
> summary(fit)
Call:
cm(formula = ~state, data = hachemeister, ratios = ratio.1:ratio.12,
    weights = weight.1:weight.12, method = "iterative")
```

```
Structure Parameters Estimators

  Collective premium: 1688.895

  Between state variance: 64366.51
  Within state variance: 139120026

Detailed premiums

  Level: state
    state Indiv. mean Weight Cred. factor Cred. premium
    1     2060.921    100155 0.9788756    2053.063
    2     1511.224     19895 0.9020069    1528.635
    3     1805.843     13735 0.8640336    1789.942
    4     1352.976      4152 0.6576516    1467.977
    5     1599.829     36110 0.9435251    1604.859

> predict(fit)
[1] 2053.063 1528.635 1789.942 1467.977 1604.859
```

B.3 The Hachemeister model

```
## Fitting of a Hachemeister regression model. This requires to
## specify a vector or matrix of regressors with argument 'xreg'.
## 'xreg' must be a matrix and represents the regressor = time here.
## The boolean adjust is set at TRUE so as to adjust intercept
## and slope in the regression model. Iterative estimators.
> fit <- cm(~state, hachemeister, ratios = ratio.1:ratio.12,
+           weights = weight.1:weight.12, regformula = ~time,
+           regdata = data.frame(time=12:1), adj.intercept = TRUE, method = "iterative")
> summary(fit, newdata = data.frame(time = 0))     # 'newdata' is the future value of regressor
Call:
cm(formula = ~state, data = hachemeister, ratios = ratio.1:ratio.12,
    weights = weight.1:weight.12, regformula = ~time, regdata = data.frame(time = 12:1),
    adj.intercept = TRUE, method = "iterative")

Structure Parameters Estimators

  Collective premium: -1676.919 120.1480

  Between state variance:  71564.69    0.000
                               0.00 3954.232
  Within state variance: 49870187

Detailed premiums

  Level: state
    state Indiv. coef. Credibility matrix  Adj. coef.  Cred. premium
    1     -2062.45704  0.9930903 0.0000000 -2059.79309 2446.439
            216.96651  0.0000000 0.8873162   206.05663
    2     -1509.28059  0.9661587 0.0000000 -1514.95368 1670.793
             59.60258  0.0000000 0.6126942    83.05219
    3     -1813.40908  0.9517141 0.0000000 -1806.81853 2062.015
            150.59895  0.0000000 0.5206650   136.00276
```

```
    4     -1356.75158  0.8562847 0.0000000 -1402.76453 1617.077
             96.69745  0.0000000 0.2530276   114.21439
    5     -1598.78534  0.9810673 0.0000000 -1600.26462 1715.503
             41.29288  0.0000000 0.7448318    61.41421

> predict(fit, newdata = data.frame(time = 0))
       1        1        1        1        1
2446.439 1670.793 2062.015 1617.077 1715.503
```

B.4 The hierarchical model

```
> ## Fitting of a hierarchical model to 4-levels portfolio simulated.
> DB <- cbind(weights(pf, prefix = "weight."),
+              aggregate(pf, classif = FALSE) / weights(pf, classif = FALSE))
> fit <- cm(~sector + sector:unit + sector:unit:interm + sector:unit:interm:contract,
+            data = DB, ratios = year.1:year.6,
+            weights = weight.year.1:weight.year.6)
> fit
Call:
cm(formula = ~sector + sector:unit + sector:unit:interm + sector:unit:interm:contract,
    data = DB, ratios = year.1:year.6, weights = weight.year.1:weight.year.6)

Structure Parameters Estimators

  Collective premium: 1276.62

  Between sector variance: 1323100
  Within sector/Between unit variance: 114639.1
  Within unit/Between interm variance: 147726.0
  Within interm/Between contract variance: 104799.6
  Within contract variance: 213898.8

> predict(fit, levels = "interm")         # unit credibility premiums only
$interm
 [1]  105.7044  920.0344  627.8887  160.2404  463.9359 1406.3761 1529.3039
 [8] 2136.5317 2282.0708 2602.2903

> summary(fit, levels = "interm")         # unit portfolio summary only
Call:
cm(formula = ~sector + sector:unit + sector:unit:interm + sector:unit:interm:contract,
    data = DB, ratios = year.1:year.6, weights = weight.year.1:weight.year.6)

Structure Parameters Estimators

  Collective premium: 1276.62

  Between interm variance: 147726.0
  Within interm variance: 104799.6

Detailed premiums

  Level: interm
    sector unit interm Indiv. mean Weight    Cred. factor Cred. premium
    1      1    1        25.60131  3.7879438 0.8422590     105.7044
```

```
    1      1    2      1017.49737  2.8141406 0.7986642     920.0344
    1      1    3       662.55746  1.9331847 0.7315456     627.8887
    1      2    1       126.90763  4.7320543 0.8696275     160.2404
    1      2    2       472.89205  6.4440790 0.9008292     463.9359
    2      1    1      1215.84112  1.8112058 0.7185545    1406.3761
    2      1    2      1474.69567  4.7225688 0.8693998    1529.3039
    2      1    3      2326.86119  0.9083596 0.5614859    2136.5317
    2      1    4      2380.56744  2.8035050 0.7980547    2282.0708
    2      2    1      2640.83723  5.6065169 0.8876780    2602.2903
```

B.5 Orthogonalization of the matrix : QR method with weights

The orthogonalization of the design matrix is done thanks to the following code. In fact it is used a bit differently in the function *hache* of the **actuar** because it is adapted but this is the same idea.

```
### One works with Hachemeister data (first contract)
data(hachemeister, package = "actuar")
ratios <- hachemeister[, 2:13]
weights <- hachemeister[, 14:25]

y <- ratios[1, ]                        # answers
x <- 1:12                               # explicative variable
w <- weights[1, ]                       # weights
X <- cbind(1, x)                        # design matrix

DF <- data.frame(x = x, y = y, w = w)

### One wants to work with orthogonal weighted matrix
### e.g. Q' W Q = I, where W = diag(w).
Xwqr <- qr(X * sqrt(w))                 # QR decomposition of W^{1/2} X
Qw <- qr.Q(Xwqr) / sqrt(w)              # orthogonal weighted matrix
t(Qw) \%*\% diag(w) \%*\% (Qw)              # checking
crossprod(Qw * w, Qw)                   # same calculation, more efficient
Rw <- qr.R(Xwqr)                        # transition matrix

lm.wfit(X, y, w)$fitted                 # fitted values with X
lm.wfit(Qw, y, w)$fitted                # fitted values with Qw

x0 <- c(1, 13)          # vector of prediction
drop(x0 \%*\% lm.wfit(X, y, w)$coefficients)                # prediction with X
drop(x0 \%*\% solve(Rw) \%*\% lm.wfit(Qw, y, w)$coefficients) # prediction with Qw
```

Bibliography

Bates, D. & Maechler, M. (2008*a*), *Matrix: A Matrix package for R.* R package version 0.999375-7.
URL: *http://cran.r-project.org/package=Matrix*

Bates, D. & Maechler, M. (2008*b*), *R News.* New features and bug fixes.
URL: *http://cran.r-project.org/doc/Rnews/*

Bühlmann, H. & Gisler, A. (1997), 'Credibility in the regression case revisited', *ASTIN Bulletin* **27**, 83–98.

Bühlmann, H. & Gisler, A. (2005), *A Course in Credibility Theory and its Applications*, Springer.

Christensen, R. (n.d.), *Plane answers to complex questions*, The theory of linear models, Springer texts in statistics.

Cleveland, W. (1993), *Visualizing data*, Hobart Press.

Concordia University, D. o. M. & Statistics (2007), The hachemeister model in credibility theory. Brief resume and important results.

Cossette, H. & Luong, A. (2003), 'Generalized least squares estimators for covariance parameters for credibility regression models with moving average errors', *Insurance: Mathematics and Economics* **32**, 281–293.

Goovaerts, M. & Hoogstad, W. (1987), *Credibility theory*, number 4 *in* 'Surveys of actuarial studies', Nationale-Nederlanden N.V., Netherlands.

Goulet, V. (2007), *actuar: An R Package for Actuarial Science, version 0.9-3*, École d'actuariat, Université Laval.
URL: *http://www.actuar-project.org*

Goulet, V. (n.d.), *Mathématiques actuarielles IARD II*, Université Laval, Canada.

Hachemeister, C. A. (1975), Credibility for regression models with application to trend, *in* 'Credibility, theory and applications', Proceedings of the berkeley Actuarial Research Conference on Credibility, Academic Press, New York, pp. 129–163.

Ho Lo, C; Fung, W. & Zhu, Z. (2006), 'Generalized estimating equations for variance and covariance parameters in regression credibility models', *Insurance: Mathematics and Economics* **39**, 99–113.

Marceau, E. (n.d.), *Modèles et méthodes actuarielles pour l'évaluation quantitative des risques*, ISFA, France.

Matthews, K. (1991), *Elementary Linear Algebra*, Anton.

McCulloch, C. & Searle, S. R. (2008), *Generalized linear and mixed models*, Wiley inter science.

Moler, C. & Van Loan, C. (1978), 'Nineteen dubious ways to compute the exponential of a matrix', *Society for Industrial and Applied Mathematics* **20**, 801–836.

Moler, C. & Van Loan, C. (2003), 'Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later', *Society for Industrial and Applied Mathematics* **45**, 3–49.

Murell, P. (2005), *R Graphics*, Chapman and Hall, Netherlands.

Radhakrishna Rao, C. & Toutenburg, H. (1999), *Least squares and alternatives*, Vol. 2, Springer.