Київський національний університет імені Тараса Шевченка
Факультет комп’ютерних наук та кібернетики
Кафедра моделювання складних систем

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА
на здобуття ступеня магістра
за спеціальністю 113 «Прикладна математика»

на тему:

Моделювання технічних резервів страхової компанії

студентки 2 курсу
Лижечко Марії Сергіївни

____

Науковий керівник:
доцент, кандидат фізико–математичних наук
Розора І. В.

____

Робота заслухана на засіданні кафедри моделювання складних систем та рекомендована до захисту, протокол № 10 від 08 травня 2020 р.

Завідувач кафедри МСС   канд. фіз.-мат. наук, доц. Черній Д.І.

Київ – 2020

Taras Shevchenko National University of Kyiv
Faculty of Computer Science and Cybernetics
Department of Complex Systems Modelling

GRADUATION THESIS
for a master's degree
in the specialty 113 «Applied Mathematics»

on the topic:

Modelling of technical reserves of an insurance company

of the 2 year student
Lyzhechko Mariia Serhiivna

____

Scientific advisor:
Associate Professor, Ph. D. in Physics and Mathematics
Rozora I. V.

____

The work was heard at a meeting of the Department of Complex Systems Modelling and recommended for thesis defense, Minutes № 10 of May 8, 2020.

Head of Department
of Complex Systems Modelling
Ph. D. in Ph. and Math., Assoc. Prof.
Chernii D. I.

Kyiv – 2020

CONTENTS

INTRODUCTION

In the modern rapidly evolving society, the science and the business are facing new needs and challenges constantly. The insurance industry and its mathematical foundation, the actuarial science, are not exceptions. To keep up with the latest tendencies, actuaries have to adjust and update the existing mathematical models and create the new ones. The development of technologies that support high-speed computations on large data sets undoubtedly facilitates these tasks.

Currently, the greatest challenge that the insurance system has to cope with is the issue of the new international financial standard [1, 2] that affects the calculation of reserves among other things. So far, insurers have mainly used common classical deterministic methods. However, the new standard emphasizes the necessity of the realistic prognosis that is best achieved with stochastic modelling tools since deterministic models do not represent the uncertainty and the random nature of future possible losses. The ***relevance*** of this work is that the author has considered two common classical deterministic methods for loss reserving – the Bornhuetter-Ferguson method and the chain-ladder method [3, 4] as well as two modifications of famous stochastic models – the Mack method and the bootstrapping technique [5 – 9] and juxtapose them in order to consider the achievement of the realistic prognosis of future losses. Besides, allowing for the time value of money and the inflation impact has become obligatory in many cases [2]. This issue is also studied in this work.

The ***aim*** of this work is to calculate deterministic reserves, analyse the obtained results and find out if any additional amounts have to be reserved to reach a certain confidence level, in other words, to ensure the achievement of a certain probability level of liability fulfilment on an insurance portfolio. The ***object*** of this work is the programme in language R that projects reserves satisfying certain criteria. The results of this work can find ***practical application*** in the non-life insurance sector.

SECTION 1. THEORETICAL BACKGROUND

1.1. The insurance industry

Definition 1.1.1. [10] An *insurance system* is a mechanism for reducing the adverse financial impact of random events that prevent the fulfilment of reasonable expectations.

The economic justification for an insurance system is that it contributes to general welfare by ensuring that plans will not be frustrated by random events. It also increases total production by encouraging individuals and corporations to take on projects with possibilities of large losses [10].

In general, the insurance industry is divided in two large sectors: *life* and *non-life* (or property and casual). The *non-life* insurance provides coverage for physical damage or loss of personal property or property a client is responsible for and general liability in respect to the third party. Besides, in Ukraine the health insurance also belongs to the non-life sector. The *life* insurance covers cases of accidental death and annuities payments. Moreover, insurance companies can offer assets management services. In this work, we are going to consider the non-life sector.

Definition 1.1.2. [10, 11] An insurance organization is known as an *insurer* or an *insurance company*. The insurer would issue contracts (*policies*) that would promise to pay the owner of this policy (a *policyholder*, an *insured*) or an approved interested party (a *beneficiary*) a defined amount equal to or less than the financial loss if an ensured event occurs during the period of a policy. The process of creating insurance policies is called *underwriting*. A group of contracts subject to similar risk and managed together forms a *portfolio*. Usually, those are the contracts within a line of business.

Definition 1.1.3. [10] In return for the promise contained in the policy, the insured pays a consideration (a *premium*). The insurer would set its basic price for full insurance coverage as the expected loss that in this context is called the *pure* or *net* *premium*. In order to provide for expenses, taxes, profit and for some security against adverse loss experience, the insurer would decide to set the final premium for the policy by *loading*, adding to the pure premium [10].

Definition 1.1.4. [11] A formal request by a policyholder to an insurance company for coverage or compensation for a covered loss or policy event is called a *claim*. The insurance company validates the claim and, once approved, issues payment to the insured or the beneficiary. This payment is called a *claim payment*.

One of the most significant employees of an insurance company is an *actuary* – a professional who assesses and manages the risks of financial investments, insurance policies and other potentially risky ventures. One of responsibilities of an actuary is modelling, calculating and forming reserves.

1.2. Types of technical reserves

Definition 1.2.1. [11] ***Technical reserves*** are the amounts insurance companies set aside to fulfil their liabilities in respect to policyholders and stakeholders.

According to the Ukrainian law [12] insurers who provide non-life insurance services are obliged to form and account for the following technical reserves:

- the unearned premium reserve (UPR), which include shares of sums of insurance premiums corresponding to insurance risks that have not expired at the reporting date;
- the outstanding claims reserve (OCR), including reserved unpaid insurance sums corresponding to known claims of insureds than have not been approved or denied yet.

Besides, insurers can choose to form additional reserves such as the reserve for incurred but not reported claims (IBNR). The common practice has been to form reserves mentioned above that does not contradict the requirements of the international financial reporting standard “IFRS 4 Insurance contracts” [1] that is currently in force.

However, the new international financial reporting standard “IFRS 17 Insurance contracts” [2] comes into force on the 1st of January in 2023. According to the requirements of the new standard, an insurance company has to form and account for liability for remaining coverage (LRC) and liability for incurred claims (LIC). If a company passes a certain eligibility test, the procedure of calculating LRC remains virtually the same as previous calculation of UPR. Differences lie in allowing for expenses. However, comparison of expenses allocation methods is not in the scope of this work.

If the expected pattern of risk release is uniform, then LRC is calculated based on the passage of time (“pro rata temporis”)

$$Am = \frac{Pb(t-d)}{t},$$

$$f = \frac{d}{t},$$

where Am is amount of liability, f is an earned factor, Pb a base premium under the contract, t is a contract duration (a coverage period) and d is number of days from the start of the contract coverage till the date of liability calculation [13].

If the expected pattern of risk release during the coverage period differs significantly from the passage of time, then LRC is calculated as follows

$$Am = (1-f)Pb$$

where f differs depending on the pattern of risk release. For instance, it may be

$$f = \left(\frac{d}{t}\right)^2.$$

If an insurance company does not pass the eligibility test, then the procedure of LRC calculation is more complicated. However, we are not going to focus on it because the liability for remaining coverage is independent of the number of probabilities, in other words, its amount is clearly given by the amount of the prescribed premium and the period for which the premium is determined.

LIC calculation is more interesting for us since here we have to work with uncertainty and project the amount of claims that have already occurred but have not been reported yet using the information about the reported claims. Besides, the new standard requires the realistic prognosis, in other words, reserves amount has to be adjusted for non-financial risk to a certain confidence level that measures the probability that a company will fulfil its liability. It is done efficiently by determining the corresponding quantile of the loss distribution. However, so far insurers have calculated IBNR using mainly classical deterministic methods. Moreover, under certain conditions insurance companies will have to allow for the time value of money and the inflation impact. The justification for this requirement and the description of some stochastic and common classical deterministic methods of modelling of IBNR reserves as well as their juxtaposition are going to be considered further.

1.3. Elements of the probability theory

Definition 1.3.1. [14, 15] A *probability space* is a triple $(\Omega, \mathcal{F}, P)$ consisting of a non-empty set Ω, a class $\mathcal{F}$ of subsets of Ω which is σ-algebra (in other words, closed with respect to the set-theoretic operations executed a countable number of times) and a probability measure P on $\mathcal{F}$. The points of Ω are said to be *elementary events* while the set Ω itself is referred to as the *space of elementary events* or the *sample space*. The subsets of Ω belonging to $\mathcal{F}$ are *random events*.

Definition 1.3.2. [14] Let $(\Omega, \mathcal{F}, P)$ be a probability space. A single-valued real-valued function $X = X(\omega)$ defined on Ω is called a *random variable* if for any real x the set $\{\omega: X(\omega) < x\}$ belongs to the class $\mathcal{F}$.

Definition 1.3.3. [14] Let X be any random variable and $\mathcal{F}_X$ the class of subsets $C \subset \mathbb{R}$ for which $\{\omega: X(\omega) \in C\} \in \mathcal{F}$; this is a σ-algebra. The class $\mathcal{B}_1$ of all Borel subsets (the σ-algebra generated by the open sets) of $\mathbb{R}$ is always contained in $\mathcal{F}_X$. The measure P_X defined on $\mathcal{B}_1$ by the equation $P_X(B) = P\{\omega : X(\omega) \in B\}$, $B \in \mathcal{B}_1$, is called the *probability distribution* of X. This measure is uniquely determined by the *distribution function* of X

$$F_X(x) = P_X\{-\infty, x\} = P\{\omega : X(\omega) < x\}.$$

Definition 1.3.4. [14] If a random variable X takes a finite or countable number of pairwise distinct values $x_1, \dots, x_n, \dots$ with probabilities $p_1, \dots, p_n, \dots$ ($p_n = P\{\omega : X(\omega) = x_n\}$), then its probability distribution, which is said to be *discrete* in this case, is given by

$$P_X(A) = \sum_{x_n \in A} p_n.$$

Definition 1.3.5. [14, 16] The distribution of X is called *continuous* if there is a function $p_X(x)$ called the *probability density* such that

$$P_X(B) = \int_B p_X(x)\, dx$$

for every interval B (or equivalently, for every Borel set B) or

$$F_X(x) = \int_{-\infty}^{x} p_X(t)\, dt.$$

Definition 1.3.6. [14] A *mathematical expectation* or an *expected value* of a random variable $X(\omega)$, $\omega \in \Omega$, is defined as the Lebesgue integral with respect to a probability measure P on a given probability space $(\Omega, \mathcal{F}, P)$

$$EX = \int_\Omega X(\omega) P(d\omega),$$

provided the integral exists. The expected value (the mathematical expectation) of a real-valued random variable X can be also calculated as the following Lebesgue integral

$$EX = \int_{\mathbb{R}} x p_X(x)\, dx.$$

The *properties* of the expected value:

1. If X_1 and X_2 are random variables, then $E(X_1 + X_2) = EX_1 + EX_2$.
2. If X is a random variable and c is a real number, then $E(cX) = cE(X)$.
3. If c is a real number, then $E(c) = c$.

Definition 1.3.7. [14] A *variance* of a random variable X is a measure of the deviation of X from its expected value EX defined by the equation

$$Var(X) \overset{\text{def}}{=} E\big(X - E(X)\big)^2.$$

The *properties* of the variance:

1. $Var(X) = E(X^2) - [E(X)]^2$.
2. If c is a real number, then $Var(cX) = c^2 Var(X)$.

Definition 1.3.8. [14] Let $(\Omega, \mathcal{F}, P)$ be a probability space. The *conditional expectation* of X with respect to a σ-algebra $\mathcal{B}$, $\mathcal{B} \subseteq \mathcal{F}$, is understood as a random variable $E(X|\mathcal{B})$ measurable with respect to $\mathcal{B}$ and such that

$$\int\limits_B XP(d\omega) = \int\limits_B E(X|\mathcal{B})P(d\omega)$$

for each $B \in \mathcal{B}$.

Definition 1.3.9. [14] The probability distribution of a random variable X is called *normal* if it has the probability density

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

where a is the expected value of X and σ^2 is its variance.

Definition 1.3.10. [14] A random variable X is subject to the *logarithmic normal* (*lognormal*) distribution if its natural logarithm has the normal distribution.

Definition 1.3.11. [14] For a real random variable X with the distribution function F_X, by a *quantile* of order p, $0 < p < 1$, we understand the number K_p for which

$$F\big(K_p\big) = P\big\{X < K_p\big\} \leq p. \tag{1.3.1}$$

1.4. Loss-development data

Let $(\Omega, \mathcal{F}, P)$ be a probability space on which all random variables are defined. Assume all variables are square integrable. All equalities and inequalities involving random variables hold almost surely with respect to the probability measure P.

We consider a portfolio of contracts that are subject to similar risks and managed together. Assume that each claim of the portfolio is settled either in an *accident period* or in the following n periods that are called *development periods*. Periods are usually taken on a quarterly, monthly, half-year or annual basis.

1.4.1. Run-off triangles

The portfolio may be modelled either by incremental losses or by cumulative losses [3].

Definition 1.4.1.1. [3] Consider a family of random variables $\{Z_{i,k}\}_{i,k\in\{0,1,\dots,n\}}$. Let $Z_{i,k}$ denote the loss of an accident period i that is settled with a delay of k periods, in other words, in a development period k and in a calendar period $i+k$. $Z_{i,k}$ is called an *incremental loss* of an accident period i and a development period k.

Definition 1.4.1.2. [3] Next consider a family of random variables $\{C_{i,k}\}_{i,k\in\{0,1,\dots,n\}}$. Let $C_{i,k}$ denote the loss of an accident period i that is settled with a delay of at most k periods, in other words, not later than in a development period k and in a calendar period $i+k$. $C_{i,k}$ is called a *cumulative loss* of an accident period i and a development period k.

To model claims experience we are going to use a technique known as a *run-off triangle* (or *a delay triangle*). This technique is specifically used to estimate the future claims that will be reported based on those already reported.

Table 1.4.1.1 – Run-off triangles of incremental losses

Accident period	Development period								
	0	1	$\cdots$	k	$\cdots$	$n-i$	$\cdots$	$n-1$	n
0	$Z_{0,0}$	$Z_{0,1}$	$\cdots$	$Z_{0,k}$	$\cdots$	$Z_{0,n-i}$	$\cdots$	$Z_{0,n-1}$	$Z_{0,n}$
1	$Z_{1,0}$	$Z_{1,1}$	$\cdots$	$Z_{1,k}$	$\cdots$	$Z_{1,n-i}$	$\cdots$	$Z_{1,n-1}$	$Z_{1,n}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
i	$Z_{i,0}$	$Z_{i,1}$	$\cdots$	$Z_{i,k}$	$\cdots$	$Z_{i,n-i}$	$\cdots$	$Z_{i,n-1}$	$Z_{i,n}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$n-k$	$Z_{n-k,0}$	$Z_{n-k,1}$	$\cdots$	$Z_{n-k,k}$	$\cdots$	$Z_{n-k,n-i}$	$\cdots$	$Z_{n-k,n-1}$	$Z_{n-k,n}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$n-1$	$Z_{n-1,0}$	$Z_{n-1,1}$	$\cdots$	$Z_{n-1,k}$	$\cdots$	$Z_{n-1,n-i}$	$\cdots$	$Z_{n-1,n-1}$	$Z_{n-1,n}$
n	$Z_{n,0}$	$Z_{n,1}$	$\cdots$	$Z_{n,k}$	$\cdots$	$Z_{n,n-i}$	$\cdots$	$Z_{n,n-1}$	$Z_{n,n}$

Table 1.4.1.2 – Run-off triangles of cumulative losses

Accident period	Development period								
	0	1	$\cdots$	k	$\cdots$	$n-i$	$\cdots$	$n-1$	n
0	$C_{0,0}$	$C_{0,1}$	$\cdots$	$C_{0,k}$	$\cdots$	$C_{0,n-i}$	$\cdots$	$C_{0,n-1}$	$C_{0,n}$
1	$C_{1,0}$	$C_{1,1}$	$\cdots$	$C_{1,k}$	$\cdots$	$C_{1,n-i}$	$\cdots$	$C_{1,n-1}$	$C_{1,n}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
i	$C_{i,0}$	$C_{i,1}$	$\cdots$	$C_{i,k}$	$\cdots$	$C_{i,n-i}$	$\cdots$	$C_{i,n-1}$	$C_{i,n}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$n-k$	$C_{n-k,0}$	$C_{n-k,1}$	$\cdots$	$C_{n-k,k}$	$\cdots$	$C_{n-k,n-i}$	$\cdots$	$C_{n-k,n-1}$	$C_{n-k,n}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$n-1$	$C_{n-1,0}$	$C_{n-1,1}$	$\cdots$	$C_{n-1,k}$	$\cdots$	$C_{n-1,n-i}$	$\cdots$	$C_{n-1,n-1}$	$C_{n-1,n}$
n	$C_{n,0}$	$C_{n,1}$	$\cdots$	$C_{n,k}$	$\cdots$	$C_{n,n-i}$	$\cdots$	$C_{n,n-1}$	$C_{n,n}$

Rows of the tables 1.4.1.1 and 1.4.1.2 correspond to accident periods, columns – to development periods. The enumeration of the development periods represents delays with respect to the accident periods.

Definition 1.4.1.3. [3] An incremental loss $Z_{i,k}$ and a cumulative loss $C_{i,k}$ are said to be:

- *observable* if $i+k \leq n$;
- *non-observable* or *future* if $i+k > n$;
- *current* if $i+k = n$;
- *ultimate* if $k = n$.

Since only the elements on and above the diagonal are known (observable), the technique name comprises the word triangle. The principal task is to predict the future (non-observable) losses [4]: the ultimate losses $C_{i,n}$, the accident period reserves $C_{i,n} - C_{i,n-i}$, the future cumulative losses $C_{i,k}$, the future incremental losses $Z_{i,k}$, the calendar period reserves $\sum_{j=p-n}^{n} Z_{j,p-j}$ and the total reserve $\sum_{j=1}^{n}\sum_{l=n-j+1}^{n} Z_{j,l}$ with $i+k \geq n+1$ and $p = n+1, \ldots, 2n$. The prediction of the latest is the main purpose.

Remark 1.4.1.1. Modelling a portfolio by incremental losses is equivalent to modelling it by cumulative losses. They can be easily derived from each other

$$C_{i,k} = \sum_{l=0}^{k} Z_{i,l} \qquad (1.4.1.1)$$

and

$$Z_{i,k} = \begin{cases} C_{i,k}, k = 0 \\ C_{i,k} - C_{i,k-1}, k \neq 0 \end{cases} \qquad (1.4.1.2)$$

Remark 1.4.1.2. Already reported claims can be taken either on a paid or incurred basis.

Remark 1.4.1.3. The enumeration of accident and development periods starting with zero instead of one is not generally accepted. We are going to use it since it seems to be more natural and simplifies mathematical formulas.

1.4.2. Development patterns

The use of the run-off triangles technique is justified only under assumption that a development pattern is common for all accident periods. We are going to consider two types of such patterns that can be viewed as a primitive stochastic model for loss reserving [3].

We start with comparing the expected cumulative losses with the expected ultimate cumulative losses.

Assumption 1.4.2.1. A development pattern consists of parameters $\gamma_0, \gamma_1, \dots, \gamma_n$ (with $\gamma_n = 1$) such that the identity

$$\gamma_k = \frac{E[C_{i,k}]}{E[C_{i,n}]}$$

holds for all $k \in \{0, 1, \dots, n\}$ and for all $i \in \{0, 1, \dots, n\}$.

Definition 1.4.2.1. The parameters $\gamma_0, \gamma_1, \dots, \gamma_n$ are called *cumulative development quotas*. Assumption 1.4.2.1 means that for every development period $k \in \{0, 1, \dots, n\}$ the cumulative quotas are identical for all accident periods.

Remark 1.4.2.1. In case of a run-off triangle for paid claims, it is reasonable to assume that $0 \leq \gamma_0 \leq \gamma_1 \leq \dots \leq \gamma_n$.

Remark 1.4.2.2. It is natural to interpret cumulative quotas as percentages of reported claims.

Next, we compare subsequent expected cumulative losses.

Assumption 1.4.2.2. A development pattern consists of parameters $\varphi_1, \dots, \varphi_n$ such that the identity

$$\varphi_k = \frac{E[C_{i,k}]}{E[C_{i,k-1}]}$$

holds for all $k \in \{1, \dots, n\}$ and for all $i \in \{0, 1, \dots, n\}$.

Definition 1.4.2.2. The parameters $\varphi_1, \dots, \varphi_n$ are called *cumulative development factors* (*age-to-age factors*). Assumption 1.4.2.2 means that for every development period $k \in \{0, 1, \dots, n\}$ the factors are identical for all accident periods.

Remark 1.4.2.3. In case of a run-off triangle for paid claims, it is reasonable to assume that $\varphi_k \geq 1$ for all $k \in \{1, \dots, n\}$.

Remark 1.4.2.4. If the parameters $\gamma_0, \gamma_1, \dots, \gamma_n$ form a development pattern for quotas, then the parameters $\varphi_1, \dots, \varphi_n$ such that

$$\varphi_k = \frac{\gamma_k}{\gamma_{k-1}}$$

form a development pattern for factors.

If the parameters $\varphi_1, \dots, \varphi_n$ form a development pattern for factors, then the parameters $\gamma_0, \gamma_1, \dots, \gamma_n$ such that

$$\gamma_k = \prod_{l=k+1}^{n} \frac{1}{\varphi_l}$$

form a development pattern for quotas.

Remark 1.4.2.5. The condition of the remark 1.4.2.3 is fulfilled if and only if the condition of the remark 1.4.2.1 holds.

For estimation of the parameter γ_k of a development pattern for quotas, the run-off triangle provides the obvious empirical individual quota:

$$\hat{\gamma}_{0,k} = \frac{C_{0,k}}{C_{0,n}}.$$

Similarly, for estimation of the parameter φ_k of a development pattern for factors, we can use the empirical individual factors

$$\hat{\varphi}_{i,k} = \frac{C_{i,k}}{C_{i,k-1}}$$

with $i \in \{0, 1, \dots, n-k\}$.

Moreover, any weighted mean of these estimators

$$\hat{\varphi}_k = \sum_{j=0}^{n-k} \omega_{j,k} \hat{\varphi}_{j,k}$$

with $\sum_{j=0}^{n-k} \omega_{j,k} = 1$ is an estimator as well.

The most widely used estimator of this family is a chain-ladder factor

$$\hat{\varphi}_k^{CL} = \frac{\sum_{j=0}^{n-k} C_{j,k}}{\sum_{j=0}^{n-k} C_{j,k-1}} = \sum_{j=0}^{n-k} \frac{C_{j,k-1}}{\sum_{h=0}^{n-k} C_{h,k-1}} \hat{\varphi}_{j,k}. \qquad (1.4.2.1)$$

Similarly, the chain-ladder quotas

$$\hat{\gamma}_k^{CL} = \prod_{l=k+1}^{n} \frac{1}{\hat{\varphi}_k^{CL}} \qquad (1.4.2.2)$$

may be used to estimate the development quotas γ_k.

Example 1.4.2.1. [4] Assume we have the run-off triangle of the incremental losses represented in the table 1.4.2.1.

Table 1.4.2.1

Accident period	Development period					
	0	1	2	3	4	5
0	1001	854	568	565	347	148
1	1113	990	671	648	422	
2	1265	1168	800	744		
3	1490	1383	1007			
4	1725	1536				
5	1889					

Using (1.4.1.1), we have derived the run-off triangle of the cumulative losses that is represented in the table 1.4.2.2. In addition, we have calculated the chain-ladder factors and the chain-ladder quotas defined by (1.4.2.1) and (1.4.2.2) respectively.

Table 1.4.2.2

Accident period	Development period					
	0	1	2	3	4	5
0	1001	1855	2423	2988	3335	3483
1	1113	2103	2774	3422	3844	
2	1265	2433	3233	3977		
3	1490	2873	3880			
4	1725	3261				
5	1889					
$\hat{\varphi}_k^{CL}$		1.899	1.329	1.232	1.12	1.044
$\hat{\gamma}_k^{CL}$	0.275	0.522	0.694	0.855	0.958	1

The table 1.4.2.3 represents the known age-to-age factors.

Table 1.4.2.3

	0-1	1-2	2-3	3-4	4-5
0	1.853	1.306	1.233	1.116	1.044
1	1.889	1.319	1.234	1.123	
2	1.923	1.329	1.23		
3	1.928	1.351			
4	1.89				

1.5. Methods of loss reserving

The uncertainty in calculations comes from the uncertainty about number, size, timing of claims, exposure to them, credibility of data and reliability of a calculation model. Deterministic techniques ignores the random nature of the claims process whereas stochastic approaches allow for it, test the model reliability, estimate the potential variability in reserves and the shape of the reserve distribution.

In this part, we are going to consider the most common classical deterministic methods for IBNR reserves calculation first. Then we are going to focus on algorithms of some stochastic methods.

1.5.1. The Bornhuetter-Ferguson method

The Bornhuetter-Ferguson method is based on the assumption that there exist parameters $\alpha_0, \alpha_1, \dots, \alpha_n$ and $\gamma_0, \gamma_1, \dots, \gamma_n$ with $\gamma_n = 1$ such that the identity

$$E\left[C_{i,k}\right] = \gamma_k \alpha_i$$

holds for all $i, k \in \{0, 1, \dots, n\}$ [3]. Then we have

$$E\left[C_{i,n}\right] = \alpha_i$$

and hence

$$E\left[C_{i,k}\right] = \gamma_k E\left[C_{i,n}\right].$$

Therefore, the parameters $\gamma_0, \gamma_1, \dots, \gamma_n$ form a development pattern of cumulative quotas. Additionally, we assume that prior estimators $\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_n$ of the expected ultimate cumulative losses $E\left[C_{i,n}\right]$ and prior estimators $\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_n$ of the development pattern are given with $\hat{\gamma}_n = 1$.

These prior estimators can be obtained from various sources of internal and external information. *Internal* information is any information provided by the run-off triangle of the portfolio under consideration. It could be used, for instance, for estimating the development pattern like chain-ladder quotas or volume measures like premiums. *External* information is any information that is not contained in the run-off triangle of the portfolio. It could be obtained, for instance, from market statistics or data from portfolios viewed as similar to the analysed one. Prior estimators may be also obtained by combining internal and external information. Their choice is an important decision. The common market practice is to take chain-ladder quotas (1.4.2.2) as prior estimators $\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_n$ and the product $\pi_i \hat{k}_i$ ($i \in \{0, 1, \dots, n\}$) as the prior estimators

$\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_n$. π_i are the earned premiums (shares of sums of insurance premiums corresponding to insurance risks that have expired at the date of the liability calculation) and $\hat{k}_i$ the estimators of loss ratios $k_i = E\left[\frac{C_{i,n}}{\pi_i}\right]$ obtained, for instance, from market statistics.

The Bornhuetter-Ferguson predictors of the ultimate cumulative losses $C_{i,n}$ are calculated as follows

$$\hat{C}_{i,n}^{BF} = C_{i,n-i} + \left(1 - \hat{\gamma}_{n-i}^{CL}\right)\pi_i \hat{k}_i.$$

The Bornhuetter-Ferguson predictors of the cumulative losses $C_{i,k}$ with $i + k > n$ are defined as

$$\hat{C}_{i,k}^{BF} = C_{i,n-i} + \left(\hat{\gamma}_{k}^{CL} - \hat{\gamma}_{n-i}^{CL}\right)\pi_i \hat{k}_i.$$

In order to predict calendar period reserves $R_i = C_{i,n} - C_{i,n-i}$ we want the model equation

$$E[R_i] = (1 - \gamma_{n-i})E\left[C_{i,n}\right].$$

to be satisfied [4].

Thus, the Bornhuetter-Ferguson predictors of the reserves R_i are defined as

$$\hat{R}_i^{BF} = \left(1 - \hat{\gamma}_{n-i}^{CL}\right)\pi_i \hat{k}_i.$$

The definition of the Bornhuetter-Ferguson predictors shows that the prior estimators are dominant for young accident periods and less important for old ones.

1.5.2. The chain-ladder method

The chain-ladder method is based on the assumption that there exist parameters $\varphi_1, \dots, \varphi_n$ such that the identity

$$E\left[C_{i,k}\right] = \varphi_k E\left[C_{i,k-1}\right]$$

holds for all $i \in \{0, 1, \dots, n\}$ and $k \in \{1, \dots, n\}$ [3]. Then the parameters $\varphi_1, \dots, \varphi_n$ form a development pattern for factors. This method relies completely on the observable cumulative losses of the run-off triangle and involves no information obtained from external sources at all.

The future cumulative losses $C_{i,k}$ satisfy the model equation

$$E\big[C_{i,k}\big] = E[C_{i,n-i}] \prod_{l=n-i+1}^{k} \varphi_l$$

which is a consequence of the model assumption.

Accordingly, the chain-ladder predictors of the future cumulative losses $C_{i,k}$ with $i + k > n$ are defined as

$$\hat{C}_{i,k}^{CL} = C_{i,n-i} \prod\nolimits_{l=n-i+1}^{k} \hat{\varphi}_l^{CL} \tag{1.5.2.1}$$

where $\hat{\varphi}_l^{CL}$ are the chain-ladder factors (1.4.2.1) introduced before.

Therefore, the chain-ladder predictors of the ultimate cumulative losses $C_{i,n}$ are obtained as follows

$$\hat{C}_{i,n}^{CL} = C_{i,n-i} \prod\nolimits_{l=n-i+1}^{n} \hat{\varphi}_l^{CL}. \tag{1.5.2.2}$$

Therefore, this method consists in successive scaling of the current loss $C_{i,n-i}$ to the level of the future cumulative loss $C_{i,k}$.

As a consequence of the relation (1.4.2.2), $\hat{C}_{i,k}^{CL}$ can be written as follows

$$\hat{C}_{i,k}^{CL} = \hat{\gamma}_k^{CL} \frac{C_{i,n-i}}{\hat{\gamma}_{n-i}^{CL}}.$$

The estimators of reserves R_i are calculated as follows

$$\hat{R}_i^{CL} = \hat{C}_{i,n}^{CL} - C_{i,n-i} = C_{i,n-i} \prod_{l=n-i+1}^{n} \hat{\varphi}_l^{CL} - C_{i,n-i} = C_{i,n-i} \cdot$$

$$\cdot \left(\prod\nolimits_{l=n-i+1}^{n} \hat{\varphi}_l^{CL} - 1\right) \tag{1.5.2.3}$$

It is remarkable that the chain-ladder predictors involve, via the chain-ladder factors, all cumulative losses of the run-off triangle.

1.5.3. The Mack method

The Mack method is a statistical model underlying pure chain ladder. The aim is not only to estimate the non-observable or future losses $C_{i,k}$ with $i + k > n$ and the claims reserves R_i for accident periods $i \in \{1, \dots, n\}$ but also to calculate the standard error of the chain ladder reserve estimates as a measure of the uncertainty contained in the data [5].

The model has three explicit assumptions:

1.5.3.1. The expected value of cumulative losses at the development period k is equal to cumulative losses at the development period $k - 1$ multiplied by the development factor

$$E\left(C_{i,k} \middle| C_{i,0}, C_{i,1}, \dots, C_{i,k-1}\right) = C_{i,k-1}\varphi_k$$

where $i \in \{0, 1, \dots, n\}$ and $k \in \{1, \dots, n\}$.

1.5.3.2. The cumulative losses are independent between accident periods for all development periods, in other words $\{C_{i,0}, C_{i,1}, \dots, C_{i,n}\}$ and $\{C_{j,0,} C_{j,1}, \dots, C_{j,n}\}$ with $i \neq j$ are independent.

1.5.3.3. The variance of the cumulative losses at the development period k is proportional to the cumulative losses at the development period $k - 1$, in other words,

$$Var\left(C_{i,k} \middle| C_{i,0}, C_{i,1}, \dots, C_{i,k-1}\right) = C_{i,k-1}\sigma_k^2$$

where $i \in \{0, 1, \dots, n\}$ and $k \in \{1, \dots, n\}$.

We are going to prove the theorem that shows that the assumptions 1.5.3.1 and 1.5.3.2 are the implicit assumptions of the chain-ladder method.

Theorem 1.5.3.1. [5] Let $D = \{C_{i,k} | i + k \leq n\}$ be the set of losses observed so far. Under the assumptions 1.5.3.1 and 1.5.3.2 we have

$$E\left(C_{i,n} | D\right) = C_{i,n-i} \prod_{l=n-i+1}^{n} \varphi_l.$$

Proof. We denote

$$E_i(X) = E\left(X \middle| C_{i,0}, C_{i,1}, \dots, C_{i,n-i}\right).$$

Then the application of 1.5.3.2 and the repeated application of 1.5.3.1 yield

$$E(C_{i,n}|D) = E_i(C_{i,n}) = E_i\left(E\left(C_{i,n} \middle| C_{i,0}, C_{i,1}, \dots, C_{i,n-1}\right)\right) = E_i(C_{i,n-1}\varphi_n) =$$

$$= E_i(C_{i,n-1})\varphi_n = \dots = E_i(C_{i,n-i})\varphi_{n-i+1} \cdot \dots \cdot \varphi_n = C_{i,n-i} \prod_{l=n-i+1}^{n} \varphi_l.$$

This theorem shows that the estimator of the ultimate losses $\{C_{i,n}\}_{i\in\{1,\dots,n\}}$defined by (1.5.2.2) has the same form as $E(C_{i,n}|D)$ which is the best forecast of $C_{i,n}$ based on the observations D.

Further, we are going to show that estimating $\prod_{l=n-i+1}^{n} \varphi_l$ by $\prod_{l=n-i+1}^{n} \hat{\varphi}_l^{CL}$ from (1.4.2.1) is reasonable.

Theorem 1.5.3.2. [5] Under the assumptions 1.5.3.1 and 1.5.3.2 the estimators of the development factors $\{\varphi_k\}_{k\in\{1,\dots,n\}}$ defined by (1.4.2.1) are unbiased and uncorrelated.

Proof. We denote

$$B_k = \{C_{i,j} | j \le k, i + j \le n\}$$

where $k \in \{0, \dots, n-1\}$.

Then the assumptions 1.5.3.2 and 1.5.3.1 yield

$$E(C_{i,k}|B_{k-1}) = E\left(C_{i,k} \middle| C_{i,0}, C_{i,1}, \dots, C_{i,k-1}\right) = C_{i,k-1}\varphi_k.$$

Thus, we have

$$E(\hat{\varphi}_k^{CL}|B_{k-1}) = \frac{\sum_{j=0}^{n-k} E(C_{j,k}|B_{k-1})}{\sum_{j=0}^{n-k} C_{j,k-1}} = \frac{\sum_{j=0}^{n-k} C_{j,k-1}\varphi_k}{\sum_{j=0}^{n-k} C_{j,k-1}} = \varphi_k$$

that leads to the unbiasedness of the factors estimates

$$E(\hat{\varphi}_k^{CL}) = E\left(E(\hat{\varphi}_k^{CL}|B_{k-1})\right) = \varphi_k$$

where $k \in \{1, \dots, n\}$.

The $\{\hat{\varphi}_k^{CL}\}_{k\in\{1,\dots,n\}}$ are uncorrelated because for $j < k$

$$E\left(\hat{\varphi}_k^{CL}\hat{\varphi}_j^{CL}\right) = E\left(E\left(\hat{\varphi}_k^{CL}\hat{\varphi}_j^{CL}|B_{k-1}\right)\right) = E\left(\hat{\varphi}_j^{CL}E\left(\hat{\varphi}_k^{CL}|B_{k-1}\right)\right) = E\left(\hat{\varphi}_j^{CL}\right)\varphi_k =$$

$$= E\left(\hat{\varphi}_j^{CL}\right)E\left(\hat{\varphi}_k^{CL}\right).$$

The fact that the chain-ladder estimates of the development factors are uncorrelated is not obvious since $\hat{\varphi}_k^{CL}$ and $\hat{\varphi}_{k-1}^{CL}$ depend on the same data $C_{0,k-1}$, $C_{1,k-1}$, …, $C_{n-k,k-1}$. This result can be easily extended to arbitrary products of pairwise different elements of the set $\left\{\hat{\varphi}_k^{CL}\right\}_{k\in\{1,\dots,n\}}$.

Accordingly, the estimators of the reserves $\{R_i\}_{i\in\{1,\dots,n\}}$ defined by (1.5.2.3) are unbiased.

The unbiased estimators of σ_k^2 are given as

$$\hat{\sigma}_k^2 = \frac{1}{n-k}\sum_{i=0}^{n-k} C_{i,k-1}\left(\frac{C_{i,k}}{C_{i,k-1}} - \hat{\varphi}_k\right)^2$$

where $k \in \{1, \dots, n-1\}$.

If $\hat{\varphi}_n^{CL} = 1$ so the claims development is believed to be finished after $(n-1)^{th}$ period, we can put $\hat{\sigma}_n = 0$. If not, we extrapolate the series $\hat{\sigma}_1, \dots, \hat{\sigma}_{n-2}, \hat{\sigma}_{n-1}$ by one additional member, for instance by loglinear regression or more simply by requiring that

$$\frac{\hat{\sigma}_{n-2}}{\hat{\sigma}_{n-1}} = \frac{\hat{\sigma}_{n-1}}{\hat{\sigma}_n}$$

holds at least as long as $\hat{\sigma}_{n-2} > \hat{\sigma}_{n-1}$. This last possibility leads to the equality

$$\hat{\sigma}_n^2 = \min\left(\frac{\hat{\sigma}_{n-1}^4}{\hat{\sigma}_{n-2}^2}, \min(\hat{\sigma}_{n-2}^2, \hat{\sigma}_{n-1}^2)\right).$$

In order to consider the mean squared errors further, we are going to prove the auxiliary lemma first.

Lemma 1.5.3.1. Let X be a random variable and a be a real number, then

$$E(X-a)^2 = Var(X) + (E(X)-a)^2.$$

Proof. Using properties the properties of the expected value and the variance definition, we get the following

$$E(X-a)^2 = E(X^2-2aX+a^2) = E(X^2) - 2aE(X) + a^2 = E(X^2) \pm$$
$$\pm[E(X)]^2 - \ 2aE(X) + a^2 = (E(X^2) - [E(X)]^2) + ([E(X)]^2 - 2aE(X) +$$
$$+a^2) = Var(X) + (E(X) - a)^2.$$

The mean squared error $mse\left(\hat{C}_{in}\right)$ of the estimator $\hat{C}_{in}$ of C_{in} is defined as

$$mse\left(\hat{C}_{i,n}\right) = E\left(\left(\hat{C}_{i,n} - C_{i,n}\right)^2 | D\right)$$

where $D = \left\{C_{i,k} | i + k \le n\right\}$ is the set of losses observed so far.

We see that

$$mse\left(\hat{R}_i\right) = E\left(\left(\hat{R}_i - R_i\right)^2 | D\right) = E\left(\left(\hat{C}_{i,n} - C_{i,n}\right)^2 | D\right) = mse\left(\hat{C}_{i,n}\right).$$

Because of the result obtained in lemma 1.5.3.1 we have

$$mse\left(\hat{C}_{i,n}\right) = Var\left(C_{i,n} | D\right) + \left(E\left(C_{i,n} | D\right) - \hat{C}_{i,n}\right)^2$$

so the mean squared error is the sum of the stochastic error (the process variance) and of the estimation error.

Now we are going to state and prove the main result of this part.

Theorem 1.5.3.3. [5] Under the assumptions 1.5.3.1, 1.5.3.2 and 1.5.3.3 the mean squared error $mse\left(\hat{R}_i\right)$ can be estimated by

$$\widehat{mse\left(\hat{R}_\iota\right)} = \hat{C}_{i,n}^2 \sum_{k=n-i+1}^{n} \frac{\hat{\sigma}_k^2}{(\hat{\varphi}_k^{CL})^2} \left(\frac{1}{\hat{C}_{i,k-1}} + \frac{1}{\sum_{j=0}^{n-k} C_{j,k-1}}\right)$$

where $\hat{C}_{i,k}^2$ are the estimated values defined by (1.5.2.1) for $i + k > n$.

Proof. We denote

$$E_i(X) = E\left(X \middle| C_{i,0}, C_{i,1}, \dots, C_{i,n-i}\right),$$

$$Var_i(X) = Var\left(X \middle| C_{i,0}, C_{i,1}, \dots, C_{i,n-i}\right).$$

Above we have shown that

$$mse(\hat{R}_i) = Var(C_{i,n}|D) + \left(E(C_{i,n}|D) - \hat{C}_{i,n}\right)^2.$$

The repeated application of assumptions 1.5.3.1 and 1.5.3.3 yields for the first term of $mse(\hat{R}_i)$

$$Var(C_{i,n}|D) = Var_i(C_{i,n}) = E_i\left(Var\left(C_{i,n} \middle| C_{i,0}, C_{i,1}, \dots, C_{i,n-1}\right)\right) +$$

$$+Var_i\left(E\left(C_{i,n} \middle| C_{i,0}, C_{i,1}, \dots, C_{i,n-1}\right)\right) = E_i(C_{i,n-1})\sigma_n^2 + Var_i(C_{i,n-1})\varphi_n^2 =$$

$$= E_i(C_{i,n-2})\varphi_{n-1}\sigma_n^2 + E_i(C_{i,n-2})\sigma_{n-1}^2\varphi_n^2 + Var_i(C_{i,n-2})\varphi_{n-1}^2\varphi_n^2 = \dots =$$

$$= C_{i,n-i} \sum_{k=n-i+1}^{n} \varphi_{n-i+1} \dots \varphi_{k-1}\sigma_k^2\varphi_{k+1}^2 \dots \varphi_n^2$$

because $Var_i(C_{i,n-i}) = 0$.

According to the theorem 1.5.3.1, we can obtain the second term of $mse(\hat{R}_i)$

$$\left(E(C_{i,n}|D) - \hat{C}_{i,n}\right)^2 = C_{i,n-i}^2 \left(\prod_{l=n-i+1}^{n} \varphi_l - \prod_{l=n-i+1}^{n} \hat{\varphi}_l^{CL}\right)^2.$$

Now we have to estimate these two terms. For the first term it can be done by replacing the unknown parameters φ_k and σ_k^2 with their estimators $\hat{\varphi}_k^{CL}$ and $\hat{\sigma}_k^2$, in other words, the estimator of $Var(C_{i,n}|D)$ is defined as

$$C_{i,n-i} \sum_{k=n-i+1}^{n} \hat{\varphi}_{n-i+1}^{CL} \cdots \hat{\varphi}_{k-1}^{CL} \hat{\sigma}_k^2 \hat{\varphi}_{k+1}^{CL} \cdots \hat{\varphi}_n^{CL} =$$

$$= \sum_{k=n-i+1}^{n} \frac{C_{i,n-i}^2 \prod_{l=n-i+1}^{n} \left(\hat{\varphi}_l^{CL}\right)^2 \hat{\sigma}_k^2}{C_{i,n-i} \prod_{l=n-i+1}^{k-1} \hat{\varphi}_l^{CL} \, (\hat{\varphi}_k^{CL})^2} = \hat{C}_{i,n}^2 \sum_{k=n-i+1}^{n} \frac{\frac{\hat{\sigma}_k^2}{(\hat{\varphi}_k^{CL})^2}}{\hat{C}_{i,k-1}}.$$

However, if we replace φ_k with $\hat{\varphi}_k^{CL}$ in the second term, we will get zero. Therefore, we are going to use another approach. We denote

$$\Phi = \prod_{l=n-i+1}^{n} \varphi_l - \prod_{l=n-i+1}^{n} \hat{\varphi}_l^{CL} = \sum_{l=n-i+1}^{n} C_l$$

where $C_k = \hat{\varphi}_{n-i+1}^{CL} \cdot \ldots \cdot \hat{\varphi}_{k-1}^{CL} \cdot \left(\varphi_k - \hat{\varphi}_k^{CL}\right) \varphi_{k+1} \cdot \ldots \cdot \varphi_n$.

Thus,

$$\Phi^2 = \left(\sum_{l=n-i+1}^{n} C_l \right)^2 = \sum_{k=n-i+1}^{n} C_k^2 + 2 \sum_{j<k} C_j C_k.$$

Afterwards we replace C_k^2 with $E(C_k^2 | B_{k-1})$ and $C_j C_k$, $j < k$, with $E\left(C_j C_k | B_{k-1}\right)$. This means that we approximate C_k^2 and $C_j C_k$ by averaging over as little data as possible, such that as many values $C_{i,k}$ as possible from the observed data are fixed. In the proof of the theorem 1.5.3.2, we have shown that $E\left(\varphi_k - \hat{\varphi}_k^{CL} | B_{k-1}\right) = 0$. Because of this $E\left(C_j C_k | B_{k-1}\right) = 0$ for $j < k$. Since

$$E\left(\left(\varphi_k - \hat{\varphi}_k^{CL}\right)^2 | B_{k-1}\right) = Var\left(\hat{\varphi}_k^{CL} | B_{k-1}\right) = \frac{\sum_{j=0}^{n-k} Var\left(C_{j,k} | B_{k-1}\right)}{\left(\sum_{j=0}^{n-k} C_{j,k-1}\right)^2} = \frac{\sigma_k^2}{\sum_{j=0}^{n-k} C_{j,k-1}},$$

we have

$$E(C_k^2|B_{k-1}) = \frac{\left(\hat{\varphi}_{n-i+1}^{CL}\right)^2 \cdot \ldots \cdot \left(\hat{\varphi}_{k-1}^{CL}\right)^2 \cdot \sigma_k^2 \cdot \varphi_{k+1}^2 \cdot \ldots \cdot \varphi_n^2}{\sum_{j=0}^{n-k} C_{j,k-1}}.$$

Hence, we replace Φ^2 with $\sum_k E(C_k^2|B_{k-1})$ and we can replace all unknown parameters σ_k^2, φ_{k+1}, …, φ_n with their unbiased estimators $\hat{\sigma}_k^2$, $\hat{\varphi}_{k+1}^{CL}$, …, $\hat{\varphi}_n^{CL}$. Altogether, we get the following estimator of Φ^2

$$\sum_{k=n-i+1}^{n} \frac{\left(\hat{\varphi}_{n-i+1}^{CL}\right)^2 \cdot \ldots \cdot \left(\hat{\varphi}_{k-1}^{CL}\right)^2 \cdot \hat{\sigma}_k^2 \cdot \left(\hat{\varphi}_{k+1}^{CL}\right)^2 \cdot \ldots \cdot (\hat{\varphi}_n^{CL})^2}{\sum_{j=0}^{n-k} C_{j,k-1}} = \prod_{l=n-i+1}^{n} \left(\hat{\varphi}_l^{CL}\right)^2 \cdot$$

$$\cdot \sum_{k=n-i+1}^{n} \frac{\frac{\hat{\sigma}_k^2}{(\hat{\varphi}_k^{CL})^2}}{\sum_{j=0}^{n-k} C_{j,k-1}}.$$

After substituting this result in the second term, we have

$$\left(E\left(C_{i,n}|D\right) - \hat{C}_{i,n}\right)^2 = C_{i,n-i}^2 \prod_{l=n-i+1}^{n} \left(\hat{\varphi}_l^{CL}\right)^2 \cdot \sum_{k=n-i+1}^{n} \frac{\frac{\hat{\sigma}_k^2}{(\hat{\varphi}_k^{CL})^2}}{\sum_{j=0}^{n-k} C_{j,k-1}} = \hat{C}_{i,n}^2 \cdot$$

$$\cdot \sum_{k=n-i+1}^{n} \frac{\frac{\hat{\sigma}_k^2}{(\hat{\varphi}_k^{CL})^2}}{\sum_{j=0}^{n-k} C_{j,k-1}}.$$

Finally, we obtain the result from the statement of the theorem.

Definition 1.5.3.1. The square root of an estimator of the mean squared error is defined to be the standard error of $\hat{R}_i$ ($s.e.\left(\hat{R}_i\right)$).

Corollary 1.5.3.1. [5] The mean squared error of the total reserve estimate $\hat{R} = \sum_{i=1}^{n} \hat{R}_i$ can be estimated by

$$\left(s.e.\left(\hat{R}\right)\right)^2 = \widehat{mse\left(\hat{R}\right)} = \sum_{i=1}^{n}\left\{\left(s.e.\left(\hat{R}_i\right)\right)^2 + \hat{C}_{i,n}\left(\sum_{j=i+1}^{n}\hat{C}_{j,n}\right)\sum_{k=n-i+1}^{n}\frac{2\frac{\hat{\sigma}_k^2}{(\hat{\varphi}_k^{CL})^2}}{\sum_{t=0}^{n-k}C_{t,k}}\right\}.$$

Proof. We have

$$mse\left(\sum_{i=1}^{n}\hat{R}_i\right) = E\left(\left(\sum_{i=1}^{n}\hat{R}_i - \sum_{i=1}^{n}R_i\right)^2\middle| D\right) = E\left(\left(\sum_{i=1}^{n}\hat{C}_{i,n} - \sum_{i=1}^{n}C_{i,n}\right)^2\middle| D\right) =$$

$$= Var\left(\sum_{i=1}^{n}C_{i,n}\middle| D\right) + \left(E\left(\sum_{i=1}^{n}C_{i,n}\middle| D\right) - \sum_{i=1}^{n}\hat{C}_{i,n}\right)^2.$$

The independence of the accident years yields

$$Var\left(\sum_{i=1}^{n}C_{i,n}\middle| D\right) = \sum_{i=1}^{n}Var\left(C_{i,n}|D\right).$$

We have already calculated these summands in the proof of the theorem 1.5.3.3.

When we look at the second term, we obtain

$$\left(E\left(\sum_{i=1}^{n}C_{i,n}\middle| D\right) - \sum_{i=1}^{n}\hat{C}_{i,n}\right) = \left(\sum_{i=1}^{n}\left(E\left(C_{i,n}|D\right) - \hat{C}_{i,n}\right)\right)^2 =$$

$$= \sum_{i,j=1}^{n}\left(E\left(C_{i,n}|D\right) - \hat{C}_{i,n}\right)\cdot\left(E\left(C_{j,n}|D\right) - \hat{C}_{j,n}\right) = \sum_{i,j=1}^{n}C_{i,n-i}C_{j,n-j}\Phi_i\Phi_j$$

where

$$\Phi_i = \prod_{l=n-i+1}^{n}\varphi_l - \prod_{l=n-i+1}^{n}\hat{\varphi}_l^{CL}.$$

In the proof of the theorem 1.5.3.3, we have shown with that

$$mse\left(\hat{R}_i\right) = Var\left(C_{i,n}|D\right) + \left(C_{i,n-i}\Phi_i\right)^2.$$

Therefore,

$$mse\left(\sum_{i=1}^{n} \hat{R}_i\right) = \sum_{i=1}^{n} mse\left(\hat{R}_i\right) + \sum_{1\le i<j\le n} 2 \cdot C_{i,n-i} \cdot C_{j,n-j}\Phi_i\Phi_j.$$

In the proof of the theorem 1.5.3.3, we have found the estimator for Φ^2. Using the similar approach, we get the estimators for $\Phi_i\Phi_j$, $i < j$,

$$\sum_{k=n-i+1}^{n} \frac{\hat{\varphi}_{n-j+1}^{CL} \cdot \ldots \cdot \hat{\varphi}_{n-i}^{CL} \cdot \left(\hat{\varphi}_{n-i+1}^{CL}\right)^2 \cdot \ldots \cdot \left(\hat{\varphi}_{k-1}^{CL}\right)^2 \cdot \hat{\sigma}_k^2 \cdot \left(\hat{\varphi}_{k+1}^{CL}\right)^2 \cdot \ldots \cdot (\hat{\varphi}_n^{CL})^2}{\sum_{t=0}^{n-k} C_{t,k-1}}.$$

Thus, we have

$$\sum_{1\le i<j\le n} C_{i,n-i} \prod_{l=n-i+1}^{n} \hat{\varphi}_l^{CL} \cdot C_{j,n-j} \prod_{l=n-j+1}^{n} \hat{\varphi}_l^{CL} \sum_{k=n-i+1}^{n} \frac{\frac{\hat{\sigma}_k^2}{(\hat{\varphi}_k^{CL})^2}}{\sum_{t=0}^{n-k} C_{t,k-1}} =$$

$$= \sum_{i=1}^{n} \hat{C}_{i,n} \left(\sum_{j=i+1}^{n} \hat{C}_{i,n}\right) \sum_{k=n-i+1}^{n} \frac{2\frac{\hat{\sigma}_k^2}{(\hat{\varphi}_k^{CL})^2}}{\sum_{t=0}^{n-k} C_{t,k-1}}.$$

This completes the proof.

Having calculated the estimators for $\hat{R}$ and $s.e.\left(\hat{R}\right)$, we can use them as parameters to build the loss distribution. The question is which distribution to choose. Two most popular loss distributions are normal and lognormal [17]. For instance, in the case of the lognormal distribution we approximate the unknown distribution of $\hat{R}$ by a lognormal distribution with parameters μ and σ^2 such that mean values as well as variances of both distributions are equal, in other words, such that

$$\exp\left(\mu + \frac{\sigma^2}{2}\right) = \hat{R}$$

$$\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1) = \left(s.e.\left(\hat{R}\right)\right)^2.$$

This leads to $\sigma^2 = \ln\left(1 + \frac{\left(s.e.(\hat{R})\right)^2}{\hat{R}^2}\right)$ and $\mu = \ln\left(\hat{R}\right) - \frac{\sigma^2}{2}$.

1.5.4. The bootstrapping method

Generally speaking, bootstrapping is a resampling method of independent sampling with replacement from an existing sample data with same sample size n, and performing estimation of the sampling distribution based on these resampled data [8, 9].

The bootstrapping technique that is going to be considered in this work involves the following steps:

1. We calculate the age-to-age factors from a run-off triangle of cumulative losses using the data observed so far. These factors represent the ratio of loss amounts from one valuation date to another, and they are intended to capture growth patterns of losses over time.

2. The previous step results in an upper triangle of age-to-age factors. Then a lower triangle is built using a technique known as “resampling with replacement”. When an age-to-age factor is sampled, it is not removed from the upper triangle and can hence be sampled again. Age-to-age factors in each column of the lower triangle are sampled from the known age-to-age factors in the same column of the upper triangle. The possibility to choose any age-to-age factor in a column is uniform.

3. The age-to-age factors in the lower triangle are then applied to the corresponding values of the run-off triangle of cumulative losses to project the future claims. Then the reserve estimates $\hat{R}_i = \hat{C}_{i,n} - C_{i,n-i}$, $i \in \{1, \dots, n\}$, are found.

The results of this step are saved as the results produced by one simulation.

4. Further simulations are produced by repeating steps 2 and 3.

In the end, we can build the distribution of future losses.

1.6. The time value of money and the inflation impact

The time value of money is the concept that money a person has currently is worth more than the identical sum in the future due to its potential earning capacity [11]. Discounting is the process of determining the present value of payments that are to be received in the future. According to the new standard [2], insurers are obliged to adjust the estimates of reserves to reflect the time value of money if the claim settlement exceeds 12 months. Besides, the impact of inflation that measures the rate at which the purchasing power of currency is falling has to be taken into account.

Suppose we have projected the values $\hat{C}_{i,k}$ with $i + k > n$ using one of the modelling methods mentioned above and we have estimated the *inflation rate* r^{infl} as well as the discrete values of *discount rates* for future years $\left\{r_j^{disc}\right\}_{j\in\{1,\dots,m\}}$ where m is a number of future years in the run-off triangle. Then $\hat{Z}_{i,k}$ with $i + k > n$ can be derived as defined in (1.4.1.2). In order to adjust the incremental losses for the time value of money and the inflation impact we can apply the following formula

$$\hat{Z}_{i,k}^{Adj} = \hat{Z}_{i,k} \cdot \left(\frac{1 + r^{infl}}{1 + r_j^{disc}}\right)^{\frac{y}{Y}}$$

where $i + k > n$, $j \in \{1, \dots, m\}$, $\hat{Z}_{i,k}$ is one of the incremental losses in the future year j, Y is a number of days in the year (usually 365) and y is the difference in days between the valuation date and the future date. The future date can be taken, for instance, as the middle of the last month of the future period.

SECTION 2. REALISATION IN R

2.1. The programming language R

R is a programming language and environment for statistical computing and graphics, which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. There exists the integrated development environment for R – RStudio [18]. R provides a wide variety of statistical techniques (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering etc.). One of the greatest strengths of R is the ease with which we can create high-quality graphics. Furthermore, there are packages in R specifically created for the actuarial work. One of them, "ChainLadder", has built-in functions dedicated to the Mack method and several bootstrapping techniques [19]. However, they do not allow for the time value of money and the inflation impact. Therefore, we have used own functions.

2.2. The input data

The idea is to build the basic and discounted loss distributions of reserves for LIC using the Mack and the bootstrapping methods and find the order of quantiles of these distributions to which the reserves amounts obtained by the chain-ladder and Bornhuetter-Ferguson methods correspond. According to the definition of a quantile (1.3.1), that shows us the probability of a company fulfilling its liability in case any of deterministic approaches are chosen for reserves projection. We have also set the desired probability of a liability fulfilment as 80%. Thus, in cases when the reserves amounts obtained with the deterministic approaches are not enough, we compute the additional sums that have to be reserved in order to reach the desired confidence level.

As the input run-off triangles of the cumulative losses, we have taken some built-in datasets of R that are usually used as demonstration material [19]:

- GenIns (a run off triangle of accumulated general insurance claims data)

Table 2.2.1

Accident period	Development period									
	0	1	2	3	4	5	6	7	8	9
0	357,848	1,124,788	1,735,330	2,218,270	2,745,596	3,319,994	3,466,336	3,606,286	3,833.515	3,901,463
1	352,118	1,236,139	2,170,033	3,353,322	3,799,067	4,120,063	4,647,867	4,914,039	5,339,085	
2	290,507	1,292,306	2,218,525	3,235,179	3,985,995	4,132,918	4,628,910	4,909,315		
3	310,608	1,418,858	2,195,047	3,757,447	4,029,929	4,381,982	4,588,268			
4	443,160	1,136,350	2,128,333	2,897,821	3,402,672	3,873,311				
5	396,132	1,333,217	2,180,715	2,985,752	3,691,712					
6	440,832	1,288,463	2,419,861	3,483,130						
7	359,480	1,421,128	2,864,498							
8	376,686	1,363,294								
9	344,014									

- MCLpaid (a run-off triangle based on a fire portfolio)

Table 2.2.2

Accident period	Development period						
	0	1	2	3	4	5	6
0	576	1,804	1,970	2,024	2,074	2,102	2,131
1	866	1,948	2,162	2,232	2,284	2,348	
2	1,412	3,758	4,252	4,416	4,494		
3	2,286	5,292	5,724	5,850			
4	1,868	3,778	4,648				
5	1,442	4,010					
6	2,044						

- MedMal$MedMalPaid (a run-off triangle of medical malpractice insurance)

Table 2.2.3

Accident period	Development period							
	0	1	2	3	4	5	6	7
0	125,000	406,000	1,443,000	2,986,000	4,467,000	8,179,000	12,638,000	15,815,000
1	43,000	529,000	2,016,000	3,641,000	7,523,000	14,295,000	18,983,000	
2	295,000	1,147,000	2,479,000	5,071,000	11,399,000	17,707,000		
3	50,000	786,000	3,810,000	9,771,000	18,518,000			
4	213,000	833,000	3,599,000	11,292,000				
5	172,000	1,587,000	6,267,000					
6	210,000	1,565,000						
7	209,000							

- RAA (a run-off triangle of automatic facultative business in general liability)

Table 2.2.4

Accident period	Development period									
	0	1	2	3	4	5	6	7	8	9
0	5,012	8,269	10,907	11,805	13,539	16,181	18,009	18,608	18,662	18,834
1	106	4,285	5,396	10,666	13,782	15,599	15,496	16,169	16,704	
2	3,410	8,992	13,873	16,141	18,735	22,214	22,863	23,466		
3	5,655	11,555	15,766	21,266	23,425	26,083	27,067			
4	1,092	9,565	15,836	22,169	25,955	26,180				
5	1,513	6,445	11,702	12,935	15,852					
6	557	4,020	10,946	12,314						
7	1,351	6,947	13,112							
8	3,133	5,395								
9	2,063									

Other input data are given in table 2.2.5.

Table 2.2.5

Type of input data	Value	Comments
Inflation rate	1.631%	European Union annual inflation in 2019 according to the World Bank [20]; applicable to all run-off triangles.
Discount rates	1 year – 0.055% 2 year – 0.045% 3 year – 0.065% 4 year – 0.095% 5 year – 0.135% 6 year – 0.174% 7 year – 0.214% 8 year – 0.254%	Risk-free rates with volatility adjustment for Euro calculated by European Insurance and Occupational Pensions Authority as of 31/03/2020 [21]; applicable to all run-off triangles.
Valuation date	31/12/2019	Applicable to all run-off triangles.
Assumption about probability distribution for the Mack method	Lognormal	Applicable to all run-off triangles.
Number of simulations for the bootstrapping method	10,000	Applicable to all run-off triangles.
Assumptions about periods type	Quarterly basis	Applicable to GenIns.
	Annual Basis	Applicable to MCLpaid.
	Half-year basis	Applicable to MedMal$MedMalPaid and RAA.

Earned premiums	2017Q3 – 3,802,596 2017Q4 – 6,993,203 2018Q1 – 7,105,451 2018Q2 – 6,907,309 2018Q3 – 5,470,946 2018Q4 – 5,247,609 2019Q1 – 5,375,851 2019Q2 – 9,143,934 2019Q3 – 6,389,883 2019Q4 – 5,980,535	Applicable to GenIns.
	2013 – 2,010 2014 – 2,429 2015 – 4,356 2016 – 5,718 2017 – 5,550 2018 – 4,518 2019 – 6,098	Applicable to MCLpaid.
	2016H1 – 15,783,433 2016H2 – 23,063,142 2017H1 – 40,280,059 2017H2 – 65,512,341 2018H1 – 77,255,603 2018H2 – 84,962,130 2019H1 – 78,039,361 2019H2 – 91,671,888	Applicable to MedMal$MedMalPaid.
	2015H1 – 21,020 2015H2 – 15,466	Applicable to RAA.

	2016H1 – 31,075 2016H2 – 31,681 2017H1 – 28,139 2017H2 – 20,001 2018H1 – 24,550 2018H2 – 22,941 2019H1 – 15,730 2019H2 – 17,526	
Loss ratios	2017Q3 – 1.016 2017Q4 – 0.767 2018Q1 – 0.747 2018Q2 – 0.757 2018Q3 – 0.878 2018Q4 – 0.964 2019Q1 – 1.043 2019Q2 – 0.732 2019Q3 – 0.873 2019Q4 – 0.821	Applicable to GenIns.
	2013 – 1.05 2014 – 0.97 2015 – 1.058 2016 – 1.071 2017 – 0.901 2018 – 1.082 2019 – 0.995	Applicable to MCLpaid.
	2016H1 – 0.992 2016H2 – 1.02	Applicable to MedMal$MedMalPaid.

	2017H1 – 0.764 2017H2 – 0.845 2018H1 – 0.853 2018H2 – 1.059 2019H1 – 1.068 2019H2 – 0.748	
	2015H1 – 0.886 2015H2 – 1.08 2016H1 – 0.765 2016H2 – 0.896 2017H1 – 1.018 2017H2 – 0.965 2018H1 – 0.713 2018H2 – 1.037 2019H1 – 1.01 2019H2 – 1.04	Applicable to RAA.

2.3. The modelling results

After the execution of the programme, the following results have been obtained.

Table 2.3.1 – GenIns results

		Stochastic methods	
		Bootstrapping	Mack
Desired reserve*		19,531,742	20,671,824
Desired reserve**		19,741,101	20,892,024
Deterministic methods			
Chain-ladder	Reserve	18,680,856	18,680,856
	Reserve**	18,889,151	18,889,151
	Order of quantile	0.656	0.526
	Order of quantile**	0.657	0.526
	Additional amount	850,886	1,990,968
	Additional amount**	851,950	2,002,873
Borhuetter-Ferguson	Reserve	18,460,879	18,460,879
	Reserve**	18,666,725	18,666,725
	Order of quantile	0.609	0.490
	Order of quantile**	0.611	0.489
	Additional amount	1,070,863	2,210,945
	Additional amount**	1,074,376	2,225,299

*the quantile of the order 0.8

**the time value of money and inflation impact

The shapes of the distribution are presented in figures 2.3.1-2.3.4.

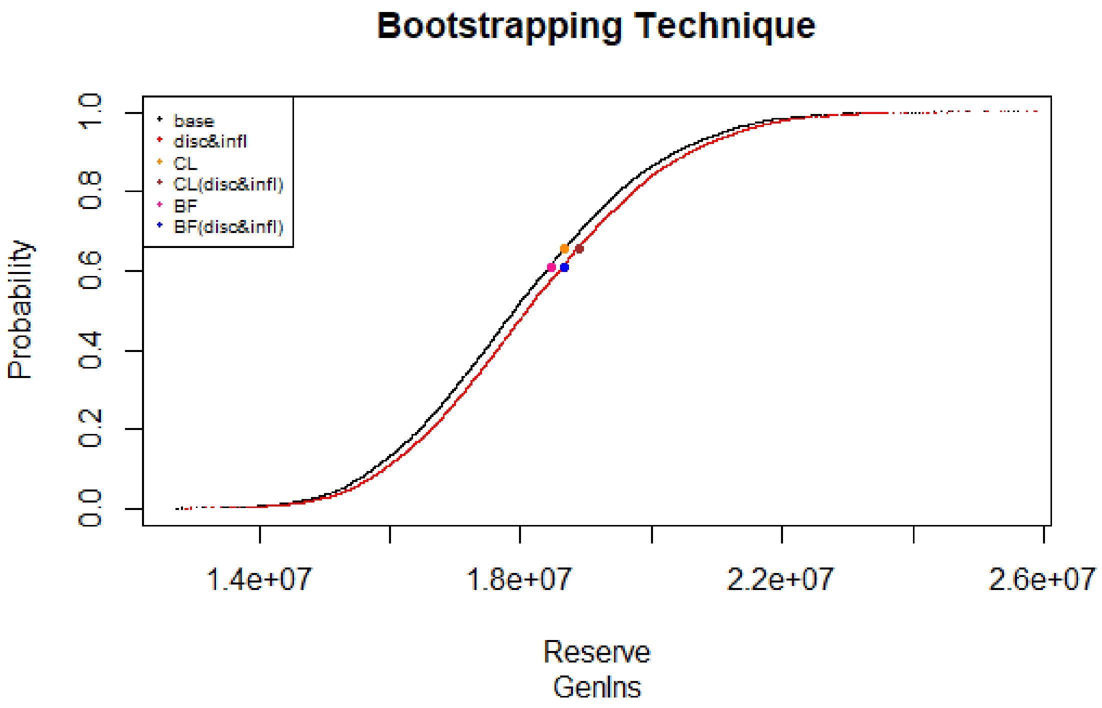


Fig. 2.3.1

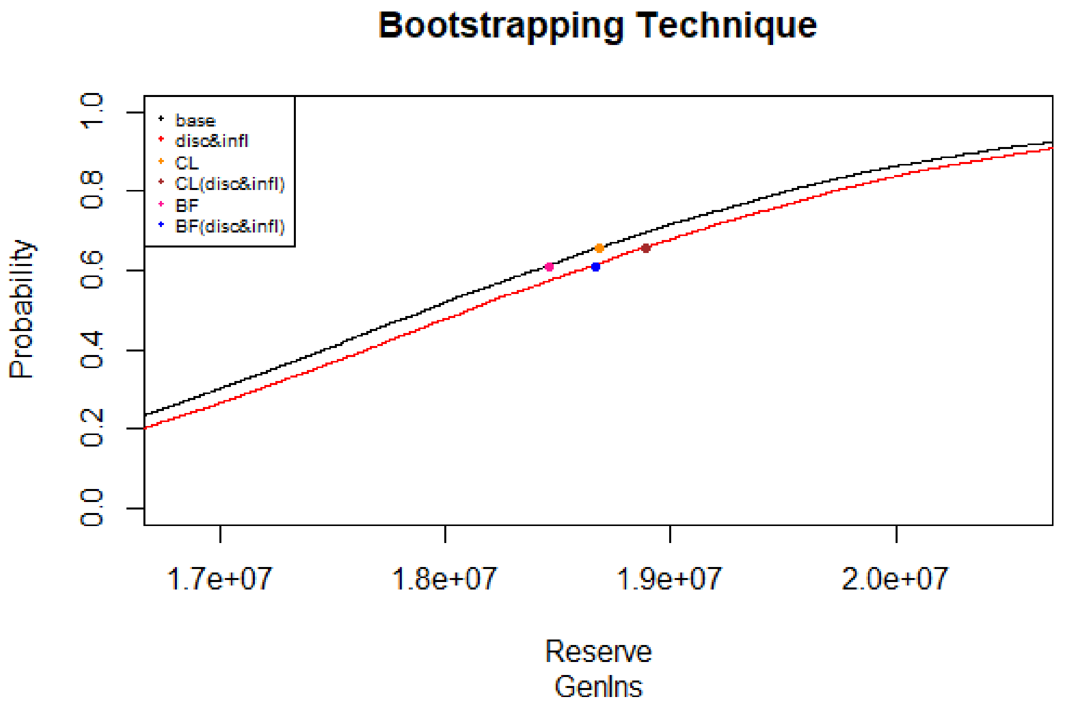


Fig. 2.3.2 – Close-up

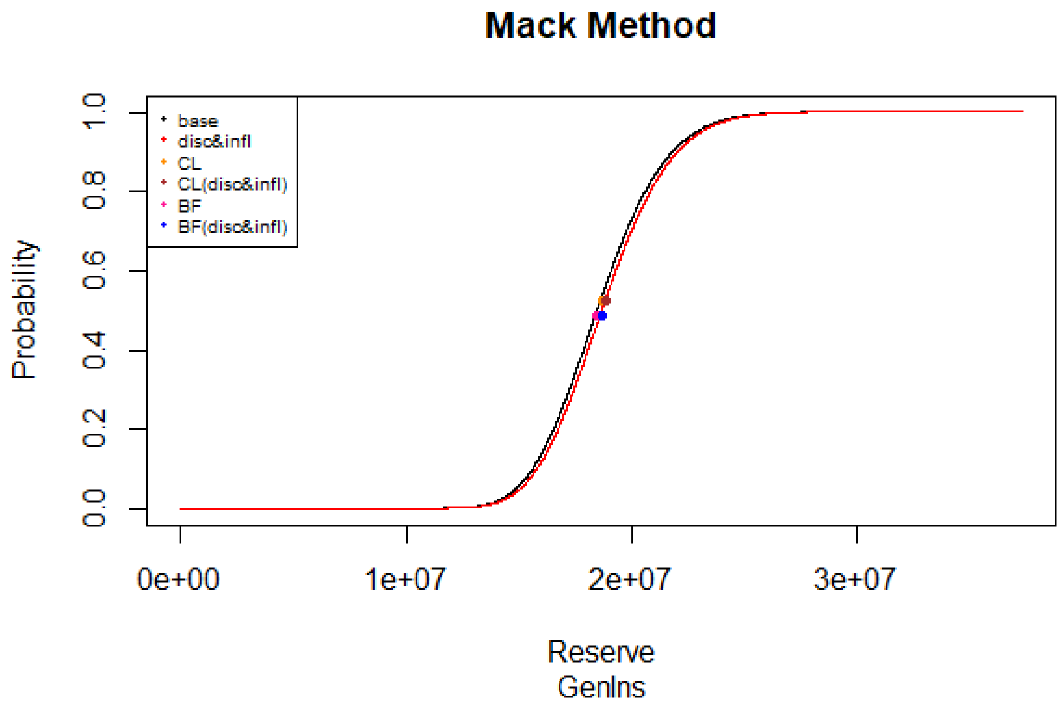


Fig. 2.3.3

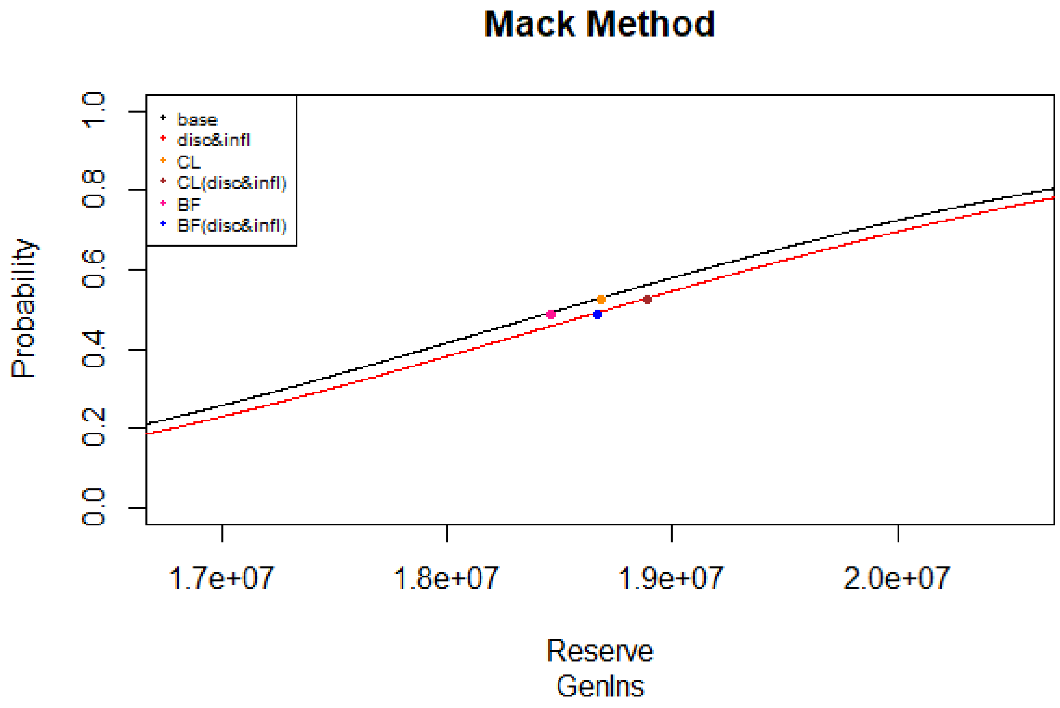


Fig. 2.3.4 – Close-up

Table 2.3.2 – MCLpaid results

		Stochastic methods	
		Bootstrapping	Mack
Desired reserve*		6,783	6,737
Desired reserve**		6,941	6,898
Deterministic methods			
Chain-ladder	Reserve	5,938	5,938
	Reserve**	6,091	6,091
	Order of quantile	0.570	0.533
	Order of quantile**	0.575	0.533
	Additional amount	845	798
	Additional amount**	850	806
Borhuetter-Ferguson	Reserve	5,880	5,880
	Reserve**	6,031	6,031
	Order of quantile	0.549	0.509
	Order of quantile**	0.555	0.509
	Additional amount	903	857
	Additional amount**	910	866

*the quantile of the order 0.8

**the time value of money and inflation impact

The shapes of the distribution are presented in figures 2.3.5-2.3.8.

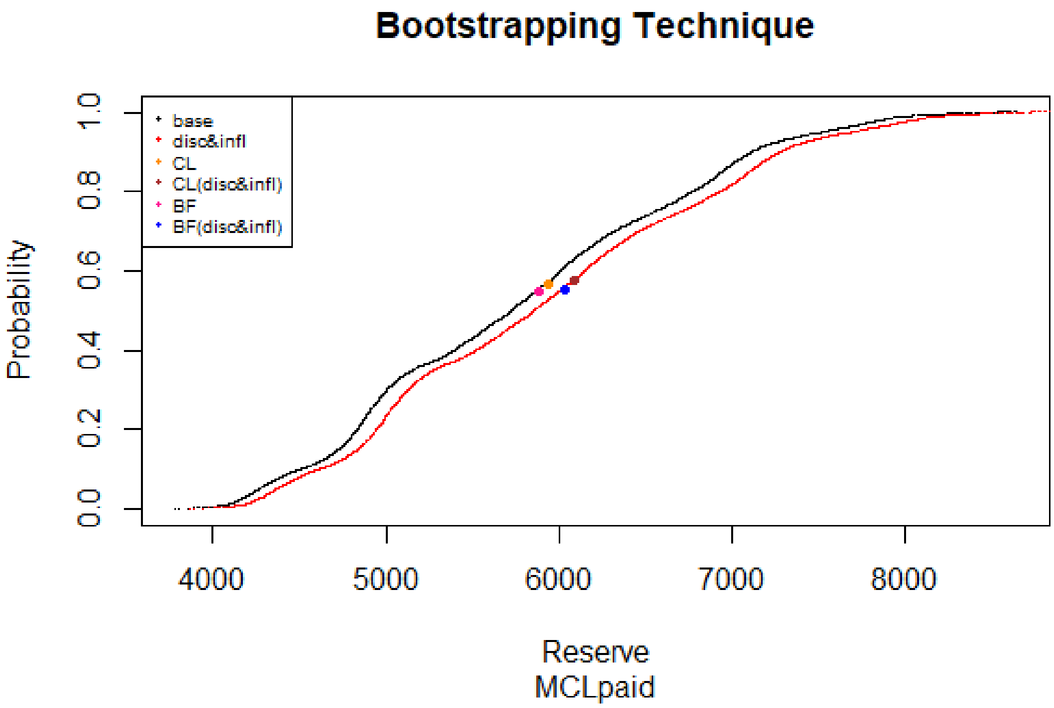


Fig. 2.3.5

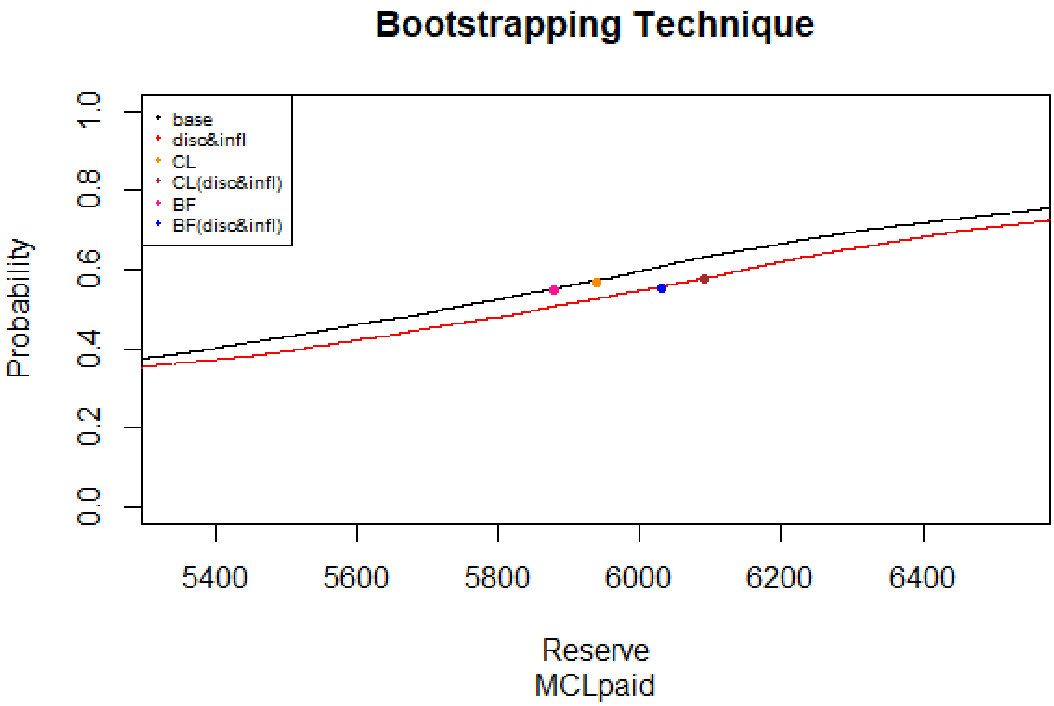


Fig. 2.3.6 – Close-up

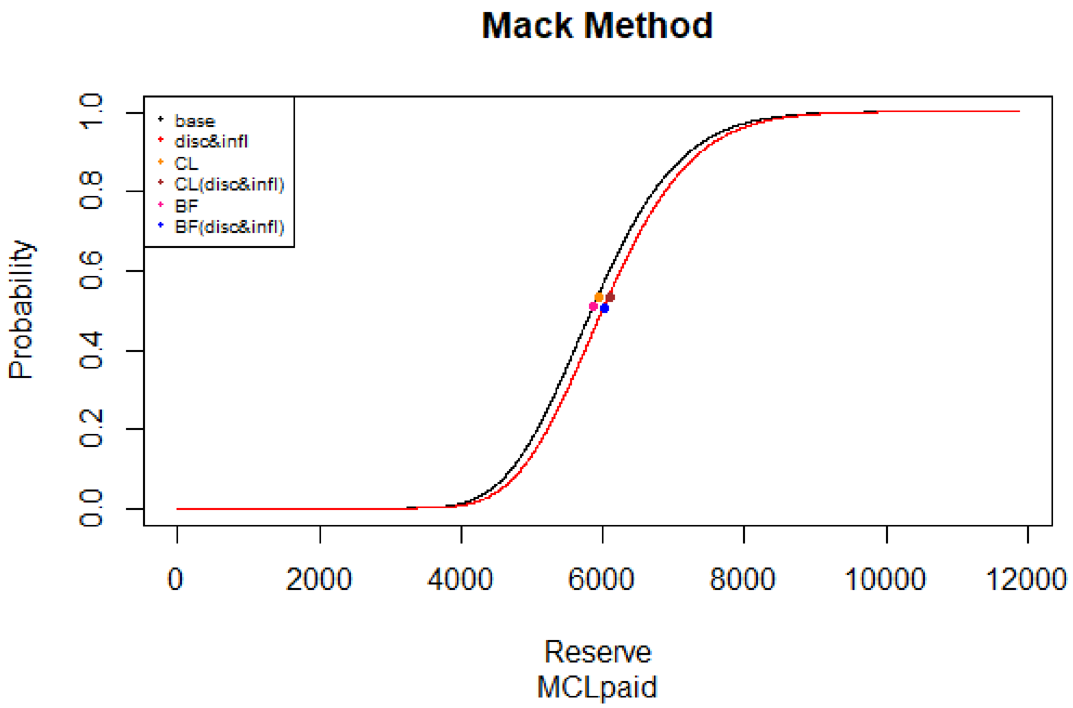


Fig. 2.3.7

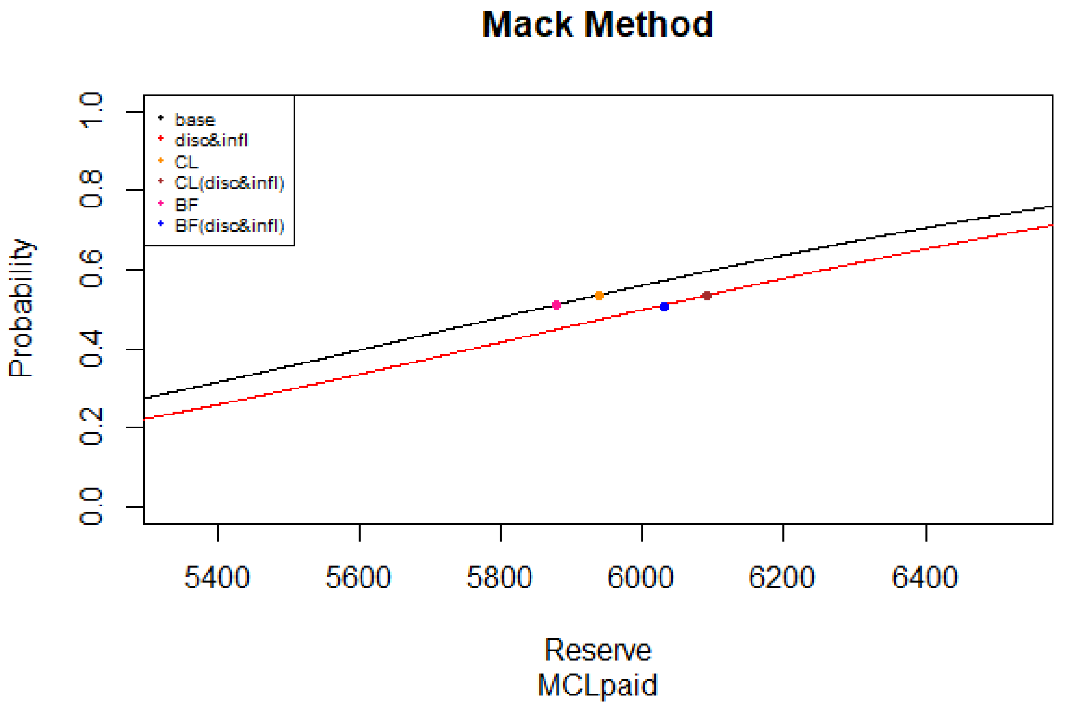


Fig. 2.3.8 – Close-up

Table 2.3.3 – MedMal$MedMalPaid results

		Stochastic methods	
		Bootstrapping	Mack
Desired reserve*		326,264,664	414,167,915
Desired reserve**		334,643,082	425,070,142
Deterministic methods			
Chain-ladder	Reserve	347,513,678	347,513,678
	Reserve**	356,864,548	356,864,548
	Order of quantile	0.865	0.549
	Order of quantile**	0.865	0.548
	Additional amount	0	66,654,237
	Additional amount**	0	68,205,594
Borhuetter-Ferguson	Reserve	343,742,237	343,742,237
	Reserve**	352,991,118	352,991,118
	Order of quantile	0.855	0.531
	Order of quantile**	0.855	0.530
	Additional amount	0	70,425,678
	Additional amount**	0	72,079,023

*the quantile of the order 0.8

**the time value of money and inflation impact

The shapes of the distribution are presented in figures 2.3.9-2.3.12.

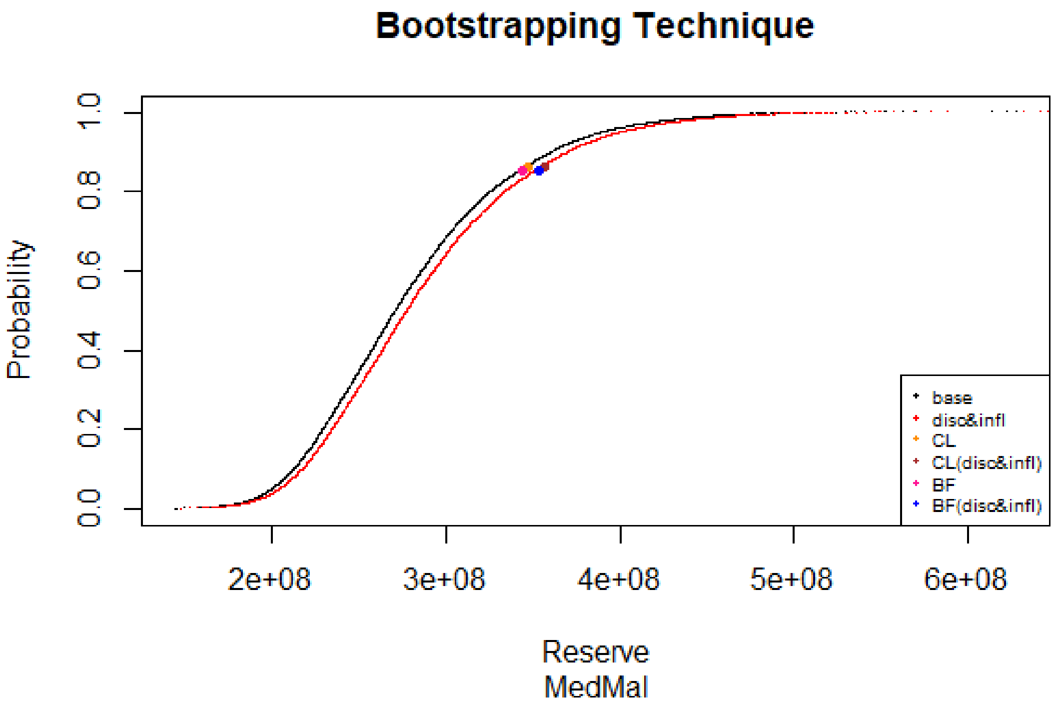


Fig. 2.3.9

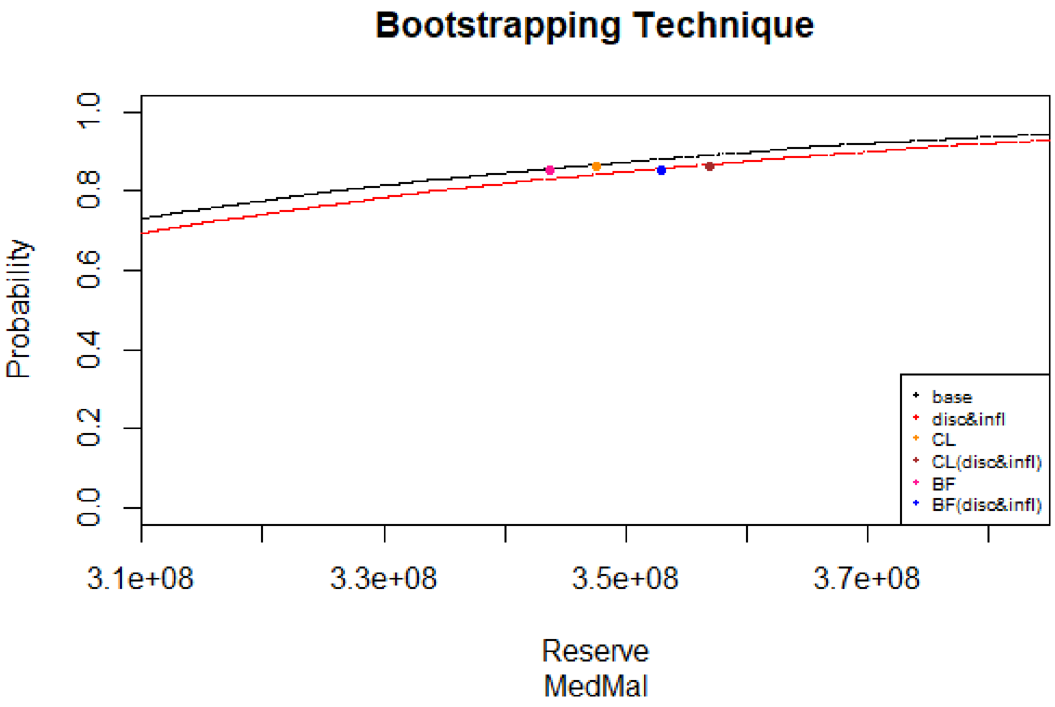


Fig. 2.3.10 – Close-up

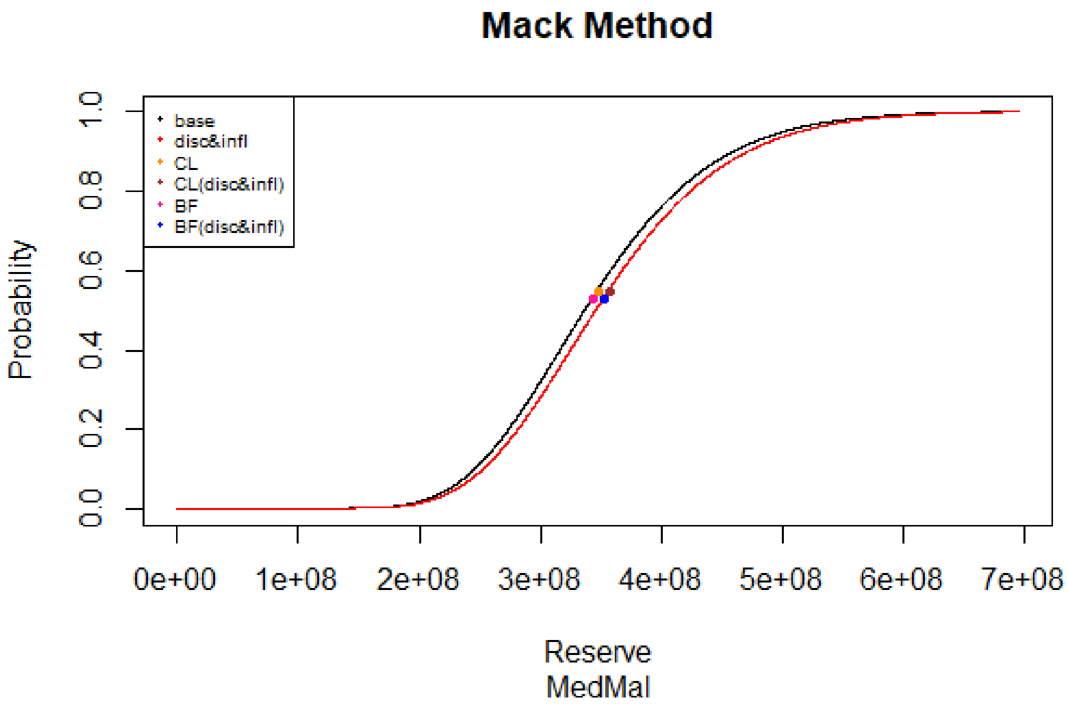
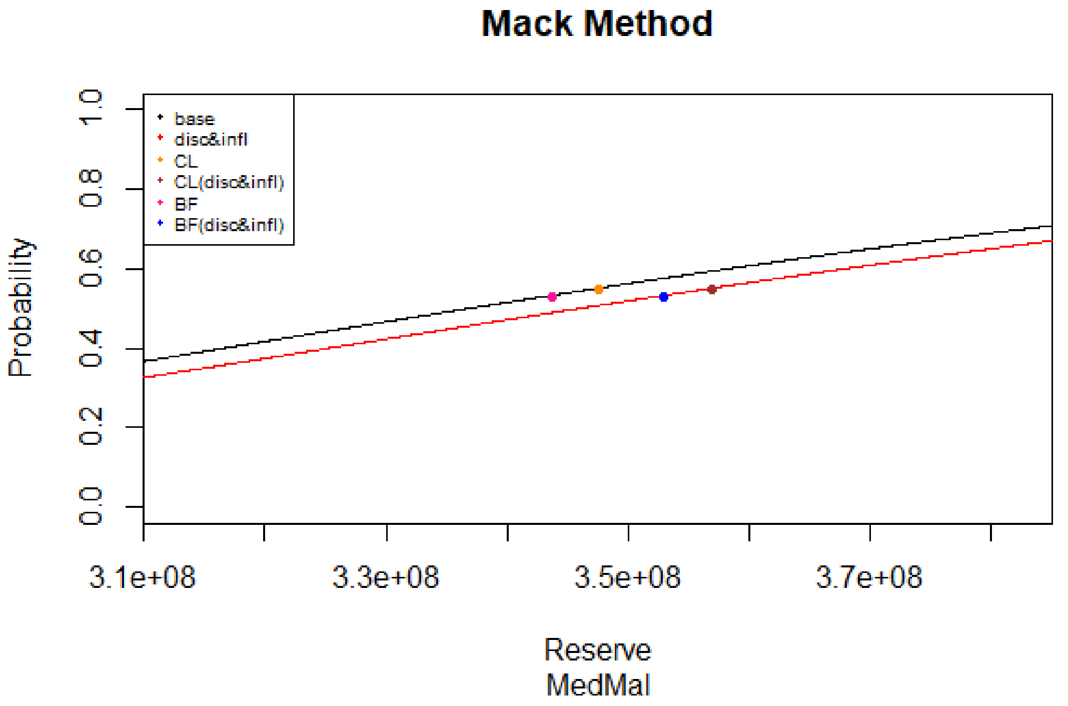


Fig. 2.3.11



Fig, 2.3.12

Table 2.3.4 – RAA results

		Stochastic methods	
		Bootstrapping	Mack
Desired reserve*		94,609	69,739
Desired reserve**		96,458	71,012
Deterministic methods			
Chain-ladder	Reserve	52,135	52,135
	Reserve**	53,160	53,160
	Order of quantile	0.312	0.596
	Order of quantile**	0.313	0.595
	Additional amount	42,474	17,604
	Additional amount**	43,298	17,852
Borhuetter-Ferguson	Reserve	51,604	51,604
	Reserve**	52,619	52,619
	Order of quantile	0.301	0.588
	Order of quantile**	0.302	0.587
	Additional amount	43,005	18,135
	Additional amount**	43,839	18,393

*the quantile of the order 0.8

**the time value of money and inflation impact

The shapes of the distribution are presented in figures 2.3.13-2.3.16.

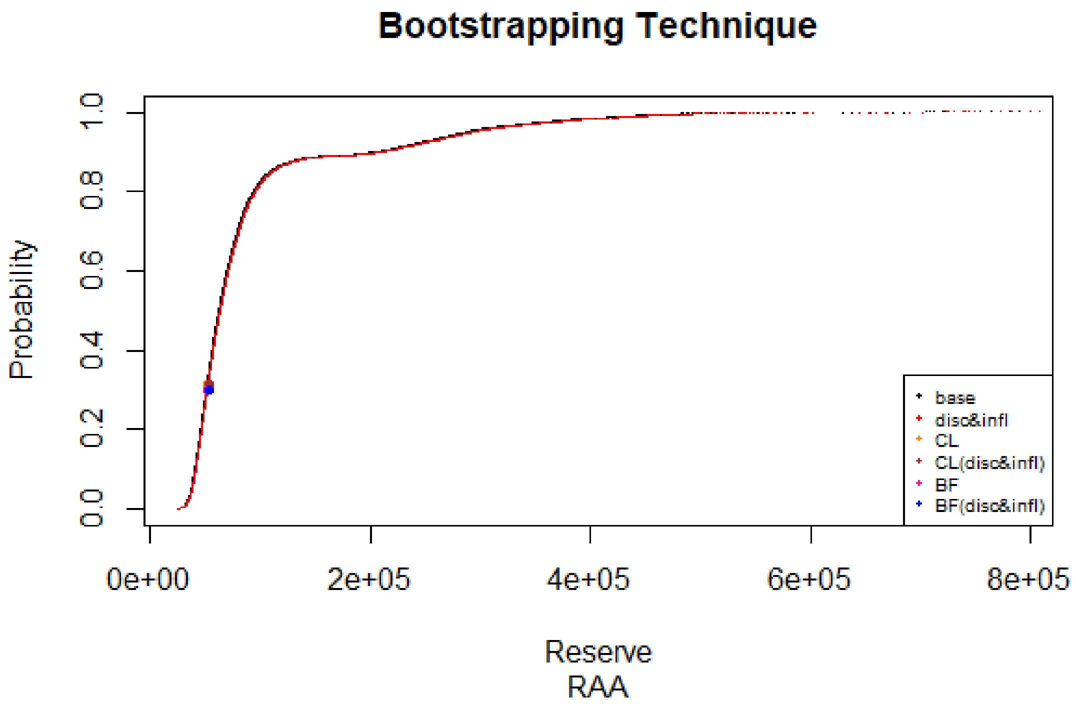


Fig. 2.3.13

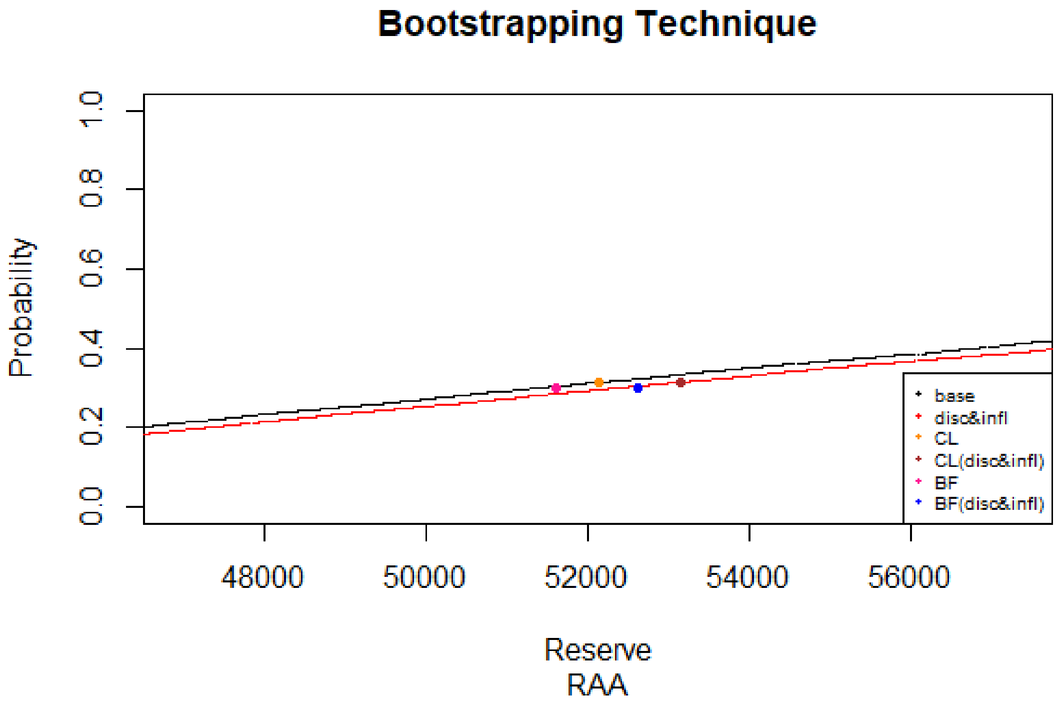


Fig. 2.3.14 – Close-up

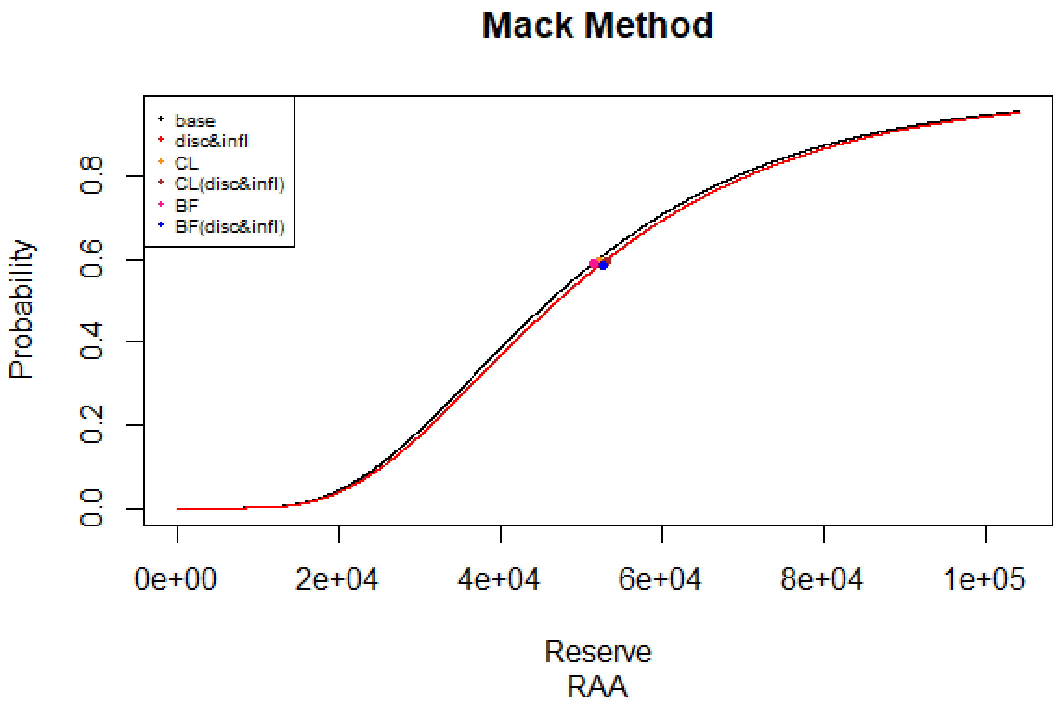


Fig. 2.3.15

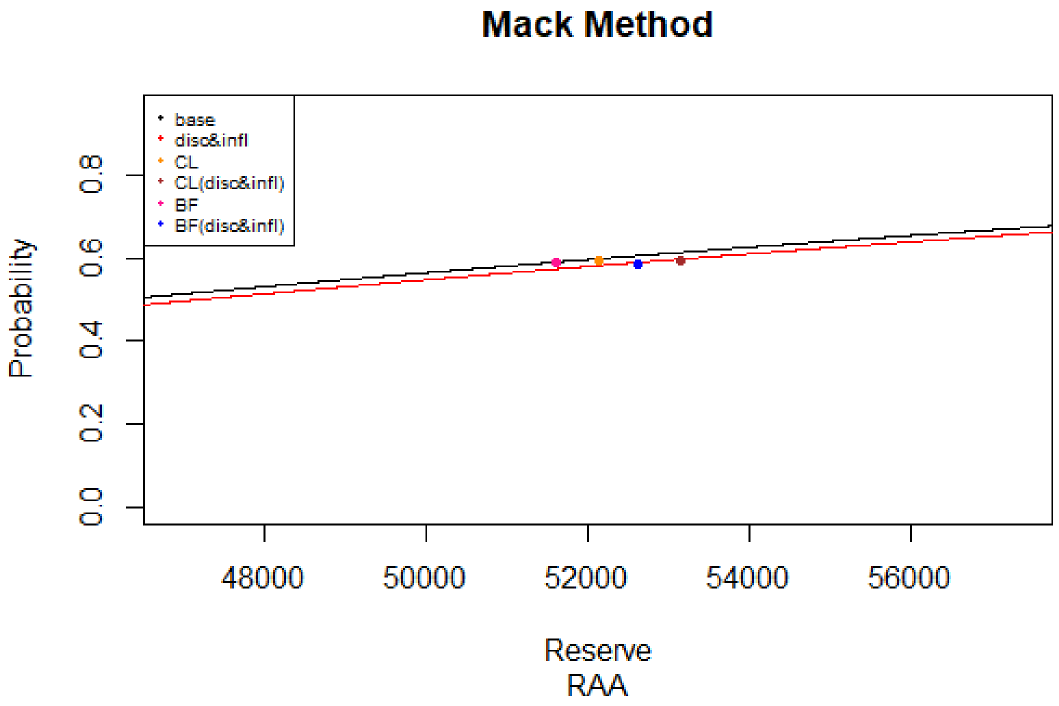


Fig. 2.3.16 – Close-up

CONCLUSIONS

The insurance industry and the actuarial science are currently facing quite a lot of fascinating challenges. One of them is new requirements for loss reserves calculations.

In this work, we have considered two common classical deterministic methods for loss reserving – the Bornhuetter-Ferguson method and the chain-ladder method as well as two modifications of famous stochastic models – the Mack method and the bootstrapping technique. In order to achieve the realistic prognosis of future losses, we have set the desired confidence level, in other words, the probability of a liability fulfilment. We have also juxtaposed deterministic and stochastic results and calculated the additional amounts that have to be reserved in order to reach the confidence level if this additional reserving has been needed. Furthermore, the time value of money and the inflation impact have been taken into account in all cases.

In order to perform these calculations and build plots that illustrate the obtained results, we have used R – the programming language and environment for statistical computing and graphics (IDE: RStudio).

Since the topics mentioned and considered in this work are highly relevant, the further research in this area, creating of new models and applications are advisable.

REFERENCES

1. IFRS 4 Insurance Contracts: https://www.ifrs.org/issued-standards/list-of-standards/ifrs-4-insurance-contracts/

2. IFRS 17 Insurance Contracts: https://www.ifrs.org/issued-standards/list-of-standards/ifrs-17-insurance-contracts/

3. Klaus D. Schmidt. Methods and Models of Loss reserving Based on Run-Off Triangles: A Unifying Survey. Casualty Actuarial Society Forum, Fall 2006: https://pdfs.semanticscholar.org/685f/fe447207da2da251a2be5b604cd6731774fc.pdf

4. Klaus D. Schmidt & Mathias Zocher. Bornhuetter-Ferguson as a General Principle of Loss Reserving. ASTIN Conference, Manchester, July 14-16, 2008: http://www.actuaries.org/ASTIN/Colloquia/Manchester/Presentations/Schmidt.pdf

5. Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. Astin Bulletin, Vol. 23, No. 2, 1993, pp. 213-225.

6. Sloma, Przemyslaw. Generalized Mack Chain-Ladder Model of Reserving with Robust Estimation. Variance 12:2, 2019, pp. 226-248.

7. Thomas Mack. The standard error of chain ladder reserve estimates: recursive calculation and inclusion of a tail factor. Astin Bulletin, Vol. 29, No 2, 1999, pp. 361-366.

8. Scott. P. Weinstein. Practical Applications of the Mack and Bootstrapping Methods when Estimating Reserve Ranges. 2007 Casualty Loss Reserve Seminar, September 10, 2007: https://www.slideserve.com/guinevere/scott-p-weinstein-fcas-maaa-ash-ruparelia-2007-casualty-loss-reserve-seminar-september-10-2007

9. Tim Hesterberg, David S. Moore, Shaun Monaghan, Ashley Clipson, and Rachel Epstein. Bootstrap Method and Permutation Tests: https://web.archive.org/web/20060215221403/http://bcs.whfreeman.com/ips5e/content/cat_080/pdf/moore14.pdf

10. Newton L. Bowers Jr., Hans U. Gerber, James C. Hickman, Donald A. Jones, Cecil J. Nesbitt. Actuarial Mathematics. The Society of Actuaries, 1997.

11. https://www.investopedia.com/

12. Ukrainian law on non-life reserves: Про затвердження Методики формування страхових резервів за видами страхування, іншими, ніж страхування життя: https://zakon.rada.gov.ua/laws/show/z0019-05

13. Christopher L. Pulp. Structured Finance and Insurance. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

14. Encyclopedia of Mathematics: https://www.encyclopediaofmath.org/

15. A.N. Kolmogorov. Foundations of the theory of probability. Chelsea, reprint, 1950 (Translated from Russian).

16. Yu.V. Prohorov, Yu. A. Rozanov. Probability theory, basic concepts. Limit theorems, random processes. Springer, 1969.

17. Richard E. Sherman. Estimating the Variability of Loss Reserves: https://www.casact.org/pubs/forum/98fforum/sherman.pdf

18. https://www.r-project.org/about.html

19. https://cran.r-project.org/web/packages/ChainLadder/ChainLadder.pdf

20. Inflation, consumer prices (annual %) – European Union: https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=EU

21. Risk-free interest rate term structures: https://www.eiopa.europa.eu/tools-and-data/risk-free-interest-rate-term-structures-0_en#ExtraordinaryRFRweeklycalculations

APPENDIX

Listing of the programme

```
library(openxlsx)
library(lubridate)
library(ChainLadder)

#reading data from the file
readData <- function(path, sheetName, rows, cols, rNames = FALSE, cNames = FALSE){
  x <- read.xlsx(path, sheetName, rowNames = rNames, colNames = cNames, rows = rows,
                  cols = cols)
  as.matrix(x)
}

#writing results of calculation in the file
writeData <- function(path, x){
  hs <- createStyle(textDecoration = "BOLD", fontColour = colours()[1], fontSize = 12,
                  fontName  =  "Calibri",  numFmt  =  "TEXT",  border  =  c("top",  "left",
                  "right"),  fgFill  =  colours()[102],  halign  =  "center",  valign  =
                  "center")
  write.xlsx(x, path, col.names = TRUE, row.names = TRUE, gridLines = FALSE, borders =
             "all", headerStyle = hs, colWidths = "auto")
}

#calculating the factors
develFactTable <- function(cumLosses, periods){
  develFacts1 <- matrix(nrow = periods, ncol = periods - 2)
  for (j in 2:(periods - 1)){
    develFacts1[seq(periods - j + 1), j - 1] <- cumLosses[seq(periods - j + 1), j] /
      cumLosses[seq(periods - j + 1), j - 1]
  }
  develFacts2 <- cbind(develFacts1, c(cumLosses[1, periods] / cumLosses[1, periods - 1],
      rep(NA, periods - 1)))[-periods,]
  weightedDevFacts <- c(0, periods - 1)
  for (j in seq(periods - 1)){
    weightedDevFacts[j] <- sum(cumLosses[seq(periods - j), j + 1])/
```

```
                                                sum(cumLosses[seq(periods - j), j])
  }
  return(list("develFacts1" = develFacts1, "develFacts2" = develFacts2,
      "weightedDevFacts" = weightedDevFacts))
}

#calculating the total reserve
countReserve <- function(cumLosses, periods){
  Reserves <- rep(0, periods - 1)
  for (j in 2:periods){
    Reserves[j - 1] <- cumLosses[j, periods] - cumLosses[j, periods - j + 1]
  }
  sum(Reserves)
}

#the Bornhuetter-Ferguson method
BornFerg <- function(cumLosses, periods, weightedDevFacts, earnPrems, lossRatios){
  inverse <- c(1/weightedDevFacts,1)
  quotas <- rev(cumprod(rev(inverse)))
  for (j in 2:periods){
    for (i in (periods-j+2):periods)
      cumLosses[i,j] <- cumLosses[i, periods + 1 - i] +
                          (quotas[j] - quotas[periods + 1 - i]) *
                          earnPrems[i] * lossRatios[i]
  }
  Reserve <- countReserve(cumLosses, periods)
  list("Reserve" = Reserve, "cumLosses" = cumLosses)
}

#one iteration of the bootstrapping method
bootstrapProc <- function(cumLosses, develFacts, randval, ind, periods){
  for (j in seq(periods - 2)){
    f <- develFacts[seq(periods - j),j]
    len <- length(which(is.na(f) | is.infinite(f) == TRUE))
    if (len != 0){
      f <- f[-which(is.na(f) | is.infinite(f) == TRUE)]
    }
    len <- length(f)
```

```
    parts <- seq(1 / len, 1, by = 1 / len)
    for (i in periods:(periods - j + 1)){
      develFacts[i, j] <- f[which(parts > randval[ind])[[1]]]
      ind <- ind + 1
      cumLosses[i, j + 1] <- cumLosses[i, j] * develFacts[i, j]
    }
  }
  cumLosses[, periods] <- cumLosses[, periods - 1]
  Reserve <- countReserve(cumLosses, periods)
  return(list("CumLosses" = cumLosses, "DevelFactors" = develFacts, "Reserve" = Reserve))
}

#forming the distribution
fullDistribution <- function(ResList, numOfSimuls){
  Reserve <- sort(unlist(ResList))
  newMin <- Reserve[1] - (Reserve[2] - Reserve[1]) / 2
  newMax <- Reserve[numOfSimuls] + (Reserve[numOfSimuls] - Reserve[numOfSimuls - 1]) / 2
  Reserve <- c(newMin, Reserve, newMax)
}

#calculating the Mack parameters
MackParam <- function(ultimateTail, procVarMult, estimErrorFacts, periods){
  procSD <- sqrt(procVarMult * ultimateTail)
  procRisk <- sqrt(sum(procSD ^ 2))
  covar <- matrix(nrow = periods - 1, ncol = periods - 1)
  for (j in seq(periods - 1)){
    covar[j:(periods - 1), j] <- ultimateTail[j:(periods - 1)] *
       ultimateTail[j]*estimErrorFacts[j:(periods - 1), j]
    covar[j,] <- covar[, j]
  }
  paramRisk <- sqrt(sum(covar))
  MackSD <- sqrt(procRisk ^ 2 + paramRisk ^ 2)
}

#the Mack method
MackMethod <- function(cumLosses, develFacts2, weightedDevFacts, periods){
  LDF <- cumprod(c(1,rev(weightedDevFacts)))
  errorTerms <- matrix(nrow = periods - 1, ncol = periods - 1)
```

```
var <- rep(0, periods - 1)
for (j in seq(periods - 1)){
  errorTerms[1:(periods - j), j] <- cumLosses[1:(periods - j), j]*
     (develFacts2[1:(periods - j), j] - weightedDevFacts[j]) ^ 2
}
for (j in seq(periods - 1)){
  er <- errorTerms[1:(periods - j), j]
  len <- length(which(is.na(er) | is.infinite(er) == TRUE))
  if (len != 0){
    er <- er[-which(is.na(er) | is.infinite(er) == TRUE)]
  }
  len <- length(er)
  if (len != 1){
    var[j] <- sum(er) / (len-1)
  }
  else{
    if (max(var[j - 1], var[j - 2]) == 0){
      var[j] <- 0
    }
    else{
      var[j] <- var[j - 1] * min(var[j - 1], var[j - 2]) / max(var[j - 1], var[j - 2])
    }
  }
}
procVarMultAnnual <- (rev(LDF)[-periods]) * var / weightedDevFacts ^ 2
procVarMult <- cumsum(rev(procVarMultAnnual))
otherDiag <- rep(0, periods)
for (i in seq(periods)){
  otherDiag[i] <- cumLosses[i, periods - i + 1]
}
ultimate <- LDF * otherDiag
ultimateTail <- ultimate[-1]
paramVarMultAnnual <- rep(0, periods - 1)
for (j in seq(periods - 1)){
  tmp <- cumLosses[, j]
  paramVarMultAnnual[j] <- (var[j] / weightedDevFacts[j] ^ 2) /
     sum(tmp[1:(periods - j)])
}
```

```
  paramVarMult <- cumsum(rev(paramVarMultAnnual))
  estimErrorFacts <- matrix(nrow = periods - 1, ncol = periods - 1)
  for (j in seq(periods - 1)){
    estimErrorFacts[j:(periods - 1), j] <- rep(paramVarMult[j], periods - j)
    estimErrorFacts[j,] <- estimErrorFacts[, j]
  }
  MackSD <- MackParam(ultimateTail, procVarMult, estimErrorFacts, periods)
  for (j in 2:periods){
    for (i in periods:(periods - j + 2)){
      cumLosses[i,j] <- cumLosses[i, j - 1]*weightedDevFacts[j - 1]
    }
  }
  Reserve <- countReserve(cumLosses, periods)
  return(list("Reserve" = Reserve, "CumLosses" = cumLosses, "MackSD" = MackSD,
       "procMult" = procVarMult,   "ErrorFacts" = estimErrorFacts,  "Latest" = otherDiag))
}

#calculating the incremental triangle
lowerIncrLossesTable <- function(cumLosses, periods){
  incrLosses <- matrix(nrow = periods - 1, ncol = periods - 1)
  for (j in 2:periods){
    incrLosses[(periods - 1):(periods - j + 1), j - 1] <-
       cumLosses[periods:(periods - j + 2), j] –
       cumLosses[periods:(periods - j + 2), j - 1]
  }
  return(incrLosses)
}

#the discounting and inflation impact
reservesInflDisc <- function(incrLosses, valDate, inflation, discounts, typeOfPeriods,
                             periods, daysInYear = 365){
  Reserve <- 0
  ReservesAnnual <- rep(0, periods - 1)
  mon <- month(valDate)
  quart <- ceiling(mon / 3)
  halfyear <- ceiling(mon / 6)
  y <- year(valDate)
  orderOfYear <- 0
```

```
for (d in seq(periods - 1)){
  if (typeOfPeriods == "quart"){
    if (quart == 4){
      quart <- 1
      y <- y + 1
    }
    else{
      quart <- quart + 1
    }
    futDate <- as.Date(paste(y, "-", quart * 3, "-", 15, sep = ""))
    orderOfYear <- ceiling(d / 4)
  }
  else if (typeOfPeriods == "month"){
    if (mon == 12){
      mon <- 1
      y <- y + 1
    }
    else{
      mon <- mon + 1
    }
    futDate <- as.Date(paste(y, "-", mon, "-", 15, sep = ""))
    orderOfYear <- ceiling(d / 12)
  }
  else if (typeOfPeriods == "halfyear"){
    if (halfyear == 2){
      halfyear <- 1
      y <- y + 1
    }
    else{
      halfyear <- halfyear + 1
    }
    futDate <- as.Date(paste(y, "-", halfyear * 6, "-", 15, sep = ""))
    orderOfYear <- ceiling(d / 2)
  }
  else{
    y <- y + 1
    futDate <- as.Date(paste(y, "-", 12, "-", 15, sep = ""))
    orderOfYear <- d
```

```
    }
    numOfDays <- ceiling(as.numeric(futDate - valDate))
    for (i in (periods - 1):d){
      incrLosses[i, periods - 1 + d - i] <- incrLosses[i, periods - 1 + d - i] *
        ((1 + inflation) / (1 + discounts[orderOfYear])) ^ (numOfDays / daysInYear)
      Reserve <- Reserve + incrLosses[i, periods - 1 + d - i]
    }
  }
  for (i in seq(periods - 1)){
    ReservesAnnual[i] <- sum(incrLosses[i, (periods - i):(periods - 1)])
  }

  return (list("IncrLosses" = incrLosses, "Reserve" = Reserve,
               "ResAnnual" = ReservesAnnual))
}

#the bootstrapping method
bootstrapMethod <- function(cumLosses, develFacts1, randVal, numOfUnkns,
                            valDate, inflRate, discRates, typeOfPeriods,
                            periods, ReserveCL, ReserveCLID, ReserveBF, ReserveBFID,
                            name, confLevel){
  results <- list()
  resultsID <- list()
  Res <- rep(0, numOfSimuls)
  ResID <- rep(0, numOfSimuls)
  for (i in seq(numOfSimuls)){
    results[[i]] <- bootstrapProc(cumLosses, develFacts1, randVal,
       (i - 1) * numOfUnkns + 1, periods)
    Res[i] <- results[[i]]$Reserve
    resultsID[[i]] <-
      reservesInflDisc(lowerIncrLossesTable(results[[i]]$CumLosses, periods), valDate,
                       inflRate, discRates, typeOfPeriods, periods)
    ResID[i] <- resultsID[[i]]$Reserve
  }
  Reserve <- fullDistribution(Res, numOfSimuls)
  ReserveID <- fullDistribution(ResID, numOfSimuls)
  quantCL <- (max(which(Reserve <= ReserveCL)) + min(which(Reserve >= ReserveCL)))/
    (2 * numOfSimuls)
```

```
quantCLID <- (max(which(ReserveID <= ReserveCLID)) + min(which(ReserveID >=
     ReserveCLID)))/(2 * numOfSimuls)
quantBF <- (max(which(Reserve <= ReserveBF)) + min(which(Reserve >= ReserveBF)))/
  (2 * numOfSimuls)
quantBFID <- (max(which(ReserveID <= ReserveBFID)) + min(which(ReserveID >=
     ReserveBFID)))/(2 * numOfSimuls)
ReserveDes <- (Reserve[confLevel * numOfSimuls] +
     Reserve[confLevel * numOfSimuls + 1]) / 2
ReserveDesID <- (ReserveID[confLevel * numOfSimuls] +
     ReserveID[confLevel * numOfSimuls + 1]) / 2
Prob <- seq(numOfSimuls + 2) / (numOfSimuls + 2)
for (b in c(FALSE, TRUE)){
  if (b){
    plot(Reserve, Prob, xlab = "Reserve", ylab = "Probability", main = "Bootstrapping
           Technique", sub = name, pch = 46, xlim = c(0.9 * ReserveCL, 1.1*ReserveCL))
  }
  else{
    plot(Reserve, Prob, xlab = "Reserve", ylab = "Probability",
         main = "Bootstrapping Technique", sub = name, pch = 46)
  }
  points(ReserveID, Prob, col = "red", pch = 46)
  points(ReserveCL, quantCL, col = "darkorange", pch = 20)
  points(ReserveCLID, quantCLID, col = "brown", pch = 20)
  points(ReserveBF, quantBF, col = "deeppink", pch = 20)
  points(ReserveBFID, quantBFID, col = "blue", pch = 20)
  pos <- "topleft"
  if (name == "RAA" | name == "MedMal"){
    pos <- "bottomright"
  }
  legend(pos, legend = c("base", "disc&infl", "CL", "CL(disc&infl)",
     "BF", "BF(disc&infl)"), col = c("black", "red", "darkorange", "brown", "deeppink",
     "blue"), pch = rep(20, 6), cex = 0.6)
}
data.frame(BM = c("ReserveCL" = ReserveCL, "ReserveCLID" = ReserveCLID,
                  "ReserveBF" = ReserveBF, "ReserveBFID" = ReserveBFID,
                  "quantCL" = quantCL, "quantCLID" = quantCLID,
                  "quantBF" = quantBF, "quantBFID" = quantBFID,
                  "ReserveDes" = ReserveDes, "ReserveDesID" = ReserveDesID,
```

```
                    "addCL" = ReserveDes - ReserveCL,
                    "addCLID" = ReserveDesID - ReserveCLID,
                    "addBF" = ReserveDes - ReserveBF,
                    "addBFID" = ReserveDesID - ReserveBFID))
}

#the main function
calculation <- function(cumLosses, name, inflRate, discRates, valDate, typeOfPeriods,
                        numOfSimuls, seedRand, confLevel){
  periods <- ncol(cumLosses)
  dF <- develFactTable(cumLosses, periods)
  develFacts1 <- dF$develFacts1
  develFacts2 <- dF$develFacts2
  weightedDevFacts <- dF$weightedDevFacts
  MCL <- MackMethod(cumLosses, develFacts2, weightedDevFacts, periods)
  ReserveCL <- MCL$Reserve
  SE <- MCL$MackSD
  m <- log(ReserveCL ^ 2 / sqrt(SE ^ 2 + ReserveCL ^ 2))
  s <- sqrt(log(SE ^ 2 / ReserveCL ^ 2 + 1))
  FinalTr <- MCL$CumLosses
  MCLID <- reservesInflDisc(lowerIncrLossesTable(FinalTr, periods), valDate,
                            inflRate, discRates, typeOfPeriods, periods)
  ReserveCLID <- MCLID$Reserve
  ultimateTail <- MCLID$ResAnnual + MCL$Latest[-1]
  procVarMult <- MCL$procMult
  estimErrorFacts <-  MCL$ErrorFacts
  SEID <- MackParam(ultimateTail, procVarMult, estimErrorFacts, periods)
  mID <- log(ReserveCLID ^ 2 / sqrt(SEID ^ 2 + ReserveCLID ^ 2))
  sID <- sqrt(log(SEID ^ 2 / ReserveCLID ^ 2 + 1))
  lossRatios <- sample(seq(0.70, 1.1, by = 0.001), periods, replace = TRUE)
  earnPrems <- unname(FinalTr[, periods]) / lossRatios
  print(earnPrems)
  lossRatios <- lossRatios + sample(c(-1, 1), 1, replace = TRUE) * 0.01
  print(lossRatios)
  BF <- BornFerg(cumLosses, periods, weightedDevFacts, earnPrems, lossRatios)
  ReserveBF <- BF$Reserve
  BFID <- reservesInflDisc(lowerIncrLossesTable(BF$cumLosses, periods), valDate,
                           inflRate, discRates, typeOfPeriods, periods)
```

```
ReserveBFID <- BFID$Reserve
quantCL <- plnorm(ReserveCL, meanlog = m, sdlog = s)
quantCLID <- plnorm(ReserveCLID, meanlog = mID, sdlog = sID)
quantBF <- plnorm(ReserveBF, meanlog = m, sdlog = s)
quantBFID <- plnorm(ReserveBFID, meanlog = mID, sdlog = sID)
ReserveDes <- qlnorm(confLevel, meanlog = m, sdlog = s)
ReserveDesID <- qlnorm(confLevel, meanlog = mID, sdlog = sID)
x <- seq(0, ReserveCL * 2, length.out = 10000)
for (b in c(FALSE, TRUE)){
  if (b){
    plot(x, plnorm(x, meanlog = m, sdlog = s), xlab = "Reserve", ylab = "Probability",
         main = "Mack Method", sub = name, pch = 46,
         xlim = c(0.9 * ReserveCL, 1.1*ReserveCL))
  }
  else{
    plot(x, plnorm(x, meanlog = m, sdlog = s), xlab = "Reserve", ylab = "Probability",
         main = "Mack Method", sub = name, pch = 46)
  }
  points(x, plnorm(x, meanlog = mID, sdlog = sID), col = "red", pch = 46)
  points(ReserveCL, quantCL, col = "darkorange", pch = 20)
  points(ReserveCLID, quantCLID, col = "brown", pch = 20)
  points(ReserveBF, quantBF, col = "deeppink", pch = 20)
  points(ReserveBFID, quantBFID, col = "blue", pch = 20)
  legend("topleft", legend = c("base", "disc&infl", "CL", "CL(disc&infl)",
                               "BF", "BF(disc&infl)"),
         col = c("black", "red", "darkorange", "brown", "deeppink", "blue"),
         pch = rep(20, 6), cex = 0.6)
}
set.seed(seedRand)
numOfUnkns <- (periods ^ 2 - periods) / 2 - ( periods - 1)
randVal <- runif(numOfUnkns * numOfSimuls, 0, 1)
BM <- bootstrapMethod(cumLosses, develFacts1, randVal, numOfUnkns,
                      valDate, inflRate, discRates, typeOfPeriods,
                      periods, ReserveCL, ReserveCLID, ReserveBF, ReserveBFID,
                      name, confLevel)
cbind(BM, data.frame("Mack" = c("ReserveCL" = ReserveCL, "ReserveCLID" = ReserveCLID,
                                "ReserveBF" = ReserveBF, "ReserveBFID" = ReserveBFID,
                                "quantCL" = quantCL, "quantCLID" = quantCLID,
```

```
                                  "quantBF" = quantBF, "quantBFID" = quantBFID,
                                  "ReserveDes" = ReserveDes,
                                  "ReserveDesID" = ReserveDesID,
                                  "addCL" = ReserveDes - ReserveCL,
                                  "addCLID" = ReserveDesID - ReserveCLID,
                                  "addBF" = ReserveDes - ReserveBF,
                                  "addBFID" = ReserveDesID - ReserveBFID)))
}

pathR <- "C:\\Users\\mlyzhechko\\Documents\\D\\data.xlsx"
sheetNameR <- "disc_infl"
pathW <- "C:\\Users\\mlyzhechko\\Documents\\D\\results.xlsx"
dataName <- c("GenIns", "MCLpaid", "MedMal", "RAA")
inflRate <- readData(pathR, sheetNameR, 2, 2)
discRates <- readData(pathR, sheetNameR, 4:11, 2)
valDate <- as.Date("2019-12-31")
type <- c("quart", "annual", "halfyear", "halfyear")
numOfSimuls <- 10000
seedRand <- 0.5
confLevel <- 0.8
data("GenIns")
data("MCLpaid")
data("MedMal")
data("RAA")
Triang <- list(GenIns, MCLpaid, MedMal$MedMalPaid, RAA)
resCalc <- list()
for (i in seq(length(Triang))){
  print(dataName[i])
  resCalc[[i]] <- calculation(Triang[[i]], dataName[i], inflRate, discRates,
                              valDate, type[i], numOfSimuls, seedRand,
                              confLevel)
  print(resCalc[[i]])
}
finalRes <- data.frame("GenIns" = resCalc[[1]],
                       "MCLpaid" = resCalc[[2]],
                       "MedMal" = resCalc[[3]],
                       "RAA" = resCalc[[4]])
writeData(pathW, finalRes)
```