Master in Actuarial and Financial Sciences

MASTER THESIS

Title: An introduction to the Generalized Multivariate Chain-Ladder Model: The use of the "ChainLadder" package in R.

Author: Goran Ehsan Pernía

Advisors: Eva Boj del Val and Teresa Costa Cor

Academic year: 2022-2023

UNIVERSITAT DE BARCELONA
Facultat d'Economia i Empresa

Màster de Ciències Actuarials i Financeres

Faculty of Economics and Business
Universitat de Barcelona

Master thesis
Master in Actuarial and Financial Sciences

An introduction to the Generalized Multivariate Chain-Ladder Model: The use of the "ChainLadder" package in R.

Author: Goran Ehsan Pernía

Advisors: Eva Boj del Val and Teresa Costa Cor

Abstract

The aim of this work is to deepen into a stochastic variant of the classic Chain-Ladder model to calculate claims reserves of an insurance company. Specifically, we will focus on the stochastic Generalized Multivariate Chain-Ladder model (GMCL).

First, the classic deterministic method will be explained. Then, we will introduce the stochastic method and develop a practical example by using both methods to see the differences in the estimations.

The function *MultiChainLadder* of the *ChainLadder* package for R (R Development Core Team, 2023) will be used.

Key words: *Chain-Ladder, Reserving, Multivariate, Non-Life, R*

Index

1. Introduction

One of the most important objectives of an insurance company is to estimate properly their future reserves, even more since the establishment of Solvency II, that obligates companies to make this estimation more accurate. Since we are focused on the non-life insurance, we will talk about claim reserves. We can define claim reserves as the amount of money that the insurance company set aside in order to ensure that they will be able to pay policyholders that have filed or are expected to file legitimate claims on their policies.

The idea is that insurance companies must keep technical reserves for IBNR (Incurred But Not Reported) claims. IBNR claims refers to the fact that during a certain year, many accidents occur, but some of them are not reported until the next periods. Therefore, the company must consider these possible future obligations, and calculate them correctly, since they will represent a major part of the liabilities of the current and the coming years in its accounting balance.

In this master thesis we will see how the classic Chain-Ladder method and the Generalized Multivariate Chain-Ladder method can be developed. The objective will be to prove if there are benefits in using the stochastic Chain-Ladder method instead of the classic one. The GMCL method has been chosen because we wanted to explore and understand a different method than the ones that we have seen in the master's program.

Regarding the structure of the work, it has been divided into four distinct parts. First, the operation of the classic Chain-Ladder method is explained. After this, the stochastic GMCL is introduced and described. Once the methodologies of both models are explained, a practical exercise is conducted to determine if there are differences in the obtained results. Finally, after analyzing the results, the claims reserve is calculated for the best method of our case.

2. Reserving methods

We can differentiate between deterministic and stochastic reserving methods. The principal difference is that the stochastic methods use random variables, and the deterministic ones don’t.

Some deterministic methods could be the classic Chain-Ladder method, the de Vylder least squares method, or the arithmetic and geometric separation method of Taylors.

In the case of stochastic methods, we can draw attention to the Mack model, the Bornhuetter-Ferguson model, as well as generalized linear models. These models have been studied during the master's program, and in this work, we present the Generalized Multivariate Chain-Ladder model, which is an extension of the deterministic Chain-Ladder model.

One advantage of the stochastic methods that we will be working with is that in addition to estimating a value for the reserve, these methods can also provide a measure of how much the results can vary around that estimation.

2.1. Run-off triangles

Before starting with the explanation of the reserving methods, it will be introduced the concept of the so-called development triangles, and the role they play in the Chain-Ladder methodology.

To perform the calculations of the IBNR loss reserves, these run-off triangles are commonly used.

The basic structure of these triangles (Weindorfer, 2012) can be visualized in the following table:

	Development year j					
Accident year i	**0**	**1**	**2**	**…**	$\boldsymbol{k-1}$	$\boldsymbol{k}$
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	$\cdots$	$C_{0,k-1}$	$C_{0,k}$
1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	$\cdots$	$C_{1,k-1}$	
…	$\cdots$	$\cdots$	$\cdots$	$\cdots$		
$\boldsymbol{k-2}$	$C_{k-2,0}$	$C_{k-2,1}$	$C_{k-2,2}$			
$\boldsymbol{k-1}$	$C_{k-1,0}$	$C_{k-1,1}$				
$\boldsymbol{k}$	$C_{k,0}$					

Table 1: Run-off triangle structure. [Source: own elaboration]

Where:

$i \in (0, \dots, k)$: Years of occurrence of the loss event.
$j \in (0, \dots, k)$: Development year.
$C_{i,j}$: Cumulative amounts paid for loss events occurred in the year of occurrence i, up to the development year j (including this one).

If the amounts were not cumulative, then the nomenclature $c_{i,j}$ would be used, which represents the amount of losses of claims occurred in the year i and paid only in the development year j, without taking into account the payments of previous development years.

Understanding the concept of accumulated and non-accumulated losses, we can infer that non-accumulated amounts are useful for calculating accumulated amounts, as shown above, where:

$\sum_{j=0}^{4} c_{0,j} = C_{0,4}$.

That is, $C_{0,4}$ will be equivalent to the sum of all those non-accumulated amounts of claims occurred in the origin year 0 and paid until the development year 4.

To see clearly what we are doing in these development triangles, we can look at Table 2:

	Development year *j*					
Accident year *i*	**0**	**1**	**2**	**…**	$\boldsymbol{k-1}$	$\boldsymbol{k}$
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	…	$C_{0,k-1}$	$C_{0,k}$
1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	…	$C_{1,k-1}$	
…	…	…	…	…		
$\boldsymbol{k-2}$	$C_{k-2,0}$	$C_{k-2,1}$	$C_{k-2,2}$			
$\boldsymbol{k-1}$	$C_{k-1,0}$	$C_{k-1,1}$				
$\boldsymbol{k}$	$C_{k,0}$					

Table 2: Run-off triangle and amounts to be estimated. [Source: own elaboration]

Starting from initial data, such as those in the triangle of Table 1, the aim is to estimate the values of the cells marked in gray in Table 2.

These cells would be the amounts that the company should be accumulating for the different development years j, to cope with the payments for claims occurred in origin year i.

The term C' , that we will see below, refers to the values that are estimated from the original data.

That is, the last cell in Table 3, $C'_{k,k}$, would correspond to the total amount that the company expects to pay in the development year k for claims in origin year k.

	Development year *j*					
Accident year *i*	***0***	***1***	***2***	**…**	***k-1***	***k***
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	…	$C_{0,k-1}$	$C_{0,k}$
1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	…	$C_{1,k-1}$	
…	…	…	…	…		
$\boldsymbol{k-2}$	$C_{k-2,0}$	$C_{k-2,1}$	$C_{k-2,2}$			
$\boldsymbol{k-1}$	$C_{k-1,0}$	$C_{k-1,1}$				
$\boldsymbol{k}$	$C_{k,0}$	$C'_{k,1}$	$C'_{k,2}$	…	$C'_{k,k-1}$	$C'_{k,k}$

Table 3: Estimated provisions for accident year k *. [Source: own elaboration]*

If we do the same procedure to obtain the final cells for all the different accident years, we will obtain a final column that we usually call ultimate loss (see Table 4).

Each value of this column would be the total amount that the company expects to pay in the future due to claims occurred in their respective origin year i.

	Development year *j*					
Accident year i	***0***	***1***	***2***	**…**	***k-1***	***k***
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	$\cdots$	$C_{0,k-1}$	$C_{0,k}$
1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	$\cdots$	$C_{1,k-1}$	$C'_{1,k}$
…	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\boldsymbol{k-2}$	$C_{k-2,0}$	$C_{k-2,1}$	$C_{k-2,2}$	$\cdots$	$C'_{k-2,k-1}$	$C'_{k-2,k}$
$\boldsymbol{k-1}$	$C_{k-1,0}$	$C_{k-1,1}$	$C'_{k-1,2}$	$\cdots$	$C'_{k-1,k-1}$	$C'_{k-1,k}$
$\boldsymbol{k}$	$C_{k,0}$	$C'_{k,1}$	$C'_{k,2}$	$\cdots$	$C'_{k,k-1}$	$C'_{k,k}$

Table 4: Ultimate loss. [Source: own elaboration]

Up to this point, we have seen what run-off triangles are, what they are made of, and what they are used for, but we have not yet discussed how these future year amounts are calculated.

It should be mentioned, that depending on the method that we use to calculate these provisions, the calculations will be made in one way or another.

In the following sections we will introduce the considered methods in this work and see how the calculations are performed.

3. Chain-Ladder method

The most common method for claims reserving is the Chain-Ladder method. It is one of the most used in the companies due to the facility of its implementation and its lack of probabilistic assumptions, with the exception that the proportion of claims that are reported to the company from one period to the next one, remains constant.

This method uses an accumulated run-off triangle, and with the so-called development factors, that we will define below, it is possible to calculate the estimation of future losses for different development periods.

The problem is that on the one hand, the estimations depend on these development factors, and on the other hand, these development factors are calculated based on historical data.

Due to this, the effectiveness of the estimations will decay over time, because we will be calculating the development factors with less and less data as we progress over the time.

We can refer to Boj et al. (2020) to see the derivation of the formulas for the Chain-ladder method from an annual triangle.

The estimator or development factor is obtained as:

$$\widehat{m_h} = \frac{\sum_{i=0}^{k-h-1} C_{i,h+1}}{\sum_{i=0}^{k-h-1} C_{i,h}}.$$

If we want to obtain the development factor m_0, which allow us to obtain the reserves estimation from development year 0 to 1, we can use the following formula:

$$\widehat{m_0} = \frac{\sum_{i=0}^{k-1} C_{i,1}}{\sum_{i=0}^{k-1} C_{i,0}}.$$

The way in which we obtain these development factor can be seen more clearly in Table 5, where it will simply be dividing the sum of the cells in the dark grey rows by the sum of the cells in the light grey rows.

	Development year *j*					
Accident year *i*	***0***	***1***	***2***	**…**	***k-1***	***k***
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	…	$C_{0,k-1}$	$C_{0,k}$
1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	…	$C_{1,k-1}$	$C'_{1,k}$
…	…	…	…	…	…	…
$\boldsymbol{k-2}$	$C_{k-2,0}$	$C_{k-2,1}$	$C_{k-2,2}$	…	$C'_{k-2,k-1}$	$C'_{k-2,k}$
$\boldsymbol{k-1}$	$C_{k-1,0}$	$C_{k-1,1}$	$C'_{k-1,2}$	…	$C'_{k-1,k-1}$	$C'_{k-1,k}$
$\boldsymbol{k}$	$C_{k,0}$	$C'_{k,1}$	$C'_{k,2}$	…	$C'_{k,k-1}$	$C'_{k,k}$

Table 5: Example to obtain development factors. [Source: own elaboration]

The same procedure would be followed to obtain the different development factors of the different periods, where we will obviously use the data of the corresponding periods.

Once we have obtained the development factors, we can obtain the estimation of future expenses as:

$$C'_{i,h+1} = \widehat{m}_h \cdot C_{i,h}.$$

For run-off triangles with the same number of origin and development periods (years in our case), one less data will always be taken as we increase the number of the development year.

4. General Multivariate Chain-Ladder model (GMCL)

As we have mentioned before, although the classical Chain-Ladder method is the most common and widely used in companies, some variants of it are becoming more important for calculating reserves due to the implementation of more restrictive regulations.

One of these stochastic variants is the Multivariate Chain-Ladder method (MCL).
The difference with the classic method is that in the multivariate case we work with more than one run-off triangle simultaneously to make the estimates, which allows us to model their correlation.

As we can see in Table 6, in the multivariate case we will work with N loss triangles with the same structure as used in the classical method. We will explain the model based in Zhang (2010).

	Development year k				
Accident year i	***1***	***2***	**…**	$I-1$	I
1	$Y_{1,1}$	$Y_{1,2}$	…	$Y_{1,I-1}$	$Y_{1,1}$
2	$Y_{2,1}$	$Y_{2,2}$	…	$Y_{2,I-1}$	
⋮	…	…	…		
i	$Y_{i,1}$	…			
⋮	…	…			
$I-1$	$Y_{I-1,1}$	$Y_{I-1,2}$			
I	$Y_{I,1}$				

Table 6: Structure of the N run-off triangles [Source: own elaboration]

In this case:

$i = (1, \dots, I)$: accident year.

$k = (1, \dots, I)$: development year.

Denote $Y_{i,k}$ que as an $N \times 1$ vector of the cumulative losses in accident year i and development year k. The superscript (n) refers to the nth triangle. The symbol $'$, at the end of the vector denotes the transpose of the vector.

$$Y_{i,k} = (Y_{i,k}^{(1)}, \dots, Y_{i,k}^{(N)})'.$$

The inclusion of more triangles is very important because there may be structural connections between them, so the development of one triangle may depend on past information from others. Therefore, it is possible to make estimates that take into account the correlation between triangles, which could lead to an improvement in the estimations compared to the univariate case.

The next step is to look for a suitable model in order to make a correct estimation of our parameters.

This will be carried out by making several modifications to a baseline model.

4.1.Baseline Model

We can formulate the GMCL model as:

$$Y_{i,k+1} = \boldsymbol{B_k} \cdot Y_{i,k} + \epsilon_{i,k} \; . \qquad (1)$$

This model is a simply natural generalization of the multivariate model.

We will use the model (1) as our baseline model. In this case, it will be for the development year k, that is, from development year k to $k+1$.

Where,

$$\boldsymbol{B_k} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1N} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \cdots & \beta_{NN} \end{pmatrix}$$

$\boldsymbol{B_k}$ is an $N \times N$ development matrix in development period k, and the nth row contains the development parameters for the nth triangle.

Basically, the matrix B_k is a matrix where each row contains a vector with the different development factors of the different triangles. For example, the first row would correspond to the vector of development factors of the first triangle in the development year k.

This non-diagonal development matrix B_k allows the development of one triangle in period k to depend directly on the loss information of the other triangles in the same period k.

One problem is that the fully parameterized model has N^2 development parameters in each period, which can make parameters estimation difficult or even impossible. Additionally, in the latest development periods, the parameters may be difficult to estimate even under constraints.

One possible solution to this problem is to use data in trapezoid form to ensure that there are enough observations in the tail of the distribution.
This can improve the quality of parameter estimations and allow the model to be more accurate in predicting future losses.

And to conclude, the last term represents the residuals of the model. Residuals are the differences between actual observations and model predictions.

Some assumptions are made for the model (1):

$$E(\epsilon_{i,k} | D_{i,k}) = 0 \qquad (1.1)$$

$$cov(\epsilon_{i,k} | D_{i,k}) = D(Y_{i,k})^{1/2} \cdot \textstyle\sum_k \cdot D(Y_{i,k})^{1/2} \qquad (1.2)$$

losses from different accident years are independent (1.3)

$\epsilon_{i,k}$ are symmetrically distributed. (1.4)

Where:

$$\sum_k = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \dots & \sigma_{NN} \end{pmatrix}$$

is a symmetric positive definite $N \times N$ matrix.

$\sum_k$ is a variance-covariance matrix that allows us to model the relationship between errors in different development periods and loss triangles.

4.2. Introduction to Seemingly Unrelated Regressions (SUR)

Seemingly Unrelated Regressions (SUR), as stated in Mazuelos (2013) are a type of multivariate regression model that allows multiple regression equations to be jointly modeled, where each equation can have different independent variables and a common dependent variable.

In the conventional multivariate linear regression models, it is assumed that all independent variables are related to each other, but in SUR models the relationships between the independent variables are allowed to vary from one equation to another.

Then, SUR models allow us to model situations in which it is expected that the relationships between the independent variables will be different across different equations.

The objective of SUR is to estimate the regression coefficients of each equation independently but also consider the shared information among the equations to improve the precision of the estimates.

To achieve this, the variance-covariance matrix $\sum_k$ (that we have already introduced) is used to capture the shared information among the equations.

In our case, since none of the dependent variables $Y_{i,k+1}^{(n)}$ appears as an explanatory variable in the other equations, and since all these equations individually are statistically related with the non-diagonal matrix, the use of SUR is suggested for parameter estimation.

It can be verified that we can easily switch from our base model to model 2

$$\begin{pmatrix} Y_{\leq,k+1}^{(1)} \\ Y_{\leq,k+1}^{(2)} \\ \vdots \\ Y_{\leq,k+1}^{(N)} \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & X_N \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} \qquad (2)$$

by applying SUR.

Here, $n = 1,2, \dots, N$.

The first term in model (2), that represents the dependent variables of the model, corresponds to a $(I-k) \times 1$, vector of the first $(I-k)$ losses in the development year $k+1$ of the nth triangle.

So, if we want to visualize the first term of that vector, we have:
$Y_{\leq,k+1}^{(1)} = (Y_{1,k+1}^{(1)}, \dots, Y_{I-k,k+1}^{(1)})'$.

That would be a vector with the first $(I-k)$ losses in the development year $k+1$, of the first triangle.

For the example that we will see later, where two triangles are developed simultaneously, the first term would be composed of $Y_{\leq,k+1}^{(1)}$ and $Y_{\leq,k+1}^{(2)}$.

The second term of the new model, composed of the elements $X_1, X_2, \dots, X_N$ on the diagonal, is an $(I-k) \times N$ matrix of the first $(I-k)$ observations in development year k of each of the triangles.

The condition $X_1 = X_2 = \cdots = X_N$ holds, where, on one hand, $X_n = (Y_{<,k}^{(1)}, \dots, Y_{<,k}^{(N)})$ and on the other hand, $Y_{<,k}^{(n)} = (Y_{1,k}^{(n)}, \dots, Y_{I-k,k}^{(n)})'$, that, as we have said, is a vector of the first $(I-k)$ losses of the nth triangle in the development year k.

Finally, we have two vectors composed of the development factors and error terms.
The first case is a $N \times 1$ vector composed of $\beta_1, \ \beta_2, \dots, \ \beta_N$, where
$\beta_n = (\beta_{n1}, \dots, \beta_{nN})'$ is a $N \times 1$ vector of development factors in the nth equation.

For example, $\beta_1 = (\beta_{11}, \dots, \beta_{1N})'$ would be the vector of development factors for the first equation.

And as we have mentioned, a final vector appears composed of the error terms $\epsilon_1, \ \epsilon_2, \dots$, ϵ_N, where $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{n,I-k})$ is an $(I-k) \times 1$ vector of the error terms of the nth equation.

4.3.First transformation to the SUR model

Once we have understood how the model is structured with the application of SUR, we need to modify it again in order to reach the appropriate model for our parameter's estimation. This modification will consist in vectorizing the matrices.

The concept of vectorizing refers to the process of transforming a matrix A of size $m \times n$, where $A = (a_1, \dots, a_n)$, and where each element of this matrix is A is a $m \times 1$ vector, into a vector of size $m \cdot n \times 1$.

Equation (3) shows the structure of the new model that we obtain by vectorizing.

$Y = X \cdot \beta + \epsilon$. (3)

Once again, we need to define the terms of the new model (3).

The first term, $Y = vec\left(Y^{(1)}_{\leq,k+1}, \dots, Y^{(N)}_{\leq,k+1}\right)$, is an $N \cdot (I-k) \times 1$ vector of response variables. We are vectorizing a matrix with the same structure as the X_n terms in the SUR matrix, so when we vectorize a $(I-k) \times N$ matrix, we obtain the structure of the Y term.

The second term, $X = I \otimes X_n$, is a $N \cdot (I-k) \times N^2$ diagonal matrix, where $\otimes$ represents the Kronecker product operator of matrices. The diagonal is composed by the $X_1 = X_2 = \dots = X_N$ terms.

The third term, $\beta = vec\,(\beta_1, \beta_2, \dots, \beta_N)$, is a $N \cdot N \times 1$ vector of development parameters. Since we have defined in the SUR model that each element β_n is a vector with $N \times 1$ structure, and we have up to β_N elements, then we will have a $N \times N$ matrix, and by vectorizing an $N \times N$ matrix, we obtain a $N \cdot N \times 1$ vector.

Finally, we have a vector of error terms $\epsilon = vec\,(\epsilon_1, \epsilon_2, \dots, \epsilon_N)$, where each element ϵ_n is a vector with dimensions $(I-k) \times 1$, obtaining a matrix $(I-k) \times N$. Vectorizing that matriz we obtain a $N \cdot (I-k) \times 1$ vector.

4.4. Final model

To arrive at the model that we will use to obtain our parameters, we need to make one last change to the model shown in (3).

We introduce a new term $W = vec\,(Y^{(1)}_{<,k}, Y^{(2)}_{<,k}, \dots, Y^{(N)}_{<,k})$, where each $Y^{(n)}_{<,k}$ is a $(I-k) \times 1$ vector. As there are up to $Y^{(N)}_{<,k}$, we will be vectorizing a $(I-k) \times N$ matrix, which will result in a vector of dimensions $N \cdot (I-k) \times 1$.

W is a vector of the first $(I-k)$ losses in the development year k.

From (1.3) and (1.4), we arrive at $cov(\epsilon) = D(W)^{\frac{1}{2}} \cdot (\sum_k \otimes \boldsymbol{I}) \cdot D(W)^{\frac{1}{2}}$, where $\boldsymbol{I}$ is the identity matrix of order $(I-k) \times (I-k)$, and D is the diagonal operator, defined as an $N \times N$ matrix with the N-dimensional vectors $\boldsymbol{a} = (a_1, \dots, a_n)'$ along the diagonal.

If we multiply by $D(W)^{-\frac{1}{2}}$ the model (3), we get:

$$D(W)^{-\frac{1}{2}} \cdot Y = D(W)^{-\frac{1}{2}} \cdot X \cdot \beta + D(W)^{-\frac{1}{2}} \cdot \epsilon,$$

that can be expressed as:

$$Y^* = X^* \cdot \beta + \epsilon^*. \qquad (4)$$

Where:

$$Y^* = D(W)^{-1/2} \cdot Y$$

$$X^* = D(W)^{-1/2} \cdot X$$

$\epsilon^* = D(W)^{-1/2} \cdot \epsilon$.

Using the structure $cov(\epsilon) = D(W)^{\frac{1}{2}} \cdot (\sum_k \otimes \boldsymbol{I}) \cdot D(W)^{\frac{1}{2}}$, previously obtained, we can arrive to $cov(\epsilon^*) = D(W)^{-\frac{1}{2}} \cdot cov(\epsilon) \cdot D(W)^{-\frac{1}{2}} = \sum_k \otimes \boldsymbol{I}$.

Now, this variance-covariance structure is consistent with the typical SUR assumption introduced by Zellner (1962).

For this reason, we will use model (4) as our final model to obtain the parameters. Applying Aitken's Generalized Least Squares (GLS) to the final model, we can obtain the best linear unbiased estimator (BLUE) of β as:

$\hat{\beta}_G = (X^{*'} \cdot (\sum_k \otimes \boldsymbol{I})^{-1} \cdot X^*)^{-1} \cdot X^{*'} \cdot (\sum_k \otimes \boldsymbol{I})^{-1} \cdot Y^*$

And the variance-covariance matrix of this estimator $\hat{\beta}_G$ looks like:

$$V(\hat{\beta}_G) = (\boldsymbol{X}^{*'} \cdot (\sum_k \otimes \boldsymbol{I})^{-1} \cdot \boldsymbol{X}^*)^{-1}.$$

Finally, it should be mentioned that there is a problem when we are obtaining the BLUE estimator of β and its variance-covariance matrix, and that is that we are assuming that the matrix $\sum_k$ is known, which is generally not the case.

In cases where the matrix $\sum_k$ is unknown, we need to use an estimator of it to calculate $\hat{\beta}_G$ and $V(\hat{\beta}_G)$.

Thus, if we use the estimator $\widehat{\sum}_k$ in the previous equations, we obtain a Feasible Generalized Least Squares (FGLS) estimator of β such that:

$\hat{\beta}_{FG} = (X^{*'}(\widehat{\sum}_k \otimes \boldsymbol{I})^{-1} \cdot X^*)^{-1} \cdot X^{*'}(\widehat{\sum}_k \otimes \boldsymbol{I})^{-1} \cdot Y^*$.

Although there exist many estimators for the matrix $\sum_k$, one of the most commonly used, which is consistent and unbiased, is the shown below.

$\widehat{\sum_o} = \frac{1}{I-k}(\widehat{\epsilon_1^*}, \widehat{\epsilon_2^*}, \dots, \widehat{\epsilon_N^*})'(\widehat{\epsilon_1^*}, \widehat{\epsilon_2^*}, \dots, \widehat{\epsilon_N^*})$.

Considering that the objective of this model is to obtain optimal estimators, there will be a loss of efficiency due to the estimation of the matrix $\sum_k$.

5. Practical example

In this section, a practical example of a GMCL model will be developed with the help of the R program, using the *ChainLadder* package. To be familiarized with all the practical applications of the package, we have look on Zhang (2023).

We will use the database that we have named *AUTO*, obtained by applying a multiplicative factor to another database. *AUTO* contains a list with two accumulated run-off triangles called *AUTOPAID* and *AUTOINCURRED*.

AUTOPAID and *AUTOINCURRED* correspond to two development triangles composed of paid and incurred losses of an insurance company from the origin year 2013 to the year 2022.

In addition to the GMCL model, we will also perform calculations with the classical Chain-Ladder method, the SCL and the MCL, with or without intercept terms, in order to see how the results vary depending on the methodology applied.

First, we will introduce each of these models in a generic way, and in the final part we will make a comparison of the obtained results.

We can define the SCL as a multivariate method that ignores correlations between triangles, which ends up being the same as developing the univariate model, where the triangles will be developed independently.

On the other hand, the MCL model is a direct multivariate generalization of the classical method, where correlations between triangles are considered, but the average development in each triangle is only estimated based on its historical values, without using direct information from other triangles. To achieve this, the MCL restricts the matrix B_k to be diagonal.

And finally, the GMCL extends the MCL model by allowing development dependencies between different triangles and the inclusion of regression intercepts.

According to Zhang (2010), the structure of the GMCL model in the development year k can be seen as:

$$Y_{i,k+1} = A_k + B_k \cdot Y_{i,k} + \epsilon_{i,k} \ . \qquad (5)$$

This structure may be familiar to us since it is the same as the one used in our baseline model, but with the inclusion of a new term A_k, which we can define as a vector of intercept terms.

Barnett and Zehnwirth (2000) pointed out that both the classical Chain-Ladder method and the multivariate one are not suitable, as they tend to overestimate large values and underestimate small ones. This, results in a decreasing trend in the residuals plot.
The reason why both models fail in their residual plot is that they do not include an intercept term. For this reason, we will now incorporate this term A_k.

5.1.ChainLadder package

Although we will not explain in detail the *ChainLadder* package (Gessman et al, 2023), we must mention that it includes two functions for predicting insurance loss payments based on several cumulative losses development triangles: *MultiChainLadder* and *MultiChainLadder2*.

Both options permit to obtain the estimations of the reserves from different models such as the Separate Chain-Ladder model, the Multivariate Chain-Ladder model or the Generalized Multivariate Chain-ladder.

In our case, we will develop the practical example using the second option, as it allows us to make the robustness of the tail easily.

We should understand that if we employ a multivariate method, it will be in a context where we have evidence that the triangles used are correlated.

For the latest periods, where we have limited data available, running a multivariate model often produces extremely volatile estimates or may even fail.

That is why we will divide the data into two distinct periods. In the first period, we will apply the multivariate method, and in the second period, which represents the tail, we will apply the SCL method, where we do not take correlations into account.

This process is called tail robustification and allows us to make more accurate estimates.

As mentioned, with *MultiChainLadder2* (unlike *MultiChainLadder*), we can directly partition the data, so it will be the function that we will use in our example.

The number of periods included in the tail is selected by us and is relevant in the estimation process, as it significantly affects the results.

In our case, we conducted the study of different models by robustifying the tail using three and four periods. Finally, we selected three periods as we obtained better paid-incurred ratios.

5.2.Classic Chain-Ladder model

After understanding the first part of the work, which explains in detail the procedure for calculating reserves using the Chain-Ladder method, we complete both run-off triangles after obtaining the development factors, which can be visualized in Table 7.

	Development Factors m_h								
h	0	1	2	3	4	5	6	7	8
AUTOPAID	1,98999	1,28515	1,13664	1,06395	1,03114	1,01668	1,00635	1,00409	1,00065
AUTOINCURRED	1,00380	1,00030	0,99309	0,99505	1,00071	1,00122	1,00119	0,99989	1,00106

Table 7: Development factors univariate Chain-Ladder. [Source: own elaboration]

The purpose will be to verify that we obtain the same results with the classic and the SCL method.

As we are doing the calculations with a univariate method, we work with one triangle at a time, so we need to calculate the complete run-off triangles *AUTOPAID* and *AUTOINCURRED* separately.

For the classic model we obtain the following results:

	Development year j									
Accident Year i	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	126406	262401	333272	381384	409812	425836	435538	438991	441691	441980
1	128176	254016	325846	378978	411165	426185	434166	437266	438154	438440
2	143665	284630	372650	431928	459700	472499	479514	481530	483501	483818
3	143065	284701	376340	425836	449974	461560	466656	469620	471543	471851
4	144496	304514	394019	443112	465470	478422	486403	489493	491497	491818
5	159700	324270	408719	457225	483406	498458	506773	509992	512080	512415
6	169520	327868	408858	459196	488564	503776	512180	515433	517543	517881
7	158971	305311	397465	451774	480667	495634	503901	507102	509178	509511
8	160789	308504	396475	450649	479469	494399	502646	505839	507910	508242
9	157860	314140	403718	458881	488229	503431	511828	515079	517188	517526

Table 8: AUTOPAID Full Triangle. [Source: own elaboration]

	Development year j									
Accident Year i	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	406779	420532	432576	434658	438744	441998	443496	443781	443732	444204
1	404534	424084	430634	436619	438798	439479	440062	440289	440241	440709
2	448012	482912	482105	480874	484598	484942	485675	486795	486742	487259
3	506649	495801	489791	481024	476142	475204	474632	475198	475146	475651
4	542581	536639	527726	511652	492692	491002	491601	492186	492133	492655
5	521472	527884	516858	508389	508129	508489	509108	509715	509659	510201
6	498661	498484	497525	500675	498199	498551	499159	499754	499699	500230
7	473442	451371	461660	458472	456204	456527	457083	457628	457578	458064
8	438851	419384	419511	416614	414553	414847	415352	415847	415802	416244
9	411545	413108	413234	410380	408350	408639	409137	409625	409580	410015

Table 9: AUTOINCURRED Full Triangle. [Source: own elaboration]

5.3.Separate Chain-Ladder model (SCL)

As we have mentioned, in this second case, we are working with a multivariate method where we develop two triangles simultaneously. In a SCL, OLS[1] is used to estimate the parameters. Here, the estimation of development factors for each triangle is independent of the others.

Correlation Coefficients SCL								
0	0	0	0	0	0	0	0	0

Table 10: Correlation Coefficients SCL. [Source: own elaboration]

In summary, although we are using a multivariate method, by applying OLS, we do not assume correlation between the triangles, so it is like developing a univariate model for each triangle.

This can be easily verified as we obtain correlation coefficients equal to zero in all periods and there are no differences observed in the results obtained between the classical case and the SCL:

	Development year j									
Accident Year *i*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	126406	262401	333272	381384	409812	425836	435538	438991	441691	441980
1	128176	254016	325846	378978	411165	426185	434166	437266	438154	438440
2	143665	284630	372650	431928	459700	472499	479514	481530	483501	483818
3	143065	284701	376340	425836	449974	461560	466656	469620	471543	471851
4	144496	304514	394019	443112	465470	478422	486403	489493	491497	491818
5	159700	324270	408719	457225	483406	498458	506773	509992	512080	512415
6	169520	327868	408858	459196	488564	503776	512180	515433	517543	517881
7	158971	305311	397465	451774	480667	495634	503901	507102	509178	509511
8	160789	308504	396475	450649	479469	494399	502646	505839	507910	508242
9	157860	314140	403718	458881	488229	503431	511828	515079	517188	517526

Table 11: AUTOPAID Full Triangle. [Source: own elaboration]

	Development year j									
Accident Year *i*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	406779	420532	432576	434658	438744	441998	443496	443781	443732	444204
1	404534	424084	430634	436619	438798	439479	440062	440289	440241	440709
2	448012	482912	482105	480874	484598	484942	485675	486795	486742	487259
3	506649	495801	489791	481024	476142	475204	474632	475198	475146	475651
4	542581	536639	527726	511652	492692	491002	491601	492186	492133	492655
5	521472	527884	516858	508389	508129	508489	509108	509715	509659	510201
6	498661	498484	497525	500675	498199	498551	499159	499754	499699	500230
7	473442	451371	461660	458472	456204	456527	457083	457628	457578	458064
8	438851	419384	419511	416614	414553	414847	415352	415847	415802	416244
9	411545	413108	413234	410380	408350	408639	409137	409625	409580	410015

Table 12: AUTOINCURRED Full Triangle. [Source: own elaboration]

[1] OLS: Ordinary Least Squares

5.4.MCL with SUR

By specifying the SUR methodology (previously defined), we allow the inclusion of correlations between triangles. In this case, they are modeled simultaneously.

As we add this correlation the technical reserves will either tend to increase or decrease. This will clearly depend on the sign and magnitude of the correlation.

We can observe that the correlation coefficients are different from zero now, unlike in the SCL case.

Correlation Coefficients SUR								
0,326	-0,010	0,597	0,711	0,857	0,928	0	0	0

Table 13: Correlation Coefficients SUR. [Source: own elaboration]

The last three values are zero because, as we explained in the previous section, we are robustifying the tail (data partitioning) to obtain more precise estimates. So for the last segment, we apply OLS, ignoring the correlations.

The following are the results obtained:

	Development year j									
Accident Year *i*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	126406	262401	333272	381384	409812	425836	435538	438991	441691	441980
1	128176	254016	325846	378978	411165	426185	434166	437266	438154	438440
2	143665	284630	372650	431928	459700	472499	479514	481530	483501	483818
3	143065	284701	376340	425836	449974	461560	466656	469620	471543	471851
4	144496	304514	394019	443112	465470	478422	486399	489489	491493	491814
5	159700	324270	408719	457225	483406	498457	506768	509987	512075	512409
6	169520	327868	408858	459196	488526	503736	512134	515387	517497	517836
7	158971	305311	397465	451719	480571	495534	503796	506996	509071	509404
8	160789	308504	396476	450595	479375	494300	502542	505734	507804	508136
9	157860	314117	403689	458793	488097	503293	511685	514935	517043	517381

Table 14: AUTOPAID Full Triangle. [Source: own elaboration]

	Development year j									
Accident Year *i*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	406779	420532	432576	434658	438744	441998	443496	443781	443732	444204
1	404534	424084	430634	436619	438798	439479	440062	440289	440241	440709
2	448012	482912	482105	480874	484598	484942	485675	486795	486742	487259
3	506649	495801	489791	481024	476142	475204	474632	475198	475146	475651
4	542581	536639	527726	511652	492692	491002	491598	492184	492130	492653
5	521472	527884	516858	508389	508129	508484	509100	509707	509651	510193
6	498661	498484	497525	500675	498146	498494	499099	499693	499639	500169
7	473442	451371	461660	458413	456097	456416	456969	457514	457464	457950
8	438851	419384	419512	416561	414457	414747	415250	415745	415699	416141
9	411545	412900	413026	410121	408049	408334	408830	409317	409272	409707

Table 15: AUTOINCURRED Full. [Source: own elaboration]

Regarding the SCL model, we observe a decrease in the reserves when we apply SUR. At the beginning, we might think that if, for example, there is a high positive correlation between both triangles, the reserves should increase compared to the SCL model.

However, this is not always the case. In some cases, a positive correlation between triangles can lead to a decrease in estimated reserves. This happens when the multivariate model accurately captures the correlation and the information from one triangle can help to predict the development of the other, resulting in more precise estimates and reducing the bias in the reserves.

5.5.GMCL model without intercepts

As we have mentioned, the GMCL extends the MCL model by allowing dependencies between triangles. We will first develop a model without including the intercept terms, and then another one incorporating them.

The idea is to visualize the effect of including this dependency between triangles individually and jointly with the intercepts.

On the one hand, the correlation coefficients obtained for the GMCL-int model are the following:

Correlation Coefficients GMCL-int								
0,411	0,337	0,877	0,980	0,680	0,925	0	0	0

Table 16: Correlation Coefficients GMCL without intercepts. [Source: own elaboration]

And on the other hand, the future reserves estimate is:

	Development year j									
Accident Year *i*	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	126406	262401	333272	381384	409812	425836	435538	438991	441691	441980
1	128176	254016	325846	378978	411165	426185	434166	437266	438154	438440
2	143665	284630	372650	431928	459700	472499	479514	481530	483501	483818
3	143065	284701	376340	425836	449974	461560	466656	469620	471543	471851
4	144496	304514	394019	443112	465470	478422	484529	487607	489603	489924
5	159700	324270	408719	457225	483406	494420	499653	502827	504886	505216
6	169520	327868	408858	459196	485111	494720	499019	502189	504245	504574
7	158971	305311	397465	443200	465526	471014	472672	475674	477622	477934
8	160789	308504	385913	427686	447507	450374	450375	453236	455092	455389
9	157860	298236	373317	413964	433306	436301	436448	439220	441018	441307

Table 17: AUTOPAID Full Triangle. [Source: own elaboration]

	Development year j									
Accident Year *i*	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	406779	420532	432576	434658	438744	441998	443496	443781	443732	444204
1	404534	424084	430634	436619	438798	439479	440062	440289	440241	440709
2	448012	482912	482105	480874	484598	484942	485675	486795	486742	487259
3	506649	495801	489791	481024	476142	475204	474632	475198	475146	475651
4	542581	536639	527726	511652	492692	491002	491050	491635	491581	492103
5	521472	527884	516858	508389	508129	506099	505829	506432	506377	506915
6	498661	498484	497525	500675	508148	505256	504709	505311	505256	505792
7	473442	451371	461660	468018	483047	478059	476819	477387	477335	477842
8	438851	419384	432661	441874	461391	455168	453514	454055	454005	454487
9	411545	407419	419946	428584	447019	441124	439565	440088	440040	440508

Table 18: AUTOINCURRED Full Triangle. [Source: own elaboration]

In this GMCL model without intercepts, we can see that the ultimate loss of the first and second full triangle is now much closer compared to the MCL model applying SUR.

Two reasons could be that there are higher correlations now and that we have included the dependence between triangles.

5.6.GMCL model with intercepts

Finally, we will develop the GMCL model with intercept terms. As we said before, there are cases where the multivariate method fails due to the exclusion of these terms, which allow us to obtain more precise estimates.

In our case, we have applied intercepts in all periods of the first part of the partition in the robustification.

The obtained correlation coefficients are:

Correlation Coefficients GMCL with intercepts								
0,248	0,384	0,723	0,947	0,602	1	0	0	0

Table 19: Correlation Coefficients GMCL with intercepts. [Source: own elaboration]

And the estimated reserves are:

	Development year j									
Accident Year *i*	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	126406	262401	333272	381384	409812	425836	435538	438991	441691	441980
1	128176	254016	325846	378978	411165	426185	434166	437266	438154	438440
2	143665	284630	372650	431928	459700	472499	479514	481530	483501	483818
3	143065	284701	376340	425836	449974	461560	466656	469620	471543	471851
4	144496	304514	394019	443112	465470	478422	483973	487047	489041	489361
5	159700	324270	408719	457225	483406	494336	498839	502007	504063	504392
6	169520	327868	408858	459196	485758	496014	500184	503362	505422	505753
7	158971	305311	397465	448428	478590	488640	492985	496117	498148	498473
8	160789	308504	386678	437915	470320	480535	485232	488314	490313	490634
9	157860	298777	375500	426920	459640	470613	475965	478988	480949	481263

Table 20: AUTOPAID Full Triangle. [Source: own elaboration]

	Development year j									
Accident Year *i*	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	406779	420532	432576	434658	438744	441998	443496	443781	443732	444204
1	404534	424084	430634	436619	438798	439479	440062	440289	440241	440709
2	448012	482912	482105	480874	484598	484942	485675	486795	486742	487259
3	506649	495801	489791	481024	476142	475204	474632	475198	475146	475651
4	542581	536639	527726	511652	492692	491002	490973	491557	491504	492026
5	521472	527884	516858	508389	508129	505981	505605	506207	506152	506690
6	498661	498484	497525	500675	509165	506937	506413	507017	506961	507500
7	473442	451371	461660	475680	500885	499125	498605	499199	499144	499674
8	438851	419384	435911	457822	492246	490984	490548	491133	491079	491601
9	411545	410415	427940	449599	482322	481650	481430	482004	481951	482463

Table 21: AUTOINCURRED Full Triangle. [Source: own elaboration]

By adding intercepts, we can observe that the columns of ultimate loss for both full triangles remain very similar but slightly increase compared to the model without intercepts.

In the next section, we will explain the importance of obtaining larger or smaller reserves, although we can already get an idea if we have understood the concept of reserve.

5.7.Paid-Incurred ratios

The paid-incurred ratio is used to evaluate the adequacy of claims reserves. It allows us to measure the evolution of claims costs over time and assess the quality of existing reserve estimates. This ratio is defined as the proportion of total claims paid and the total claims incurred.

If we obtain a ratio greater than one, it means that the actual costs of claims have exceeded the initial estimates, indicating a need of additional reserves as we have underestimated the costs.

On the other hand, if the ratio is less than one, it signifies that the costs of claims have been lower than expected, indicating that we have set aside too high reserves.

If this ratio is close to one, it suggests that the reserve estimates are accurate, and the actual payments align with the expected payments. In the case of ratios deviating significantly from one, it indicates that the estimates obtained are highly imprecise, and adjustments to the model must be made to reflect the actual costs of claims better.

In Table 22, we can observe a table of the ratios obtained in each model over time:

		Model		
Period	**SCL**	**MCL**	**GMCL-int**	**GMCL**
1	99,50%	99,50%	99,50%	99,50%
2	99,49%	99,49%	99,49%	99,49%
3	99,29%	99,29%	99,29%	99,29%
4	99,20%	99,20%	99,20%	99,20%
5	99,83%	99,83%	99,56%	99,46%
6	100,43%	100,43%	99,66%	99,55%
7	103,53%	103,53%	99,76%	99,66%
8	111,23%	111,24%	100,02%	99,76%
9	122,10%	122,11%	100,20%	99,80%
10	126,22%	126,28%	100,18%	99,75%

Table 22: Paid-Incurred Ratios [Source: own elaboration]

We can conclude that the most accurate model is the GMCL with intercepts. Both the SCL and MCL models show ratios for the last periods over 120%.

Lastly, when we introduce the GMCL without intercepts, we see that the model adjusts much better, and when we add these intercepts, we reach values slightly below one, indicating a correct estimation of reserves.

5.8. Residual plots of the error terms

Another way to check the model's goodness of fit is by examining the error residuals plots. As we already know, residuals error are the differences between the observed values and the values predicted by the model.

If it is a good model, we expect the residuals error to be distributed around zero. If the residuals exhibit patterns or trends, it may indicate that the model is not capturing all the relevant information in the data, or that a different specification is needed.

Lastly, another advantage of visualizing the residuals plot is that it allows the detection of outliers. This is particularly helpful as the GMCL model is highly sensitive to these outlier values Peremans (2018), and appropriately handling them can greatly improve the model's fit.

The plots of the residual errors obtained for the different models are shown:

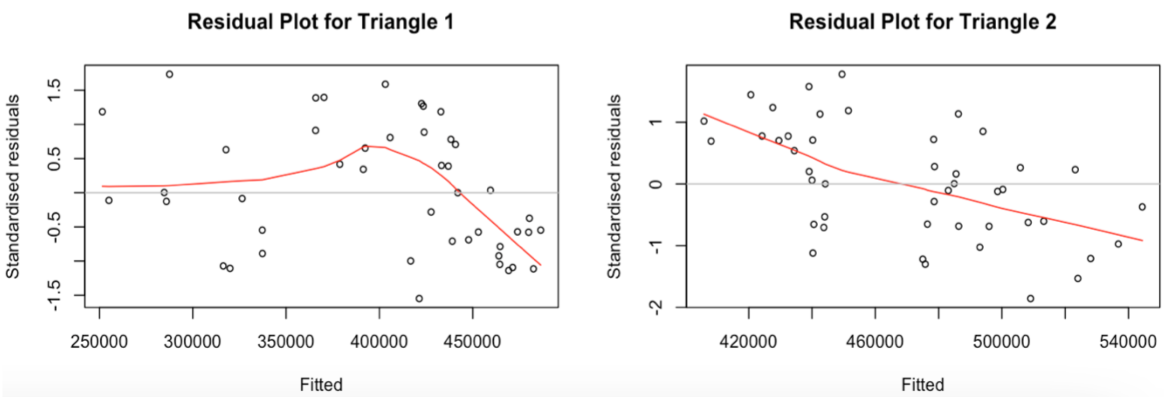

Figure 1: Residuals Plot with SCL model. [Source: R script]

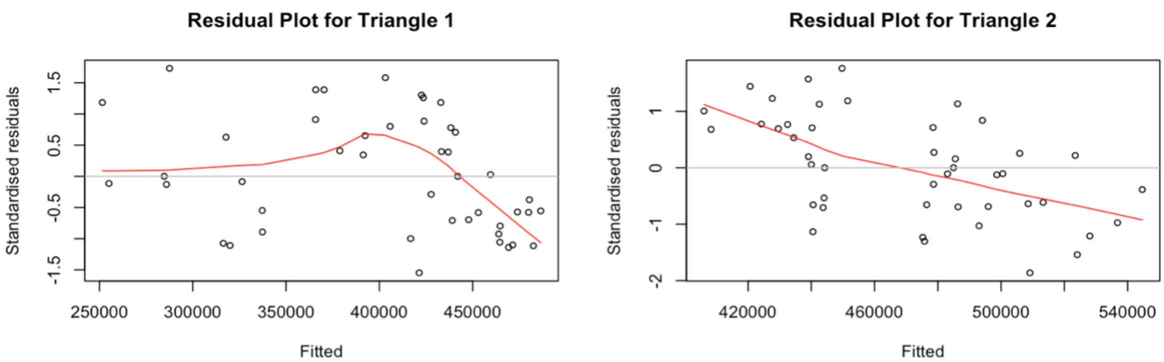

Figure 2: Residuals Plot with MCL model. [Source: R script]

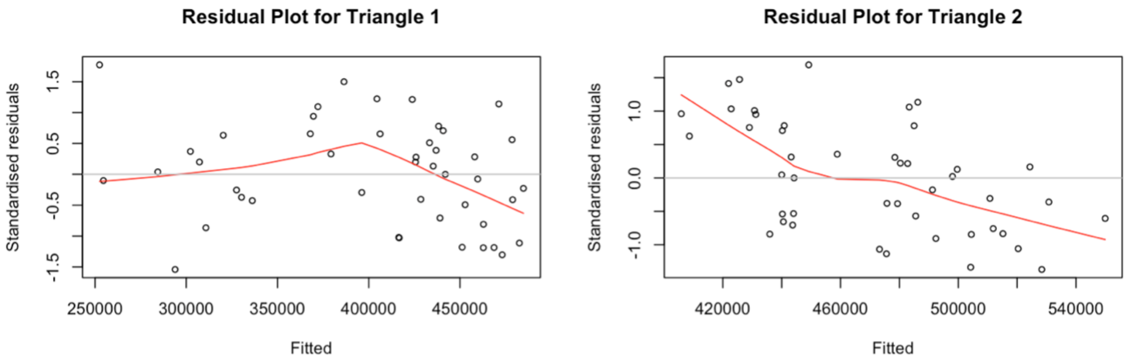

Figure 3: Residuals Plot with GMCL-int model. [Source: R script]

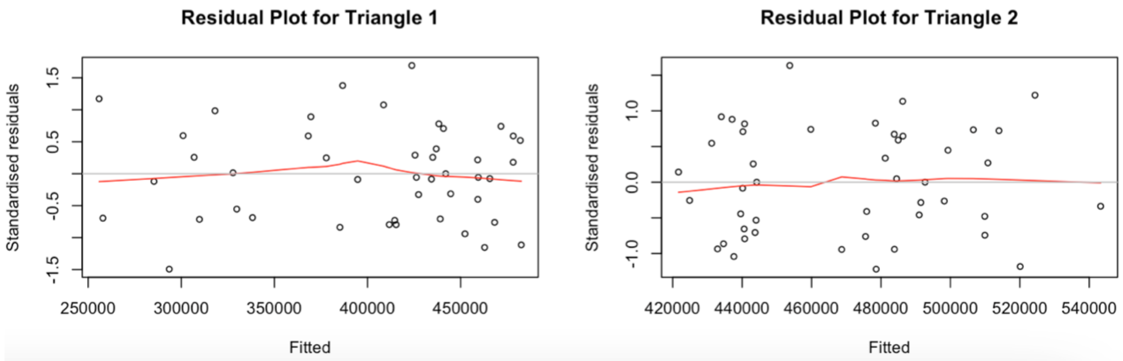

Figure 4: Residuals Plot with GMCL with intercepts model. [Source: R Script]

It is clear from the plots that all the models without intercepts exhibit a clear decreasing trend in their error residuals plots, indicating an overestimation of large values and an underestimation of small values.

Therefore, the model that would provide the best fit in this case would once again be the GMCL with intercepts, where the residuals are distributed around zero in both cases, without any apparent trend or pattern.

5.9.EIOPA Risk Free Rate (RFR) application

The results that we will present in this section have been updated using the interest rate curve provided by EIOPA (European Insurance and Occupational Pensions Authority).

Specifically, we will use the European interest rate curve published by EIOPA without volatility adjustment (volatility adjustment, VA) for April 2023, as stated in EIOPA (2023). We will not consider other factors such as inflation.

Table 23 presents a summary table of the interest rates to be used for the first ten years, as provided by EIOPA.

Year	Interest Rate
1	3,673%
2	3,362%
3	3,128%
4	2,998%
5	2,932%
6	2,893%
7	2,872%
8	2,865%
9	2,866%
10	2,875%

Table 23: Annual EIOPA Interest Rate, May 2023. [Source: own elaboration]

The purpose of updating the reserves with the EIOPA curve is to adjust the value of the technical reserves to their present value.

Before delving into how reserves are updated, we need to explain what the diagonals of triangles are and their importance when performing this update.

	Development year j					
Accident year i	**0**	**1**	**2**	**…**	$k-1$	k
0	$c_{0,0}$	$c_{0,1}$	$c_{0,2}$	$\cdots$	$c_{0,k-1}$	$c_{0,k}$
1	$c_{1,0}$	$c_{1,1}$	$c_{1,2}$	$\cdots$	$c_{1,k-1}$	
…	$\cdots$	$\cdots$	$\cdots$	$\cdots$		
$k-2$	$c_{k-2,0}$	$c_{k-2,1}$	$c_{k-2,2}$			
$k-1$	$c_{k-1,0}$	$c_{k-1,1}$				
k	$c_{k,0}$					

Table 24: Diagonals of a natural year. [Source: own elaboration]

If we look at Table 24, we can see that a claim occurred in the origin year $k-1$ but paid in the development year $j=2$, in terms of calendar year, has been realized in $k-1+2$, or in other words, in year $k+1$.

Similarly, if we look at the origin year k, but this time at the development year $j=1$, we will obtain the exact same calendar year as before.

By repeating this process, we end up with a diagonal of losses (as marked in gray in Table 24) that correspond to the same calendar year.

These diagonals correspond to the vectors of future payments for each calendar year. We will update these vectors of future payments to obtain the present value of the reserves for each calendar year.

In summary, we will take the values from each diagonal (corresponding to the same calendar year), sum them, and bring them to present value using the EIOPA curve.

To make this update we will treat them as an income, and we will apply the classic formula used in the financial sector:

$$\sum_{t=k+1}^{2k} R_t \cdot (1 + EIOPA_{RFR})^{-(t-k)}.$$

Where R_t are the total reserves for each natural year t, that is, the sum of the diagonal of losses of each natural year. On the other hand, $t-k$ is the number of years elapsed since the calendar year k. In $k+2$, it will be $k+2–k$, which is 2 years, and so on.

This update is important because it allows insurance companies to comply with regulatory and accounting requirements, which obligates the technical reserves to be valued at their present value using an appropriate discount rate. The EIOPA curve is considered a reliable and recognized source for determining the appropriate discount rate for insurance valuations in the European context.

Lastly, it can provide a better understanding of the current financial position of the company and can help in the risk management.

In Table 25, is shown the vector of future payments and its present value of the two triangles of the GMCL model, since it has been the selected model.

Vectors of future payments

	Year	**2023**	**2024**	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**	**2031**	**Present Value**
GMCL+int	*AUTOPAID*	320748	181955	107653	56681	24554	13333	5380	2294	317	667185
	AUTOINCURRED	37339	64079	55997	33139	1005	1658	1052	470	513	180397

Table 25: Vectors of future payments and their present value. [Source: own elaboration]

Based on Table 25, we can conclude that the company should have total reserves of 847,582 to meet its obligations in the coming years.

6. Conclusions

With the completion of this work, the objective of introducing and understanding a new model for reserve calculation that we had not worked with during the master's program, the GMCL model, and the advantages that it offers over the classical SCL or MCL models has been achieved.

To accomplish this, a practical exercise was conducted in R, where the correlations between triangles and the reserves obtained in each model were visualized.

Subsequently, paid-incurred ratios and error residuals plots were calculated to assess the model fit quality. An SCL model was developed as it ignores correlations between triangles and provides the same results as the classical method, allowing us to compare the results obtained from the classical method against those obtained from the stochastic method.

Then correlations between triangles were included by applying the SUR methodology to the MCL model, and finally, a GMCL model was applied both with and without intercepts. It should be noted that the difference between an MCL model and a GMCL model is that in the GMCL model, in addition to correlation, dependencies between triangles are also allowed.

The GMCL model with intercepts emerged as the clear appropriate, as it does not underestimate reserves and its residual errors are distributed close to zero.

While there are various resources available to assess the quality of a model, there is no established criterion for selecting models, such as the Akaike criterion in statistical models. In this case, the knowledge and experience of the actuary play a crucial role in model selection.

Although in our practical exercise we concluded that the GMCL model with intercepts had better fit quality, we cannot generalize that applying a GMCL model will always be better to the classical method.

This multivariate methodology provides greater efficiency gain only when the triangles present high correlation, which, as we know, tends to decrease in the later periods.

If this is not the case, it may be more advisable to apply the classical method, as the application of a multivariate model is much more complex.

We should mention that an existing problem in developing multivariate methods is the over-parametrization, which refers to the inclusion of too many parameters, resulting in imprecise estimates.

Furthermore, unlike the classical method, in the multivariate approach, the development factors cannot be easily obtained since they are included in the matrix B_k. Estimating this matrix generally reduces the efficiency of the model and is one of the main reasons why we cannot guarantee that a multivariate model is always better than the univariate one.

As the purpose of this work was to introduce and gain a solid understanding of the model, a practical exercise was conducted without adding excessive complexity to the calculations.

These models can become much more complex by adding parameter constraints or studying the optimal inclusion of intercepts in specific periods. In our case, we did not delve into these aspects, leaving them open for future research.

7. Bibliography

Barnett, G., Zehnwirth, B. (2000). Best estimates for reserves. https://www.casact.org/sites/default/files/database/forum_98fforum_zehnwirth.pdf

Boj, E., Claramunt, M. M., Costa, T. (2020). *Tarificación y provisiones (tercera edición)*. Depósito Digital de la Universidad de Barcelona, colección OMADO. https://diposit.ub.edu/dspace/handle/2445/149241

Carrato A., Concina F., Gesmann M., Murphy D., Wüthrich M., Zhang W. (2023). *ChainLadder: Claims reserving with R.* https://cran.rstudio.com/web/packages/ChainLadder/vignettes/ChainLadder.html#multivariate-chain-ladder-using-seemingly-unrelated-regressions

EIOPA (2023). *Risk-free interest rate term structures,* EIOPA, European Insurance and Occupational Pensions Authority Westhafenplatz 1, 60327 Frankfurt am Main, April, 2023.

https://www.eiopa.europa.eu/tools-and-data/risk-free-interest-rate-term-structures_en

Gesmann, M., Murphy, D., Zhang, Y., Carrato, A., CRupi, G., Dutang, C., Lacoume, A., Charpentier, A., Withrich, M., Concina, F., Dal Moro, E., Kryvayych, Y., Goulet, V., De Virgilis, M., Spina, M. (2023). ChainLadder, R package in Cran, version 0.2.17. https://cran.r-project.org/web/packages/ChainLadder/index.html

Mazuelos, G. (2013). Stochastic Chain Ladder models and their applications in the calculation of reserves for insurance companies. http://hdl.handle.net/20.500.12404/6176

Peremans K., Van AElst S., Verdonck T. (2018). A Robust General Multivariate Chain Ladder Method. *Risks*, 6, 4, 108.

R DEVELOPMENT CORE TEAM (2023). *R Development Core Team, R: a language and environment for statistical computing*. Vienna, Austria. https://www.r-project.org/

Weindorfer, B. (2012). *A practical guide to the use of the Chain-ladder method for determining technical provisions for outstanding reported claims in non-life insurance.* University of Applied Sciences (Vienna).

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *American Statistical Association*, 57, 348-368. https://doi.org/10.1080/01621459.1962.10480664

Zhang, Y. (2010). A general multivariate chain ladder model. *Insurance Mathematics & Economics 03-13.* https://doi.org/10.1016/j.insmatheco.2010.03.002

Appendix A: R Script

A.1. Install packages and database

```
install.packages("ChainLadder")
library(ChainLadder)

install.packages("systemfit")
library(systemfit)

install.packages("Metrics")
library(Metrics)

database<-list(a,b)
database<-as(database,"triangles")

database

triangle<-database[[1]]
triangle

triangle2<-database[[2]]
triangle2
```

A.2. Development Factors

```
n <- 10
f <- sapply(1:(n-1),
        function(i){
          sum(triangle[c(1:(n-i)),i+1])/sum(triangle[c(1:(n-i)),i])
        }
)
f

n <- 10
f2 <- sapply(1:(n-1),
        function(i){
          sum(triangle2[c(1:(n-i)),i+1])/sum(triangle2[c(1:(n-i)),i])
        }
)

f2
```

A.2. Obtaining Full Triangles

```
full1<- cbind(triangle);
for(k in 1:n) {
  full1[(n-k+1):n, k+1] <- full1[(n-k+1):n,k]*f[k]}

A<-round(full1)
```

```
colnames(A)<-c(0:9)
rownames(A)<-c(2013:2022)

A

getLatestCumulative(triangle)

full2<- cbind(triangle2);
for(k in 1:n) {
  full2[(n-k+1):n, k+1] <- full2[(n-k+1):n,k]*f2[k]
}

B<-round(full2);B

colnames(B)<-c(0:9)
rownames(B)<-c(2013:2022)

B
```

A.4. EIOPA RFR

```
EIOPA<c(3.673,3.362,3.128,2.998,2.932,2.893,2.872,2.865,2.866,2.875,2.890,2.896,2.
903)
EIOPA<-EIOPA/100
EIOPA


# obtain VPF

a<-matrix(c(rep(0,dim(C)[1]),mch$FullTriangle),nrow=dim(C)[1],ncol=dim(C)[1]);a
noncumFullTriangle<-mch$FullTriangle-a; noncumFullTriangle

# Obtenemos el vector de pagos futuros

vpf <- rep(0, dim(C)[1] - 1)
for (k in 1:dim(C)[1] - 1) {
  future <- row(noncumFullTriangle) + col(noncumFullTriangle) - 1 == dim(C)[1] + k
  vpf[k] <- sum(noncumFullTriangle[future])
}
vpf


#UPDATE VPF#

EIOPA
i.renta<-numeric(length(EIOPA))
for (i in 1:length(EIOPA)) {i.renta[i]<- (1+EIOPA[i])^(-i)}; i.renta
prov.renta<-sum(vpf*i.renta); prov.renta
```

A.4. Separate Chain-Ladder

```
# Multivariate chain-ladder using Separate Chain-Ladder ignoring correlations #

fit.1 <- MultiChainLadder(database,fit.method = "OLS")
fit.1


C<-round(fit.1$FullTriangles[[1]])
D<-round(fit.1$FullTriangles[[2]])

par(mfrow = c(2, 2))
plot(fit.1, which.plot = 3:4)

round(unlist(residCor(fit.1)), 3)
```

A.5. SUR

```
fit.2 <- MultiChainLadder2(database, fit.method = "SUR",last=3)
fit.2

round(unlist(residCor(fit.2)), 3)

E<-round(fit.2$FullTriangles[[1]])
F<-round(fit.2$FullTriangles[[2]])

#PLOT#
par(mfrow = c(2, 2))
plot(fit.2, which.plot = 3:4)
round(unlist(residCor(fit.2)), 3)
```

A.6. GMCL without intercepts

```
fit.3<-MultiChainLadder2(database,type = "GMCL-int", last = 3)
fit.3

round(unlist(residCor(fit.3)), 3)

G<-round(fit.3$FullTriangles[[1]])
H<-round(fit.3$FullTriangles[[2]])


par(mfrow = c(2, 2))
plot(fit.3, which.plot = 3:4)
```

A.7. GMCL with intercepts

```
fit.4<-MultiChainLadder2(database, type = "GMCL", last = 3)
fit.4

round(unlist(residCor(fit.4)), 3)

I<-round(fit.4$FullTriangles[[1]])
J<-round(fit.4$FullTriangles[[2]])

par(mfrow = c(2, 2))
plot(fit.4, which.plot = 3:4)

fit.1 <- MultiChainLadder(database,fit.method = "OLS")
fit.1

fit.2 <- MultiChainLadder2(database, fit.method = "SUR",last=4)
fit.2

fit.3<-MultiChainLadder2(database,type = "GMCL-int", last = 4)
fit.3

fit.4<-MultiChainLadder2(database, type = "GMCL", last = 4, int)
fit.4
```

A.8. Plots summary

```
par(mfrow = c(4, 2)) #scl#
plot(fit.1, which.plot = 3)

par(mfrow = c(2, 2))

par(mfrow = c(2, 2)) #mcl#
plot(fit.2, which.plot = 3)

par(mfrow = c(2, 2)) #gmcl-int#
plot(fit.3, which.plot = 3)

par(mfrow = c(2, 2)) #gmcl
plot(fit.4, which.plot = 3)

par(mfrow = c(2, 2)) #mcl + interceptos
plot(fit.5, which.plot = 3:4)
```

A.9. PAID-INCURRED ratios

```
#RATIO 1: MCL-OLS = SCL#
ultimateC<-C[,10]
ultimateD<-D[,10]
```

```
ratio1<-ultimateC/ultimateD;ratio1

#RATIO 2: MCL-SUR = MCL#
ultimateE<-E[,10]
ultimateF<-F[,10]

ratio2<-ultimateE/ultimateF;ratio2

#RATIO 3: GMCL-int #
ultimateG<-G[,10]
ultimateH<-H[,10]

ratio3<-ultimateG/ultimateH;ratio3

#RATIO 4:GMCL #
ultimateI<-I[,10]
ultimateJ<-J[,10]

ratio4<-ultimateI/ultimateJ;ratio4

ratios<-cbind(ratio1,ratio2,ratio3,ratio4)*100;ratios
colnames(ratios)<-c("SCL","MCL","GMCL-int","GMCL")
ratios
```