UNIVERSITY OF LIVERPOOL

STOCHASTIC CLAIMS RESERVING IN INSURANCE

WITH REGRESSION MODELS

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor of Philosophy
in Mathematical Sciences by

Luís Portugal

13th September 2018

Contents

Motivation

The Chain-Ladder (CL) is the claims reserving method most used by actuaries with data in triangle format. Several references in the literature highlight its role, for example, Wüthrich and Merz (2008) and Marcuson (2013). There is also a survey from the International Actuarial Association, IAA (2017), confirming the CL as the method most used by actuaries.

However, a CL bias has been identified by Halliwell (2007) and this may create problems to the insurer's management: the CL may not be the optimal solution to match the data and very often presents high prediction errors (the square root of the mean square error of prediction). We explain these limitations in the following two paragraphs.

At least since Straub (1988), we know that the CL is not the exact solution but only an approximation to have the minimization of the sum of the square of the errors, when a linear regression is applied to the claims reserving triangles - and the CL is a linear regression, as Straub (1988) also showed this. Also, we know from Mack (1993a) that the CL does not minimize the sum of the square of the errors but minimizes the weighted sum of the square of the errors, which means that it assumes errors as heteroscedastic (with non-constant variance). With regression techniques, heteroscedasticity is a feature from cross-section models; see for example Fomby et al. (1984). Cross-section models have data coming from the same period (with claim's triangles would be the same origin year) and from several entities (with claim's triangles could be more than one development year). We also have cross-section models inside panel data models: cross-section models from several years. With claim's triangles, this should correspond to estimate a complete triangle with regression techniques. Panel data models may show heteroscedasticity if the model parameters are assumed the same for all the entities (the same loss development factors for all the triangle columns). However, when we estimate the loss development factors they are not equal or similar between all the triangle columns. This means that we should not expect to find in claim's triangles, in most of the cases, heteroscedasticity. It is possible that heteroscedasticity arises when data is irregular, as in such cases it will be more difficult to do prediction and the variance of the error is probably not going to be constant. The same may happen if we consider several lines of business with the same development factors, but that is not a common procedure in multivariate claims reserving literature, see for example Zhang (2010).

It is also common that actuaries mention the existence of high prediction errors when the CL is applied. This is confirmed, as an actuary, by the author of this thesis and may be also seen in several papers that present prediction errors for the stochastic CL or for the models that replicate the CL. One example may be seen in Mack (1993a, 1993b, and 1994). Having a high prediction error on a model, which is a simplification of the reality, does not give confidence on its results. When the prediction error is high the predictions are not close to the actual experience. Probably other measures for model selection, such as the errors analysis and the back-testing, will also present poor results.

Apparently, the CL seems to show a paradox. It is known, as stated by Straub (1988), that the regression models minimize the sum of the square of the errors and may be used in claims reserving. But according to Taylor (1978), the regression techniques were always seen with suspicion by actuaries, even though most of them use the CL, which is a weighted-regression model.

Our first motivation in this thesis is to present a method that assumes the use of regression techniques and that minimizes the sum of the square of the errors. We expect this approach to have better predictions than the traditional CL. Then we want this new method to be general and as such less dependent on the triangle considered, because it should have enough flexibility to adapt to each situation. As a general method, it should be able to replicate the results of several known methods like the CL.

Having this general method will bring us a solution for two other important insurer's problems: the claim's reserves estimation dependency on payments speed (inside each triangle) and the estimation of several triangles at the same time (with an accurate method for all the triangles). These issues are very significant. The first is the recognition of a stylized fact of claims reserving: if we increase/decrease the speed of payments on one development year we decrease/increase the payments in the following years. The second oblige us to have accurate methods for claims reserving estimation: if the CL is not accurate to one triangle, applying it to several triangles at the same time will be even worse.

For this purpose, we will start by developing a first stochastic model like the one from Mack (1993a, 1993b, 1994) but considering a loss development factor, the Vector Projection (VP), which is a regression through the origin between two adjacent columns in the triangle. We expect this to bring better predictions in most triangles, when compared with the traditional CL. This method is based on the Mack's framework but changes two things. Firstly, and as we said before, it considers the VP loss development factors, instead of the ones from the CL. Secondly, it considers the claim's payments variance proportional to the square of the payments, as they are the weights from the VP loss development factors. The motivation of this method is to show that a small change in the Mack assumptions allows for a better prediction in most of the triangles: the VP method changes CL loss development factors but maintains the heteroscedastic feature from CL. However, we will change this VP last feature on the generalized models (see next paragraphs).

We will develop a second method using regression methods, the generalized link ratios method that should consider the VP and the CL as particular cases. With this general method we will also be able to generate other methods like the Simple Average (SA). This general approach should improve the VP predictions even further as this new VP method will be homoscedastic. Due to the reasons presented in previous paragraphs, we believe that heteroscedasticity is not a common feature of insurer's triangles, unless claim's payments are irregular over the years.

Having this generalized link ratios method, we will consider a third method with a multivariate approach. However, the latter will be different from the ones on the current literature of multivariate claims reserving. It will just have one triangle but with all the regressions contemporaneously correlated, giving us a multivariate regression.

We will finish this thesis with a fourth and fifth method on portfolio data, which means several triangles estimated at the same time. Here we are going to have contemporaneous correlations between each triangle as happens in several methods from the literature, see for example Zhang (2010). However, we will introduce two differences for the methods presented in the literature. We use multivariate regression not just between triangles but also inside each triangle. Also, we do not use the CL to estimate the triangles loss development factors. Instead, we use either the generalized link ratios or the multivariate generalized link ratios, the methods two and three. This shall improve the accuracy of the methods when compared with the traditional univariate methods.

We want to develop for all the five method's non-recursive formulas that will give the analytical solution of the mean squared error of prediction. Having the square root of that we get the prediction error. This will allow us to compare several methods using this criterion. It is an important criterion due to its relationship with the gap between prediction and experience. We will also try to understand if the prediction error is related with conclusions from other techniques for method selection, as error's analysis and back-testing.

Finally, we also want to see if some regression techniques tests (as the heteroscedasticity test, the serial correlation test, the equation's correlations test and the pooled data test) can be useful on method selection.

To summarize, we want to develop the following five methods:

Methods	**Number of Triangles**		**Correlations Between**	
	One	**Several**	**Equations**	**Triangles**
VP	Yes	--	No	--
GLR	Yes	--	No	--
MGLR	Yes	--	Yes	--
PGLR	--	Yes	No	No
PMGLR	--	Yes	Yes	Yes

Abstract

The insurance business and the claims process features, as two sources of uncertainty and risk, are presented and the importance of a correct reserving to overcome this is highlighted.

The claims reserving framework with data in triangle format is summarized and the main method used for reserving, the Chain-Ladder (CL), is explained. The assumptions of the latter are also presented and criticized. An attempt is made to explain why actuaries use the CL.

The current methods for claims reserving are also summarized, both the deterministic and the stochastic. The relation of some of these methods with regression analysis is highlighted and an historical summary of the use of regression models in claims reserving is presented. The definition of the prediction error (as the square root of the mean square error of prediction) and the formulas for the confidence intervals are shown. Some conclusions about reserving models and the use of regression techniques are summarized.

A first method, as an alternative to the traditional stochastic CL, Mack (1993a, 1993b, 1994), is presented. The new method, the stochastic Vector Projection (VP), is based on regression techniques with heteroscedastic errors and is shown, on the survey conducted, to produce lower prediction errors. A numerical analysis with regular and irregular data is performed and a method selection is done with errors inspection and back-testing calculations. The conclusions from these two tools for method selection are compared with the obtained prediction errors.

A second method is presented, the stochastic generalized link ratios (GLR). The latter can replicate the VP, the CL, and the Simple Average (SA), as cases with a specific parameter. It also shows that other methods may be obtained through this specific parameter. The parameter is defined so that we get the method with the lowest prediction error. The method is also able to show an alternative to the prediction error estimation from the stochastic CL, from Mack (1993a). This GLR method presents the prediction errors with an analytical formula, not recursive, as was traditional with some similar approaches, such as the one from Murphy (1994).

The GLR method highlights the importance of the heteroscedasticity assumption (non-constant variance of the errors) in some claims reserving methods. A homoscedastic (constant variance of the errors) GLR is also developed, the homoscedastic VP.

Using this GLR method a third method is presented with stochastic multivariate regressions inside the claims triangle, the multivariate generalized link ratios (MGLR). This method considers the contemporaneous correlations between all the regressions inside the triangle and brings light to other issues known in practice, such as the speed of payments that affects reserve estimation. This approach contrasts with the methods on multivariate claims reserving that estimate several triangles at the same time with the traditional CL, see for example Prohl and Schmidt (2005), Wüthrich and Merz (2007b), and Zhang (2010). With MGLR, we just have one triangle and the multivariate approach comes from the contemporaneous correlations considered inside that triangle. Using a specific parameter (as in the GLR), the MGLR will also present, in particular cases, the multivariate versions from VP, CL and SA. Other multivariate methods may be obtained for other values from this parameter.
Numerical results are presented for irregular and regular datasets and a survey of 114 triangles is summarized. Heteroscedasticity tests are conducted as well as tests on the correlations between triangle equations. Serial correlation inside each equation is also analysed.

In a fourth and fifth method, GLR and MGLR are extended, and we will consider the estimation of several triangles at the same time. The new methods, the portfolio generalized link ratios (PGLR) and the portfolio multivariate generalized link ratios (PMGLR) consider the estimation of several triangles at the same time. The PMGLR allow the consideration of contemporaneous correlations between those triangles and between equations inside each triangle. The PGLR and the MPGLR will also present, in particular cases, the portfolio versions (univariate and multivariate) for VP, CL, and SA. As with GLR and MGLR, a specific parameter is used to identify these methods. Other portfolio methods may be obtained for other values from this parameter, following the same procedures used with GLR and MGLR.
Numerical results are presented using the three triangles considered in this thesis, either as portfolio data (the three triangles estimated at the same time with their correlations) or as aggregated data (the three triangles sum in just one triangle). A test for the possibility of having pooled data is also conducted.

Finally, several general conclusions are presented about the thesis. Most of them respect the CL, the five alternative methods presented, and the decrease of the prediction errors when the latter is considered. The absence of heteroscedasticity in most insurers' triangles is emphasized. The existence of heteroscedasticity in irregular data triangles is not excluded.

The relation between prediction errors and two other method selection techniques, errors analysis and back-testing, is emphasized and the importance of some regression tests to help for method selection is also highlighted.

The need to consider multivariate regressions in claims reserving, with correlations between the equations, is explained, as well as the advantage of working with portfolio data and triangle's correlations.

Acknowledgments

The support from the University of Liverpool to develop this research was very important. My supervisors, Dr. Athanasios Pantelous and Dr. Hirbod Assa, provided the right direction to the papers produced and the thesis presentation. I am grateful to both for the support given.

Despite this, any eventual mistakes in this Thesis are my full responsibility.

Illustrations

List of Tables

List of Figures

Notations

$\hat{a}$	Estimated constant on regression model.
a	True constant on a regression model.
$a_{v,j}$	True parameter $v = 1, \ldots$ on a multiple regression model equation j.
AC	Average Cost method.
AD	Additive method.
b	Real slope on a regression model.
$\hat{b}$	Estimated slope on a regression model.
b_j	Real loss development factor at column j.
$b_{x,j}$	Generic parameter from regression on column j. If $x = 0$, it is a constant.
$\hat{b}_j$	Estimated loss development factor at column j.
$\hat{b}_j^{XX}$	Estimated loss development factor at column j on method xx.
BC	Best case link ratio method.
BF	Bornhuetter-Ferguson method.
BH	Benktander and Hovinen method.
C_i	Cumulative payments on origin year i. Same meaning as $C_{i,j}$
$C_{i,j}$	Cumulative payments on origin year i and development year j.
${C^*}_{i,j}$	Cumulative payments on origin year i and development year j scaled by a volume.
$\bar{C}_{i,j}$	Average payments on origin year i and development year j.
$\hat{C}_{i,j}$	Expected cumulative payments, origin year i and development year j.
$C_{i,T-i+1}$	Payments done so far on origin year i from triangle with T years.
$C_{i,T}$	Real ultimate costs on origin year i in triangle with T years.
$\hat{C}_{i,T}$	Estimated ultimate costs on origin year i in triangle with T years.
${\hat{C}_{i,T}}^{xx}$	Estimated ultimate costs on origin year i on xx method.
CC	Cape Code method.
CL	Chain-Ladder method.
CLR	Complementary Loss Ratio method.
Cov	Covariance operator.
c_j	Triangle column parameter on development year j.
$\hat{c}_j$	Triangle estimated column parameter on development year j.

d_z	Triangle diagonal parameter on diagonal z.
df	Degrees of freedom.
$diag$	Diagonal operator, transforms vector in diagonal matrix.
D_l	Set of projected claims, $\{C_{i,j}: i+j-1>T\}$.
D_u	Set of the history of claims, $\{C_{i,j}: i+j-1\leq T\}$.
E	Units of exposure.
$\mathbb{E}$	Expected Value.
$\hat{f}_j$	Estimated ultimate factor (ultimate loss development factor) at column j.
$\hat{f}_{\bar{C},j}$	The ultimate factor (ultimate loss development factor) of the average payments on column j.
$\hat{f}_{nP,j}$	The ultimate factor (ultimate loss development factor) of the number of claims settled on column j.
$\tilde{F}$	Test statistic for pooled data test.
$F_{(df_1,df_2)}$	Distribution F with df_1 and df_2 degrees of freedom.
$F_{i,j+1}$	Link ratios between two adjacent cells in the same row i of the triangle.
g_j	Grossing-up factor on column j.
$\hat{g}_j$	Estimated grossing-up factor on column j.
$g_{i,j}$	Grossing-up factor on triangle cell i,j.
GLM	Generalized Linear Model.
GLR	Stochastic Generalized Link Ratios method.
GLS	Generalized Least Squares.
GU	Grossing-up method.
h (Time)	Independent variable Time with relation with the dependent variable given by operator functional form h.
$iC_{B,i}$	Incurred claims benchmark for origin year i.
$iC_{i,j}$	Incurred claims on triangle cell i,j.
$i\bar{C}_{i,j}$	Average incurred claims on triangle cell i,j.
$I_{i,j}$	Incremental payments on incremental payments triangle cell i,j.
$I^*{}_{i,j}$	Incremental payments on incremental payments triangle cell i,j.
$\hat{I}_{i,j}$	Estimated incremental payments on incremental payments triangle cell i,j.
i	Index for origin years $i=1,\ldots,T$.

I_q	Identity matrix with size $q \times q$.
$\boldsymbol{I_q}$	Identity matrix from portfolio data models with size $\times\ q$.
j	Index for development years, $j = 1, \dots, T$.
l	Likelihood function.
LM	LM statistic from White test.
LLR	Last link ratio method.
LR	Loss ratio method.
lr	Loss Ratio, the incurred claims divided by the earned premiums (if origin year is the origin year) or by the premiums (if the origin year is the underwriting year).
lr^{xx}	Loss Ratio of xx method.
${lr_i}^{ult}$	Ultimate Loss Ratio at origin year i.
lr_j	Ultimate Loss Ratio at development year j.
LRT	Link Ratio method.
M	Idempotent matrix.
$MGLR$	Stochastic Multivariate Generalized Link Ratios method.
MD	Median link ratio method.
m	Number of observations on the upper triangle.
max	Maximization operator.
$msep$	Mean square error of prediction.
min	Minimization operator.
$nC_{i,j}$	Number of cumulative notified claims on cell i, j.
$nP_{i,j}$	Cumulative number of claims with payments on cell i, j.
OLS	Ordinary Least Squares.
P	Premiums if underwriting year and Earned Premiums if origin year.
$PGLR$	Stochastic Generalized Link Ratios method on a Portfolio of triangles.
$PMGLR$	Stochastic Multivariate Generalized Link Ratios on a Portfolio of triangles.
$\bar{p}$	Average premium per unit of exposure.
$p(y)$	Density function of random variable y.
$p(y\|x)$	Conditional density function of random variable y conditioned by x.
pe	Prediction error, also called standard error. Both are the square root of *msep*.
q	Year from a set Q of years included on loss development factors calculations.
Q	Set of years considered in the calculations of the loss development factors.

$\hat{R}_i$ Estimated reserve for origin year i.

$\hat{R}$ Estimated reserve for all origin years.

r_i Triangle row parameter on origin year i.

$\hat{r}_i$ Triangle estimated row parameter on origin year i.

$s_{jj'}$ Variance-covariance between multivariate regressions *j* and *j'*.

$\hat{s}_{jj'}$ Estimated variance-covariance between multivariate regressions *j* and *j'*.

$s_{t,lj}$ Variance-covariance between triangle t and multivariate regressions *l* and *j*.

$\hat{s}_{t,l,j}$ Estimated variance-covariance between triangle *t* and multivariate regressions *l* and *j*.

SA Simple Average link ratio method.

SSR Sum of the square of the errors.

SSR_j Sum of the square of the errors on regression j.

T Number of origin years and development years from the triangle.

T_j Number of origin years and development years from column j.

tr Trace from matrix.

TR Trend link ratio method.

U_j Proportion of incremental payments on column j.

$\hat{U}_j$ Proportion of incremental payments on column j.

v_i Error (errors) observation i from equation that explains the errors (errors).

V Volume measure, an exposure measure which may be in physical units or monetary values.

Var Variance operator.

VP Stochastic Vector Projection method.

WC Worst case link ratio method.

W GLR weights matrix.

W_{xx} GLR weights matrix from link ratios method xx.

W_F GLR future weights matrix.

$W_{F,xx}$ GLR future weights matrix from link ratios method xx.

$\boldsymbol{W}$ Heteroscedasticity matrix for portfolio of triangles.

$\boldsymbol{W}_{xx}$ Heteroscedasticity matrix from link ratios method xx for portfolio of triangles.

$\boldsymbol{W}_F$ Heteroscedasticity future weights matrix for portfolio of triangles.

$\boldsymbol{W}_{F,xx}$	Heteroscedasticity future weights matrix from link ratios method xx for portfolio of triangles.
x	Generic independent variable.
$x_{i,j}$	Independent variable on a regression of triangle with cells i,j.
X	Matrix of independent variables of specific dimension.
X_k	Vector with observations from column k.
X_F	Matrix of future value of independent variables of specific dimension.
$\boldsymbol{X}$	Matrix of independent variables of specific dimension for a portfolio of triangles.
$\widetilde{\boldsymbol{X}}$	Matrix of independent variables of specific dimension for a portfolio of triangles in the restricted model.
$\boldsymbol{X}_{\boldsymbol{F}}$	Matrix of future values of independent variables of specific dimension for a portfolio of triangles.
y	Generic dependent variable.
$y_{i,j}$	Dependent variable on a regression of triangle with cells i,j.
$y^*_{i,j}$	Transformed $y_{i,j}$
Y	Block vector of dependent variables of specific dimension.
Y_F	Block vector of estimated future dependent variables of specific dimension.
Y_k	Vector with observations from column $k+1$.
$\boldsymbol{Y}$	Matrix of dependent variables of specific dimension for a portfolio of triangles.
Z_i	Variable observation i that explain the errors evolution.
z	Index for calendar years.
α	Parameter from loss development factor general formula that defines several claims reserving methods. Also used in the W matrix to define the method and identify the level of heteroscedasticity.
β	Vector of loss development factors.
$\boldsymbol{\beta}$	Vector of loss development factors on portfolio of triangles.
$\hat{\beta}$	Estimated vector of loss development factors.
$\widehat{\boldsymbol{\beta}}$	Estimated vector of loss development factors on portfolio of triangles.
$\widetilde{\boldsymbol{\beta}}$	Vector of loss development factors on portfolio of triangles for restricted model.
ε_i	Error (or residual) from observation i.

$\varepsilon_{i,j}$	Error (or residual) from a model in the triangle cell i, j.
${\varepsilon^P}_{i,j}$	Pearson error (or residual) from a model in the triangle cell i, j.
$\hat{\varepsilon}_{i,j}$	Estimated error (or estimated residual) in the triangle cell i, j.
ε	Vector of random errors (or errors) from a model.
$\boldsymbol{\varepsilon}$	Vector of random errors from a model on a portfolio of triangles.
$\hat{\varepsilon}$	Vector of estimated errors (or estimated errors).
$\hat{\boldsymbol{\varepsilon}}$	Vector of estimated random errors on a portfolio of triangles.
ε_F	Vector of future errors (future errors) from a model.
σ^2	Variance parameter or variance diagonal block-matrix.
${\sigma^2}_j$	Variance on regression j or triangle column j or variance diagonal block-matrix.
${\hat{\sigma}^2}_j$	Variance estimated parameter on triangle column j.
${\sigma^2}_{j,k}$	Variance on regression or triangle column j in cell k.
$\boldsymbol{\sigma}^2$	Variances vector.
Σ	MGLR errors variance and covariances.
Σ_{xx}	MGLR errors variance and covariances for claims reserving method xx.
$\boldsymbol{\Sigma}$	PMGLR errors variance and covariance.
Σ_{xx}	PMGLR errors variance and covariance for claims reserving method xx.
Σ^F	MGLR future errors variance and covariance.
$\boldsymbol{\Sigma}^F$	PMGLR future errors variance and covariance.
$\boldsymbol{\Psi}_{xx}$	PGLR errors variance-covariance matrix for claims reserving method xx.
$\boldsymbol{\Psi}$	PGLR errors variance-covariance matrix.
$\boldsymbol{\Psi}_F$	PMGLR future errors variance covariance matrix.
$\eta_{i,j}$	Linear predictor of cell i, j in GLM model.
$\mu_{i,j}$	Mean of cell i, j.
$\hat{\mu}_{i,j}$	Estimated mean of cell i, j.
$\mu_{i,LN}$	Location parameter from the lognormal distribution on origin year i.
θ	Location coefficient in GLM models.
ϕ	Disp1ersion coefficient in GLM models.
$\rho_{jj'}$	Coefficient of correlation between j and j'.
ω	Known constant.
γ	Power parameter of the claims variance on GLM models.

δ	Murphy (1994) errors variance power-parameter to define several claims reserving methods.
χ^2	Chi-square statistic.
$'$	First derivative. When used with matrices means the transpose of the matrix.
$''$	Second derivative.

1. Introduction

This first chapter summarizes the content of all the chapters of this thesis.

The second chapter will start with the presentation of the insurance business and the claims reserving problem. After that the insurers methodology for claims reserving is summarized and the main method in use by actuaries is presented, the Chain-Ladder (CL). We will mention CL´s main limitations and reasons for being the method most used in practice.

The third chapter presents the most common deterministic methods developed in the literature and the fourth chapter extends this analysis to stochastic methods. In both chapters, the relation of some of the methods with regression models is emphasized. In the end of the third chapter and as a transition to the fourth chapter, we present a summary of the use of regression techniques in claims reserving. The definition of prediction error (the square root of the mean square error of prediction) is also introduced here and formulas for confidence intervals are also shown. We also present some conclusions about, the deterministic and stochastic methods on the literature and the use of regression models.

In the fifth chapter, we will start by developing a stochastic method like the Mack (1993a, 1993b, 1994) method, but considering as loss development factor, the Vector Projection (VP), which is the equation parameter from a regression through the origin between two adjacent columns in the triangle. This VP method is based on the Mack mentioned CL framework and is also heteroscedastic. However, the VP considers the variance on payments proportional to the square of the payments (the weights of the VP link ratios). These weights are different from the ones of the CL (which are the payments, the weights of the CL link ratios). With this approach, we consider the VP as heteroscedastic, a feature shared with the CL. Formulas for the VP prediction error are also developed.

We also present a survey with 114 triangles, comparing VP and CL results when applied to these triangles. We will see that the VP has lower prediction errors in most triangles, when compared with the CL.

This chapter also presents several issues to be considered in method selection, as the errors analysis and the back-testing, and includes numerical examples with regular and irregular data from triangles used in the literature of claims reserving. It is also explained what the criterion was to classify a triangle as regular or irregular. Some conclusions are presented in respect of the CL and the VP results.

In chapter six, we will develop a regression framework that will be used to derive the Generalized Link Ratios (GLR) and the Multivariate Generalized Link Ratios (MGLR), the second and third methods from this thesis. It will also be useful to develop the portfolio data methods from chapter seven. The GLR and MGLR methods are presented in chapter six with their assumptions, parameters estimation and prediction errors formulas. The MGLR method is like the GLR method but considers the existence of contemporaneous correlations between the triangle equations. By contemporaneous correlations between equations, from the same triangle, we mean that the error terms are correlated at the same point in time. The same point in time in claims reserving triangles, in the context of regression models, means the same origin year.

Methods as the VP, CL and SA, are also presented as particular cases from the GLR, and correspond to a specific value from one of the parameters. The same is done using the MGLR, and cases will be obtained for the multivariate VP, CL and SA.

Numerical results are presented for regular and irregular triangles and a survey is also conducted with 114 triangles. Tests are performed to study, inside each triangle, the heteroscedasticity, the correlations between the equations and the serial correlation. Conclusions are presented in the end of chapter six, with emphasis on the GLR and MGLR methods' flexibility and the advantages of considering the equations contemporaneous correlations (at the MGLR method).

Chapter seven will present two methods for portfolio data. By portfolio data modelling, we mean several triangles estimated at the same time. This will be done using the GLR and the MGLR methods from chapter six. The Portfolio Generalized Link Ratios (PGLR) is the GLR applied to portfolio data. The Portfolio Multivariate Generalized Link Ratios (PMGLR) is the MGLR applied to portfolio data. In the PMGLR method we assume contemporaneous correlations between the triangles. By contemporaneous correlations between triangles, we mean that the triangle's error terms are correlated at the same point in time, which means the same origin year. In PMGLR, we will also assume contemporaneous correlations between the equations inside each triangle. Method's assumptions are defined, and parameters estimation is presented. Prediction error formulas are developed. A test on the use of pooled data is performed (test on the hypothesis of the loss development factors from each triangle being equal, when the development year is the same).

Chapter seven finishes with some conclusions about the use of aggregate data (the sum of all the triangles in one triangle) compared with portfolio data (triangles estimated together). We also analysed the results obtained with the use of the PGLR and the PMGLR methods.

Chapter eight will present the general conclusions taken from this thesis in respect to the CL and the alternative methods presented on chapters five, six and seven. We will emphasize the decrease of the prediction errors when the alternative methods are considered, and the absence of heteroscedasticity in most insurers' triangles. The existence of heteroscedasticity in irregular data triangles is not excluded.

The relation between the prediction errors and the two other method selection tools (errors analysis and back-testing) is also presented, namely the lower prediction errors when we have lower errors and stable results. The importance of some regression tests to help for method selection is also highlighted. We will also summarize the benefits of multivariate methods and portfolio data methods and the flexibility of the generalized link ratios approach, whatever its variant.

2. Claims Reserving in Insurance

This chapter is an introduction to the subject of this thesis and will allow the reader to understand even further the motivation beyond the development of a new approach to claims reserving.

First, we define insurance and its main features in respect to risk and uncertainty. The latter will be extended with an explanation of the claims process and the importance of reserves for insurance companies. We will also refer to what should be the insurer's level of reserves according to the current standards.

The methodology for estimating reserves with data in triangular format is presented and the main method used by actuaries is explained: the CL technique. We will also see here the main limitations of the method.

Having seen CL limitations, we will try to understand why actuaries are using it. For that, we will see the roots of claims reserving with triangle techniques and the CL method.

We finish this introduction by describing our motivation to introduce a flexible claims reserving approach that produces lower prediction errors.

2.1 The Insurance Business and the Claims Process

The Oxford Dictionary (2014) defines insurance as an arrangement by which a company or the state undertakes to provide a guarantee of compensations for specified loss, damage, illness, or death in return for payment of a specified premium. We may improve this statement with other technical sources, in accordance with some authors, for example Rejda (2005). There are several definitions of insurance. This may be shown when we compare the insurance definition done by other professions that deal with insurance.

Economists, Zweifel and Eisen, (2012), see insurance as the exchange of an uncertain loss of unknown magnitude for a small and known loss, the premium. Similarly, risk managers Vaughan and Vaughan (1995), define insurance as an economic device whereby the individual substitutes a small certain cost, the premium, for a large uncertain financial loss, the contingency insured against, that would exist if there is no insurance. Legislators consider several contractual aspects of insurance, but in many jurisdictions, we do not find a definition of insurance. Instead, it is accepted, since many years, that there is no insurance without risk and the latter must be managed by insurers who use statistical laws with appropriate techniques (Moitinho de Almeida, 1971).

Finally, actuaries, Bowers and Nesbitt (1986), state that insurance is a mechanism for reducing adverse financial impact of random events that prevent the fulfilment of reasonable expectations.

Putting all these definitions together we find some common features:

- The insured pay a certain insurance premium in advance with the start of the contract, to cover a future random risky event.
- As the event is not certain, we do not know if it is going to happen.
- And if it happens we do not know when it does and how much it will cost.

This means that the insurance sector has an inverted production cycle:

- The insurers receive the premium in advance, and just posteriori they will know if they need to pay something due to that.
- It is not possible to know priori if this premium is enough to cover claims and expenses, because of the existence of risk and uncertainty on all this process.

However, as we are going to see, the claims process also brings some extra source of randomness to this.

A claim is a demand for compensation from an insured entity or a third party. it is the most visible side of insurance and should correspond to what is reasonable for the insured to expect, as compensation.

Following the Institute of Actuaries (1989, 1999) we may see that a claim has several phases:

- With the policy inception, an insured or a third party becomes eligible to claim against the insurer if the event and its notification are according to the contract wording. The latter gives us information about the claims included and excluded on policy cover after a certain date, and the amount to be paid from the loss.

- The wording may have some waiting period within which the insurer is not liable for any payment. This means that a second date must be considered, the end of the waiting period.
- After this period, if it exists, there may be an event, for example an accident that in the beginning is only known by the insured. This means that there is information asymmetry as the insured knows more about the claim than the insurer. But even the policyholder may not have all the information after the claim. Indeed, it may take some time to recognize that an event arose, for example, a health problem that was the consequence of that accident that happened some days, months or even years ago.
- With some delay after the claim occurrence the insured will report the accident to the insurance company (directly or maybe through an agent or a broker). The insurer will create a reserve and eventually will start making some payments, as soon it has the necessary information and conditions for that.
- After a certain period, all the payments will be made, and the claim is settled and closed. According to the information available, the insured decides to close the claim because he does not expect to pay or receive anything else.
- However, it might happen that the claim needs to be re-opened some time after, or because there are new payments to be made or due to the need to register some reimbursements (negative payments or amounts that are received by the insurer). Both cases represent new information that arrived and that was not available before.
- In theory the claim is closed after some time, but it is theoretically possible that the claim will be reopened again due to new demands from the policyholder, a beneficiary, a provider of services or a third party.

When a claim is opened, the insurer has an expectation about what its total cost might be. It is the initial cost of the claim that may be adjusted in the future according with the new information that will come, such as, the degree of severity of the injured people. However, what matters to the insurer is to know the ultimate cost, the one that will arise after the claim has been completely closed.

To know the claims ultimate cost, we need to go over all these phases, and just by coincidence the initial cost will be equal to the ultimate cost. This happens because there are time gaps between all the phases on claims settling: the occurrence, the recognition, the notification, the initial valuation, payments, the reimbursements, the eventual reopening and its final valuation. These gaps will depend on several factors, but mainly on the line of business, the insurer´s claims policy and the staff´s technical capabilities to anticipate the ultimate cost.

For death cover, of life insurance and personal accident insurance, it is easy to define the amount to be paid: it will be the sum insured. This type of liability should be completely settled very fast. However, in some circumstances the insurers will decline the payment, e.g. the death was by suicide or self-injured action and the policy does not consider it as an eligible claim before a certain period. In these cases, the insurer may define the reserve as zero but later, due to a court action or some lawyer intervention, he may realize that he will be obliged to pay the sum insured. These cases should not be an important percentage of the total reserves.

On disability cover, such as life insurance and personal accident, the valuation is sometimes more difficult due to the extra need to recognize, define and agree upon the percentage of disability. The scope for disagreement and discussion is very large, and sometimes some litigation arises, which postpones the definition of the ultimate cost.

On property cover, like homeowner's insurance, fire, and business interruption insurance, some covers should be easier to evaluate because we have a sum insured previously defined. However, there are several factors that might complicate the valuation:

- The sum insured may be higher or lower than the true economic value of the building´s reconstruction cost or of its content, which may bring some discussion. The insurer will want to restrict the payment to the true economic loss and may want to apply the average clause (a proportional rule that restricts the claim payment to the claim value adjusted by a percentage, of the sum insured in respect to the sum that should have been insured, usually larger than the former).

- On some claims, the time to arrive to the ultimate cost may be longer, due to investigations and litigations that may arise. Court decisions may be more difficult to anticipate when there is a fraud but it is difficult to prove it.

Some covers, like crop insurance, are even more difficult to settle and to reserve. Indeed, some of the affected goods may recover from the claim damages until the harvest. This means that a recovery percentage must also be estimated to define the true claim. The latter must be agreed with the farmer and some discussion may arise, which may end with litigation.

On liability covers of insurance like motor, marine, aviation, land transports and general liability insurance, the amount of the claim is much harder to define. It will depend on a subjective valuation and on a negotiation with a third party. In many circumstances, it will be defined in court, with several subjective factors and sometimes with the influence of the public opinion. This also means that it might take several years to get the ultimate cost. The more complicated the cases are, the longer the time will be for the courts to decide. Two examples show this: The courts need to define if there is a liability covered by the insurance contract (also called the policy), that is, the liability of the insured may exist, but the policy may exclude those events from the cover. For example, a product liability from a policy may exclude events in United States of America. Finally, in case of liability and policy cover, the courts also need to confirm the losses to be paid by the insurer. Some of these losses are personal moral damages, something which is hard to value and highly discussable. An example of these damages is the suffering of a family due to the loss of life of a son.

Other lines of business, such as credit insurance, where reimbursements are very important, are also hard to quantify. The insurer pays the outstanding debt for its policyholder but will have subrogation rights against the debtor. Insurers usually take some years to recover this money but that will reduce the ultimate cost, sometimes significantly. The liabilities should not take too much time to value (it will be a percentage of the debt), but it is more difficult to know how many reimbursements will arrive in the following years. The latter may take many years. The macroeconomic condition will also determinate how much the insurer will be able to reimburse. For example, when the economy performs very well, it is more likely that the debtors who produced a claim to an insurer in the past will be more able to pay their debts in the future and the insurer will be reimbursed.

While the claims payments are not totally finalised insurers need to have a reserve for those amounts and they have several ways to estimate it:

- Doing a subjective valuation of the claim, using its experts.
- Applying an average cost to each claim. This value might come from an actuarial analysis of the previous year's average costs or by some subjective valuation.
- Or mixing both previous approaches. One common standard is to use the average cost during a certain period and then moving to a subjective case reserve if the claim is still outstanding after that period.

This reserve is called the case reserve because it is on the file. It corresponds to claims that had been notified to the insurer, but which are not yet settled. These reserves are defined by the insurer claims department on a per claim basis or using an average cost system. Sometimes, the case reserve is also called outstanding claims reserve. However, the outstanding claims reserve very often includes the case reserve and other accounting reserves not allocated to the files. In this thesis, we will follow this last definition.

Whatever the approach, there is always a scope for differences between the initial cost estimation and the ultimate cost:

- The claims have already occurred, but the notification did not arrive to the insurance company. This means that there is no information on the claim, but the claim already exists.
- Sometimes an insurer is notified of a claim but there may be a delay before it is included in the information system of the company. This is a hidden cost, more important in valuations done during the year. Indeed, at the end of the year, insurers try to register all the claims in the information system.
- The claims may be reserved but new information on the claims might come up that may oblige to change the reserve, increasing or decreasing the ultimate cost estimate. For example, the new information shows that the severity of the claim is worse than initially expected.
- The claims may be reopened, which probably implies extra payments or reimbursements. This happens because the insurers do not usually anticipate completely

that possibility. The ultimate cost will be higher or lower depending on the reason for the claim reopening.

All these cases oblige some sort of estimation to get an approximation of the ultimate cost from the claims. That may be done for each individual claim or for a group of claims. This means that when insurers receive a claim notification they must calculate a reserve for that. The latter added to any eventual payments done, gives us the incurred claims, but we may have a long way to arrive to the final ultimate cost. At the same time, the insurer also needs to reserve for claims that were not yet notified but have already occurred.

To anticipate all these situations the insurer needs to have case reserves and two additional reserves:

- The Incurred But Not Reported (IBNR) reserve, that will be used to cover two types of claims: those that have occurred but have not been notified to the insurer (the pure IBNR) and claims that were notified but not introduced in the insurer´s information system, the Reported But Not Registered claims (RBNR).
- And, the Incurred But Not Enough Reported (IBNER) reserve, that will be used to cover new valuations on claims and the reopening of some claims.

Usually these reserves, IBNR and IBNER, are not in the claims file but in the company's accounts. This is done to have an actuarial estimate of the correct reserves, a need for the management and a requirement from regulation and accounting rules. It may also happen that an insurer creates a virtual claim to account for some IBNR and IBNER reserves. In these situations, a claim that never occurred is created on the insurer´s information system and a reserve is created. This reserve will be the IBNR and IBNER reserve of all the claims. However, this is not the best management approach. This means that the outstanding claims reserves that we consider in this thesis corresponds to the sum of the case reserves with the IBNR and IBNER reserves registered by the company on the accounts.

The insurance business has an inverted production cycle: insurers receive the premiums from policyholders but at that time they do not know how much costs they are going to incur with those customers. The problem is that the claims process might increase the risk and uncertainty of this feature: the insurers need to estimate the ultimate costs which are different from the initial estimated costs.

Due to this inverted production cycle, several risks may be generated for the insurer:

- An underwriting risk: The amount collected from the insured, the premium, may not be enough to pay the claims and the expenses, which will bring a loss. The calculation of the premiums depends on the claims data, which includes the level of reserves of those claims.
- A performance risk: If the reserves are much higher than needed, the costs will seem higher, and this may oblige the insurer to have higher premiums to cover them, which will produce lower sales.
- A tax risk: If we underestimate the reserves, the profits will be higher and the same will happen with the taxes. This means that we may pay taxes on profits that never existed.
- A reputational risk: The companies that underestimate the reserves may create suspicions in the market about their management.
- A reserve risk: There are IBNRs and IBNERs that need to be estimated to have an estimated ultimate cost, and there is some uncertainty on the calculations.
- And, a solvency risk: The lack of good management of all the risks may produce losses and the deterioration of the insurer capital.

2.2 Reserving Importance

As we saw in section 2.1, there are differences between claims initial cost and the claims ultimate cost. This means that reserving is a critical issue in insurance. This conclusion includes non-life insurance (also called general insurance or property and casualty insurance) and life insurance (mainly covers for death and disability).

As such differences, between initial costs and ultimate costs, are lower in life insurance, we may conclude that claims reserving is more critical in non-life insurance. Market figures also confirm that.

Stakeholders at an insurance company are the entities, individuals or not, with an interest on the company. The Chief Risk Officer is supposed to check that the risk implicit on the insurer´s strategy is compatible with the interests of all the stakeholders of the company. One of the risks is the insurer's reserves.

There are several stakeholders in an insurance company and all of them with an interest on the insurer's reserves. For example:

- Regulators that need to know insurer´s solvency and financial strength to fulfil its mission.
- Shareholders and managers who have a concern on the level of results.
- Internal actuaries that must be sure that reserves reported are properly calculated, following the professional guidance and the legislation.
- Financial analysts that consider financial statements for their analysis.
- Bondholders, who invest in the company giving it a loan, need to know if the company is solvent before making the investment.
- Derivatives buyers also need to understand the insurer´s financial strength to be sure that the insurer will be able to pay any future obligations.
- Reinsurers that need to understand how the insurer manages the company to renegotiate the reinsurance treaties and to anticipate their share of the claims.
- Actuaries, that need to certify the reserves and to calculate several actuarial figures that depend on incurred costs, e.g. the price of each product.
- Auditors, that needs to certify the accounts. The latter are heavily dependent on the figures of claims reserves.
- Board members, that must manage and define a strategy for the company.
- Employees, Channels of Distribution and External Providers of goods and services, that want to know if the company has good financial strength to honour its obligations, salaries, benefits, commissions, claims paid, and goods and services acquired.

With so many stakeholders involved it seems clear that the reserves play a critical role in insurance.

2.3 The Reserves Level

During many years (and even today but to a less extent) there was some lack of harmonization between countries, on the claims reserves that insurers should have on their accounts. That was recognized by a European Community document (European Community, 1999) that established, as a priority, the harmonization of the insurer's technical reserves calculation on the framework of a new solvency regime, Solvency II.

That regime was defined by a European Union directive (European Union, 2009) that defined as the required level of reserves, the fair value; that means the amount that would allow the transfer of this liability to another insurer or reinsurer on an arms-length transaction between two willing parties. The directive has been in force since the 1st of January 2016, but it has already influenced the claims reserving strategy in several countries for some time.

In several countries, such as the United States and the United Kingdom, there were already similar risk-based capital systems in force for some years, requiring insurers, as what happens currently in the European Union, to regularly perform an Own Risk and Solvency Assessment (ORSA). This obliges insurance companies to issue their own assessment of their current and future risk through an internal risk self-assessment process, and it allows regulators to form an enhanced view of an insurer ability to withstand financial stress. The assessment also includes the claims reserves.

In this European Union directive, it is defined that the fair value of the reserve has two components:

- The best estimate that corresponds to expected value (the mean) of all the future cash-flows.
- And, a risk margin that allows for the inherent fluctuation of the best estimate.

A term structure of risk-free interest rates should be applied to discount all these cash-flows, from the best estimate and from the risk margin. The term structure of risk-free interest rates corresponds to a set of interest rates for the different maturities and in the same currency of the liabilities. This means that in the Solvency II balance sheet we will have discounted best estimates for the reserves. Just in the IFRS (International Financial Reporting Standard) balance sheet the best estimates are not discounted. This means that currently insurers have two balance sheets according with the compulsory reporting rules: the IFRS and the Solvency II.

The risk margin corresponds to the additional value of the best estimate to arrive to the liability fair value. However, the legislator decided to follow a non-actuarial approach to its calculation, using the cost of capital concept. Accordingly, with the so-called standard method of the Solvency II regime, insurers must use the method of the cost of capital to calculate this risk margin. This methodology calculates the risk margin as the cost of the capital necessary to cover the best estimate volatility. The latter corresponds to 6% of the sum of all the future capital requirements, discounted by the risk-free interest rate. The future capital requirements are the capital at risk in the future years (mainly underwriting risk, credit risk and operational risk and non-diversifiable market risk) calculated until the maturity of all the liabilities in accordance with the Solvency II rules. Some proxies are also allowed to simplify these calculations, for more details see European Union (2015).

If the regulator approves an internal model specific for the insurer, to avoid the standard model which is equal to all the insurers, it is possible to calculate this risk margin using stochastic methods, the actuarial approach to arrive to the fair value. However, this does not mean that stochastic claims reserving is just useful for internal models. Indeed, they are fundamental for the best estimate calculations of standard and internal models, because they provide for both a very important indicator, the prediction error (the square root of the mean square error of prediction).

The current solution in Europe for claims reserving is not very different from the methodologies in force until the start of the Solvency II regime. Before this date most companies registered in their accounts the best estimate without discounting it (with an interest rate). Some companies calculated the risk margin using actuarial stochastic methods but just for management purposes, most of them were not considering this risk margin on their accounting reserves.

Whatever the approach and system, it is fundamental to calculate an appropriate central estimate of the reserves, the best estimate when we apply discounting to the projected cash-flows. It is important because:

- It is the reserves expected value and the value that insurers should have in their accounts under the current solvency system.

- Its value will impact the calculation of the fair value. The latter will be a best estimate plus the risk margin. If the best estimate underestimates or overestimates the true ultimate cost, the fair value may also underestimate or overestimate the true fair value.
- And finally, because it will also impact the calculation of the Solvency II capital requirements: the market risk, the underwriting risks, the credit risk and the operational risk depend on the claims reserves level.

The legislation on risk-based capital and solvency, in Europe, the United States, and other countries (such as the United Kingdom and Australia among the others), not only obliges insurers to assess all the liabilities from all lines of business, but also requires better estimation for the total reserves on the entire portfolio. Practically, this means that it is desirable to have estimates of reserves with as low prediction error as possible.

2.4 Methodology to get the Best Estimate

The techniques most in use to get best estimates (and risk margins) aggregate data on homogeneous groups of claims to produce a triangle of past information. Data in this triangle is used to estimate another triangle with the estimates of the future evolution of claims.

The first year of information from this triangle must be closed, that is, its ultimate cost must be known. If that does not happen the actuary is obliged to consider a tail factor that allows the close of the year. For example, if the last cumulative payments are 1000 and the tail factor is 1.10, the first-year ultimate cost will be 1100. The calculation of the tail factor may be done subjectively or objectively with the use of a logarithmic function or with the application of several types of smoothers. See for example, in this respect, Booth et al. (2005).

The idea of the data in triangle format is to aggregate all the information on claims on a table were the rows are the origin year and the columns are the development years. We will lose information on each claim, but we will have an overall view about all the claims.

By homogenous group we mean similar cover with claims that behave in a similar way in respect to, notification and settling features: the liability nature (if related or not to the inflation), the possibility of reimbursements or reopening and the claim duration.

For example, it should be convenient to split motor insurance claims between material damages and bodily injury claims. Comparing with the former, the latter takes more time to be reported to the insurer, depends more on claims inflation (sometimes the courts inflation, which may be higher than the country´s inflation), and takes more time to develop. Sometimes, we are not able to create these homogeneous groups like we would like, due to the triangle´s low number of claims, which may produce unstable projections. In those cases, it may be better to work together, in the same triangle, material damages and bodily injury claims or to use other subjective techniques.

The triangle content may be the cumulative payments or the incurred claims. The payments are the amounts paid to the insured and to third parties that received compensation for the claim, and to the providers that participate on the claims settling (hospitals, doctors, lawyers, experts, loss adjusters and courts). These amounts are directly allocated to each of the claims. The incurred claims are the sum of the claims cumulative payments with the claims case reserves (the IBNR and the IBNER accounting reserves should not be considered in the triangles).

It is also possible to create triangles to estimate the number of open claims, the number of closed claims, the number of claims outstanding or even the reimbursements evolution. We may also use incremental payments instead of cumulative payments and indeed some methods require this. Some practitioners argue that the use of cumulative payments makes estimates more dependent on early experience (Booth et al., 2005). An example with cumulative payments could be the following Table 2.1, used by several authors in the literature of claims reserving, e.g. (Mack 1993a, 1993b and 1994).

Table 2.1: Triangle of Cumulative Payments, Mack (1993a, 1993b, 1994)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	5012	8269	10907	11805	13539	16181	18009	18608	18662	18834
2	106	4285	5396	10666	13782	15599	15496	16169	16704	
3	3410	8992	13873	16141	18735	22214	22863	23466		
4	5655	11555	15766	21266	23425	26083	27067			
5	1092	9565	15836	22169	25955	26180				
6	1513	6445	11702	12935	15852					
7	557	4020	10946	12314						
8	1351	6947	13112							
9	3133	5395								
10	2063									

The cumulative payments $C_{i,j}$, where each row i represents an origin year with $i = 1, \dots, T$, and each column j gives us the development year with $j = 1, \dots, T$. In this example, $T = 10$. The origin year in this example is an origin year but other criteria may be used in other cases, like the underwriting year or the notification year of the claim.

This upper triangle D_u represents the past history of claims, with $D_u = \{C_{i,j}: i + j - 1 \leq T\}$ and the technique assumes that we may use it to forecast the future and estimate the lower triangle D_l given by $D_l = \{C_{i,j}: i + j - 1 > T\}$. Putting D_l together with D_u we will get a matrix that joins together the two triangles. The last column of the matrix gives us the ultimate costs of each origin year.

Each diagonal represents the calendar year $i + j - 1$. The content of each cell $C_{i,j}$ represents the cumulative payments made so far for on the cell of origin year i and development year j.

We are assuming, for simplification, that the first origin year is closed and that no more claims or payments (including reimbursements) will arise in the future.

The objective is to have the right level of reserves in the insurer´s balance sheet, the claim's reserves best estimates. To have these best estimates, we will estimate the last column of the triangle to get the ultimate costs per origin year. Subtracting to these ultimate costs the cumulative payments done so far (the last diagonal of the triangle), we get the estimated claim's reserves. These estimated reserves will be the best estimate reserves if they are the

reserves expected value. This means that the model that produced them should match the insurer´s experience and is probably the one with the lowest prediction error.
If the best estimate reserves are higher than the ones that the company already has on the case reserves (on the files), an accounting reserve should exist for IBNR and IBNER that fills the gap. In that case the best estimate reserve will be equal to the sum of the case reserves and the IBNR and IBNER reserves.

When the triangle consists of incurred claims, which means the cumulative payments plus the case reserves, the procedure is the same but with a different interpretation. We also estimate the ultimate costs per origin year. But when we subtract from them the last diagonal of the triangle we get the emerging reserves, and not the best estimate reserves. The emerging reserves, positive or negative, will be the variation we need to have on the current level of case reserves to obtain the claim's reserves best estimates. This means that the reserves best estimates will be the sum of the current level of case reserves with the emerging reserves.

To have these best estimates, we need to estimate the lower triangle, $D_l = \{C_{i,j}: i + j - 1 > T\}$. One of the most used methods for that is with the calculation of the loss development factors. This will allow us to estimate each cell, on the lower triangle, as the product of this factor by the previous cell value in the same row. Before that, it is useful to see the link ratios (also called age-to-age factors) between adjacent cells on the upper triangle.

For that we use our table 2.1 with cumulative payments. Using this matrix, we can calculate the link ratios $F_{i,j+1}$ between two adjacent cells in the same row.

$$F_{i,j+1} = \frac{C_{i,j+1}}{C_{i,j}} \tag{2.1}$$

We get in our example from Table 2.1 the link ratios presented in the Table 2.2

Table 2.2: Link Ratios arising from Table 2.1

	1	2	3	4	5	6	7	8	9
1	1,650	1,319	1,082	1,147	1,195	1,113	1,033	1,003	1,009
2	40,425	1,259	1,977	1,292	1,132	0,993	1,043	1,033	
3	2,637	1,543	1,163	1,161	1,186	1,029	1,026		
4	2,043	1,364	1,349	1,102	1,113	1,038			
5	8,759	1,656	1,400	1,171	1,009				
6	4,260	1,816	1,105	1,226					
7	7,217	2,723	1,125						
8	5,142	1,887							
9	1,722								

Having this triangle, we need, for each column, to summarize all the link ratios on a single number, the loss development factor. The latter will allow us to estimate a cell on the lower triangle as the product, of the previous cell in the same row by the loss development factor $\hat{b}_j$ with $j = 1, \dots, T$. The loss development factor is a statistic that summarizes, for each development year (each column from the triangle), the link ratios. To get them, we need a statistical method, such as the CL. The CL summarizes all the link ratios from one development year in one loss development factor. The CL calculates the weighted average of all the link ratios from that column, using the payments on each cell as weights.

As we assumed that the first year is closed, we just need to estimate *T-1* loss development factors. Hence, $\hat{b}_{10} = 1$.

The loss development factor *j* represents, for any origin year, the evolution between two development years, *j* and *j+1*. This means that the technique considers this evolution at column *j* as the same, whatever the origin year is.

With these loss development factors, we may estimate each cell of the lower triangle using the following relations

$$\begin{aligned} \hat{C}_{i,j} &= \hat{C}_{i,j-1}\hat{b}_{j-1} \quad j > T - i + 2 \\ \hat{C}_{i,j} &= C_{i,T-i+1}\hat{b}_{j-1} \quad j = T - i + 2 \end{aligned} \qquad (2.2)$$

We may also calculate the ultimate factor (also called ultimate loss development factor or age to ultimate factor). The ultimate factor $\hat{f}_i$, a figure that is multiplied by the cumulative payments of each origin year, the last diagonal of our triangle, gives us the ultimate cost. The ultimate factor, for each row i, will be the product of all the loss development factors that give the evolution from column $j = T - i + 1$ until column T.

$$\hat{f}_i = \prod_{j=T-i+1}^{T} \hat{b}_j \qquad (2.3)$$

Having the ultimate factor and the last diagonal of the upper triangle with the cumulative payments, we have the estimated ultimate cost, $\hat{C}_{i,T}$ for all the origin years given by

$$\hat{C}_{i,T} = C_{i,T-i+1}\hat{f}_i \qquad (2.4)$$

With the estimated ultimate cost and the cumulative payments, in the last diagonal, we may immediately get the estimated claims reserve for each origin year, $\hat{R}_i$

$$\hat{R}_i = C_{i,T-i+1}(\hat{f}_i - 1) \tag{2.5}$$

In (2.5), we may see that we need to have a yardstick that allows us to summarize a set of link ratios on a loss development factor to get an ultimate factor. Having this, we will have the future payments and then we will get the insurer claims reserves. The yardstick will vary depending on the method applied and the decision on that is not always straightforward. According to Brown (1993):

"Setting loss reserves is not a job of a technician, but of a professional actuary. We cannot enter data into a computer software package, press a button, and accept the reserve estimate that results. A considerable degree of judgement is required".

The same author refers that the actuary should be able to use various methods and to reconcile and explain the differences. In the same line we may find other references, like Wüthrich and Merz (2008) that state that:

"Only an experienced reserving actuary is able to tell us which is an accurate/good estimate for future liabilities for a specific data set, and which method applies to which data set".

Despite all this, it is very common to find that many people always use the same method, the CL, or methods that are CL based: such as the deterministic CL, the Bornhuetter-Ferguson (BF) with the CL, the Stochastic CL (non-parametric or parametric), and the Bootstrap CL (also CL based). Sometimes, it is mentioned that the Bootstrap is not a claims reserving method. Indeed, we may apply the Bootstrap technique to any method but it is very common to see in the literature the reference to the Bootstrap as a claims reserving method, see for example Hindley (2018).

Sometimes, the CL is also used in decomposing the cumulative payments triangles, or the incurred claims triangle on frequency and severity triangles. We will have two triangles instead of one triangle. If we use payments, the severity triangle is the average payments (payments divided by the number of claims with payments) and the frequency triangle is the number of claims with payments. When we use the incurred claims, the severity is given by the average incurred claims (the incurred claims divided by the number of claims notified) and the frequency is the number of notified claims. In this frequency-severity approach actuaries also apply the CL to all these triangles.

We may confirm this generalized use of the CL by looking at several statements on the practical and theoretical literature. For instance, in the 2013 discussions about these techniques at the Institute and Faculty of Actuaries (Marcuson, 2013) it is explicitly written:
"…there is a reason why established techniques such as the CL and BF are so well-entrenched in actuarial reserving…it is because they are robust (certainly the BF, but with suitable care the CL as well), common-sense approach to a problem. They apply to aggregate data, which means we can overcome some data deficiencies, and, most importantly, they are relatively easy to communicate to non-actuaries".

Also, some theoretical books (Wüthrich and Merz, 2008) go in the same direction stating:
"The CL and BF methods belong to the easiest claims reserving methods. Their simplicity makes the CL and the BF methods the most commonly used techniques in practice. Though they are simple, they often give surprisingly accurate results".

And (Straub, 1988):
"The oldest IBNR method and by the large still the most often used one is a straightforward extrapolation called the CL method".

From these statements it seems one of two things:

- Either the CL method is so flexible that it may be adjusted to any line of business and to any set of data.
- Or a method is being imposed to the data when it should be the opposite, that is, the data should oblige to a specific method that better matches it and the professional reserving actuary should be able to choose the most appropriate one.

Before choosing the appropriate conclusion, we will first look at the CL method and the assumptions beyond it.

2.5 The Chain-Ladder Estimation

This CL method estimates the loss development factors $\hat{b}_j^{CL}$ as the ratio of the sum of two adjacent columns (using just the common origin years to get the same number of cells on the numerator and denominator). We will get the following estimator for $j = 1, \dots, T-1$

$$\hat{b}_j^{CL} = \frac{\sum_{i=1}^{T-j} C_{i,j+1}}{\sum_{i=1}^{T-j} C_{i,j}} \tag{2.6}$$

Using (2.1) this is a weighted average of the triangle link ratios, where the weights are the payments from year j.

$$\hat{b}_j^{CL} = \frac{\sum_{i=1}^{T-j} C_{i,j} F_{i,j+1}}{\sum_{i=1}^{T-j} C_{i,j}} \tag{2.7}$$

There is an important assumption when we use this method. According to the Institute of Actuaries exam's manuals on General Insurance (Acted, 2000):

"The key assumption is that, for each origin year, the expected amount of claims, in monetary terms, paid in each development year is a constant proportion of the total claims, in monetary terms, from that origin year."

We may see this assumption defining the grossing-up factors $g_{i,j}$ the amount of payments on cell with row i and column j in proportion to the ultimate costs. For the lower triangle cells, $D_l = \{C_{i,j} : i + j - 1 > T\}$, this factor, $\hat{g}_{i,j}$, is estimated by the ratio of $\hat{C}_{i,j}$, expected payments on a cell, for $\hat{C}_{i,T}$, the expected ultimate cost.

$$\hat{g}_{i,j} = \frac{\hat{C}_{i,j}}{\hat{C}_{i,T}} \quad \text{with} \quad \hat{C}_{1,T} = C_{1,T} \tag{2.8}$$

Hence for the lower triangle D_l we have

$$\hat{g}_{i,j} = \frac{\hat{C}_{i,j}}{\hat{C}_{i,j} \hat{f}_j} \tag{2.9}$$

This means that we get a relation between the grossing-up factor per development year and the estimated ultimate factor.

$$\hat{g}_j = \frac{1}{\hat{f}_j} \tag{2.10}$$

We conclude that if the j is the same, whatever the origin year i, the proportion $\hat{g}_j$ will have the same value. The following table illustrates this situation using data on table 1 with the CL method to project the lower triangle. We get the same proportions of cumulative payments in respect to the ultimate cost in the lower triangle columns. Also, these proportions, in each column, are the same as the ones obtained, on that column, for the last diagonal. The Table 2.3 presents all these results. These results may also be obtained if we consider incremental payments instead of cumulative payments on the calculation of these proportions.

For the upper triangle cells $D_u = \{C_{i,j}: i + j - 1 \leq T\}$, we may obtain these grossing-up factors using the same framework but considering $\hat{g}_{i,j}$, as the ratio of $C_{i,j}$ the payments on a cell, over $\hat{C}_{i,T}$, the expected ultimate cost.

Table 2.3: Chain-Ladder Grossing-Up Factors

	1	2	3	4	5	6	7	8	9	10
1	27%	44%	58%	63%	72%	86%	96%	99%	99%	100%
2	1%	25%	32%	63%	82%	93%	92%	96%	99%	100%
3	14%	37%	58%	67%	78%	92%	95%	97%	99%	100%
4	20%	40%	55%	74%	82%	91%	94%	97%	99%	100%
5	4%	33%	55%	77%	90%	91%	94%	97%	99%	100%
6	8%	33%	60%	66%	81%	91%	94%	97%	99%	100%
7	3%	23%	62%	69%	81%	91%	94%	97%	99%	100%
8	6%	29%	55%	69%	81%	91%	94%	97%	99%	100%
9	20%	34%	55%	69%	81%	91%	94%	97%	99%	100%
10	11%	34%	55%	69%	81%	91%	94%	97%	99%	100%

However, in our opinion, this is not the main assumption of the CL method. The latter assumes that the triangle of payments is stable over time. Let us see this with another example.

Assume that we have a steady growth of claims every year with a constant development factor, but with some diagonal effects that are cumulative year over year (e.g., claims inflation), see Table 2.4. In this table, we follow the same data triangle format of the previous examples, where the rows are the origin years and the columns are the development years.

We get a triangle where the cumulative payments for the same development year, increase every origin year.

Table 2.4: Triangle with a Trend on Payments Increase

	1	2	3	4	5	6	7	8	9	10
1	1000	1980	4039	8402	17643	37051	77066	157214	311283	591438
2	1100	2376	5251	11762	26465	59281	131012	282985	591438	
3	1320	3089	7351	17643	42344	100778	235821	537671		
4	1716	4324	11027	28229	71984	181401	448059			
5	2402	6486	17643	47990	129572	344661				
6	3603	10378	29993	86381	246187					
7	5766	17643	53988	164124						
8	9802	31758	102578							
9	17643	60340								
10	20063									

Then, we get the link ratios presented on Table 2.5.

Table 2.5: Link Ratios obtained from (2.1) for Table 2.4

	1	2	3	4	5	6	7	8	9
1	1,98	2,04	2,08	2,10	2,10	2,08	2,04	1,98	1,90
2	2,16	2,21	2,24	2,25	2,24	2,21	2,16	2,09	
3	2,34	2,38	2,40	2,40	2,38	2,34	2,28		
4	2,52	2,55	2,56	2,55	2,52	2,47			
5	2,70	2,72	2,72	2,70	2,66				
6	2,88	2,89	2,88	2,85					
7	3,06	3,06	3,04						
8	3,24	3,23							
9	3,42								

It is clear from Table 2.5 that there is no stability, and trends appear on all the columns. Thus, the CL method estimates 3.12 for the first development factor. As we can see this estimate is more according with the past and does not reflect the current evolution (the same is happening to the other loss development factors). Now, if there is no reason to believe that the factor tends to the past average, it is difficult to sustain the use of CL. This also means that if the insurer increases (or decreases) their claims payments velocity, then more (or less) reserve is estimated by CL method. Obviously, it is exactly the opposite of what we should expect, and it is difficult to trust in CL estimated reserves. This situation happens in practice very often, and it is due to several reasons, such as the speed of paying and settling the claims, the changes in underwriting and claims policies. All of this obliges the consideration of the most recent experience, and not so much the one from the past.

It is known, from Straub (1988), that the CL is just an approximation to the least square solution which means that its loss development factors does not minimize the square of the errors. As the CL is a regression, it is not the best fit to the data because it does not minimize the square of the errors. Indeed, we know from Mack (1993a) that the CL does not minimize the sum of the square of the errors but minimizes the weighted sum of the square of the errors, which means that assumes errors as heteroscedastic (with non-constant variance). With regression techniques heteroscedasticity is a feature from cross-section models, see for example Fomby et al. (1984). This means models with the data coming from the same period (with triangle claims, the same origin year and the same development year) and from several entities (for example, more than one line of business). It may also mean data from several origin years (to the same development year) and from more than one line of business: these are called panel data models (multivariate models in claims reserving literature), which mix time-series and cross-section models. Due to this last feature, panel data models may also show heteroscedasticity if the model parameters are assumed the same for all the entities (the same loss development factors for all the triangles in the multivariate claims reserving models). However, when we use regression techniques, insurer's triangles are time-series equations per development year with data coming from one line of business (they are not cross-section or panel data models).

Due to this, when we analyse one triangle, we should not expect to find, in most of the cases, heteroscedasticity. It is possible that heteroscedasticity arises when data is irregular, as in such case it will be more difficult to predict, and the variance of the error is probably not going to be constant. The same may happen if we consider several lines of business with the same development factors, but that is not a common procedure in multivariate claims reserving literature, see for example Zhang (2010).

The CL will consider the future a weighted average of the past, but the weights are the payments. Straub (1988) showed that the best weight to minimize the sum of the square of the errors is the square of the payments.

In the example provided, where the link ratios increase with the origin year, it is easy to see that if the weights of the link ratios where the square of the payments, instead of the payments, the loss development factor would have been higher than 3.12 and closer to the most recent years.

It is also very common on insurers data, in triangle format, to have the payments increasing with the origin years, not only due to the existence of claims inflation but mainly due to the growth of the insurers business, which brings more claims. In those cases, the CL will be slower to adjust to the more recent years because of the use of the payments as the weight of the link ratios. When we use the square of the payments, in those cases, the loss development factor relies more on the more recent year's link ratios. Indeed, it is not only the most recent years that matter to claims reserving but usually they are very important, as they are closer to the way the insurer is reserving.

This CL bias is known by actuaries, and Halliwell (2007) writes *"over the past twenty years many actuaries have claimed and argued that the CL method of loss reserving is biased…nearly everyone who acknowledges this bias believes it to be upward".* It is also interesting to see that the same author writes that "*to resolve this issue (the bias) basic regression theory will suffice, specifically the much-misunderstood concept of regression toward the mean".*

The CL key assumption it is very strong and can only work with very stable data. If data does not have this feature, the prediction errors could be very high. If prediction errors are high, CL predictions do not match the experience and it is difficult to accept CL results. The reserve best estimate is not trustable and the same will happen to the fair value reserve. The latter depends on the best estimate, and even the risk margin may be calculated using the results from the best estimate. As we saw before, the bias with best estimates and risk margins has a tremendous impact on several insurer's issues, and we cannot rely anymore on its financial statements.

Other authors, such as Barnett and Zehnwirth (2000), say something similar in respect to methods that use the link ratios (where the CL is a special case, the most applied):
"Most loss arrays do not satisfy the assumptions of standard link ratios techniques." And if this happens the prediction errors will be high, and we cannot say that we have the best

estimate of our reserves. If this is the case, why is CL being used and surviving for so many years? We need to see something about its history to understand this remarkable issue.

2.6 Chain-Ladder History

Some authors, for example Straub (1988), refer to the CL as being the oldest method in claims reserving but we cannot find an official reference for the "birth" of the method.

In Tarbell's (1934), the author did not use triangles but stated that this problem of claims reserving was essentially actuarial or statistical; it should be the experience of the immediate past to guide the calculation of the IBNR reserve. Brosius (1992) also mention loss reserving methods dating back to the 50's. We know also from Masterson (1962) that in the 60's the triangles were already used as a reporting tool in the United States, and that non-life insurers were required to do some official reserving tests on the annual statements. This author proposed the application of a method, based on this reporting framework, to estimate the reserves. It was based on the incurred claims and splitting the triangle between the number of claims and the average cost.

Another method appeared in 1965 with R. Beard, Taylor (1986). It was based on risk theory and considered that claims amount could be represented by an exponential polynomial, and that the period of settlement was given by polynomials with negative indices (Kupper, 1967).

Benedikt (1969) refers to a method like the CL. Indeed, it is almost the same. The only difference relies on the use of a simple average, instead of the weighted average that we have in the CL. It was applied to the incurred claims triangles. He also explains that his method has a big advantage when compared with the Masterson (1962) and Beard (Taylor, 1986) mentioned methods: it is based on what he calls the chain relatives and is much easier to apply. The latter methods are based on the analysis of economic time series from Davis (1941). Here, the method of link relatives, that was widely used to study time series and to summarize all the chain relatives with averages or medians, was presented. The chain relative was just the ratio of two adjacent figures. It corresponds to the link ratio summarized in (2.1).

At that time, there were several presentations of statistical approaches to claims reserving and another method came from Beard (1969); a study of twenty-eight companies showed that the split of reliable data between claims frequency and average costs was a practical method of estimating insurer's reserves.

According to Bornhuetter and Ferguson (1972), claims reserving had little attention in the literature of insurance and it was not common to see papers on the subject at that time. Indeed, if we look at the Astin Bulletin's since it started in 1958 until the beginning of 70's, we just find few papers on the subject, and the Astin is the colloquium of non-life insurer´s actuaries. Usually the literature on its bulletins was concentrated on risk theory, pricing and solvency. With this paper, Bornhuetter and Ferguson (1972) also published a new method for claims reserving.

Also, in 1972, another publication came with triangles and statistical methods. It was the paper from Verbeek (1972) about the estimation of the ultimate number of claims. According to Verbeek (1972), moving from claims number to claims costs should not be difficult. The model assumed a Poisson distribution for the claims counts and required the maximum likelihood estimation of the parameters. A statistical method was also presented by Fisher and Lang (1973) without triangles and with a reporting-year base.

At that time, a kind of competition was also emerging between statisticians and actuaries to approach the problem of claims reserving, A.D.W. (1974) – the paper was signed just with these initials:

"The statistician begins with a formula and then looks for numbers to fit it, whereas the actuary begins with numbers and looks for a formula to fit them"

Beard (1974) seems to be the first paper to present the CL on the 22nd of May 1974 at the Institute of Mathematics and its Applications Symposium. The methodology was similar to the one from Benedikt (1969) five years ago, but he used a weighted average instead of the simple average. With the publication of the symposium papers, the problem of estimating claims reserves has been comprehensively aired, for the first time in Britain, A.D.W. (1974).

At that time, there was the need of reconciliation between statisticians and actuaries and the CL did this because it was considered to provide an actuarial and statistical approach to the problem. More importantly, the Department of Trade considered the CL approach to be the best one and the one to be chosen for statutory regulations, A.D.W. (1974).

Taylor (1977) presents a new method of estimation, the separation technique, and summarizes some of the problems in applying the CL method:
"In the absence of exogenous influences such as monetary inflation, changing rate of growth of the fund, changing mix of business in a fund, the distribution of delays between the incident giving rise to the claim and the payment of that claim remains relatively stable in time. In this case the columns (or rows) of the run-off triangle are, apart from random fluctuation, proportional to one another. It is crucial to the logic of the underlying CL method that the exogenous influences should not be too great. If this assumption does not hold, then the conclusion that the columns of the run-off triangle are proportional goes awry too, and the CL method give misleading results".

However, this method was harder to apply, and even the CL was not easy, due to the absence of microcomputers in that time. Before the advent of the computer, loss reserving was a time-consuming and tedious process that resulted in frequent miscalculations and errors (Fallquist and Jones, 1987). They stressed in this paper that only with microcomputers that it is feasible to start using triangles and actuarial methods for claims reserving.

The first microcomputer came with Apple in 1976 but in the beginning, it had limited use, even corporations were still relying in mainframes with terminals. After Apple, several others followed like the Commodore, the Heathkit and the Radio Schack but only in 1981 we had the IBM microcomputer with the Microsoft operating system MS-DOS (Brookshear, 2013). Just after that, we saw the spread of personal computers everywhere. This means that in the 80's the penetration of these methods on claims reserving was not very high and some actuaries were using calculators or doing some programs to apply the CL. An example of this is the APL language, also called in the actuarial profession as the Actuaries Programming Language. With the spread of personal computers inside insurance companies, in the second half of the 80's, actuaries started using the CL more and the technique won world recognition.

At the same time, actuaries and statisticians started struggling with a technical question, that being resolved, would allow them to calculate risk margins and confidence intervals to their best estimates and to know the prediction error of their calculations: which stochastic method was implicit in the CL calculation?

The first answer was given by Thomas Mack (1993a, 1993b, 1994). The method proposed was independent of any probability distribution and the market acceptance was very high. Finally, actuaries could have the best estimate, and calculate the risk margin and the prediction errors. This will also give them confidence intervals to a previously specified degree of confidence. Even if the CL was being criticized, this non-parametric stochastic method gave CL a new life.

The same happened four years later with a new parametric approach to the CL from Renshaw and Verrall (1994 and 1998). They showed that the CL stochastic method was an Over-dispersed Poisson model. It was an important step ahead because it was an application of Generalized Linear Models (GLM) to reserving when the same was already being done with pricing. The two important issues, pricing and reserving, were using the same technique and it was a second recognition to the CL. It was an important jump because this application of GLM models shows the stochastic method beyond the CL, but also that we may have other alternatives to the CL (with different a different probability distribution and with other link functions and variance functions).

One year later, the same authors published an application of the Bootstrap technique to claims reserving, once again using the CL (England and Verrall, 1999). The technique may be applied to any method and due to this, it is common to hear that it is not a claims reserving method. However, some literature calls the Bootstrap a model, see for example Hindley (2018). The technique/method had even more acceptance than the Over-dispersed Poisson and became a non-parametric alternative to the Mack method. The spread of these contributions accelerated with the publication of a practical paper on the Institute of Actuaries (England and Verrall, 2002).

All this indirect recognition of the CL promoted its practical application in several countries, and the existence of a stochastic CL allowed many actuaries to adjust the reserves and changing the confidence level of the estimate.

In this CL history, it is also important to mention the BF method published by Bornhuetter and Ferguson (1972), just before CL presentation in 1974. The method is a mixture of external information, a benchmark, with the insurer internal data to calculate the loss development factors. The latter were calculated using averages on the link ratios of several origin years. However, with the implementation of the CL, this method starts substituting the simple average on loss development factors calculations by the former. We may see that in several presentations of the BF method in the literature. For instance, the Institute of Actuaries manual on claims reserving (1989), mentions that the authors used the Link Ratio/CL approach. Indeed, they used the Link Ratios approach with a simple average, but they never said, in the original paper, that they were using a weighted average or even the CL.

This means that the professional practice and the introduction of the CL changed the original BF. As the above manual says, Institute of Actuaries (1989), the term CL is sometimes used to describe a method which uses any kind of average. It is not obligatory to use the CL method with the BF, but it is very common to see several studies always using the CL. Also, Wüthrich and Merz (2008) explain this fact very well when they state that:
"In most practical applications one deviates from the path of the pure BF method and estimates the still-to-come factor from data with the CL estimates".

As we can see the CL was an important tool to actuaries. It started by substituting the more common standard at that time, the simple average link ratio, and became the standard for some regulators and "changed" the original BF that, in the original paper of Bornhuetter and Ferguson (1972), was using the simple average link ratio. It became a standard that in some cases solved an important problem to some actuaries: which method shall we apply to this triangle? The answer of the market in most cases was: use the CL or maybe the BF, which is an extension of the CL.

3. Deterministic Methods

Actuarial methods for claims reserving may be deterministic or stochastic. The former is simpler to apply and produce a best estimate for the reserves. The stochastic methods, more complex, give us a best estimate, a confidence level to our best estimate and the reserves prediction error.

Seeing the history of claims reserving, the deterministic methods were the first to be applied. Actuaries took some time to become the recognized professionals to claims reserving, and the only way to achieve that, with the technology of the 70's, was by using clear and easy to apply methods. At the same time, to have the spread of actuarial methods for claims reserving, a simple technique was also important which the deterministic methods had this feature.

Even nowadays the deterministic methods are the most used by actuaries to calculate claims reserves. There reason for that is because most legislations require best estimate reserves on accounts, which is the output from deterministic techniques. Even the Solvency II directive, European Union (2009), went in the same direction, imposing the deterministic methods to the best estimates calculation. This means that deterministic methods are important.

We will present here the most well-known deterministic methods, and will show the relation they may have with regression techniques. To cover that objective, we will summarize:

- The methodologies presented by the UK Institute of Actuaries (now Institute and Faculty of Actuaries) on its manual of claims reserving (1989 and 1999) and Acted (2016) and by the Casualty Actuarial Society on its non-life insurance examinations, CAS (2017).
- Some variants of these methods.
- And some other methods presented in the literature.

We will mention the connection between most of the techniques that are going to be presented with regression models. Especial emphasis will be done to the methods that were developed with the explicit use of regression techniques.

3.1 Link Ratios

As we saw before in section (2.5), we may get the insurer´s reserve by using the payments done so far multiplied by the ultimate factor minus one. The ultimate factor is the product of the loss development factors, which means that for having the estimated reserves, we just need two things:

- To know the cumulative payments done by the insurer per origin year, that is, the last diagonal of the upper triangle.
- And to have the loss development factors, which means data to estimate them, the upper triangle of cumulative payments (complete or not, as it is possible that the actuary decides to have just a subset of this triangle to calculate the loss development factors).

We also saw how CL estimates these loss development factors in (2.6) and (2.7): doing a weighted average of the past link ratios. It was also clear from the history of claims reserving, in section 2.6, that before the birth of the CL, loss development factors were already calculated using simple averages. Examples may be seen with Masterson (1962) and Benedikt (1969).

This estimation is just a statistical problem. We saw in Table 2.5 that a triangle of link ratios may be calculated from the triangle of cumulative payments, presented in Table 2.4. When we look at the columns of Table 2.5, we see that each of these columns is a time series of link ratios. This means a series of data points in time order, taken at successive equally spaced points in time. What the link ratios methods do, is estimate a statistic that summarizes all the link ratios in one number: the loss development factor.

This means that the loss development factor can be any statistic: a weighted average with payments made in the past as the weights (the CL suggestion), a weighted average with subjective weights, a simple average of the link ratios, the lowest link ratio from all the link ratios in the column (an optimistic case), the highest link ratio from all the link ratios in the column (the worst case scenario), the last link ratio (the most recent link ratio), an average of the last two or three link ratios (an average of the most recent years), the median (the link ratio of the middle from a series of ordered link ratios) or a trend (the adjustment of a linear or log linear regression against time). We may also use any of these methods excluding certain link ratios that are considered outliers.

This shows that the link ratios method allows actuaries to choose one of several alternatives to estimate the loss development factors. This flexibility permits actuaries to adjust the method to the insurer's circumstances: inflation, speed of claims settlement, different future behaviours, trends, outliers, legislation and even subjective feelings about the future. The following methods are examples of this flexibility.

To get the CL, we do a weighted average using the payments as the weights, see (2.6) and (2.7). To have the Simple Average (SA) method, we just need to do the simple average of the past link ratios.

$$\hat{b}_j^{SA} = \frac{\sum_{i=1}^{T-j} F_{i,j+1}}{T-j} \tag{3.1}$$

We may follow the same approach to get the Last Link Ratio (LLR) method:

$$\hat{b}_j^{LLR} = F_{T-j,j}$$

Using ordinal statistics from the link ratios triangle, see for example Table 2.2, we may also get other alternatives:

- The Best-Case (BC) scenario, a lower bound, would be:

$$\hat{b}_j^{BC} = \min_j(F_{i,j})$$

- The Worst-Case (WC) scenario, an upper bound, would be:

$$\hat{b}_j^{WC} = \max_j(F_{i,j})$$

- The Median (MD) scenario would be:

$$\hat{b}_j^{MD} = \operatorname*{med}_j(F_{i,j})$$

All these variants of the link ratios are just statistics that summarize a set of link ratios. The actuary needs to select one of these statistics per column. Having that, he will have the ultimate factors using (2.3) and the reserve best estimate using (2.5). Usually actuaries apply the same variant of the link ratios to all columns but that is not compulsory, and a variant may be applied to each column.

Whatever the method is, the idea is always the same; with some judgment and some analysis of the results the actuary may apply these methodologies. As it is not automatic to define which one is the best, some methods, like the CL, became the most used. Having defined that "everybody uses the CL", a lot of time is saved, and the problem becomes simple. This approach is dangerous, due to the lack of accuracy it may involve, but it is indeed a practice in several countries. A good and recent evidence of this may be seen at IAA (2017).

The link ratios approach does not give the calculation of the prediction error of any of its variants. However, for the upper triangle, we may see the estimated errors (also called estimated residuals), $\hat{\varepsilon}$. These are the differences between the estimated payments $\hat{C}_{i,j}$ and the real payments, $C_{i,j}$ (that we got in the past). They are calculated backwards, starting in the last column of the triangle. An example may be seen in Booth et al. (2005).

$$\hat{\varepsilon}_{i,j} = C_{i,j} - \hat{C}_{i,j} \qquad j \leq T - i + 1 \tag{3.2}$$

The following variant of the link ratios is different. It considers the errors (also called residuals) in estimating the loss development factor and does not apply the same loss development factor to each cell of the lower triangle. It does one regression to each column from the triangle. A trend link ratio would be the result of a regression of the link ratios, $F_{i,j+1}$. The latter will be explained by a constant and by time with a certain functional form, $h(Time)$. Time would be the origin year $i \leq T - j + 1$. The regression will use the least squares technique to estimate the method parameters: we will get the estimated constant $\hat{a}$ and the estimated slope $\hat{b}_j^{TR}$

$$F_{i,j+1} = \hat{a} + \hat{b}_j^{TR}.h(Time)$$

This trend will be a linear trend if the functional form is linear

$$h(Time) = Time$$

And will be a log linear trend if the functional form is logarithmic

$$h(Time) = \log(Time)$$

In these two cases we, are having a linear model and we estimate the regression parameters using least squares. The loss development factor (and the constant if included) will be estimated after minimizing the sum of the square of the errors or considering, if appropriate, weights for each observation. Such a procedure of using weights is very common when the regression shows heteroscedasticity. See for example Formby et al. (1984).

Other results may be obtained by changing the functional form, the equation structure (adding other variables, such as the speed of paying claims or omitting the constant) and the estimation method (using other alternatives to the minimization of the sum of the square of the errors, for example, the generalized least squares).

In these two cases, we had a regression model using link ratios but it is also possible to show that the CL and the SA may also be seen as regression models, see for example Straub (1988), Murphy (1994) and Barnett and Zehnwirth (2000). This is done using weighted regressions with the appropriate weights. It will mean that there is heteroscedasticity in the model.

This also means that it is possible to do the same for the last link ratios method, just giving weights to the most recent link ratio. And the same could also be done, using quantile regressions, to the ordinal statistics link ratios, as the median link ratio. The median link ratio will correspond to the loss development factor and may be replicated with the minimization of the sum of the absolute errors (instead of the square of the errors as in the linear regressions). The median is not so much sensitive to outliers.

We may conclude that the Link Ratios (LRT) methods may be presented by a regression model if we use a quantile regression. The quantile regression is out of the scope of this thesis, but we will replicate in chapters 5 and 6 the Link Ratios methods, the CL and the SA, using regression models.

3.2 Grossing-Up

The Grossing-Up (GU) methodology uses the grossing-up factor presented in (2.8). We saw already, in (2.10), that the grossing-up factor is the reciprocal of the ultimate factor. Looking at equation (2.5), we immediately see that we may obtain the same reserve as the link ratios if we calculate the GU method with the same statistic that we used for the link ratios, for example, a simple average (some small differences arise due to rounding up, as the number of operations done is not the same between the two methodologies, for example, the Excel only has 16 decimal places).

Even though the results are similar, the GU methodology may be a good help for actuaries to explain their calculations to non-actuaries. The latter will understand much better a grossing-up rather an ultimate factor. This grossing-up factor is the percentage of the payments in respect to the ultimate cost in a certain development year. People in the claims department will understand this statement and they will know that the grossing-up factor is between 0% and 100%. The ultimate factor is the number that when multiplied by the current cumulative payments gives the ultimate cost. It is a figure more difficult to understand by the claims department staff.

This means that with the GU method, we will get reserves as the proportion of the estimated ultimate cost to be paid in the future.

$$\hat{R}_i = C_{i,T-i+1}.\left(\frac{1}{\hat{g}_j} - 1\right) = C_{i,T-i+1}.\frac{1}{\hat{g}_j}.\left(1 - \hat{g}_j\right) = \hat{C}_{i,T}.\left(1 - \hat{g}_j\right) \tag{3.3}$$

As we can see in (2.10), we need an estimate of the ultimate cost for each year to be able to calculate each of the grossing-up factors. This means that we must do the calculation having the ultimate cost from all the years, not just the first one. This is done by recursion:

- We have the first origin year closed. Using (2.8) we get, for that year, the grossing-up factors from all development years $j = 1, \dots, T-1$.
- Using the grossing-up factor from year $i = 1$ and development year $T-1$ we estimate the ultimate cost of the second origin year, also using (2.8). Having this, we get the grossing-up factors for the second origin year to $j = 1, \dots, T-2$.
- In the third year and the ones that follow, we repeat the procedure. There is just one difference: now we have more than one grossing-up factor (from previous origin years and to the same development year) to estimate the ultimate cost. We need a statistic that summarizes them, for example, the simple average. Having this statistic, we proceed as before.

As the grossing-up factor is the reciprocal of the ultimate factor, see (2.10), it is also the reciprocal of the product of several loss development factors, see (2.3). As the loss development factors may be obtained by a regression, see for example Murphy (1994), the grossing-up factors are the product of the outcome of several regressions. They are just a different way of presenting the same.

3.3 Average Costs

We will see here two ways of considering the Average Costs (AC) method. Firstly, we will see the traditional method, also called the frequency-severity method. Secondly, we will see the AC using the operational time.

The traditional approach, see for instance Hindley (2018), divides the cumulative payments triangle, into two triangles: one with the number of claims with payments $nP_{i,j}$ and another one with the average payment per claim with payments, $\bar{C}_{i,j}$. This average is calculated dividing each cell of the cumulative payments by the cumulative number of claims with payments. Now we have two triangles, and our initial triangle is decomposed in the following way.

$$C_{i,j} = \bar{C}_{i,j}.nP_{i,j}$$

The method will give, as output, the ultimate number of claims with payments and the ultimate average costs, both per origin year. Multiplying the two outputs, per origin year, gives the ultimate costs per origin year.

Sometimes actuaries use the number of claims settled instead of the number of claims with payments, but in that case, to be consistent, the average cost should be calculated per claim settled.

It is also possible to use the incurred claims triangle to perform a frequency-severity analysis. The procedure is similar to the one from the cumulative payments. However, the number of claims must be the ones that were notified and the average cost, $\overline{iC}_{i,j}$, should be obtained with the division of the incurred claims by the number of cumulative notified claims $nC_{i,j}$. This happens because the incurred claims are the cumulative payments plus the case reserves. As before with cumulative payments, now we have two triangles, and our initial triangle is decomposed in the following way.

$$iC_{i,j} = \overline{iC}_{i,j}.nC_{i,j}$$

Having done the decomposition, whether with paid claims or with incurred claims, all the methods and variants of the LRT and of the GU methods are applicable and they do not need to be the same in each component of the decomposition done, for example, we may apply the CL to the frequency triangle and the SA to the severity triangle.

Using the LRT method and the paid claims, the reserve on each origin year i will be given using the ultimate factors of each component: $\hat{f}_{\bar{C},j}$ the ultimate factor of the average payments on column j and $\hat{f}_{nP,j}$ the ultimate factor of the number of claims settled at column j. All these ultimate factors are obtained in the same way we did before in (2.3). For the case where we use the paid claims we get,

$$\hat{R}_i = \left(\bar{C}_{i,T-i+1}.\hat{f}_{\bar{C},j}\right).\left(nP_{i,T-i+1}.\hat{f}_{nP,j}\right) - C_{i,T-i+1} \qquad (3.4)$$

As the traditional AC method is based on the LRT or the GU methods, we may conclude that it can also be presented with the use of regressions.

We may also use the AC method in a different way, by changing the time scale to an operational time. The definition is done for closed claims, but sometimes the claims with payments are also used depending on the data available. The operational time is the proportion of claims closed and ranges between 0 (no claims closed) and 1 (when all the claims are closed). If the operational time is 0.1 this means we have 10% of the claims closed. The concept was introduced by Reid (1987) and the most known method was developed by Wright (1990). Its main assumption is that the average cost of claims depends on the order of settlement: small claims are closed faster than bigger claims. This happens because of the complexity of the bigger claims. See for example Wright (1990) or Booth et al. (2005).

According to Wright (1990), we may use the method when we have at least one of the following sets of triangles (in each of the sets we have a triangle for the number of claims or payments and a triangle for the payments):

- The number of claims closed and payments on all claims closed.
- The number of payments and the paid claims.
- The number of claims closed and the paid claims.

Also, Lowe (1994) mentions that the method may be useful when the main source of uncertainty comes from the individual claims amounts. This may be the case with bodily injury claims where we have more volatility in the case estimates. Lowe (1994) also states that when the speed of claims is changing, this method is more prepared to adapt to these changes than traditional methods.

Even if the original method was stochastic, we may also see the operational time in the context of the deterministic average cost. For example, if we have the triangle of the number of payments and the paid claims triangle, we may proceed as follows:

- The development years will be the percentiles of claims settled, for example 20%, 40%, 60%, 80% and 100%. That is our previous j, defined in section 2.4, is now a percentile.
- The average paid claims, for example on cell $j = 20\%$, will be the first 20% of all claims paid divided by the number of payments.
- The number of payments will be the number of payments that corresponds to those 20% of claims.

Having done this data change we may apply the traditional AC method described above. This means that changing the triangles´ data allows us to apply a regression model: the AC method is the product of two regression models, as we saw in this section.

Another example of the use of regression models with the operational time is the one from Booth et al. (2005), using an average paid claims triangle and a number of claims triangle:

- The number of claims is estimated using traditional LRT techniques. This will give the estimates for the future number of claims and the ultimate number of claims (the number of claims lower triangle).
- They calculate the past operational time, as the ratio of the average number of claims (in two development years in the past) divided by the ultimate number of claims. With this, they get a triangle of the past operational time.
- A regression is then estimated between the average paid claims (dependent variable) and the operational time (independent variable), using all the cells from the upper triangles for the average paid claims and for the past operational time (obtained above).
- Using the lower and upper triangle of the number of claims, the future operational time triangle is obtained (with the same methodology as with the past operational time).
- Inserting the future operational time on the regression obtained before we get an estimate of the future average payments.

- Finally, multiplying the latter by the number of claims outstanding to be paid we get the ultimate cost.
- The reserve, as before, will be the ultimate cost minus the cumulative payments.

3.4 Loss Ratios

We started this summary about claims reserving deterministic methods, with one triangle and with two methods of estimating the reserves, the LRT and the GU methods. Then we split the paid claims or the incurred claims in its components, the average costs and the number of claims and used the AC method. Another approach is to bring together more information that we may have in respect to each origin year, specifically the loss ratio. These are the Loss Ratio (LR) methods.

The loss ratio is a measure of the quality from the business written. It divides the claims per the exposure in a certain period. We are giving examples in this thesis with yearly periods, but other periods may be considered, for instance quarterly periods. The calculation may be done using different criterions, see for example (Portugal, 2007). The most used are the following:

- If we collect our claims data per accident year (also called occurrence year), the claims of that year are the payments from claims occurred on that year plus the claim's reserves from claims occurred on that year. The claim's reserves are the case reserves added of any eventual IBNR and IBNER reserves, all of these, in respect of claims, occurred on that year. The exposure of that year is the earned premiums from that year. The earned premiums are the premiums of that year plus the premiums of the previous years that were at risk in that year less the part of the premiums of that year that will be at risk on the following years.
- We may also collect the data by the underwriting year. In that case, the claims will be the payments from claims covered by contracts underwritten on that year plus the claim's reserves from claims covered by contracts underwritten on that year. The claim's reserves are the case reserves added of any eventual IBNR and IBNER reserves, all of these in respect to claims arising from contracts underwritten on that year. The exposure of that year is the premiums from that year.
- Another alternative is to collect data by calendar year. The claims of that year are the payments from claims occurred on any year plus the claim's reserves variation between

two consecutive years. The claim's reserves variation will be the difference on claims reserves between two consecutive calendar years. For the calculation of the reserve of each calendar year, we consider the case reserves plus IBNR and IBNER reserves from that calendar year.

The exposure of that year is the earned premiums from that year.

This distinction is important because we may have the data organized in triangle format with any of the above criterions and the loss ratio must be calculated accordingly, to match claims with premiums, in respect of the amounts at risk. The triangle analysis by origin year and by underwriting year is the most common. The calendar year criterion is used on financial statements.

Indeed, very often actuaries know that the line of business loss ratio should be around a certain value. For example:

- A value coming from market statistics produced by the regulator, the association of insurers or a private provider of information.
- Subjective information from the insurance company underwriters.
- Or even a more objective actuarial estimation, done with some data and statistical models.

If that information exists, it may be very useful for years of origin where there is a lot of uncertainty in respect to the ultimate costs, probably the ones with a lower level of payments. It may even be more important if we do not have any payments in respect of one or more years of origin.

With the loss ratios, we add more information to the claim's reserves estimation. The LR method puts together two sources of information:

- One is objective, the claims paid or the incurred claims triangles from the insurer.
- The other one is subjective, the loss ratio benchmark added to the estimation.

We have more information, but we also have a second source of error; the benchmark information may not match the reality. We will need $T-1$ benchmarks, one per origin year (not yet closed). We assumed before that the first year of origin is closed.

The use of the benchmark may oblige the actuary to do some adjustments to the premiums. The latter are the product of the average premium per exposure, $\bar{p}$, multiplied by the number of units exposed to the risk, E.

$$P = \bar{p} . E$$

If the average premium from the benchmark is higher than the one we have, in the insurer being analysed, the benchmark will be distorted. Indeed, the loss ratio, lr, depends on the incurred claims, iC, but also on the premiums, P.

$$lr = \frac{iC}{P}$$

However, before using the loss ratio, we need to standardize the benchmark premiums using the insurer average premium. This means that we need data for that: the average premium from the insurer and the average premium from the benchmark. The former will substitute the latter to get the benchmark standardized premiums.

The benchmark may also reflect different claims reserving policy and we may have a higher/lower loss ratio just because there are too high/low reserves inside it.

The big advantage of the method is that defining a loss ratio per origin year and multiplying it by the earned premiums of that accident year gives us an estimate of that year´s ultimate cost.

The LR methods are recommended when the triangle shows:

- Scarce or inexistent data on payments and incurred claims.
- Immature data not representative of the reality.
- Experience not significant for the future.
- Possibility of latent claims, notified with delays, that heavily transforms the level of reserves when they are recognized.

There are several methods that use the loss ratios: Bornhuetter-Ferguson (BF), Cape Code (CC), and the Benktander-Hovinen (BH). The next paragraphs summarize each of them.

As they mix the LRT method (or the GU method) with some information about the loss ratios, these methods may be seen as a Bayesian regression. As we referenced before, in the sections 3.1 and 3.2, the LR and the GU methods are regressions and adding a priori information to the latter give us a Bayesian regression, see for example Fomby et al. (1984), O'Hagan (1994) or Lee (1997). The only difference between all the loss ratios methods will be the way the loss ratio a priori information is calculated. The development of this Bayesian regression is out of the scope of this thesis.

The BF method was developed by Bornhuetter and Ferguson (1972). The method produces a reserve estimate per origin year. Firstly, it selects a benchmark for the incurred claims for each origin year, $iC_{i,B}$. Secondly, it estimates the ultimate cost of each year, $\hat{C}_{i,T}$, using the LRT method, usually the CL. Thirdly, the BF ultimate cost of each year, ${\hat{C}_{i,T}}^{BF}$ will be a weighted average from the $iC_{i,B}$ and the $\hat{C}_{i,T}$. The weights are the estimated grossing up factors obtained using (2.10), $\hat{g}$ and *1-*$\hat{g}$. The higher the estimated proportion of payments in respect to the ultimate cost $\hat{g}$, the higher will be the weight given to the link ratios estimate, and the lower the weight given to the benchmark incurred claim.

$$ {\hat{C}_{i,T}}^{BF} = \hat{g}_i . \hat{C}_{T,i} + (1-\hat{g}_i). iC_B \qquad (3.5) $$

The higher the ultimate factor and the more recent the year is, the lower the correction to the benchmark estimation will be; which means that the benchmark incurred claim is considered more useful when the amount of payments is low. The opposite happens if we have a low ultimate factor; we will rely more on the payments done so far.

The reserve will be similar to the one from the CL, see (2.5). However, the payments that are multiplied by $(\hat{f}_j - 1)$ are substituted by the payments implicit on the incurred claims benchmark, $\frac{iC_{B,i}}{\hat{f}_i}$. These implicit payments are estimated using the link ratios ultimate factor, estimated by the LRT method (most probably CL).

$$ \hat{R}_i = iC_{B,i}.\left(1-\frac{1}{\hat{f}_i}\right) = \frac{iC_{B,i}}{\hat{f}_i}.(\hat{f}_i - 1) \qquad (3.6) $$

In the link ratio methodology, we use an objective figure to calculate the reserves, the claims paid, $C_{i,T-i+1}$. In the BF, we use a subjective benchmark, the iC_B standardized by the estimated ultimate factor (which means, the level of estimated payments on the benchmark).

The CC method was developed to overcome some of the CL problems. Following (Straub, 1988) the latter is very sensitive to changes on a single number (it is not robust), does not consider the information given by the earned premiums (the exposure) and assumes that payments of one year do not influence payments of the following years (not correlated).

The method is the same as the BF but with a different estimate of the benchmark loss ratio. The method was used by Swiss Re actuaries in the South African town of Cape Code and was presented by Hans Buhlmann, on an actuarial Summer School in 1983.

The reserve calculation is the same as the one presented to the BF algorithm in (3.6). However, the benchmark ultimate cost is estimated in a different way, as the product of the earned premiums by a loss ratio that comes from the triangle data and the LRT method calculations (and not from an external estimate):

$$\hat{R}_i = lr^{CC}{}_i \times P_i \left(1 - \frac{1}{\hat{f}_i}\right) \tag{3.7}$$

$$lr^{CC}{}_i = \frac{\sum_{q \in Q_i} C_{qq}}{\sum_{q \in Q_i} \hat{g}_q P_{qq}} \tag{3.8}$$

This means that the CC loss ratio, the internal estimate, will be the ratio of, the known total payments (from the Q_i years decided as appropriate to each origin year i) by the total used earned premiums (from the same Q_i years). The meaning of the latter is that we just consider the proportion of the premiums that corresponds to the payments done so far. To have this proportion, we use the grossing up factors, estimated by the LRT (or even the GU method), to get the proportion of the known premiums that correspond to current payments.

In the CC, it is crucial to define the set Q_i of information. It may be all the origin years, a partial set of these years or even just the origin year i. In the latter case, just one origin year considered, it is easy to see that the estimator of the reserves corresponds to the Link Ratio estimate.

$$\hat{R}_i = C_{i,T-i+1}\left(1-\frac{1}{\hat{f}_i}\right) \tag{3.9}$$

This means that the LRT method is a particular case of the CC when we just use as proxy, for each origin year loss ratio, the payments divided by the used earned premiums of that year. The estimate of the latter is done with the ultimate factor of the LRT method.

In its pure form, the method presents the loss ratio as considering all the years of the triangle.

Another approach was also developed after the BF and introduced by Benktander (1976) and Hovinen (1981). Independently, each of these two authors developed a method that gives the same estimated loss amount (Wüthrich and Merz, 2008). There are also references (Mack, 2000) about the possibility of another actuary had also developed this method without knowing of its publication as the BH method. The method, when used in practice, is also known as the Iterated BF method.

As we saw before, the Link Ratio family relies heavily on the payments registered in the last year. This is a weak point if the latter are zero, as may happen with excess loss reinsurance claims that just include claims above a certain threshold. It will also be a weak point if the payments are above or below the normal value, due to anticipations or delays on payments. To avoid this, the BF method substitutes the diagonal payments of the triangle by the benchmark payments. This may help to overcome the CL problem but relies heavily on the quality of the benchmark.

The BH tries to get the advantages of both the CL and BF and relies on the mixture of the information of ultimate costs from the BF ${\hat{C}_{i,T}}^{BF}$ and the CL ${\hat{C}_{i,T}}^{CL}$ (or any of the other Link Ratios alternatives) to estimate the ultimate costs. The weights will be given by the estimated grossing-up factors, $\hat{g}_i$.

$${\hat{C}_{i,T}}^{BH} = C_{i,T-i+1} + (1-\hat{g}_i).\left[\hat{g}_i.{\hat{C}_{i,T}}^{CL} + (1-\hat{g}_i).{\hat{C}_{i,T}}^{BF}\right]$$

Developing this expression (Wüthrich and Merz, 2008), we get that the BH estimator of the ultimate cost is a mixed linear combination of the BF and CL ultimate costs.

$${\hat{C}_{i,T}}^{BH} = \hat{g}_i{\hat{C}_{i,T}}^{CL} + (1-\hat{g}_i).{\hat{C}_{i,T}}^{BF} = [1-(1-\hat{g}_i)^2].iC_{B,i}$$

The grossing-up factors, $\hat{g}_i$, are estimated with the CL (or any other of the Link Ratio methods). The BH method does the smoothing between the CL and the BF estimate (also dependent on the latter). Therefore, the BH method will give an intermediate result between the BF and the CL. As the BF may be seen as a Bayesian regression and the CL as a weighted regression, we may say that the BH is the mixture of two regressions.

Two other methods related with loss ratios are the Additive method (AD), also called the incremental loss ratio method and the Complementary Loss Ratio (CLR) method.

The AD method works with incremental loss ratios per development year. These ratios are obtained by summing all the payments per development year and dividing them by the correspondent premiums. The latter are the premiums sum from the origin year of those payments. However, it is just a case of the BF and the CC methods, see for example Schmidt (2006b). The differences for the BF and the CC are:

- The loss ratio benchmark is not obtained using some external benchmark or the grossing-up factors, but only considers the sum of the incremental loss ratios in all the development years.
- And the proportion of the payments outstanding is not given by the grossing-up factors but by the variation of the incremental loss ratios.

The incremental claims $I_{i,j}$ are given by

$$\begin{aligned} I_{i,j} &= C_{i,j} - C_{i,j-1} \qquad & j = 2, \dots, T \\ I_{i,j} &= C_{i,j} \qquad & j = 1 \end{aligned} \tag{3.10}$$

And incremental loss ratios are obtained as follows

$$lr^{AD}{}_j = \frac{\sum_{i=1}^{T-j+1} I_{i,j}}{\sum_{i=1}^{T-j+1} P_i}$$

Having this ratio, the incremental payments on every cell from the lower triangle will be estimated as

$$\hat{I}_{i,j} = lr^{AD}{}_j \times P_i \qquad j > T - i + 1$$

The reserves on each origin year will be the sum of the incremental payments of that year. The ultimate cost on each origin year will be given by the sum of the cumulative payments with the estimated incremental payments.

$$\hat{C}_{i,T}{}^{AD} = C_{i,T-i+1} + \sum_{j>T-i+1}^{N} \hat{I}_{i,j}$$

The CLR method was developed by Buhlmann et al. (1983) and Straub (1988). The method calculates the loss ratios in the same way as the additive method, however, they get the cumulative loss ratios instead of the incremental loss ratios. This happens because the CLR method uses a triangle of cumulative payments, instead of the incremental payments. These cumulative loss ratios by development year j consider all the payments and premiums until $i = T - j + 1$.

Having obtained the loss ratios by development year, a statistical technique or the "underwriting judgement", Straub (1988), is considered to have the ultimate loss ratios per origin year, $lr^{ult}{}_i$. For example, if the year of origin 3 has a loss ratio of 50% at the development year 8, and if we know that the origin year 1 increased the ratio 10%, between development years 8 and 10, it may be reasonable to assume that the origin year 3 loss ratio is going to be 55%.

If a certain year of origin i developed until development year $j = T - i + 1$, the loss ratio from that year i is the one obtained until j, $lr^{CLR}{}_i$. Having the ultimate loss ratio per origin year, the loss ratio developed until the last payment and the premiums (earned if by origin year and gross if by underwriting year) we get the estimated reserve, $\hat{R}_i$ per origin year i.

$$\hat{R}_i = \left(lr^{ult}{}_i - lr^{CLR}{}_i\right)P_i \tag{3.11}$$

3.5 Separation Method

According to the Institute of Actuaries (1989), the first development of this model was done by Verbeek (1972) with a stochastic approach. Some years later, Taylor (1977) applied the separation method as a deterministic method to the estimation of the average cost. The idea was to understand the effect of inflation on claims. This means that with this method the stochastic algorithm was developed first, and only later that the deterministic approach appeared (the opposite of the majority of all the methods).

The method was introduced by the UK supervisors during the early years of the CL. The reasons for that were the high inflation at that time and some reluctance to correct the CL from past inflation evolution (Taylor, 2000). If we see for example the Institute of Actuaries (1989 and 1997) we may see that it is easy to correct inflation effects on the CL or on any other LRT method. However, this inflation adjusted methods were not in use at that time.

Any payments (or incurred claims) triangle may split each cell $C_{i,j}$ into three components:

- The row effect, r_i, where we mainly have the exposure effect.
- The column effect, c_j, which considers mostly the speed of claims settling.
- And the calendar year effect, d_z, where we may have several chocks on payments, like the claims inflation in the diagonals of our triangle.

This means that for every cell of our matrix of payments we have a multiplicative model

$$C_{i,j} = r_i c_j d_z$$

However, we may avoid the row effect using average payments or loss ratios, if we divide the paid claims by the number of claims opened or by the earned premiums. As these exposure measures, V, are usually the same over each row we will have

$$C_{i,j}{}^{*} = \frac{C_{i,j}}{V_i} = c_j d_z$$

This means that our matrix is totally defined by the column effect and the calendar year effect and if we can estimate all these effects on each cell, we have our ultimate average costs or loss ratios, depending on the volume measure used. According to Taylor (2000b), it is assumed that both effects are totally independent. Also, the column effect can be seen as the past inflation which means that the model is also an attempt to get the implicit claims inflation. The separation method represents an attempt to let the data speak for itself about inflation.

If our matrix is an average cost or a loss ratio, applying the Separation technique or the CL will produce the same results. And if we want to get back to original payments, we are going to have, for each cell, the replication of the CL result.

The interest of the Separation technique just arises when we have different inflation over the years and when we need an estimate of it. The CL allows the treatment about inflation but obliges us to have an estimate of the past and future inflation per calendar year. The Separation method also estimates for us the implicit past inflation. It may also be shown that the Separation model corresponds to the CL adjusted to the inflation if we correct the past inflation accordingly (Boot et al. 2004).

A summary of all the steps to implement the separation technique may be seen in Hossack et al. (1993), and according to Goovaerts (1990) the model may also be seen as a log-log regression.

The relation with the CL method means that if the CL assumption do not match our problem we will not have good estimates from the Separation technique and the implicit inflation will not be accurate. However, the Separation technique may be improved if we estimate the calendar year effects with a regression model. That is, we estimate the z's with a regression where we are going to have explanatory variables to a have a better prediction. For instance, in credit insurance we may use the real GDP growth as an explanatory variable, because the economy evolution will explain the credit insurance claims (defaults on payments) and the credit insurance reimbursements (money paid back by debtors). Having this relation estimated, the future calendar year effects (the lower triangle values) will be estimated with some assumptions about the future values of these explanatory variables (Luís Portugal, unpublished paper of 2002). A similar improvement with inflation was applied to the Over-dispersed Poisson by Verrall and Brydon (2009).

3.6 Regression Models

Regression models involve a dependent random variable that we want to study, with several observations i, for instance the claims paid on a triangle column j, $C_{i,j}$. This dependant variable is a function of a certain number of independent variables (assumed very often as fixed, non-random) and unknown parameters (also fixed) that we need to estimate.

If the relation between the variables is linear and we just have one independent variable $C_{i,j-1}$, the payments on the column $j-1$, a regression model may be summarized by the following relation, where a and b are the parameters to be estimated and $\varepsilon_{i,j}$ is the error (or residual) on each observation i from the triangle column j.

$$C_{i,j} = a + b.C_{i,j-1} + \varepsilon_{i,j} \tag{3.12}$$

There are several regression models, depending on the assumptions used. Ordinary Least Squares (OLS), see Fomby et al. (1984), assumes that:

- The conditional expected value of the errors (on independent variables) is zero, $\mathbb{E}(\varepsilon_{i,j}|C_{i,j-1}) = 0$.
- The independent variables $C_{i,j-1}$ are not random, which means $\mathbb{E}(\varepsilon_{i,j}|C_{i,j-1}) = \mathbb{E}(\varepsilon_{i,j})$ and so $\mathbb{E}(\varepsilon_{i,j}) = 0$.
- The dependent variable is explained by a relation as (3.12), that may have more than one independent variable, and due to the previous assumptions, $\mathbb{E}(C_{i,j}) = a + b.C_{i,j-1}$.
- The errors inside each equation j for observations $i \neq i'$ are not correlated, $Cov(\varepsilon_{i,j}, \varepsilon_{i',j}) = 0$.
- And the errors variance is constant on each equation j for all observations i (homoscedastic errors), $Var(\varepsilon_{i,j}) = \sigma^2$.

Eventually it may be assumed that the errors are normally distributed about their mean, $\varepsilon_{i,j} \sim N(0, \sigma^2)$. This assumption is needed if we want to do statistical inference and if we use a likelihood approach. If the errors are normally distributed, the same will happen to the dependent variable, in this case $C_{i,j}$. We will develop the models in chapter 5, 6 and 7 without using this assumption. One of the models to be presented in chapter 6, the Vector Projection (VP) with homoscedastic errors, is an OLS model.

Generalized Least Squares (GLS) assumes that the errors can be either not constant (heteroscedastic) or correlated. GLS models may have a known heteroscedastic or correlated structure (the Aitken model) or an unknown heteroscedastic or a correlated structure that needs to be estimated (the feasible model). See Fomby et al. (1984) for more details. Some models presented in chapter 6 are GLS models with a known heteroscedastic structure, as the CL and the SA.

We may also have Seemingly Unrelated Regressions (SUR) models. They will assume contemporaneous correlations between the equations inside each triangle. By contemporaneous correlations between equations, from the same triangle, we mean that the error terms are correlated in the same point in time. The same point in time in claims reserving triangles, in the context of regression models, means the same origin year. Chapter 6 (with one triangle) will use this SUR model and we will call them multivariate claims reserving models.

These SUR models are estimated with three steps, see for example Hill et al. (2012):

- First the OLS is applied separately to each equation.
- The OLS errors are used to estimate the variance and covariances of the errors.
- Then, with these estimates, the GLS is applied to all the equations at the same time.

Finally, we may have more than one triangle where we may also apply the GLS or the SUR models. With the latter model, we may also assume contemporaneous correlations between several triangles. By contemporaneous correlations between triangles, we mean that the error terms of each triangle are correlated in the same point in time. The same point in time in claims reserving triangles, in the context of regression models with more than one triangle, means the same origin year.

We saw already in the previous sections that there are several known claims reserving models that may be seen as regression models. It was the case of the Link Ratios, in section 3.1, of the Grossing-Up factors in section 3.2, of the Averages Costs in section 3.3, of the BF, CC and BH in section 3.4 and of the Separation technique in section 3.5.

We introduce here, explicitly, the use of regression models. This section could be in the stochastic models' chapter 4. The reason why it is included here is the way the regression models were used in claims reserving when they started; they were used as a deterministic model, giving the algorithm to the calculation of the loss development factors.

In chapters 6 and 7, models are developed as regression models. The model from chapter 5 considers the Mack (1993a, 1993b, 1994) framework with some changes using the regression through the origin to estimate the loss development factor.

3.6.1 Applications to Claims Reserving

The first known application of regression models to estimate insurer's reserves seems to be from Simon (1957), but according to Taylor (1978) more than twenty years later, *"regression models were not prevalent among actuaries"*.

Several events are going to contribute to change this:

- Theoretical applications of regressions to claims reserves, Kamreiter and Straub (1973), Lemaire et al. (1981) and Kremer (1984). They have a feature, not common in claims reserving model's literature, they implicitly assume serial correlation of the errors, that is, origin years are not independent.
- The use of regression models by some known actuaries, see for example, Benjamin and Eagles (1986), Straub (1988), Brosius (1992), Murphy (1994), Christofides (1997) and Barnett and Zehnwirth (2000).
- The new approach of bivariate and multivariate models, calculating the reserves of more than one triangle at each time. Some of them were based on regression models. See for example Zhang (2010).
- The application of the Hachemeister regression, Hachemeister (1975), to claims reserving, which considers the standard regression as a particular case, Wüthrich, M. and Merz, M. (2008).

- The conscious that some of the most known models, like the stochastic CL from Mack (1993a, 1993b, 1994), were indeed regression models, see for example Free (2010) and even Mack (1993a)
- And the development of some formulas to have the reserves prediction error, as most of these models were regression models, see for example, Zehnwirth (1985), Renshaw (1989), Verrall (1991a) and Christofides (1997).

Kamreiter and Straub (1973) presented two models, using multivariate times series to explain the cumulative payments per development year. According to Kremer (1984), the approach needed to be more developed and was not considered in practice.

One of the first attempts to use regression models in claims reserving was done by Jean Lemaire et al. (1981). They pretended to *"avoid the main criticism directed against all the other methods (including the chain-ladder and separation methods): clearly the problem is of stochastic nature, and yet all methods are essentially deterministic. Time-series analysis introduces a new dimension in the problem by considering the payments as observations of a random process. This is in our opinion certainly a step in the right direction: the stochastic element of the process has to be introduced in the model if one hopes to be eventually able to compute variances and confidence intervals for the estimates of the provision. Only at that point can actuaries and statisticians hope to have their techniques implemented by the insurance companies and accepted by the control authorities."*

They developed two models. The first one using only the triangle information and another one considering also information from other years, not included in the triangle. In both cases they tried to select the model that best fitted the data. A prediction error formula was not developed but parameters were estimated to minimize the square of the errors and an errors analysis was performed. The best models obtained were autoregressive, in the sense that the cumulative payments on any column j were dependant, not only on the cumulative payments of the column $j-1$ (for the same origin year i) but also on the cumulative payments from the column j for the origin years before: $i-1$ and $i-2$. This means a model with two lags.

The equations were defined as autoregressive for each column $j = 1, \dots, T-1$ with $a_{0,j}$, $a_{1,j}$, $a_{2,j}$, $a_{3,j}$ being parameters to be estimated and $\varepsilon_{i,j}$ the error.

$$C_{i,j} = a_{0,j} + a_{1,j}.C_{i,j-1} + a_{2,j}.C_{i-1,j} + a_{3,j}.C_{i-2,j} + \varepsilon_{i,j}$$

Kremer (1984) extends the autoregressive models to the general case of several years of lags. However, it was a pure autoregressive model as the cumulative payments on column j did not depend on the cumulative payments from the column $j-1$. Also, the cumulative payments were standardized by a general exposure measure, not defined by Kremer.

The method of estimation was not any more a linear regression, as with Lemaire et al. (1981), but a weighted regression where the weights were the exposure measure.

The methodology presented by Benjamin and Eagles (1986) was applied in the London Market. It was a request from Lloyd's to have a method that could help insurers to establish their minimum level of reserves. The paper developed two methods for the calculation that became known as the methods from the London Market.

The idea of the first method was to estimate a regression for each line of business and for a specific development year j. An annual time-series with i observations is obtained for the ultimate loss ratio ${lr^{ult}}_i$ and the loss ratio at the development year j, lr_i. A linear regression is defined, where ε_i is the random error with the classical properties from ordinary least squares, see for example Fomby et al. (1984).

$$ {lr^{ult}}_i = a + b.lr_i + \varepsilon_i \tag{3.13} $$

Benjamin and Eagles (1986) suggest that the ordinary least squares should be enough to estimate the model parameters, a and b. Having them, a simple formula may be applied to obtain the ultimate loss ratio of any year of origin as a function of a known statistic, the loss ratio from development year j. For example, for the Marine line of business, they obtained the following relation between the estimated ultimate loss ratio $\widehat{lr}^{ult}$ and the loss ratio after the first year of development, lr.

$$ \widehat{lr}^{ult} = 33.367 + 3.385.lr $$

Having the lr immediately they obtained the ultimate loss ratio. The method was also more flexible and accurate than the existing rule of that time. This rule was doing something similar but without a statistical estimation of the parameters, more than that, it was assuming that the slope of the equation was always equal to one. The method was also attractive due to its capability of calculating the confidence interval of the prediction, at that time by visual graphic analysis.

The second method from Benjamin and Eagles (1986) was an extension from the first method. The model was again a regression, as in (3.13), but now the ultimate loss ratio from the insurer is substituted by the ultimate loss ratios from all the Lloyd's syndicates. In both methods, after having the estimated ultimate loss ratios and the premiums, it is straightforward to have the estimated reserves. Using (3.11) and considering lr_i, instead of the CLR method loss ratio, we obtain the estimated reserves.

Straub (1988) showed that the CL was a linear regression through the origin. However, he also demonstrated that despite this the CL was not minimizing the sum of the error's squares. He also presented other alternatives of regressions. Those regressions were explaining the cumulative payments on a specific development year j as a function of the cumulative payments on development year $j-1$ and with a constant a.
He called the estimated regression presented in (3.14) the London Chain approached to straight lines.

$$\hat{C}_{i,j+1} = \hat{a} + \hat{b}.C_{i,j} \tag{3.14}$$

He suggested the use of the ordinary least squares to estimate the parameters a and b.

$$\min \sum_{i=1}^{T-j+1} \left(C_{i,j+1} - a_j - b_j.C_{i,j}\right)^2 \tag{3.15}$$

Straub (1988) also shows that the CL method may be obtained by this methodology considering the model without intercept if an approximation is done to the ordinary least square solution.

$$\hat{C}_{i,j+1} = \hat{b}C_{i,j} \tag{3.16}$$

He also shows that in this case the CL estimator is not the solution of the (3.15). Instead the solution for (3.13) is a regression through the origin, also called a Vector Projection (VP), see Gentle (2007) for the equivalence between the two concepts. The parameter $\hat{b}$ is estimated by (here relabel as $\hat{b}_j^{VP}$)

$$\hat{b}_j^{VP} = \frac{\sum_{i=1}^{T-j} C_{i,j+1}C_{i,j}}{\sum_{i=1}^{T-j} {C_{i,j}}^2} = \frac{\sum_{i=1}^{T-j} {C_{i,j}}^2 F_{i,j+1}}{\sum_{i=1}^{T-j} {C_{i,j}}^2} \tag{3.17}$$

Straub (1988) did not give a lot of importance to this estimator, and he considers it more interesting to study other straight lines, not around the origin (without the constant), but around other generic pivots $(-a, a)$.

Then he minimizes the following expression to get the estimators of the parameters.

$$\min \sum_{i=1}^{T-j+1} \left(C_{i,j+1} - a_j - b_j\left(C_{i,j} + a_j\right)\right)^2$$

In the 1992 exam study kit for the Casualty Actuarial Society examinations Brosius (1992) shows that the link ratios can be seen as a regression model.

He also concludes that regression models *"is often proof to be the right tool for the job, although a non-linear Bayesian loss development function is, in theory, preferable in some cases"*.

He mentions the use of regression models in claims reserving since the 50's.

Daniel Murphy (1994) made an extensive summary of the use of regression models to select the best link ratio method. He presented four possible models.

Model 1 was equivalent to Straub (1988) London Chain presented in (3.14). He called the model the Least Squares Linear method. The loss development factors are obtained from least squares estimates and are considered the best estimates from Gauss Markov theorem. See Murphy (1994) or Fomby and Johnson (1984).

In the data that Murphy (1994) used, the fitted straight line was close to zero and that motivated him to a model 2. In this model, Murphy (1994) considered a regression through the origin as Straub (1988) did with (3.16) and he obtained the same loss development factor presented in (3.17). Murphy (1994) called this model the Least Squares Multiplicative.

Finally, he presents two more models. In model 3, the dependent variable will be the link ratios defined in (2.5). The link ratios are assumed to be dependent just of a constant. He calls this model the Simple Average Development. This happens because this model will replicate the simple average link ratios model that we presented with (3.1). In model 4, he presents a multiplicative model. The latter may be seen as the log-log version of model 1 and will give a geometric average of the link ratios. Despite the presentation of these two models, 3 and 4, the whole paper is developed just with the models 1 and 2.

He develops recursive formulas for the unbiased estimates of the ultimate losses and claim's reserves per origin year and multiple years. He also presents recursive formulas for the prediction errors per origin year and multiple years.

The conclusions he gets are that "*loss development predictions can be improved by the use of least square estimators"* and that *"the weighted average link ratio estimator (for example, the CL), is always inferior to an alternative the least squares estimator"*. The simulation study done made him recommend the use of a constant in the regression model. He expects the model 1 to perform better, when compared with model 2.

The motivation of Christofides (1997) was to solve two CL problems:
"these models are often over-parameterised and adhere to closely to the actual observed data. This process ...and can lead to a high degree of instability in the values predicted from the model as the close adherence to the observed values results in parameter estimates which are very sensitive to small changes in the observed values".

He started by presenting the traditional CL with a more formal model using regression analysis. The results obtained for this model were very similar to CL. He presents other models and shows that within the class of log-linear models, changing the model just involves a change in one of the matrices used. In the conclusions, he highlights that regression techniques were beginning to dominate developments in claims reserving methodology.

In the same line of thought, Barnett and Zehnwirth (2000) concluded that the link ratios can be regarded as weighted regressions, however, according to them, sometimes data does not match the linear regression assumptions. To solve this, they presented an extended link ratios family using regression models. In their paper, we see that modelling the error variance allows us to get several regression models, including the CL.

They start with Murphy (1994) Least Squares Linear method which is presented with incremental paid claims. Then they add a new explanatory variable, the trend in each origin year i. The variance of the error, $Var(\varepsilon_i)$, is a function of a weighting parameter δ. For each origin year i the model is defined by

$$C_{i,j+1} - C_{i,j} = a_0 + a_1 . i + (b-1)C_{i,j} + \varepsilon_{i,j} \tag{3.18}$$

$$Var(\varepsilon_i) = \sigma^2 {C_{i,j}}^{\delta} \quad (3.19)$$

They mention that with this model also represents several other methods like CL ($a_0 = a_1 = 0, \delta = 1$) and the CC ($a_1 = 0$ and $b = 1$).

They also suggested the use of the use of the logarithm of the incremental payments to deal with changing trends.

It seems that regression models, even if not explicitly used in practice by actuaries, are having more interest and are implicit in their work. However, some of the new developments done with regression models moved in another direction: mixing more than one triangle to produce the reserves. That may be the case when we want:

- To estimate the reserves with simultaneous information from paid claims and incurred claims, like in the Munich Chain-Ladder from Quarg and Mack (2004a and 2004b).
- And to estimate the reserves from several triangles at the same time, the so-called in the literature of multivariate models.

The latter includes several approaches and most of them do not use regression models: the aggregation of unpaid loss distributions (Brehm, 2002), the calculation of correlated triangles using rank correlation and bootstrapping (Kirschner et al., 2002), the prediction error calculation, (Braun, 2004), the multivariate Chain-Ladder (Prohl and Schmidt, 2005), the optimal and additive reserving for dependent lines of business (Schmidt, 2006b), the approximate bounds for bivariate Chain-Ladder (Hurlimann, 2005), the Bayesian multivariate claims counts (Mildenhall, 2006), the multivariate additive modelling (Hess et al., 2006), the prediction errors of multivariate reserving with additive models (Merz and Wüthrich, 2007a) and with the Chain-Ladder (Merz and Wüthrich, 2007b) and the bootstrapping of correlated lines of business (Taylor and McGuire, 2007).

Despite this, Zhang (2010) generalizes the multivariate Chain-Ladder with correlations and structural connections between triangles using regression models. These models are called multivariate because they have correlations between triangles. However, as we shall see in the chapter 8, it should be more appropriate to assume that most of them are portfolio models, which are not necessarily multivariate models.

Hachemeister (1975) developed a general regression model using credibility theory for ratemaking. The model considers the standard regression as a special case and was also implemented in claims reserving. However, the model was considered difficult to apply, see for example Wüthrich, M. and Merz, M. (2008). Another theoretical example from this technique may also be seen in De Vylder (1982). According to Shi and Hartman (2014), the credibility theory has not been applied very often to claims reserving, even though the BF method may be seen as a credibility formula.

Mack (1993a) mentioned the CL connection to the weighted least squares regression, and Free (2010) explicitly shows that the CL is a weighted regression. Using a weighted regression means transforming the variables from the original regression using weights, see Fomby et al. (1984). These weights allow the use of the OLS model, as they remove the heteroscedasticity. We do not minimize the square of the errors anymore, but instead the weighted square of the errors. It is an alternative approach to the GLS model, where the variables are the originals, and the minimization is done with the square of the errors and assuming the heteroscedasticity. Both approaches produce the same results, but the analysis of the errors is not done in the same way.

In chapters 5 and 6 we will follow the GLS model.

3.6.2 Prediction Error Calculation

The first attempt to estimate the prediction errors (the square root of the mean square error of prediction, to be defined below) comes from Taylor and Ash (1983) and used regression models. Their purpose was to consider *"the lack of methodology for obtaining second moments of outstanding claims in non-life insurance and give some arguments as to why this lack of generality of the models and methods currently in use in claims analysis, and we suggest further that some unification might be achieved through regression analysis. The claims analysis problem is formulated and solved in terms of regression methods."*

All the following publications presented the formulas to the prediction error using regression models. It was the case of Zehnwirth (1985), Renshaw (1989), Christofides (1997) and Verrall (1991a). These models considered incremental claims following a lognormal distribution. Wright (1990) used generalized linear models to estimate the prediction errors, a development of regression models that allows distributions from the exponential family. Mack (1991) also used a distribution from the exponential family, the gamma distribution, to get the prediction errors.

The mean square error of prediction ($msep$) for an origin year i predicted incremental payment, that is, the reserve $\hat{R}_i$ is the expected squared difference between the true reserve R_i and the predicted reserve.

$$msep_i(\hat{R}_i) = \mathbb{E}\left[\left(R_i - \hat{R}_i\right)^2\right] \qquad (3.20)$$

Developing (3.20), we get

$$\mathbb{E}\left[\left(R_i - \mathbb{E}(R_i)\right)^2\right] + \mathbb{E}\left[\left(\hat{R}_i - \mathbb{E}(\hat{R}_i)\right)^2\right] - 2.\,\mathbb{E}\left[(R_i - \mathbb{E}(R_i)).\left(\hat{R}_i - \mathbb{E}(\hat{R}_i)\right)\right]$$

The *msep* (and consequently the prediction error) is the sum of two sources of error deducted by twice the covariance of the reserve with the predicted reserve:

- If the covariance is positive the prediction error is reduced.
- Otherwise is increased.

If we consider that the covariance is null, or near that, the prediction error is presented as the sum of two sources of error:

- The process variance (or process error), because there is a volatility on the true reserves. Is the first term, $\mathbb{E}\left[\left(R_i - \mathbb{E}(R_i)\right)^2\right] = Var[R_i]$.
- And the parameter variance (or estimation error), because we need a model to estimate the reserves and the model will be based in one or more parameters. We will have an estimation error: is the second term, $\mathbb{E}\left[\left(\hat{R}_i - \mathbb{E}(\hat{R}_i)\right)^2\right] = Var[\hat{R}_i]$, assuming we have an unbiased estimator for the reserves.

Putting both sources of error together we get

$$msep_i(\hat{R}_i) \approx E\left[\left(R_i - E(R_i)\right)^2\right] + E\left[\left(\hat{R}_i - E(\hat{R}_i)\right)^2\right]$$

The $msep$ error does not consider the model error; that is, the error that may arise because the model does not accurately reflect the reality about the claims cost process. According to Hindley (2018), *"Model error is very difficult to determine, and may not always receive detailed consideration"*.

Merz and Wüthrich (2008) defines the conditional mean squared error of prediction as conditional to the available data at D_u, as the purpose of the *msep* is to see how good the predictor from the reserves is given the known data. Mack (1993a) also uses a similar *msep* and defines it in terms of the ultimate claims estimate, see (3.21). Indeed, we know from Mack (1993a) that the $msep$ of the reserves equals the $msep$ of the estimated ultimate claims.

$$msep(\hat{R}_i) = Var(C_{i,T}|D_u) + \left[\mathbb{E}(C_{i,T}|D_u) - \hat{C}_{i,T}\right]^2 \tag{3.21}$$

Other definitions of the msep were considered in the literature, as the ones presented by, Renshaw and Verrall (1994) and Taylor (2000a). The prediction error is also called as the standard error, see for example Mack (1993a) and is very often presented in percentage of the claims reserves.

In chapter 4 we will start by presenting the CL prediction error, following Mack (1993a, 1993b, 1994). In chapter 5, the prediction error is presented using (3.21) in the context of Mack (1993a, 1993b, 1994). In chapters 6 and 7 we will develop the reserves prediction error general formula in the context of regression models. It will be general in the sense that includes several models: univariate or multivariate and with one or several triangles. Several prediction errors will be generated from known models, like the CL and the SA.

In all numerical examples the prediction error, pe, will be presented as the percentage of the claim's reserves $\frac{pe(R_i)}{R_i} \times 100$.

3.6.3 Confidence Intervals

We may also use the prediction error, pe, to have a confidence interval for the reserves, for a certain confidence level. For that Mack (1993b) suggests the use of the normal distribution if the prediction error is not higher than 50% of the reserves. Otherwise Mack (1993a) suggests the use of the lognormal distribution. We remind here that the stochastic CL from Mack (1993a, 1993b, 1994) is related with a weighted-regression.

The reserves normal distribution 95% confidence interval (which means an interval between 2.5% and 97.5% and distribution percentiles of -1.96 and 1.96) is given by, for each origin year i.

$$R_i \pm 1.96\sqrt{msep(\,R_i)} \tag{3.22}$$

For the lognormal distribution, with parameters μ_i and $\sigma^2{}_i$ for each origin year i, the general formula for the 95% reserves confidence interval is given by matching the mean and variances of the unknown distribution of R_i with the lognormal distribution, see Mack (1993a).

$$R_i = \exp\left(\mu_i + \frac{\sigma^2{}_i}{2}\right)$$

$$msep(R_i) = \exp(2\mu_i + \sigma^2{}_i)\,(\exp(\sigma^2{}_i) - 1) \tag{3.23}$$

Using

$$pe(R_i) = \sqrt{msep(R_i)} \tag{3.24}$$

And solving (3.2.3), we get

$$\sigma^2{}_i = \ln\left(1 + \frac{msep(R_i)}{{R_i}^2}\right) = \ln\left(1 + \frac{pe(R_i)}{R_i}\right)$$

$$\mu_i = \ln(R_i) - \frac{\sigma^2{}_i}{2} \tag{3.25}$$

To get the percentile 97.5th, the upper bound from the confidence interval, we need the percentile from the lognormal distribution, which is $\exp(\mu_i + 1.96\,\sigma_i)$. Using (3.25) we get for this percentile

$$R_i \exp\left(1.96\, \sigma_i - \frac{\sigma^2{}_i}{2}\right) = R_i \exp\left(1.96 \sqrt{\ln\left(1+\frac{pe(R_i)}{R_i}\right)} - \frac{\ln\left(1+\frac{pe(R_i)}{R_i}\right)}{2}\right) \quad (3.26)$$

The lower bound of the confidence interval, the percentile 2.5th, will come from the development with $\exp(\mu_i - 1.96\, \sigma_i)$.

$$R_i \exp\left(1.96\, \sigma_i - \frac{\sigma^2{}_i}{2}\right) = R_i \exp\left(-1.96 \sqrt{\ln\left(1+\frac{pe(R_i)}{R_i}\right)} - \frac{\ln\left(1+\frac{pe(R_i)}{R_i}\right)}{2}\right) \quad (3.27)$$

Both the normal distribution and the lognormal distribution confidence intervals depend on the confidence level we want to the interval and on the proportion of the prediction error in respect of the reserves. The confidence level is a question of taste, or risk appetite, or of regulatory requirements. If we increase it, the confidence level will be wider, and the reserves upper percentile will be larger. However, the prediction error it will depend on the triangle data and on the stochastic claims reserving method used to forecast the ultimate costs, see for example (3.21).

3.7 Conclusions

We have the following types of deterministic methods in accordance with the information used:

- Paid claims or incurred claims, the original raw data. It is the case of the Link Ratios and the Grossing Up methods. They may be seen as regression models.
- Decomposition of original data in the number of claims and the average payments/costs. It is the average costs method and the operational time method. As they are a decomposition of the original triangle they may also be seen as regression models.
- Mixture of other sources of information with the Link Ratios. These may be done in three ways and using the grossing up factors as weights: with internal information, the CC method, with external information, the BF and with internal and external information, the BH method. Some of these methods, like the BF and the CC, may also be seen as a Bayesian regression model. As the BH is a combination of the BF with the CL it may also be seen as a regression model. The AD method may be seen as a particular case of the CC and as such it is also a regression model.
- And the decomposition of the original data in the row effect, the column effect and the calendar year effect, the separation technique. We will get the average costs when

standardized by the number of claims and the loss ratios when standardized by the premiums. The separation technique may be seen as equivalent to the CL, which means that we also have a regression model. Even if we do not use this relation it is possible to see the separation technique as a log-log regression.

As the claims reserving methods may be seen as regression models we may apply them in several contexts:

- Estimating just one equation, a univariate regression. For example, when we estimate for one triangle, independently, each of the loss development factors. That is what we saw in this chapter, implicitly, in several methods. We will show more regression models on chapter 4, about stochastic models. In chapter 5 we will develop a univariate stochastic method, the Stochastic Vector Projection, which is explicitly a regression model.
- Estimating more than one equation at the same time, a multivariate regression, for example estimating simultaneously all the loss development factors from one triangle, considering their relations. We will do that in chapter 6 about stochastic generalized link ratios and stochastic multivariate generalized link ratios.
- Or estimating the regressions on several triangles at the same time (in econometrics is called panel data). For example, estimating the loss development factors of several triangles at the same time and considering the correlations between triangles. We will do that in chapter 7 about portfolio data.

4. Stochastic Methods

Even if the best estimate is the most important figure to put in the accounts, insurers need to know its accuracy and variability. This means that the stochastic models are important because they give us an answer to the accuracy and variability of the estimates, through the calculation of the reserves prediction error and confidence interval.

4.1 First Models

The first stochastic models appear in the 70's. Erwin Straub (1971) applied the least squares to the triangles of the burning cost from excess-of-loss reinsurance (the total incurred claims, above an excess point, divided by the exposure). The model was distribution-free and minimized the mean squared error, giving unbiased estimators for the ultimate burning cost of each year. Kamreiter and Straub (1973) also suggested the use of regression models to calculate the insurer's reserves with an auto-regression.

In between, the use of probability distributions on claims reserving was proposed by Verbeek (1972). The method was applied to reinsurer data. The idea was to estimate the number of claims from reinsurance excess-of-loss, exceeding an excess point, using a triangle. He considered the number of claims as coming from a Poisson distribution with a multiplicative model that includes two parameters: the row parameter, the probability of a claim to be reported on that year and a diagonal parameter, the probability of having a claim exceeding the excess point. These parameters are estimated using the maximum likelihood and the number of expected claims is projected after the use of exponential curves over the estimated parameters. The severity was not specified.

Since 1975, actuaries knew the relation of the Poisson distribution with the CL when applied to the number of claims. That was shown at the Astin Colloquium of Portimão by Hachemeister and Stanard (1975) but it was not published in the Astin Bulletin. A first publication of this relation was done in German by Kremer (1985) and finally in English by Mack (1991). Hachemeister and Stanard (1975) showed that the CL ultimate number of pure IBNR claims could be obtained by maximizing a Poisson likelihood function.

Some years later, De Vylder (1978) presented the use of least squares to estimate the insurer's reserves, in the context of another multiplicative model. He used, as a dependent variable, the incremental payments $I_{i,j}$, as defined in (3.10). The idea was to minimize the difference between the incremental payments on each cell of the upper triangle and the estimated incremental payments given by the following multiplicative model.

$$I_{i,j} = C_i . U_j . \varepsilon_{i,j} \tag{4.1}$$

In the latter, C_i is cumulative paid claims until that year, i, U_j is the proportion of incremental payments of that column j in respect to the total claim size and $\varepsilon_{i,j}$ is a random variable with expected value one. He estimates the model parameters C_i and U_j using least squares on the upper triangle. Having these estimated parameters for the lower triangle, and using (4.1), he gets the incremental payments. The sum of the incremental payments on one origin year gives the estimated claim reserves of that year.

Also, Buhlmann, Schnieper and Straub (1980) proposed claims reserving based on a probabilistic model.

The stochastic approach began and after some years it will have more developments.

4.1.1 Kremer Model

Kremer (1982) was able to show that the reserving problem could be addressed considering it as a statistical problem and using the De Vylder (1978) model as a starting point. He transformed the multiplicative model, given by (4.1), into a linear model, applying the logarithm to the equation both sides. The estimators are derived from analysis of variance and Kremer (1982) showed that the results are related with the CL method.

The Kremer model presents the log of the incremental claims on row i and column j, $\log(I_{i,j})$, as the sum of a constant, a with a row effect, r_i, a column effect, c_j and a random effect $\varepsilon_{i,j}$. The latter was assumed, to have a zero-expected value, a constant variance and no correlation. It was also assumed that the sum of all the row effects was zero and that the sum of all the column effects was also zero. These two zero-sum constraints are common in these models to make the parameters in the model identifiable, avoiding over-parameterization (more parameters than cells).

$$\log I_{i,j} = a + r_i + c_j + \varepsilon_{i,j} \tag{4.2}$$

Using linear statistical models, Kremer (1982) gets best linear unbiased estimators for all the parameters: a, r_i and c_j. Having these estimators, Kremer (1982) obtained estimators for the multiplicative parameters from De Vylder (1978).

Kremer (1982) uses this relation to show that its model results are related to the CL. It was the first connection between the CL and a stochastic model, a log-linear two-way analysis of variance model. However, no method was presented for getting prediction errors and confidence intervals for the reserves. Indeed, Searle (1987) reproduces a statement from a communication between two famous statisticians, Ronald Fisher to George Snedecor, that helps to understand this. The former states that:

"The analysis of variance is (not a mathematical theorem but) a simple method of arranging arithmetical facts so as to isolate and display the essential features of a body of data with the utmost simplicity."

This also means that despite that Kremer (1982) developed a stochastic model, it was necessary to have more outcomes to his model to be one claims reserving stochastic model. This was the time of some statistical and rating applications to reserving with several stochastic models under development: credibility theory, Bayes theory, state space models and Kalman filter, see for example De Jong and Zehnwirth (1983a and 1983b) and Verrall, (1988, 1989 and 1990).

Also, and according to Renshaw (1989), the Kremer connection between the two-way analysis of variance model and the CL was not developed by the literature. This happened due to its high level of parameters which usually brings prediction instability, Renshaw (1989).

4.1.2 Renshaw Development

Despite this, the Kremer (1982) paper inspired some developments. Renshaw (1989) presented a development to the statistical analogue of the original CL technique with different parameter estimation and different predicted values. Renshaw (1989) also analysed the predictor instability and the possibility of correcting it. For that, Renshaw (1989) considered the empirical Bayes based on Verrall (1988) and the Kalman filter studied by De Jong and Zehnwirth (1983a). Renshaw (1989) presented results using software on generalized linear models.

Renshaw (1989) concentrated on Kremer (1982) relation presented in (4.2) but using adjusted incremental payments, ${I^*}_{i,j}$.

$$ {I^*}_{i,j} = \frac{I_{i,j} \times (Inflation\ Factor)}{Exposure} \tag{4.3} $$

The model was

$$ \log {I^*}_{i,j} = a + r_i + c_j + \varepsilon_{i,j} \tag{4.4} $$

It was assumed that errors were homoscedastic with mean zero and variance σ^2, and with two other assumptions that: $\varepsilon_{i,j}$ and $\log {I^*}_{i,j}$ are normally distributed and that the $\log {I^*}_{i,j}$ has a mean equal to the parameter a and constant variance σ^2.

The parameters are obtained by minimizing the sum of squares of the errors with two constraints on two parameters, $\hat{r}_1 = \hat{c}_1 = 0$.

$$ min \sum_{i,j} \left(Y_{i,j} - \hat{a} - \hat{r}_i - \hat{c}_i \right)^2 $$

Renshaw (1989) did not present the statistical background to have the numerical results obtained. However, the results were obtained using the incomplete design experiments technique when we have over-parameterized models. More details may be seen in Searle (1987).

4.1.3 Maximum Likelihood Approach

Verrall (1991a) introduced the theory of maximum likelihood to examine the properties of the CL parameters and to assess the accuracy of the estimate. The method uses the CL representation formulated by Kremer (1982).

Verrall (1991a) summarized three possibilities to represent CL:

- The CL framework based on the cumulative payments: the origin year i expected value from the payments on any cell (from the lower triangle) is obtained by multiplying the payments from the cell on the previous development year (and the same origin year) by the CL loss development factor, defined in (2.6).
- The multiplicative model with incremental payments, see for example De Vylder (1978) and (4.1).
- And the two-way analysis of variance model, with a logarithm of the incremental payments and an additive structure, see for example Kremer (1982), Renshaw (1989) or even (4.2) and (4.4).

Verrall (1991a) presents the relation between the parameters of the three alternatives to represent CL and show that these three models are equivalent and are re-parameterisations of the same structure. Kremer (1982) showed that the CL and the multiplicative models are equivalent and that the multiplicative model is equivalent to the additive model. Verrall (1991a) also showed the relation between the additive model and the CL. According to Verrall (1991a) "*The parameters of the first two* (CL and multiplicative models) *have physical interpretations, while the statistical analysis of the latter* (additive model) *is the more straightforward."* Verrall (1991a) considers the additive model, which can be estimated by least squares. This is equivalent to maximum likelihood estimation if the errors are assumed to be independently, normally distributed.

It was shown with the additive model (Verrall, 1991a), that the use of the maximum likelihood estimation (of the development factors and the proportions of ultimate claims), instead of the least squares, allowed a straightforward way of getting the second moments estimation. The reason lies on the invariant property of likelihood estimation under parameters transformations.

Zehnwirth (1989) also suggested the estimation of the Kremer (1982) parameters using the E-M algorithm. Zehnwirth (1989) argued that the latter is more suitable when we have incomplete data, as it is the case with paid (or incurred) claims triangles. The procedure also used the likelihood function.

4.1.4 Negative Incremental Claims

In several models, we get the ultimate costs starting with incremental claims, instead of the cumulative claims used on the traditional CL. This means that we may have some problems when, as normal, incremental claims are negative, e.g. with reimbursements.

The problem exists because the Kremer (1982) model and the Renshaw (1989) development assumed positive incremental claims. The reason for this is the logarithm of incremental claims, see for example (4.2) or (4.4), which obliges a positive $I_{i,j}$. In Kremer (1982), the assumption is qualified as not important. Kremer (1982) suggests the addition of a constant to all the incremental claims.

Verrall and Li (1993) considered a method of choosing the constant and the sensitivity of the results to the constant considered. They considered the CL linear model as given by

$$y_{i,j} = x_{i,j}\beta + \varepsilon_{i,j}$$

The approach consists of getting a maximum likelihood estimate of one threshold parameter τ together with the vector of parameters β. This can be done redefining $y_{i,j}$ through ${y^*}_{i,j}$

$${y^*}_{i,j} = log\left(x_{i,j} + \tau\right)$$

Verrall and Li (1993) shows that considering this constant is equivalent to the use of a lognormal distribution with three parameters that will be maximized like in the standard case with two parameters. The expected value of each cumulative claim will be given by the following formula ($\hat{\sigma}^2$ is the scale parameter of the lognormal distribution).

$$E\left({y^*}_{i,j}\right) = exp\left(x_{i,j}\hat{\beta} + \frac{1}{2}\hat{\sigma}^2\right) - \hat{\tau}$$

All the parameters are coming from the differentiation of a three parameters lognormal distribution. The method presented a solution to the negative incremental claims assuming that the CL method is appropriate for the data set.

4.2 The Distribution-Free Chain-Ladder

As the deterministic CL method could be applied with negative payments there was a move to show a stochastic CL that could fit with this feature (without doing any transformation on data).

It was claimed by Mack (1993a) that log-linear models were not the true stochastic model underlying the CL due to two reasons:

- It deviates from the original method on the calculations.
- And its results were more unstable.
-

It is shown by Mack (1993a) that there is a different distribution-free model underlying the CL that reproduces its original results and, in a different paper, Mack (1993b) presented the standard errors for the reserve's estimates. The first paper was also published in Mack (1994).

The main starting point was to use conditional expectations instead of expectations, something done by Schnieper (1989) that inspired Mack (1993a and 1993b) to his distribution-free model. The former produced approximations to the prediction errors but the latter deducted the prediction error of its distribution-free model. The model is also related with De Vylder (1978) model as demonstrated by Denuit and Charpentier (2005).

The method assumes that the cumulative claims of different origin years are independent and for all cells, $i = 1, .., T$ and $j = 1, \dots, T$ we will have

$$E\left(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}\right) = b_j . C_{i,j} \tag{4.6}$$

It may be show (Mack, 1993a) that for $j > T - i + 1$ and considering the past information D_u with the loss development factors estimated by (2.6) we will get

$$E\left(C_{i,j} | D_u\right) = C_{i,T-i+1} b_{T-i+1} . \ \dots \ . b_j \tag{4.7}$$

This relation replicates exactly the CL reserves and the method just has $T - 1$ parameters, which means it avoids over-parameterization problems of the log-linear models.

It was also shown in (Mack, 1994) that under the above assumptions, the loss development factors estimators are unbiased and uncorrelated. In the same paper, it was presented a formula for the reserves standard error, which is the square root from the mean squared error of prediction. The latter is equal to the ultimate cost mean squared error of prediction.

$$msep(\hat{R}_i) = E\left[(\hat{R}_i - R_i)^2 | D_u\right] = E\left[(\hat{C}_{i,T} - C_{i,T})^2 | D_u\right] = msep(\hat{C}_{i,T}) \qquad (4.8)$$

To have this standard error, another assumption is considered about the cumulative payments variance: the conditional variance on payments of the link ratios given in (2.1) is inversely proportional to the payments $C_{i,j}$. This means that for the parameters $\sigma_j{}^2$ the conditional variance of the payments will be different for every development year until $T-1$:

$$Var\left(C_{i,j+1}/C_{i,j} | C_{i,1}, \dots, C_{i,j}\right) = \sigma^2{}_j / C_{i,j}$$

$$Var\left(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}\right) = \sigma^2{}_j . C_{i,j} \qquad (4.9)$$

Mack (1993b) shows that these parameters $\sigma^2{}_j$ may be obtained by the following unbiased estimator for $1 \le j \le T-2$:

$$\hat{\sigma}^2{}_j = \frac{1}{T-j-1} \sum_{i=1}^{T-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{b}_j \right)^2 \qquad (4.10)$$

Mack (1993b) assumes that the first year of origin is closed, $\hat{\sigma}^2{}_T = 1$, and suggests the following approach to the $\hat{\sigma}^2{}_{T-1}$.

$$\hat{\sigma}^2{}_{T-1} = min\left(\frac{\hat{\sigma}^2{}_{T-2}}{\hat{\sigma}^2{}_{T-3}}, min(\hat{\sigma}^2{}_{T-3}, \hat{\sigma}^2{}_{T-2}) \right)$$

Mack (1993b) presents an estimator for the mean squared error of prediction for the reserves per origin year

$$msep(\hat{R}_i) = \hat{C}_{i,T}{}^2 \sum_{j=T-i+1}^{T-1} \frac{\hat{\sigma}^2{}_j}{\hat{b}^2{}_j{}^{CL}} \left(\frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{j=1}^{T-j} C_{j,j}} \right) \qquad (4.11)$$

And for the total reserves (the sum of all origin years), $\hat{R}$.

$$msep(\hat{R}) = \sum_{i=2}^{T} \left\{ msep(\hat{R}_i) + \hat{C}_{i,T} \left(\sum_{j=i+1}^{T} C_{j,T} \right) \sum_{j=T-i+1}^{T-1} \frac{2\hat{\sigma}^2{}_j / \hat{b}^2{}_j{}^{CL}}{\sum_{k=1}^{T-j} C_{k,j}} \right\} \qquad (4.12)$$

With this method, Mack (1993b) presented a completely non-parametric stochastic model. As non-parametric, this means that there is no probability distribution and that there is a proxy to the true underlying probability distribution beyond the CL. The prediction error formula obtained includes two sources of error, the process variance and the estimation variance.

As written before, in section 3.6, Mack (1993a) mentioned the CL connection to the weighted least squares regression and Free (2010) explicitly shows that the CL is a weighted regression.

4.3 Towards a Parametric Approach

As the Mack method was non-parametric, research continued to find a parametric stochastic model. It was already known that the lognormal approach brought some insights, but also some problems when we have negative incremental payments, which were not an issue in the traditional CL.

Another step forward was given by Renshaw and Verrall (1998), with the application of generalized linear models to claims reserving. This paper went back to Poisson distribution using the Kremer (1982) structure given in (4.2) but with a different modelling. It was assumed that the incremental claims $I_{i,j}$ followed a Poisson distribution with mean $\mu_{i,j}$ given by the same parameters as in (4.2).

$$\log \mu_{i,j} = a + r_i + c_j + \varepsilon_{i,j} \tag{4.13}$$

The model avoided the use of $\log I_{i,j}$ and the previous assumption of positive incremental claims. It just assumed that the sum of the incremental claims is positive in each column.

Renshaw and Verrall (1998) suggested that this model is more appropriate for triangles of claims numbers and presented two results for the estimates of the number of claims: the likelihood function with the Poisson distribution and the conditional likelihood function with a multinomial distribution with the row totals as the conditioning. Both approaches give the same result as CL.

Renshaw and Verrall (1998) concluded that using the quasi-likelihood function will also allow them to use this model for claim's amounts: using just the form of the Poisson model but not the distribution. Renshaw and Verrall (1998) develop it further with generalized linear models.

Meanwhile, other models where developed using the Gamma distribution to each payments triangle cell. It was the case with Mack (1991). The model was very difficult to implement and Renshaw and Verrall (1994) showed that the same results could be more easily obtained with generalized linear models and a gamma distribution.

Generalized linear models (GLM) may be seen on McCullagh, P. and Nelder, J. (1989). They extended classic linear modelling with the normal distribution to the exponential family of probability distributions. There were some non-normal models already at that time, such as the probit and logit models and the log-linear models, but GLM unified all these models in one general theory. After this, models such as the Poisson regression, the normal regression and the logistic regression become models of the GLM framework.

The GLM is an extension of the regression models. The motivation for that, see for example Taylor and McGuire (2016), is the fact that claims data exhibits skewness, with the mass of the data concentrated in smaller values, and a long right tail distribution for higher values, with more frequency than the anticipated by the normal distribution. This situation could only be overcome using logs and transforming the normal distribution on the lognormal. However, the, lognormal distribution can only be applied to positive data and sometimes insurers have negative payments, like reimbursements.

GLM can be used with any distribution from the exponential family such as the Poisson, the gamma, the normal, the inverse Gaussian and others, and we can achieve more flexibility on the dependent variable and independent variables relation using link functions. This allows us to deal with several data transformations, like the logs or the powers, and we are not restricted to the identity relation like in other models.

This general framework of GLM applies to any distribution from the exponential family (also called exponential dispersion family), and not just to the normal distribution. The former, to a generic variable y with parameters θ (location) and ϕ (dispersion), has a density function $p(y|\theta, \phi)$ of the following type

$$p(y|\theta,\phi) = exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

This expression includes several known distributions, and if we consider all of its observations (in our triangle defined by i,j) and following McCullagh and Nelder (1989), we get as expected value always a constant $\mu_{i,j}$ and as variance a value dependent on $a(\phi)$ and on $b''(\theta)$.

$$\mathbb{E}(y) = b'(\theta) = \mu_{i,j} \qquad Var(y) = a(\phi)b''(\theta)$$

The expected value $\mu_{i,j}$ is related with the linear predictor η by a monotone and differentiable function h, to each observation. The inverse of this function is called the link function, and for each observation from the claims reserving triangle we get

$$\mu_{i,j} = h(\eta_{i,j}) \qquad \eta_{i,j} = g(\mu_{i,j})$$

The $b''(\theta)$ depends on the parameter $\mu_{i,j}$ and it is called the variance function, $V(\mu_{i,j})$. The $a(\phi)$ is usually defined by the standardized dispersion coefficient, ϕ/ω, the dispersion coefficient divided by a known constant, ω.

Following Renshaw and Verrall (1998), claims reserving may be seen as a GLM problem. Using one link function for each row and column, we get the CL linear predictor $\eta_{i,j}$, where a is a constant and r_i and c_j row and column parameters.

$$\eta_{i,j} = a + r_i + c_j \tag{4.14}$$

These functions must be associated to a probabilistic model. Having the probabilistic model, and as we are going to predict the lower triangle, we know that we will have a prediction variance, Renshaw (1989). The process variance will come from the probability distribution we use. The estimation variance does not depend on the later but on the link function that we use.

England and Verrall (2002) presented the Over-Dispersed Poisson (ODP) in the context of GLM models. In this model, we know that the expected value is not equal to the variance because the latter is proportional to the former. Considering an ODP, we have

$$E(I_{i,j}) = \mu_{i,j} \tag{4.15}$$

$$Var(I_{i,j}) = \phi\ \mu_{i,j} \tag{4.16}$$

$$\eta_{i,j} = log(\mu_{i,j}) = a + r_i + c_j \qquad r_1 = c_1 = 0 \tag{4.17}$$

We also know that the quasi-likelihood from a Poisson distribution has the same structure from the likelihood function. England and Verrall (2002) get the following quasi-likelihood function l to be maximized to get the parameters vector (the loss development factors)

$$log\ \ l \propto I_{i,j}\ log(\mu_{i,j}) - \mu_{i,j}$$

To get the gamma we just need to make some slight changes. Now the model will be given by

$$E(I_{i,j}) = \mu_{i,j} \tag{4.18}$$

$$Var(I_{i,j}) = \phi\ {\mu_{i,j}}^2 \tag{4.19}$$

$$\eta_{i,j} = log(\mu_{i,j}) = a + r_i + c_j \qquad r_1 = c_1 = 0 \tag{4.20}$$

Hence, the following quasi-likelihood function is obtained to be maximized and to get the parameters vector

$$log\ l \propto -\frac{I_{i,j}}{\mu_{i,j}} - log(\mu_{i,j})$$

The ODP and Gamma models suggest a general model where we may have a general case in which the formers are included:

$$E(y_{i,j}) = \mu_{i,j} \tag{4.21}$$

$$Var(y_{i,j}) = \phi\ {\mu_{i,j}}^{\gamma} \tag{4.22}$$

$$\eta_{i,j} = log(\mu_{i,j}) = a + r_i + c_j \qquad r_1 = c_1 = 0 \tag{4.23}$$

For γ = 1 we get the ODP model and for γ = 2 we get the Gamma model. For γ = 3 we will get an Inverse Gaussian model. In the following equations and following England and Verrall (2002), we also present the mean square error of prediction in a general form

$$msep(\hat{R}) = \sum_{j>T-i+1}^{T} \phi {\hat{\mu}_{i,j}}^{\gamma} \quad + \sum_{j>T-i+1}^{T} {\hat{\mu}_{i,j}}^2\, Var(\hat{\eta}_{i,j})$$

$$+\ 2 \sum_{j>T-i+1, j_2>j_1}^{T} \hat{\mu}_{i,j_1}\hat{\mu}_{i,j_2}\, Cov(\hat{\eta}_{i,j_1}\hat{\eta}_{i,j_2})$$

As we saw before, one of the problems in claims reserving estimation with incremental data is the existence of negative numbers, due to reimbursements or recoveries that must be considered. In previous ODP and Gamma models, there are logarithms in formulas to be applied to payments and if they are negative something wrong will happen. This creates problems in most commercial software as they usually follow shortcuts to resolve the problem that does not correspond to the exact solution. For instance, it is the case of S-Plus with quasi-likelihood functions as we show in Portugal (2009). One possibility to overcome this is to replace the deviance function by the generalized Pearson statistic. The latter corresponds to the Pearson errors on the ODP model and may be used with negative numbers. The function is a good yardstick to the model accuracy (Turkman, 2000).

England and Verrall (2002) also present results for the negative binomial model that is also over-dispersed like the ODP model. With that, it is shown that the Mack (1993a) distribution-free model seems to be a normal distribution approximation to the negative binomial. Indeed, not only the loss development factors are the same (as expected), but the dispersion coefficient of the negative binomial shows to be similar to the variance parameter in the distribution-free model. The negative binomial model may be presented with incremental and cumulative claims. The results are the same (England and Verrall, 2002).

The GLM approach from England and Verrall (2002) also presents results for the lognormal and shows it is straightforward to use other linear predictors like the Hoerl curve and other smoothers. In a previous paper, it was also shown how to use generalized additive models as smoothers Verrall (1995).

4.4 Bootstrapping

Another alternative to get the prediction errors and confidence intervals to the estimated reserves is the use of the bootstrapping technique. The latter is a resampling technique that allows us to estimate the volatility of a certain variable. The resampling is done in the initial data several times and the inferences are produced from resamples done. It was introduced by Efron (1979) and developed with Efron and Tibshirani (1993).

For example, if we need to calculate the mean squared error of prediction for the reserves, we may generate several triangles, where each of the latter is the sum of our triangle with something else, like the errors produced on the reserves estimation. On each sample we obtain the reserves for that sample. All the samples together give a non-parametric distribution of the reserves and we will be able to obtain the mean squared error of prediction. The latter will be calculated doing the average of the differences between the reserves on all the samples obtained with its mean.

It was introduced in claims reserving by Ashe (1986). In this paper, three methods are analysed to measure the variability of outstanding claims: the jackknife, a parametric model for the distribution of aggregated claims and the bootstrap technique. Ashe (1986) concluded that it was not possible to qualify one technique as better than the others. Taylor (1988) also discussed this issue and his base was a paper on aggregated losses second moments from Taylor and Ashe (1983).

Lowe (1994) presented a first comparison of bootstrapping with the distribution-free technique of Mack (1993a) and the operational time of Wright (1990). Lowe (1994) concludes that the prediction errors are very similar, which was a good support to the use of this technique.

Some years later, several other papers published using the bootstrapping technique in claims reserving. Examples of this, are England and Verrall (1999) who presented the use of this technique in the context of the CL, and Pinheiro et al. (2003) who used generalized linear models and did not restrict the bootstrap to the CL.

England and Verrall (1999) use the Pearson errors ${\varepsilon^P}_{i,j}$ on each cell i,j to create another triangle where the new incremental claims amounts will be given by

$$I_{i,j}{}^{**} = {\varepsilon^P}_{i,j}\sqrt{\mu_{i,j}} + \mu_{i,j}$$

The model is refitted, and the statistic of interest calculated: the reserve R. The procedure is repeated several times. The bootstrap standard error will be the standard deviation of the reserve. The stored results from the sampling will give us the predictive distribution of reserves and its standard deviation is the prediction error.

Pinheiro et al. (2003) uses the standardized Pearson errors, instead of the Pearson errors, since only the former can be considered identically distributed (Efron and Tibshirani, 1993). With this procedure it will not be necessary to do the Pearson errors adjustment done by England and Verrall (1999).

As Wüthrich and Merz (2008) states:
“Efron’s bootstrap can essentially be applied to every stochastic claims reserving model that we have considered so far. If we have, in addition, distributional assumptions we can apply the parametric bootstrap method”

4.5 Bayesian Models

The mixture of information is used by actuarial science since 1914 (CAS, 1996) when American Actuaries started implementing experience rating models at workmen compensation premiums. It is usually said that this fact is in the origin of the Casualty Actuarial Society.

There were some applications of credibility theory to claims reserving with the reserve estimate being presented as a mix of two sources of information. One of such examples is De Vylder (1982). Mack (1990) improved this model by changing some of its assumptions and getting another estimator, which is a special case of the rating Buhlmann-Straub model (Buhlmann and Gisler, 2005).

The Bayesian theory was also applied to the log-linear models beyond CL by Verrall (1990) and as he stated:
“The empirical Bayes estimates have a credibility theory interpretation, and it is interesting to note that De Vylder (1982) obtained credibility-type estimates by applying the linear empirical Bayes theory directly to the CL technique.”

Gisler and Wüthrich (2008) will also present the CL stochastic method on a full Bayesian model. They show that if we use the exponential dispersion family of distributions, with its natural conjugate priors, the credibility estimators are the exact Bayesian estimates. It is a result similar to the one obtained on pricing (Buhlmann and Gisler, 2005). As expected, they also show that if we use a vague prior, without any information, we get the classical CL result.

We saw for BH method that we may linearly combine the CL and the BF reserve's estimates to get a new estimate. The BH credibility factor obtained was the grossing-up factor and Mack (2000) obtained the credibility factor that minimizes the mean squared error.

Verrall (2001) also presented a stochastic model closely related to the BF method using Bayesian statistics and the ODP model presented in Verrall (2000). The result shows that CL estimators may be obtained if we use improper priors to the row parameters. BF estimators will be obtained if we use strong prior information for the row parameters. These results also show a main assumption of the BF method, a complete knowledge about the ultimate benchmark behaviour. Once again, see section 3.4, this means the need for a good benchmark. Alai et al. (2009) also presented mean square errors of prediction for the BF method using the ODP model with the CL development pattern.

Schmidt (2006a) and Schmidt and Zocher (2008) generalized the BF method to what they call the Extended Bornhuetter-Ferguson (EBF). They showed that the stochastic BF may have two types of prior's estimators, one for the grossing-up factors and another one for the expected ultimate losses.

Mack (2008) presented prediction errors analytical formulas for BF using a distribution-free model. The paper also shows that the appropriate development pattern of BF is not the one of CL. For Mack (2008) the BF method has its own development pattern and may be seen as a standalone method independent from CL. The idea of questioning BF development pattern was already approached by Mack (2006), which proposed a method to overcome some of the CL difficulties.

Schmidt (2006a) has showed that many other stochastic models can be seen as BF predictors. England, Verrall and Wüthrich (2010) also concluded that Bayesian predictors may be seen as BF predictors. Saluz et al. (2011) also presented prediction errors analytical formulas for BF.

4.6 Multivariate Models

One of the recent trends in the literature on claims reserving is the development of models that consider more than one triangle estimation at the same time. It is the case of the bivariate Munich CL (Quarg and Mack, 2004) which estimates the paid and incurred claims at the same

time, using CL, and the so called multivariate models that estimate triangles from different lines of business at the same time, usually also using CL on all the triangles.
In the first case, it is imposed that CL is appropriate for paid claims and incurred claims, even when prediction errors are high. In the second case, the problem was similar but applied to different triangles from different lines of business.

At that time, it was also expected to have the aggregate CL result, from all the triangles, equal to the sum of the individual CL results from all the triangles. This problem was first studied by Holmberg (1994) who analysed the dependency between different triangles. Halliwell (1997) and Braun (2004) considered a bivariate model to the joint estimation of paid and incurred claims but estimating the loss development factor by the univariate model. Quarg and Mack (2004) Munich CL also developed a bivariate model but calculated the loss development factors within the bivariate model. Hurlimann (2006) presented bounds for the bivariate CL.

Merz and Wüthrich (2007b) developed a multivariate CL model. They estimated the CL development factors using the univariate estimate, but they presented a second paper one year later with those factors estimated in the multivariate way (Merz and Wüthrich, 2008). It was shown some years before by Prohl and Schmidt (2005) and Schmidt (2006b) that the CL univariate estimators are not optimal when we have correlated triangles, and most of the multivariate models are presented with CL. Some exceptions may be seen in Hess, Schmidt and Zocher (2008) and (Merz and Wüthrich, 2007a) for the AD method.

Zhang (2010) also presented a general multivariate CL with correlations and structural connections between the triangles and showed that the results from Prohl and Schmidt (2005) and Merz and Wüthrich (2008) models can be obtained from his model. In the same paper, Zhang also showed that the necessary and sufficient condition for the multivariate CL to be equivalent to the separate univariate CL is to have a diagonal variance-covariance matrix of the errors and proportional losses between the upper triangles of the different portfolios. He also discusses the differences between his estimators and the Munich CL results for a bivariate model.

4.7 Individual Claim Modelling

The models for individual claims aim to get information about individual claims. This information disappears with triangles, see for example Table 2.1, because they aggregate all the claims by origin year and development year. For instance, Antonio and Plat (2012) argue that the triangles are from the time of manual calculations, and that with the current computer power other more computer-intensive solutions should be developed.

Following Taylor et al. (2008), it seems that despite some papers from the 80's and the 90's the theoretical interest in individual claims reserving is recent:

"It appears that Norberg (1986 and 1993) and Jewell (1989 and 1990) were the first to attempt to lay down a comprehensive architecture for individual claims loss modelling. This framework has recently been developed by Larsen (2007). Other specific individual claim models appearing in the literature are due to Hachemeister (1980) and Haastrup and Arjas (1996)."

Most of these papers consider the claims as a Poisson process. Roselund (2012) presented a bootstrapping technique over the history of claims to generate samples that would allow reserving and mean squared error of prediction calculations. The method is very computer-intensive and, in some cases, required five hours of simulation. It also demands several dates from claim occurrence, claim reporting and finalization and payments dates and amounts. Roselund (2012) assumes that claims are identically and independently distributed, and that the historic claims used have the same distribution as the ones we want to study.

Antonio and Plat (2012) also presented a model for individual claim reserving creating a stochastic process to each claim phase using the Poisson process:

"The time of occurrence of the claim, the delay between occurrence and reporting of the claim, to the insurance company, the occurrence of payments and their size and the final settlement of the claim".

Parodi (2013) argues for the need for a triangle-free methodology, stating that the triangles are inherently inadequate to accurately model the distribution of reserves, although they may be good enough to produce a point estimate of such reserves. Parodi (2014) follows pricing methodologies splitting frequency from severity to mix them together with Monte Carlo simulation to produce the aggregate loss distribution. The claim's reserves are obtained from the latter.

These types of approaches, individual claim modelling, show several problems, and until now actuaries and the market are not using them. The reasons we see for that are the following:

- Some of the methods do not consider the IBNR claims.
- The implementation is too complex and running times of long hours.
- The methods require huge amounts of data, but despite this, they do not allow actuaries to see trends on data.
- With some exceptions, see Taylor (2008), the methods do not require information on covariates that explains the claim's cost, which is difficult to accept if we move to an individual claim level. For example, the information on the injured income will explain the level of reserves when we have disabilities, temporary or forever.
- Some of the information necessary to claim´s reserves, at the claim level, is deterministic and not stochastic, like the covariates that explain the compensation to the injured.
- They analyse something, the individual claims, that it is not necessary to calculate the overall reserve.
- The idea that the triangles come from the lack of computers is not accurate, as we saw in chapter 2, the triangles methodology just developed with the appearance of the microcomputers.
- It may be a lost battle to reserve a claim without all the information on it, just using averages or standard deviations, when compared with a loss adjuster that has several pieces of information to use.

However, and despite all this, this type of method may be useful for claims analysis. That may be the case when they are able to identify the covariates that explain the reserves and have a model for predicting some of its behaviour. Some practical applications of regression models have been done by some actuaries in some countries, using, as explanatory variables, the variables that explain the ultimate cost of the claim, very often in Motor insurance bodily injury claims. These works were never published.

The individual claims modelling may be important for individual claims analysis. However, the current state of the art ignores a lot of information from claims (everything is considered stochastic and there are several deterministic variables that are not considered, like the salary of the injured in bodily injured claims). These methods lose the forest view to just have a stochastic tree view. They are not substitutes from the triangle analysis and just a supplementary approach.

4.8 Conclusions

The start of the stochastic claims reserving moved away from the CL and several models were presented:

- Straub (1971) and the Kamreiter and Straub (1973) models, the latter using regression techniques.
- Verbeek (1972), Hachemeister and Stanard (1975) models based on the Poisson distribution.
- De Vylder (1978), multiplicative model.
- And Buhlmann and Straub (1980) probabilistic model.

However, these models did not influence a lot the developments done in the following years. The CL influence or replication will be the path of the literature and as we saw in chapter 3 the CL is a weighted regression. Examples of this are:

- Kremer (1982) relation with the CL. Kremer (1982) showed that CL and multiplicative models are equivalent and that the multiplicative model is equivalent to the additive model.
- Renshaw (1989), which is a development of Kremer (1982) model.
- Verrall (1991a) presents the relation between the parameters of the three alternatives to represent CL and shows that these three models are equivalent and are re-parameterisations of the same structure. Verrall (1991a) also showed the relation between the additive model and the CL.
- Mack (1993a and 1994) argues that his distribution-free method is the stochastic method underlying CL.
- Renshaw and Verrall (1998) development was done with the Poisson distribution and the Kremer (1982) structure within a GLM framework, which is an extension of regression models.

- England and Verrall (1999) and Pinheiro et al. (2003) developed the bootstrapping technique with CL model.
- England and Verrall (2002) showed that the ODP model gives the same results as the CL. The paper also showed that the GLM framework may also be used to have different models with the probability distributions from the exponential family.
- England and Verrall (2002) also showed that the Mack (1993a) method could be seen as a normal distribution approximation to the negative binomial distribution.
- And Bayesian stochastic models are also developed around CL, see for example Verrall (1990) and Gisler and Wüthrich (2008). Mack (2000), Verrall (2001), Schmidt (2006a), Mack (2008), Alai et al. (2009) and Saluz (2011) developed stochastic models for CL based models, like BF and BH. We already saw in chapter 3 that they are CL based.

The developments, from the last years that could be considered more important, involve the estimation of several triangles at the same time. However, in most of the papers the methods are also CL based:

- The development of dual methods, mixing paid claims with incurred claims, but again using the CL in both triangles, see for example Quarg and Mack (2004).
- The estimation of more than one triangle at the same time, with multivariate methods that consider contemporaneous correlations between the triangles. Examples may be seen in Prohl and Schmidt (2005), Schmidt (2006b), Gisler and Wüthrich (2008) and Zhang (2010). It is not clear, at this moment, that regulators will accept the reduction of the reserves from a line of business due to correlation effects between lines of business.
- The Zhang (2010) is another example of the use of regression models on claims reserving.

Finally, some models are triangle-free and use information from individual claims. This is a disruptive technique when compared with the traditional triangle approach. However, the individual claims technique is more important when we want to study the case reserves on the current files. The information from the individual files is not important when we want to have just one number for the best estimate of the total reserves. Due to this, we will concentrate on claims reserving methods that are based on triangles.

As we saw in chapters 3 and 4, the regression models are implicitly (and not so often explicitly) underlying the claims reserving models. In chapters 6 and 7, we will concentrate explicitly on developing claims reserving models that use regression techniques and minimize the prediction error.

According to Antonio and Plat (2012), what is important is to create methods that produce good best estimates for the reserves. An attempt to do this is done in chapter 5, using Mack (1993a, 1993b, 1994) framework, but with two changes that help achieve this objective.

We will use, in the following chapters, as the main indicator of good best estimates, the prediction error presented in (3.21). As we saw before, the prediction error includes the parameter error and the process error (but does not include the model error). We are aware that it is not the only indicator to consider on model selection, but it is a very important figure on the actuary decision. In the next chapter, other indicators will be presented for illustration and to complete the analysis of some of the results presented. The relation between the prediction error results and these other indicators conclusions is also presented.

5. Stochastic Vector Projection

In this chapter, we investigate that it might be possible to obtain much better results, if the development factors considered between the two columns are calculated by the Vector Projection (VP) method.

5.1 Vector Projection Fundamentals

As it is known from Straub (1988), the CL is just an approximation to the least square solution to the loss development factor estimation[1]. Straub (1988) shows that the loss development factor that arises from the minimization of the square of the errors is given by a regression without intercept.

Following Gentle (2007), this regression, through the origin, may be seen as a vector projection between two adjacent columns of our upper triangle. For example, in Table 2.1, this could be the column 2 projected using the column 1, for both columns with origin years 1 to 9.

The projection of the vector y onto the vector x is a new vector that corresponds to the y estimate, $\hat{y}$, see for example, Gentle (2007) for more details.

$$\hat{y} = \frac{<x,y>}{\|x\|^2} x \tag{5.1}$$

$< x, y >$ is the inner product of x and y, and $\| x \|^2$ is the 2-norm of x. This means that we get a new vector, which is based on the previous one, y, but now projected in the x direction.

This VP is a regression between two variables without an intercept term and it is an alternative approach to estimate any loss development factor between two development years (where we have at least two observations).

For example, the first loss development factor from Table 2.1 is given by

[1] Straub (1988) shows that the CL is just an approximation to the minimization of squares of errors. He did it without having any assumptions in respect of the errors.

	1	2			
	x	y		y . x	x^2
1	5012	8269		41444228	25120144
2	106	4285		454210	11236
3	3410	8992		30662720	11628100
4	5655	11555		65343525	31979025
5	1092	9565		10444980	1192464
6	1513	6445		9751285	2289169
7	557	4020		2239140	310249
8	1351	6947		9385397	1825201
9	3133	5395		16902535	9815689
10	2063				
			Sum	186628020	84171277
				(a)	(b)

Loss Development Factor = (a)/(b)

Following Murphy (1994) and Barnett and Zehnwirth (1999), it is known that the link ratios approach for reserving may also be considered as a regression without an intercept term for each of the loss development factors. For each observation used in the calculation, they show that we have

$$C_{i,j+1} = b_j C_{i,j} + \varepsilon_{i,j}. \tag{5.2}$$

Murphy (1994) and Barnett and Zehnwirth (1999) assume that the (unconditional) variance of each residual $\varepsilon_{i,j}$ is given by a constant σ_j^2 weighted by the $C_{i,j}^{\delta}$ on each observation,

$$Var(\varepsilon_{i,j}) = \sigma_j^2 C_{i,j}^{\delta}. \tag{5.3}$$

b_j is estimated from the data and corresponds to the loss development factor[2], and its value depends on the parameter δ and on the history of payments (if $\delta \neq 0$). Indeed, if $\delta = 1$, we get the CL loss development factor obtained in (2.6), i.e., a weighted average of the link

[2] Indeed, it makes a good sense to estimate the loss development factors as a regression through the origin, i.e., through the VP method. The loss development factors may be seen as weighted averages of link ratios and the latter are the ratios of two cumulative payments from different development years. These ratios are estimated by the slope of the line that summarizes the relation between the two cumulative payments. Straub (1988) shows that the slope that minimizes the sum of squares of errors is not the CL, but the regression through the origin, i.e., the VP in our case.

ratios, and if $\delta = 2$, we get the loss development factor from the SA method obtained in (3.1), i.e., a simple average of the link ratios.

These results have something in common; they always have *heteroscedastic*[3] errors when $\delta \neq 0$, that is the variance of the errors, (5.3), is not constant on each regression. Now, if $\delta = 0$ is assumed, the ordinary least squares regression without intercept is derived, i.e., a VP with constant variance of the errors, σ_j^2, across all the observations of each regression. This result is different from the heteroscedastic cases where $\delta \neq 0$.

Now, in VP, see Straub (1988), or Murphy (1994) or Barnett and Zehnwirth (1999), the loss development factors b_j, with $j = 1, \dots, T-1$, are estimated by $\hat{b}_j^{VP}$

$$\hat{b}_j^{VP} = \frac{\sum_{i=1}^{T-j} C_{i,j} C_{i,j+1}}{\sum_{i=1}^{T-j} C_{i,j}^2} \tag{5.4}$$

The cumulative payments may be obtained, as in the CL (Mack 1993a, 1994), using the following relation

$$\hat{C}_{i,j}^{VP} = C_{i,j-1} \hat{b}_j^{VP} \tag{5.5}$$

This means that in our example (see Tables 2.4 and 2.5), the loss development factor for the first column is 3.32, which is more consistent with the recent increase of the link ratios.

Indeed, it is known that the use of the regression models has several advantages in claims reserving, see section 3.6 and (3.12), Taylor (1978), Barnett and Zehnwirth (1999) and Frees (2010).

---

[3] In Murphy (1994) and Barnett and Zehnwirth (1999), the general regression through the origin by using assumptions to the errors is developed and they demonstrated that several models might be contacted because of it. Obviously, some of the models are heteroscedastic, such as the SA and the CL. In Murphy (1994), the VP method is introduced, but with homoscedastic errors. In the present paper, as it is clear in the text, the stochastic VP method of Murphy (1994) with heteroscedastic errors instead is introduced to compete with Mack (1993, 1994)'s stochastic heteroscedastic CL approach.

Furthermore, let us consider the situation where we use the CL method to estimate the loss development factors over a perfect triangle. In such a perfect triangle, in each column, the link ratios are always equal to the loss development factor. Because of this, after estimating the lower triangle, we get a prediction error of zero, as it is defined in Mack (1993a). A triangle like this can be the one presented in Table 5.1. In this theoretical case, the loss development factors from CL and VP are exactly the same and the prediction error, as defined in (3.21), is zero.

As we can easily verify, the VP gives the same results as the CL under some conditions. Indeed, this happens because the link ratios are totally stable on each development year. However, if this is not the case, the VP gives different results from the CL, and in some cases, it adapts better to the evolution of the link ratios (see the example given with Tables 2.4 and 2.5, where the VP is closer to the more recent link ratios).

However, as it will become clearer in the application part, we may have triangles where the VP approach may perform worse than CL. This will happen, as anticipated before, see section 2.5, if we have irregular data.

Consequently, in this chapter, see Portugal et al. (2017), we introduce an alternative to the distribution-free stochastic CL method of Mack (1993a), the stochastic VP, using also the well-known regression through the origin approach proposed by Murphy (1994), but with heteroscedastic errors instead, following similar arguments to the Mack (1993a, 1994) approach; that is, the variance is not constant over all the observations from each regression. This means that we will use the Murphy (1994) approach to estimate the loss development factors, the regression through the origin. However, we changed two things in respect of Murphy (1994) approach. Firstly, we consider the errors heteroscedastic. Secondly, we didn't consider a recursive formula to get the prediction error formula. Instead, we developed a non-recursive formula. As with Murphy (1994), the regression through the origin loss development factor comes from the least squares model. The VP is similar to CL, indeed it uses Mack framework, but with different assumptions for loss development factors estimation and the second moments. A more general class of methods is summarised in 5.2 and 5.3. The author believes that delta =2 is the best candidate and should be superior to the standard CL method which has delta = 1.

Moreover, the eminent stochastic Mack distribution-free framework may be further improved using this technique to some particular sets of data.

Table 5.1: Perfect Chain-Ladder Matrix of Cumulative Payments

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1000	1800	3060	4896	7344	10282	13366	16039	17643	17643
2	1100	1980	3366	5386	8078	11310	14703	17643	19408	
3	1331	2396	4073	6517	9775	13685	17790	21348		
4	1772	3189	5421	8674	13010	18214	23679			
5	2594	4669	7937	12699	19048	26668				
6	4177	7519	12782	20452	30678					
7	7400	13320	22645	36232						
8	14421	25958	44128							
9	30913	55643								
10	72890									

5.2 Method and Assumptions

In this section, the combined technique for estimating outstanding claims based on the VP methodology is proposed and developed in detail. The method is formulated on the Mack (1993a, 1993b, 1994) distribution-free method framework. In the first mentioned paper Mack (1993a) writes that the purpose of his method is *"to know the standard error of the chain ladder reserve estimates as a measure of the uncertainty contained in the data and in order to see whether the difference between the results of the chain ladder method and any other method is significant or not"*. Mack (1993a) didn't define how he got the CL to estimate the loss development factors. As he mentions in the above paper "the chain ladder method is probably the most popular method for estimating IBNR claims reserves" and "it seems to work with almost no assumptions". Mack didn't define the CL method; he just used the most popular method to estimate the loss development factors. Our purpose with this VP method is the same, but we use a different algorithm to get the loss development factors, the VP.

We also assume, about the second moments, that the VP has heteroscedastic errors inside each origin year, but the errors now are proportional to the square of payments. This assumption is a consequence of the way we estimate the loss development factors, i.e., a weighted average of the link ratios with weights given by the square of the cumulative payments. In the CL, the weights are given by the cumulative payments, see for example (2.7) or Mack (1993a, 1993b, 1994).

Thus, the cumulative payments $C_{i,j}$ between different origin years are independent (because the method does not consider any dependencies between origin years) and there exist some loss development factors b_j, such that $j = 1, \dots, T-1$, where we have for $1 \le i \le T$ and $1 \le j \le T$:

$$\mathbb{E}\left(C_{i,j+1}|C_{i,1}, \dots, C_{i,j}\right) = \mathbb{E}\left(C_{i,j+1}|C_{i,j}\right) \tag{5.1.1}$$

$$\mathbb{E}\left(C_{i,j+1}|C_{i,j}\right) = b_j C_{i,j}, \tag{5.1.2}$$

$$Var\left(C_{i,j+1}|C_{i,j}\right) = \sigma_j^2 C_{i,j}^2. \tag{5.1.3}$$

In this method, the way that the loss development factors are estimated is changed using the VP approach instead of the typical CL estimator. Those loss development factors are obtained as in Straub (1988) and Murphy (1994), by minimizing the sum of the square of the errors.

$$\hat{b}_j^{VP} = \frac{\sum_{i=1}^{T-j} C_{i,j} C_{i,j+1}}{\sum_{i=1}^{T-j} C_{i,j}^2}. \tag{5.1.4}$$

Additionally, the conditional variance, (5.1.3), differs with the derived results of Mack (1993a, 1993b, 1994) distribution-free method. Mack (1993a, 1993b, 1994) method considers the $Var\left(C_{i,j+1}|C_{i,j}\right) = \sigma_j^2 C_{i,j}$ because the CL loss development factors are $C_{i,j}$-weighted mean, see (2.7), of the link ratios, given by (2.1). In the VP, the loss development factors are $C_{i,j}^2$-weighted mean of the same link ratios, see (3.17). Due to that, we obtain the cumulative payments conditional variance given by (5.1.3).

To estimate σ_j^2 we obtained $\hat{\sigma}_j^{2,VP}$, the unbiased estimator for the VP method

$$\hat{\sigma}_j^{2,VP} = \frac{1}{T-j-1} \sum_{i=1}^{T-j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{b}_j^{VP}\right)^2. \tag{5.1.5}$$

The following result has initially been derived by Mack (1993a), and it is also valid in our VP approach because the result just depends on the true development factors and the cumulative payments. In other words, Lemma 5.1.1 makes it clear that (5.1.2) and the independency between the cumulative payments $C_{i,j}$ in each origin year are indeed implicit assumptions of the VP as well as the CL method.

Lemma 5.1.1 [Mack 1993] *Under the method assumption* (5.1.2) *and for* $D_u = \{C_{i,j}: i + j - 1 \leq T\}$, *a recursive algorithm is derived for the calculation of the ultimate cost on an origin year* i *based on the upper triangle information* D_u*:*

$$\mathbb{E}\left(C_{i,T}|D_u\right) = \mathbb{E}\left(C_{i,T}|C_{i,T-i+1}\right) = b_{T-1} \cdots b_{T-i+1} C_{i,T-i+1}. \tag{5.1.6}$$

■

The VP estimator of the unknown loss development factors is given by Eq. (5.1.3), and the ultimate cost estimator is

$$\hat{C}_{i,T}^{VP} = C_{i,T-i+1} \hat{b}_{T-i+1}^{VP} \cdots \hat{b}_{T-1}^{VP}. \tag{5.1.7}$$

5.3 Estimation

Several properties from the VP estimator Eq. (5.1.4) are presented in the next Lemma:

Lemma 5.1.2 *For the VP estimators (5.1.4),* $\hat{b}_j^{VP}$ *for* $j = 1, \ldots, T - 1$, *the following properties are derived.*

a) *They are unbiased.*
b) *They are uncorrelated.*
c) *Given* D_u, *the estimator of the ultimate costs is an unbiased estimator of the true value,* $\mathbb{E}\left(\hat{C}_{i,T}^{VP}|D_u\right) = \mathbb{E}\left(C_{i,T}|D_u\right)$.
d) *They are weighted average of the intermediate link ratios* $F_{i,j+1} = \frac{C_{i,j+1}}{C_{i,j}}$, *with the weights to be given by the square of the payments.*

$$\hat{b}_j^{VP} = \frac{\sum_{i=1}^{T-j} C_{i,j}^2 F_{i,j+1}}{\sum_{i=1}^{T-j} C_{i,j}^2}.$$

e) *The conditional variance of* $\hat{b}_j^{VP}$ *which has minimal condition variance among all unbiased linear combinations of the unbiased estimators* $\left(F_{i,j+1}\right)_{1\le i\le T-j}$ *for* b_j *conditional on* D_u *is given by*

$$Var\left(\hat{b}_j^{VP}|D_u\right) = \left(\sum_{i=1}^{T-j} \frac{1}{\sigma_j^2}\right)^{-1}.$$

Similarly, the covariance is given by

$$Cov\left(\left(\frac{C_{i,j+1}}{C_{i,j}}, \hat{b}_j^{VP}\right)\middle| D_u\right) = \sigma_j^2 \frac{C_{i,j}^2}{\sum_{i=1}^{T-j} C_{i,j}^2}.$$

Proof. (a) It is straightforward to show that $\mathbb{E}\left(\hat{b}_j^{VP}|D_u\right) = b_j$ *, for* $j = 1, \dots, T-1$, i.e.,

$$\mathbb{E}\left(\hat{b}_j^{VP}|D_u\right) = \frac{\sum_{i=1}^{T-j} C_{i,j}\mathbb{E}\left(C_{i,j+1}|D_u\right)}{\sum_{i=1}^{T-j} C_{i,j}^2} = b_j.$$

(b) The proof is similar to the one obtained by Mack (1993b, 1994), so it is omitted. A proof can also be found in Wüthrich and Merz (2008).

(c) Considering

$$\mathbb{E}\left(\hat{C}_{i,T}^{VP}|C_{i,T-i+1}\right) = \mathbb{E}\left(C_{i,T-i+1}\hat{b}_{T-i+1}^{VP}\cdots\hat{b}_{T-1}^{VP}|C_{i,T-i+1}\right)$$
$$= b_{T-1}\mathbb{E}\left(C_{i,T-i+1}\hat{b}_{T-i+1}^{VP}\cdots\hat{b}_{T-2}^{VP}|C_{i,T-i+1}\right) = b_{T-1}\mathbb{E}\left(\hat{C}_{i,T-1}^{VP}|C_{i,T-i+1}\right).$$

This means that $\mathbb{E}\left(\hat{C}_{i,T-1}^{VP}|C_{i,T-i+1}\right) = b_{T-2}\mathbb{E}\left(\hat{C}_{i,T-2}^{VP}|C_{i,T-i+1}\right)$.

Continuing the iteration, we get

$$\mathbb{E}\left(\hat{C}_{i,T}^{VP}|C_{i,T-i+1}\right) = b_{T-1}b_{T-2}\cdots b_1\mathbb{E}\left(\hat{C}_{i,1}^{VP}|C_{i,T-i+1}\right) = \mathbb{E}\left(C_{i,T}|D_u\right),$$

As $\mathbb{E}\left(C_{i,T}|D_u\right)$, does not depend on the loss development factor calculation, the result is also similar to the one obtained by the Mack (1993b, 1994) distribution-free method. A proof may also be found in Wüthrich and Merz (2008).

(d) $\hat{b}_j^{VP} = \frac{\sum_{i=1}^{T-j} C_{i,j}C_{i,j+1}}{\sum_{i=1}^{T-j} C_{i,j}^2} = \frac{\sum_{i=1}^{T-j} C_{i,j}^2 F_{i,j+1}}{\sum_{i=1}^{T-j} C_{i,j}^2}.$

(e) Considering the two Lemmas 3.3 and 3.4 in Wüthrich and Merz (2008), and the fact that

$$Var\left(F_{i,j+1}\middle|D_u\right) = Var\left(F_{i,j+1}\middle|C_{i,j}\right) = \sigma_j^2,$$

then

$$Var\left(\hat{b}_j^{VP}\middle|D_u\right) = \left(\sum_{i=1}^{T-j}\frac{1}{\sigma_j^2}\right)^{-1} = \frac{\sigma_j^2}{T-j}$$

and the covariance is given by

$$Cov\left(\left(\frac{C_{i,j+1}}{C_{i,j}}, \hat{b}_j^{VP}\right)\middle|D_u\right) = \frac{C_{i,j}^2}{\sum_{i=1}^{T-j} C_{i,j}^2} Var\left(\frac{C_{i,j+1}}{C_{i,j}}\middle|D_u\right) = \sigma_j^2 \frac{C_{i,j}^2}{\sum_{i=1}^{T-j} C_{i,j}^2}.$$

■

Remark 5.1.1 Based on the Gauss-Markov theorem, the VP loss development factors, $\hat{b}_j^{VP}$, are the best linear unbiased estimator, see Fomby et al. (1984). Moreover, the loss development factor's variances and covariance are a function of σ^2, and the past observations, thus, $\hat{b}_j^{VP}$ are the ones with a lower variance.

Remark 5.1.2 Now, the loss development factors in (5.1.4) can also be given considering (3.19), the general Murphy (1994) assumption for the conditional variance i.e., $Var\left(C_{i,j+1}|C_{i,j}\right) = \sigma_j^2 C_{i,j}^{\delta}$. Then, the estimator is given by $\hat{b}_j^{Mu} = \frac{\sum_i C_{i,j}^{1-\delta} C_{i,j+1}}{\sum_i C_{i,j}^{2-\delta}}$. If one sets $\delta = 0$, the VP estimator is derived, and equivalently, $\hat{b}_j^{VP} \equiv \hat{b}_j^{Mu}$.

Lemma 5.1.3 *Under the method assumptions (5.1.1), (5.1.2) and (5.1.3) the estimator which is given by* (5.1.5) *is an unbiased estimator of* σ_j^2.

Proof. For (5.1.5) to be unbiased we need

$$\mathbb{E}\left(\hat{\sigma}_j^{2,VP}|D_u\right) = \frac{1}{T-j-1}\sum_{i=1}^{T-j} \mathbb{E}\left[\left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{b}_j^{VP}\right)^2 |D_u\right].$$

Then, the expected value is provided. It may be decomposed as follows

$$\mathbb{E}\left[\left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{b}_j^{VP}\right)^2 |D_u\right]$$
$$= \mathbb{E}\left[\left(\frac{C_{i,j+1}}{C_{i,j}} - b_j\right)^2 |D_u\right] + 2\mathbb{E}\left[\left(\frac{C_{i,j+1}}{C_{i,j}} - b_j\right)\left(b_j - \hat{b}_j^{VP}\right)|D_u\right]$$
$$+ \mathbb{E}\left[\left(\hat{b}_j^{VP} - b_j\right)^2 |D_u\right].$$

Developing the first component on the right side,

$$\mathbb{E}\left[\left(\frac{C_{i,j+1}}{C_{i,j}} - b_j\right)^2 |D_u\right] = Var\left(\frac{C_{i,j+1}}{C_{i,j}}\middle| D_u\right) = \sigma_j^2.$$

Using Lemma 5.1.2, the second component is derived

$$\mathbb{E}\left[\left(\frac{C_{i,j+1}}{C_{i,j}} - b_j\right)\left(\hat{b}_j^{VP} - b_j\right)|D_u\right] = Cov\left(\left(\frac{C_{i,j+1}}{C_{i,j}}, \hat{b}_j^{VP}\right)\middle| D_u\right) = \sigma_j^2 \frac{C_{i,j}^2}{\sum_{i=1}^{T-j} C_{i,j}^2}.$$

This means that

$$-2\mathbb{E}\left[\left(\frac{C_{i,j+1}}{C_{i,j}} - b_j\right)\left(\hat{b}_j^{VP} - b_j\right)|D_u\right] = -2\sigma_j^2 \frac{C_{i,j}^2}{\sum_{i=1}^{T-j} C_{i,j}^2}.$$

Because for the last component, we also get from Lemma 5.1.2 that

$$\mathbb{E}\left[\left(\hat{b}_j^{VP} - b_j\right)^2 |D_u\right] = Var\left(\hat{b}_j^{VP}|D_u\right) = \frac{\sigma_j^2}{T-j}.$$

Adding all the three components together, we get

$$\mathbb{E}\left(\hat{\sigma}_j^{2,VP}|D_u\right) = \frac{1}{T-j-1}\sum_{i=1}^{T-j}\left(\sigma_j^2 - 2\sigma_j^2 \frac{C_{i,j}^2}{\sum_{i=1}^{T-j} C_{i,j}^2} + \frac{\sigma_j^2}{T-j}\right) =$$

$$\frac{1}{T-j-1}\sigma_j^2(T-j-2+1) = \sigma_j^2.$$

■

5.4 Prediction Error

With the following results, the calculation of the mean squared error of prediction (*msep*) is provided. First, the necessary Lemma for the connection between the *msep* of the estimated reserves and claims is given.

Lemma 5.1.4 [Mack, 1993a] *The msep of the estimated reserves and claims is equal.* ■

Theorem 5.1.1 *Under the assumptions* (5.1.1), (5.1.2) *and* (5.1.3), *where all the origin years are independent and there are unbiased estimators for the loss development factor and the variance parameter, the mean squared error for each origin year reserve, $msep(\hat{R}_i)$, can be estimated by using (3.21)*

$$\widehat{msep(\hat{R}_\iota)} = \hat{C}_{i,T}^{2,VP} \sum_{j=T-i+1}^{T-1} \frac{\hat{\sigma}_j^{2,VP}}{\hat{b}_j^{2,VP}} \left(\frac{1}{\hat{C}_{i,j}^{VP^2}} + \frac{1}{T-j} \right), \tag{5.1.8}$$

where $\hat{\sigma}_j^{2,VP}$, $\hat{b}_j^{2,VP}$ and $\hat{C}_{i,T}^{VP}$ are given by Eqs. (5.1.5), (5.1.4) and (5.1.7).

Proof. Considering (3.21), the Mack (1993a) definition of the *msep*

$$msep(\hat{R}_i) = Var(C_{i,T}|D_u) + [\mathbb{E}(C_{i,T}|D_u) - \hat{C}_{i,T}^{VP}]^2.$$

And using the following abbreviations

$$\mathbb{E}_i(X) = \mathbb{E}(X|C_{i,1}, \dots, C_{i,T-i+1})$$

$$Var(X) = Var(X|C_{i,1}, \dots, C_{i,T-i+1})$$

By the law of total variance

$$Var(C_{i,T}|D_u) = \mathbb{E}_i\left(Var(C_{i,T}|C_{i,1}, \dots, C_{i,T-1})\right) + Var_i\left(\mathbb{E}(C_{i,T}|C_{i,1}, \dots, C_{i,T-1})\right)$$

Applying (5.1.3) to the first term and (5.1.2) to the second term, we get

$$Var(C_{i,T}|D_u) = \mathbb{E}_i(C_{i,T-1}^2) + \hat{b}_{T-1}^{2,VP} Var_i(C_{i,T-1})$$

Repeated use of (5.1.6) and (5.1.7) and knowing that $\mathbb{E}_i(C_{i,T-i+1}^2) = C_{i,T-i+1}^2$

$$Var(C_{i,T}|D_u) = [\sigma_{T-2}^2 \sigma_{T-1}^2 + \hat{b}_{T-2}^{2,VP} \sigma_{T-1}^2 + \hat{b}_{T-1}^{2,VP} \sigma_{T-2}^2] \mathbb{E}_i(C_{i,T-2}^2) + \hat{b}_{T-1}^{2,VP} \hat{b}_{T-2}^{2,VP} Var_i(C_{i,T-2})$$

Knowing that $\mathbb{E}_i\left(C_{i,T-i+1}^2\right) = C_{i,T-i+1}^2$ and that $Var_i\left(C_{i,T-i+1}^2\right) = 0$

$$Var\left(C_{i,T}\middle|D_u\right) = C_{i,T-r+1}^2 \sum_{k=T-i+1}^{T-1} \hat{b}_{T-i+1}^{2,VP} \cdots \hat{b}_{k-1}^{2,VP} \sigma_k^2 \hat{b}_{k+1}^{2,VP} \cdots \hat{b}_{T-1}^{2,VP} = \hat{C}_{i,T}^{2,VP} \sum_{j=T-i+1}^{T-1} \frac{\hat{\sigma}_j^{2,VP}}{\hat{b}_j^{2,VP}} \frac{1}{\hat{C}_{i,j}^{VP^2}}.$$

Moreover, considering the Mack (1993a)'s proof of the second component (we just need to substitute the CL loss development factors estimators by the VP), we get

$$\left[\mathbb{E}\left(C_{i,T}\middle|D_u\right) - \hat{C}_{i,j}^{VP}\right]^2 = C_{T-i+1}^2\left(b_{T-i+1} \cdots b_{T-1} - \hat{b}_{T-i+1}^{VP} \cdots \hat{b}_{T-1}^{VP}\right)^2 = C_{T-i+1}^2 F^2,$$

with

$$F = b_{T-i+1} \ldots b_{T-1} - \hat{b}_{T-i+1}^{VP} \cdots \hat{b}_{T-1}^{VP} = S_{T-i+1} + \cdots + S_{T-1},$$

and

$$S_j = \hat{b}_{T-i+1}^{VP} \cdots \hat{b}_{j-1}^{VP}\left(b_j - \hat{b}_j^{VP}\right) b_{j+1} \cdots b_{T-1}.$$

Hence

$$F^2 = \sum_{j=T-i+1}^{T-j} S_j^2 + 2\sum_{i<j} S_i S_j = \mathbb{E}\left(S_j^2\middle|D_u\right) + 2\mathbb{E}\left(S_i S_j\middle|D_u\right).$$

Following Mack (1993a), as the estimator for the loss development factor is unbiased, see Lemma 5.1.2, we have that $\mathbb{E}\left(S_i S_j\middle|D_u\right) = 0$. Consequently, we get $F^2 = \mathbb{E}\left(S_j^2\middle|D_u\right)$, thus we just need to have the variance of the estimator, see Lemma 5.1.2 e)

$$\mathbb{E}\left(\left(b_j - \hat{b}_j^{VP}\right)^2\middle|D_u\right) = Var\left(\hat{b}_j^{VP}\middle|D_u\right) = \frac{\sigma_j^2}{T-j}.$$

Following Mack (1993a) generic formula, we get now

$$\mathbb{E}\left(S_j^2\middle|D_u\right) = \frac{\hat{b}_{T-i+1}^{2,VP} \cdots \hat{b}_{j-1}^{2,VP} \sigma_j^2 b_{j+1}^2 \cdots b_{T-1}^2}{T-j}.$$

Using $F^2 = \sum_{j=T-i+1}^{T-j} S_j^2$, the estimators of the loss development factors and of the variance parameter, we get

$$\left[\mathbb{E}\left(C_{i,T}|D_u\right) - \hat{C}_{i,j}^{VP}\right]^2 = \hat{C}_{i,T}^{2,VP} \sum_{j=T-i+1}^{T-1} \frac{\hat{\sigma}_j^{2,VP}}{\hat{b}_j^{2,VP}}\left(\frac{1}{T-j}\right).$$

And finally, we have

$$msep\left(\hat{R}_i\right) = \hat{C}_{i,T}^{2,VP} \sum_{j=T-i+1}^{T-1} \frac{\hat{\sigma}_j^{2,VP}}{\hat{b}_j^{2,VP}}\left(\frac{1}{\hat{C}_{i,j}^{VP\,2}} + \frac{1}{T-j}\right).$$

This leads to the estimator stated in the Theorem. ■

From Theorem 5.1.1, we observe that the mean squared error of prediction given by (5.1.8), for each origin year, is similar to the Mack (1993a) CL prediction error, i.e., it depends on the level of square of estimated ultimate claims, and on the sum of the variance estimator for each development factor weighted by the estimator of a development factor and a multiplicative factor. However, there are also some differences. Obviously, the estimators are obtained through the VP instead of the CL method, and the multiplicative factor has now two different components. The first component is the inverse from the square of the estimated cumulative payments. In Mack (1993a, 1994) method, this component is the inverse of the estimated cumulative payments. The reasoning behind this difference lies on the assumption (5.1.3) from the VP method, where the conditional variance of payments depends on the square[4] of payments. The second component is totally different between the VP and the CL methods. In the former, it does not depend on the inverse of the sum of payments, as the CL does, but depends only on the inverse of number of years to complete the evolution of each development factor, i.e., $1/(T-j)$. This is due to the conditional variance of the VP loss development factor, see Lemma 5.1.2-e.

In the CL method, this factor is the inverse of sum of future cumulative payments. The next Corollary completes the theoretical findings.

Corollary 5.1.1 *With the assumptions and notations of Theorem 5.1.1, the msep of the total reserve estimate for every origin year* i, $\hat{R} = \sum_{i=2}^{T} R_i = \hat{R}_2 + \hat{R}_3 + \cdots + \hat{R}_T$ *can be given by*

$$msep\left(\hat{R}\right) = \sum_{i=2}^{T}\left\{mse\left(\hat{R}_i\right) + \hat{C}_{i,T}^{VP}\left(\sum_{j=i+1}^{T} C_{j,T}\right)\sum_{j=T-i+1}^{T-1}\frac{2\hat{\sigma}_j^{2,VP}/\hat{b}_j^{2,VP}}{T-j}\right\}. \quad (5.1.9)$$

[4]In the Mack (1993a, 1993b, 1994)'s model, it depends on the payments.

Proof. The result comes immediately following Mack (1993) using

$$\mathbb{E}\left(S_j^2 \middle| D_u\right) = \frac{\hat{b}_{T-i+1}^{2,VP} \cdots \hat{b}_{j-1}^{2,VP} \sigma_j^2 b_{j+1}^2 \cdots b_{T-1}^2}{T-j}. \qquad \blacksquare$$

As we saw before, the $msep$ for the total reserve is similar to the Mack CL prediction error. However, the estimators of the payments, the reserves $msep$ and the ratio of variance to the square of loss development factors, are all obtained using the VP results.

Additionally, as it was the case above, the last component is totally different between the VP and CL methods.

In the former, it does not depend on the inverse of the sum of payments, like the CL does, but just on the inverse of number of years to develop i.e., $1/(T-j)$.

5.5 Numerical Examples

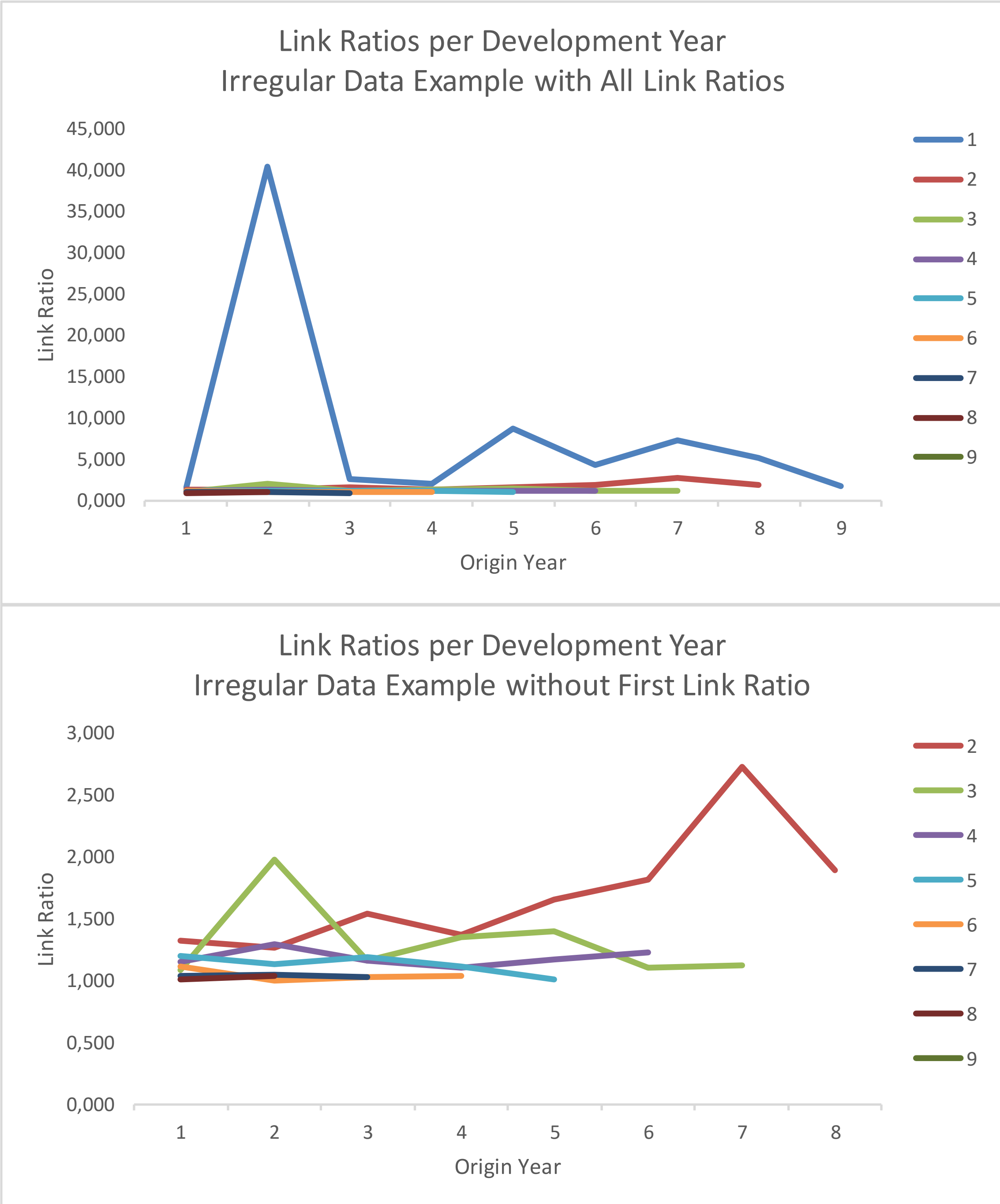
We give here two types of examples with different datasets. In one of the examples, we call the data as with irregular development.

We may see an example in Table 2.1, and in it corresponding link ratios in Table 2.2. The data is from Mack (1993a).

If we analyse each column, we see that there is a big range from link ratios on column 1: it goes from 1.650 until 40.425. In column 2, these link ratios have also some variability and they go from 1.259 until 2.723. In all the other columns, they are between 0.993 and 1.977. Also, inside column 1, the variability is very high even if we skip the 40.425 outlier: the range goes from 1.650 until 8.759 and in some origin years it is below 2.

Figure 5.1 summarizes this evolution (the second graph of the figure is equal to the first one without the first origin year).

Figure 5.1: Irregular Data Example



In the other example, we have two datasets, and both are considered has regular data: Taylor and Ashe (1983), see Table 5.2 and Taylor and McGuire (2016), see Table 5.3. We will see that these two datasets are totally different from the one with irregular data.

Table 5.2: Triangle of cumulative payments, Taylor and Ashe (1983)

	1	2	3	4	5	6	7	8	9	10
1	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
2	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	
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315		
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268			
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311				
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712					
7	440 832	1 288 463	2 419 861	3 483 130						
8	359 480	1 421 128	2 864 498							
9	376 686	1 363 294								
10	344 014									

Table 5.3: Triangle of cumulative payments, Taylor and McGuire (2016)

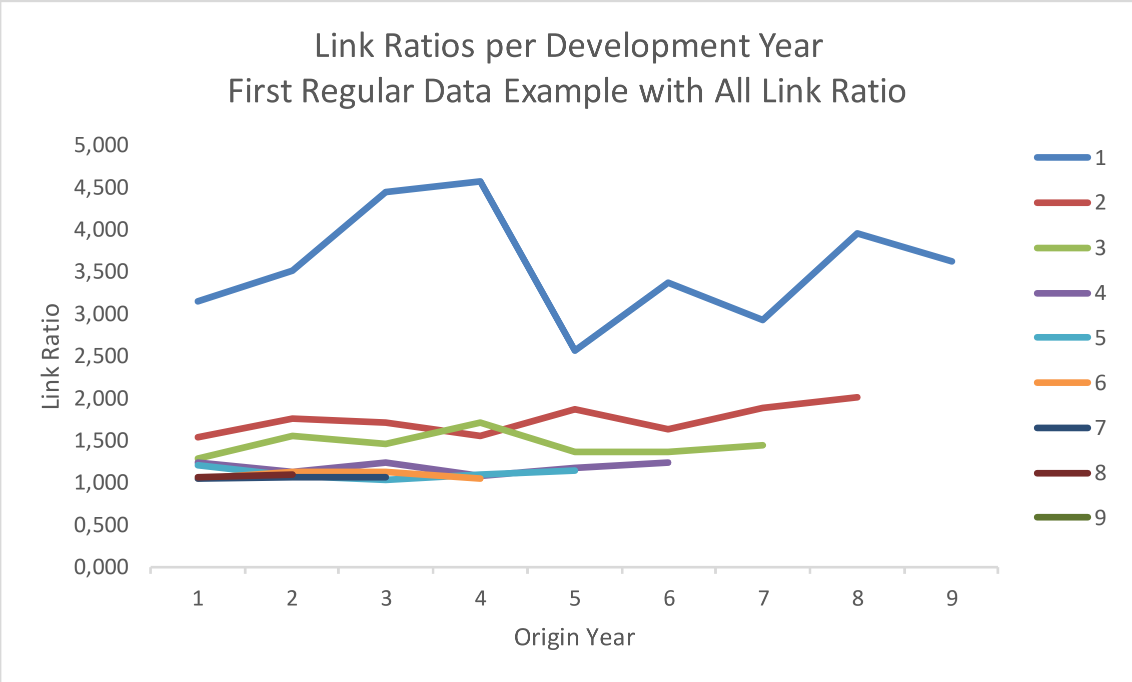
	1	2	3	4	5	6	7	8	9	10
1	45 630	68 980	71 904	73 702	75 709	76 913	78 211	78 774	79 551	80 172
2	53 025	79 491	82 320	84 068	84 800	86 224	86 623	87 160	87 500	
3	67 318	109 651	107 797	110 975	114 020	117 301	120 210	122 823		
4	93 489	130 962	138 393	145 041	149 248	155 010	156 900			
5	80 517	113 578	120 441	124 769	128 772	131 122				
6	68 690	102 621	108 266	114 444	117 923					
7	63 091	95 289	104 227	111 106						
8	64 430	96 921	105 335							
9	68 548	103 914								
10	76 013									

The same analysis done for Mack (1993a) data is now done for these two datasets in the following Figures 5.2 and 5.3. In both cases, Taylor and Ashe (1983) data in the first example, and Taylor and McGuire (2016) data in the second example, there are no outliers.

We may also see that, in both datasets, payments evolution is stable. Also, the first link ratios, in both datasets, are higher than the others because most of the payments are done in the second development year. However, the range of the first link ratios is much smaller than the one detected with irregular data (even when the outlier from irregular data is not considered). When we consider the other link ratios we may also see that in the irregular data they are between 1 and 2.7, but with the regular datasets they are between 1 and 2 and between 0.98 and 1.1.

This means that the irregular data is irregular due to the existence of an outlier but also due to the higher range of the link ratios.

Figure 5.2: First Regular Data Example



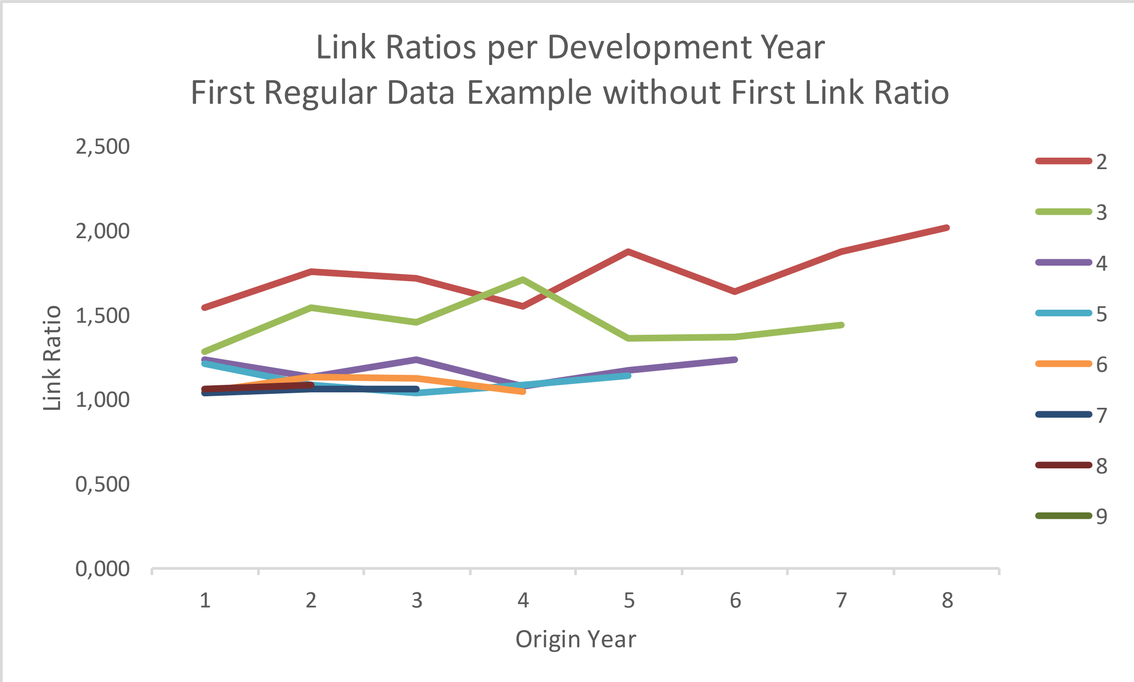
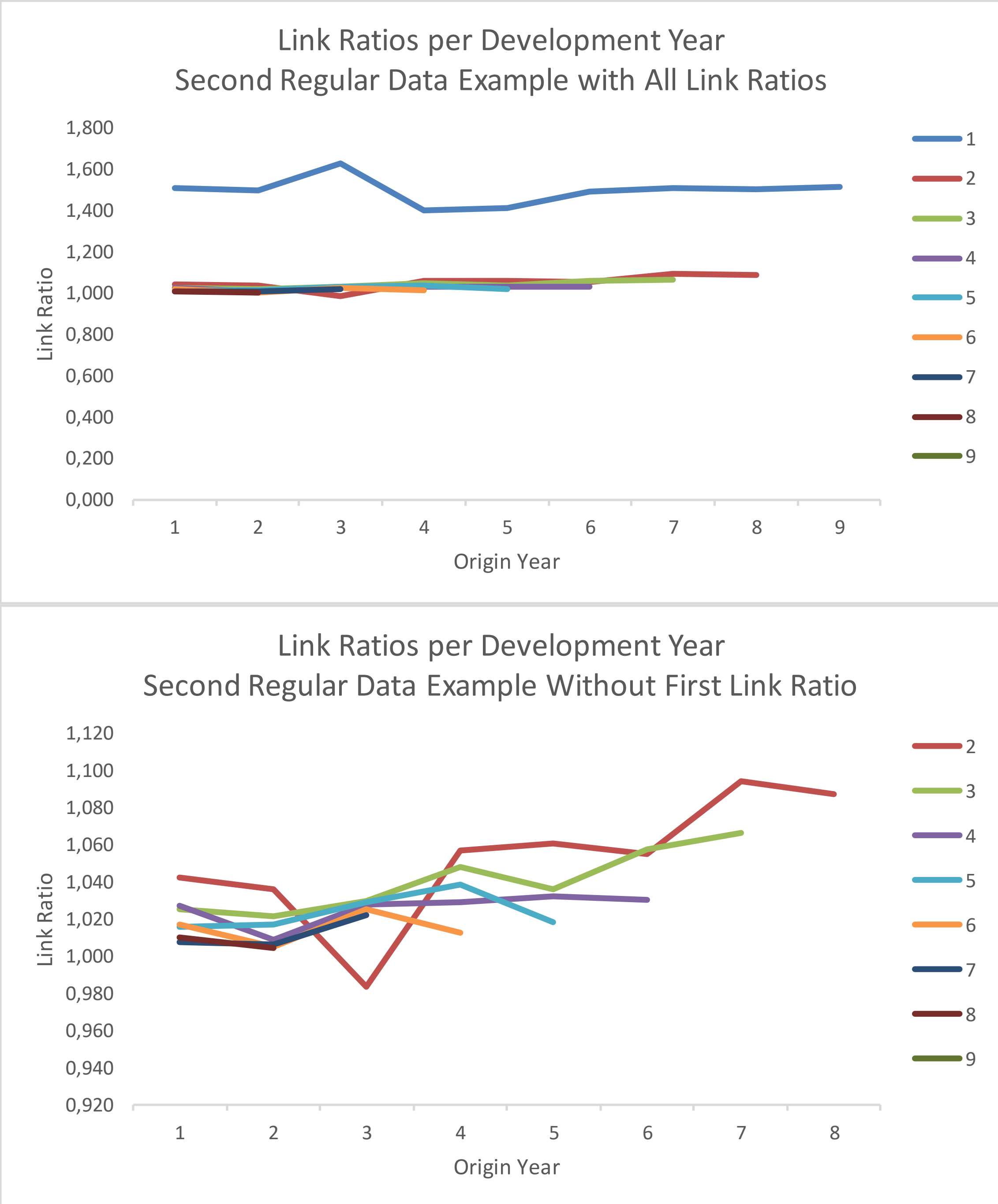
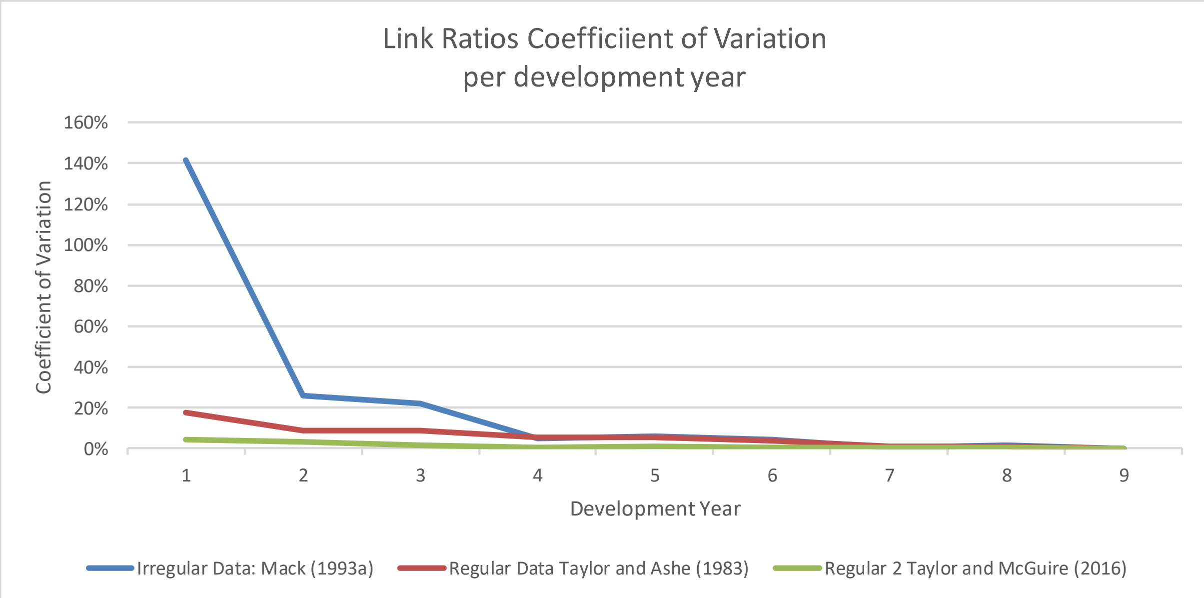


Figure 5.3: Second Regular Data Example



Finally, in Figure 5.4, we compare all the datasets using the coefficient of variation. Then we may see that the regular data has always the link ratios coefficient of variation below 20%, whatever the development year is. This does not happen with the irregular data, where just after the fourth year of development the link ratios coefficient of variation is below 20%.

Figure 5.4: Comparing Regular Data with Irregular Data



5.5.1. Irregular Development of Data

In this section, cumulative payments $C_{i,j}$ data from Table 5.1 is used to illustrate the comparison between the two claims reserving methodologies. The dataset is used by Mack (1993a).

Indeed, using Table 5.1, which has very irregular (extreme) development of data, we can observe that both the VP and CL have high prediction errors (see Tables 5.4 and 5.5). The prediction errors are, 52% for the CL and 63% for the VP.

Table 5.4: Stochastic Vector Projection with Irregular Data

Development Year	Loss Development Factors	Variance
2	2,217	192,637
3	1,569	0,243
4	1,261	0,104
5	1,162	0,005
6	1,100	0,007
7	1,041	0,003
8	1,032	0,000
9	1,016	0,000
10	1,009	0,000

Origin Year	Estimated Reserves	Prediction Error
2	154	97%
3	593	71%
4	1 577	33%
5	2 648	33%
6	3 344	26%
7	5 013	18%
8	10 151	25%
9	9 623	24%
10	10 670	250%
Total	**43 772**	**63%**

Table 5.5: Mack (1993a) Distribution-Free Method with Irregular Data

Development Year	Loss Development Factors	Variance
2	2,999	27 883,479
3	1,624	1 108,526
4	1,271	691,443
5	1,172	61,230
6	1,113	119,439
7	1,042	40,820
8	1,033	1,343
9	1,017	7,883
10	1,009	1,343

Origin Year	Estimated Reserves	Prediction Error
2	154	134%
3	617	101%
4	1 636	46%
5	2 747	53%
6	3 649	55%
7	5 435	41%
8	10 907	49%
9	10 650	59%
10	16 339	150%
Total	**52 135**	**52%**

Due to the strong irregular development of data, there is not a good fit in both methods and the VP does not improve the CL results. The CL estimates 19% more of reserves than the VP does. This is due to the features of the triangle for the first two development years:

- The payments were higher in the past (year 1, 3 and 4 are the ones with more payments, see Table 2.1) but on those years the link ratios were lower (see Table 2.2).

- The existence of higher link ratios on more recent years (see Table 2.2) is associated with an outlier in the past. Figure 5.1 shows that the most recent link ratios are higher (when compared with the link ratios from the past).

As the VP gives more weight to the older link ratios, when compared with the CL, its fit is not so good. This happens because the VP weights with the square of payments, see Lemma 5.1.2, and the CL with the payments, see (2.7).

The same happens with the link ratio outlier on origin year 2, see Figure 5.1 and Table 2.2. The VP gives a higher weight to this link ratio (due to the use of the square of the payments as weights), something that does not happen so much with the CL (as just the payments are used as weights). Practically, this means that the CL smooths more the effect of this link ratio outlier.

In the following subsections, regular development of data is used.

5.5.2 Regular Development of Data

Example 1: Data from Taylor and Ashe (1983)

In this subsection, we now consider a different set of data also used by Mack (1993a), and originally from Taylor and Ashe (1983).

With this more regular triangle, both the VP and the CL have lower prediction errors. However, the VP presents a smaller prediction error than the CL, i.e., 9% for the former and 13% for the latter, see Tables 5.6 and 5.7.

Table 5.6: Stochastic Vector Projection with Regular Data from Example 1

Parameters	
Loss Development Factors	**Variance**
3,418	0,472
1,749	0,029
1,462	0,020
1,167	0,005
1,097	0,005
1,087	0,002
1,055	0,000
1,078	0,000
1,018	0,000

Origin Year	Estimated Reserves	Prediction Error
2	94 634	63%
3	478 103	18%
4	723 104	12%
5	1 002 041	13%
6	1 408 034	14%
7	2 131 332	12%
8	3 885 296	10%
9	4 255 237	9%
10	4 501 720	10%
Total	**18 479 500**	**9%**

Table 5.7: Mack (1993a) Distribution-Free Method Regular Data Example 1

Parameters	
Loss Development Factors	**Variance**
3,491	160280,327
1,747	37736,855
1,457	41965,213
1,174	15182,903
1,104	13731,324
1,086	8185,772
1,054	446,617
1,077	1147,366
1,018	446,617

Origin Year	Estimated Reserves	Prediction Error
2	94 634	80%
3	469 511	26%
4	709 638	19%
5	984 889	27%
6	1 419 459	29%
7	2 177 641	26%
8	3 920 301	22%
9	4 278 972	23%
10	4 625 811	29%
Total	**18 680 856**	**13%**

The difference of the estimated reserves, between the VP and the CL, is of -1%.

We may use these results, see section 3.6.3, to have the reserves stochastic estimators with a certain confidence level, for example 99.5%. This allows us to have the estimated reserves added of a risk margin[5], that is we will obtain the reserve fair value for a 99.5% confidence level. For this degree of confidence, we will expect a probability of 0.5% to have the reserves lower than the future payments.

Following Mack (1993b) and section 3.6.3, we calculated the 99.5% confidence level upper bound for the VP and the CL reserves. We used the Mack (1993b) approach with the normal distribution to calculate the upper bound of the confidence interval. For a 99.5% confidence level, the VP stochastic reserve is of 22 624 853 and the same CL stochastic reserve is of 24 984 154 money-units. This means that although the VP best estimate is just -1% lower than the CL best estimate, the stochastic difference, for 99.5% of confidence level, is of +10%.

Example 2: Data from Taylor and McGuire (2016)

We consider now a different set of very regular data used recently by Taylor and McGuire (2016). It is even more regular than the previous one. Under this new triangle, the VP and the CL have lower prediction errors. However, the VP presents a smaller prediction error than the CL, 1.3% for the former and 2.9% for the latter, see Tables 5.8 and 5.9.

Table 5.8: Stochastic Vector Projection with Regular Data from Example 2

Development Year	Loss Development Factors	Variance
2	1,812	0,008
3	1,260	0,000
4	1,158	0,000
5	1,088	0,000
6	1,056	0,000
7	1,039	0,000
8	1,030	0,000
9	1,025	0,000
10	1,021	0,000

[5] The risk margin is important not only to give us a measure of uncertainty of our estimate, but also because it is one of the components of claims reserving on the new Solvency II regime when internal models are considered. The Fair Value of reserves is the sum of best estimate with a risk margin.

Origin Year	Estimated Reserves	Prediction Error
2	3 398	0,0%
3	8 155	0,1%
4	14 608	1,6%
5	22 719	1,8%
6	32 025	2,0%
7	45 870	1,9%
8	60 175	1,5%
9	80 926	1,4%
10	105 594	2,5%
Total	**373 469**	**1,3%**

Table 5.9: Mack (1993a) Distribution-Free Method Regular Data Example 2

Development Year	Loss Development Factors	Variance
2	1,815	449,408
3	1,261	22,347
4	1,158	8,575
5	1,088	7,547
6	1,055	4,294
7	1,039	1,887
8	1,030	0,576
9	1,025	0,001
10	1,021	0,000

Origin Year	Estimated Reserves	Prediction Error
2	3 398	0,0%
3	8 155	0,2%
4	14 579	2,8%
5	22 645	3,7%
6	31 865	4,3%
7	45 753	4,3%
8	60 093	3,8%
9	80 983	3,9%
10	105 874	8,7%
Total	**373 346**	**2,9%**

We may also see that the difference of estimated reserves, between the VP and the CL, is also very small, but due to the low prediction error from the VP, its stochastic reserve will also be lower than the one from CL, about -4%.

Remark 5.2.2.1 It should be mentioned here that the prediction error is not the only quantitative criterion to follow when analysing the triangle results. Other items should also be addressed, such as the errors and the back-testing, when these methods are considered. A recent and good example of this may be seen in Taylor and McGuire (2016). We will analyse these items in the following sections.

5.6 Use Test

In this section, to compare the conclusions derived by the previously described three numerical examples and to provide also a "business orientated" analysis, we select 114[6] triangles randomly[7] with paid claims and 10 years of information to be comparable directly to the previous cases. Table 5.10 reports the derived results. Additionally, Figure 5.5 provides a comparison between the prediction errors calculated based on CL and VP methods, respectively.

Among the 114 triangles studied, the VP has a lower prediction error and lower reserve estimation in 65 cases. In the 32 other cases, despite the lower prediction error, the estimated reserve of the VP is higher. The CL from Mack (1993a, 1993b, 1994) has a lower prediction error only in 17 cases and in 16 of them produced a higher level of reserves. Thus, based on this data, we can conclude that the VP has a lower prediction error in 85% (97 out of 114) of these cases. The VP reserves are lower in 71% (81 out of 114) of the cases, with an average reduction in reserves of 2%.

Table 5.10: Summary of Results of the Use Test

	VP Prediction Error Lower	CL Prediction Error Lower	Total
VP Reserves Lower	65	16	81
CL Reserves Lower	32	1	33
Total	97	17	114

Moreover, we concluded that the triangles with the lower CL prediction errors have the following features:

- In most of these 17 cases, there are some special situations, such as cells with zeros (five triangles) or cells with negative cumulative payments (nine triangles).

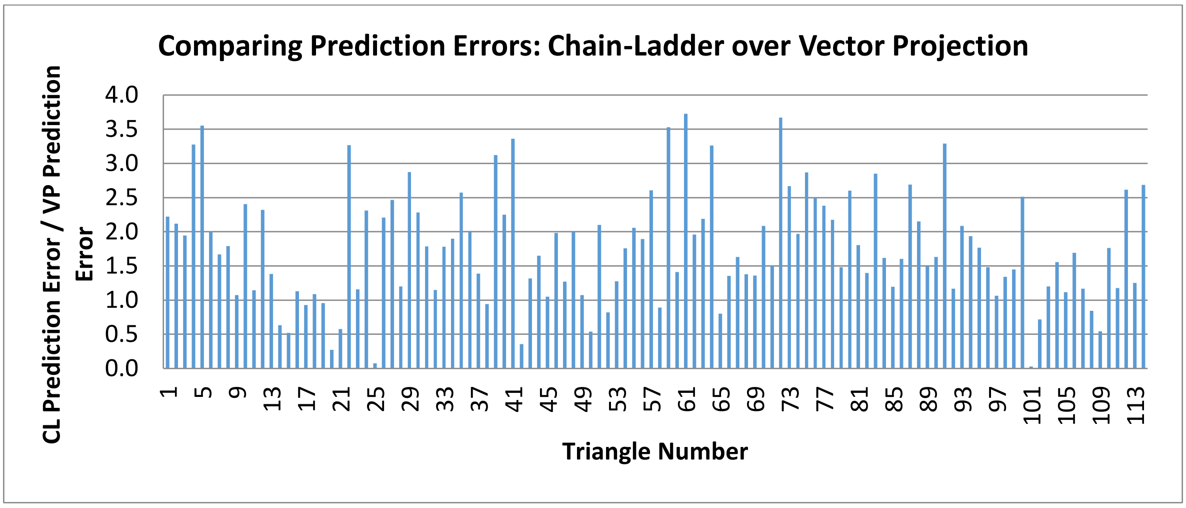
---

[6] The data set used is from the consulting company, Actuarial Group, Lisbon Portugal. Obviously, it is not possible to disclose any further information about the triangles used and their orientation. Table 5.10 and Figure 5.5 are for illustration properties and useful for our "business orientated" analysis. The interested readers and particularly the practitioners can use their own data to evaluate and to reconfirm our findings.

[7] Practically, we mean that the dataset is chosen by using different companies and for different periods of businesses.

- There are three triangles that appear to share something similar, i.e., the payments increase with the origin year until a certain point and then start decreasing.

Figure 5.5: Chain Ladder Prediction Error / Vector Projection Prediction Error



5.7 Selecting a Method

So far, we just considered the prediction error as a quantitative tool to compare the CL with the VP. Indeed, the prediction error is very important because it summarizes how well the method fits the experience. This feature is fundamental for the LRT methods (as we saw in chapter 2 and 3, the CL and the VP are LRT methods). That happens because the LRT methods assumes that the past helps to explain the future. If the prediction error from one method is lower than the prediction error from another method, this means that the former method fits better the experience than the latter method. Consequently, that method with a lower prediction error is more able to use the past to predict the future.

However, there are other considerations to be made when selecting a method to predict the reserves. We may have qualitative and quantitative tools to choose a method.

As qualitative tools we have several:

- Purpose of the analysis: if we are calculating the reserves to satisfy regulatory requirements there may be methods recommended by the regulators.

- Line of business: for example, with credit insurance it is important to be able to have a good prediction of reimbursements. If a method fails to do this, it may not be recommended for that line of business.
- Governance requirements: some companies may impose some constrains on method selection, for example, they may not accept methods not used by the market.
- Law: for example, the Solvency II requires best estimates, which means that we should not consider methods that always produce an upward (or downward) estimation of the reserves.
- Data requirements: if the method is supposed to be used annually and quarterly, and if we do not have some of the inputs for quarter reports, it may be wise to consider an annual method that may be used quarterly.
- And stability: some methods are more robust to new data than others and some insurers prefer methods that are more stable, to avoid big changes on reserves. For instance, the Last Link Ratio method presented in section 3.1 is very sensitive to new data, as it just relies on one link ratio, the most recent one. The MED method, also presented in section 3.1, is more robust, as we considered several link ratios and select the median of all of them. An additional link ratio, coming from new data, will not change necessarily the median.

  The CL weights the link ratios with the payments, see (2.7). If the latter increase, a new observation will have more weight. The same happens with the VP, but in this case, as the weights are the square of the payments, the weight of the recent observation with the payments increase is even bigger.

Some quantitative tools are also available to analyse the method's assumptions and results, see for example Mack (1993b). The first ones are the errors given by (3.2) and calculated retrospectively for the cells of the upper triangle, from development year T until development year 1 and using incremental payments (instead of cumulative payments), see for example Booth et al. (2005). This means that we need to have, for the upper triangle, the incremental payments and the estimated incremental payments. The former is obtained from (3.10) but the latter must be calculated. They are obtained using the loss development factors estimated from the claims reserving method considered for the calculations. For that, we use the following relation that allows us to calculate the estimated payments $\hat{C}_{i,j}$ for the upper triangle where $j \leq T - i + 1$

$$\hat{C}_{1,T} = C_{1,T} \qquad\qquad \hat{C}_{i,j} = \hat{C}_{i,j+1}/\hat{b}_j \quad j \leq T - i + 1 \tag{5.4.1}$$

Then we calculate the incremental payments using (3.10). We do this calculation for the observed cumulative payments and for the estimated payments given by (5.4.1). Finally, we get the errors for the upper triangle doing the difference between the incremental payments and the incremental payments calculated retrospectively.

Booth et al. (2005) writes that the errors:
"*are a basic measure which can be used to test how appropriate models are, given the underlying data".*

Errors analysis may be used to highlight regions of the upper triangle with a poor fit and is based on the upper triangle. In the following test, back-testing, we will do an analysis with the lower triangle.

The back-testing is a technique for validating internal models under Solvency II, see for example, in this respect, European Union (2015). This allows analysing discrepancies between the results provided by a model and the real observations. The errors approach, presented before, may be seen as a back-testing but does not allow us to consider a requirement from the Solvency II internal models: the stability of the results over time. A good model should not bring too much variation on the results as new information arrives, because any new events not yet in data should have been already considered by the model, European Union (2009 and 2015).

It is not common to see papers about back-testing claims reserving. Meyers and Shi (2011) mention:
"That the sparsity of studies on retrospective tests might be attributed to the unavailability of the data on realized claims".
The same paper highlights that this may be overcome with Bayesian models or with the use of the bootstrapping technique.

Following the Solvency II requirements, see European Union (2009 and 2015), we do a back-testing analysing the results stability of a specific claims reserving method within several calendar years. This means estimating the ultimate costs per origin year for different sub-sets of the triangle and seeing the evolution of the ultimate costs over time. We excluded from the analysis the first calendar year (there is just one cell in the triangle and is not possible to get any result) and the second calendar year (whatever the method the result is not significant and usually the same for all the link ratios methods).

For example, we may start back-testing with the data from the first, second and third calendar years, that is $i = 1,2,3$ and $j = 1,2,3$, and estimate the ultimate costs for the second and third origin years (we are assuming in all the triangles that the ultimate cost from the first origin year is already known). Then we do the same, adding one more calendar year and get another estimate of the ultimate costs (now including also $i = 4$). We repeat this until we have ultimate costs per origin year for all the calendar years.

5.7.1 Mack Data

Here we present the results for Mack (1993a) data using the CL and the VP. The following Table 5.11 presents the errors (with retrospective calculation) and Table 5.12 the same errors standardized by the incremental payments.

Table 5.11: Errors with CL for Mack Data

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	2 901	-964	-1 311	-1 887	-509	906	1 113	8	-257	0
2	-1 784	401	-2 423	2 777	1 108	263	-743	144	257	
3	710	184	-168	-1 293	-274	1 259	-265	-152		
4	2 437	-533	-1 807	1 255	-1 260	12	-105			
5	-2 151	1 989	206	2 055	341	-2 441				
6	-673	561	1 169	-1 651	594					
7	-1 433	-515	3 205	-1 257						
8	-1 342	212	1 129							
9	1 334	-1 334								
10	0									

Table 5.12: Standardized Errors with CL for Mack Data

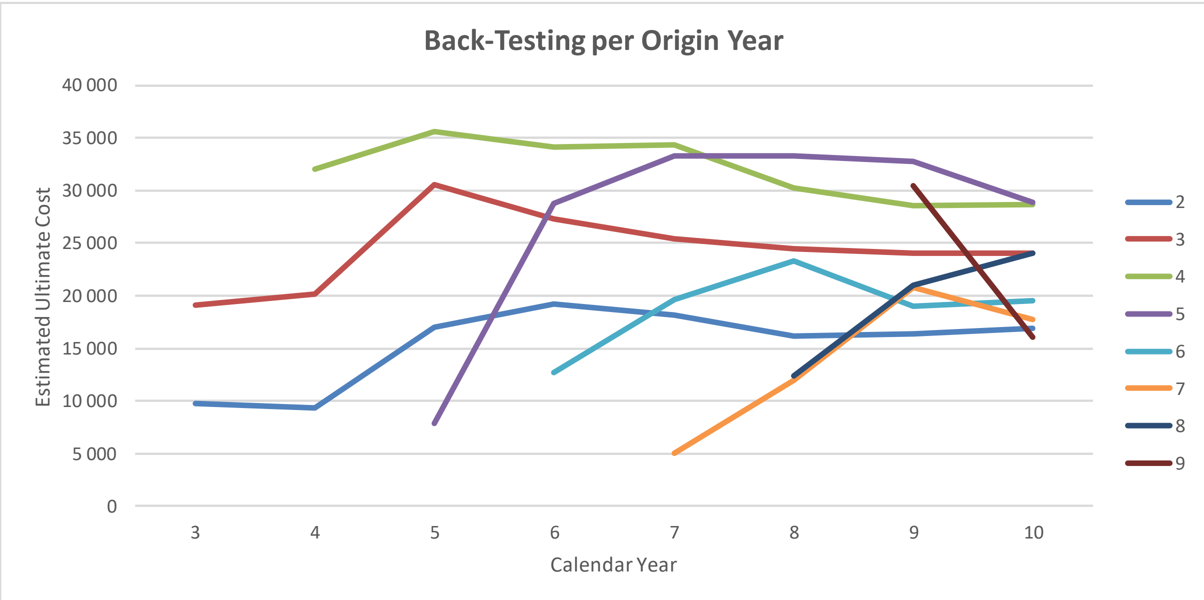
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0,579	-0,296	-0,497	**-2,101**	-0,294	0,343	0,609	0,014	**-4,756**	0,000
2	**-16,829**	0,096	**-2,181**	0,527	0,356	0,145	**7,212**	0,214	0,480	
3	0,208	0,033	-0,034	-0,570	-0,106	0,362	-0,408	-0,253		
4	0,431	-0,090	-0,429	0,228	-0,583	0,005	-0,107			
5	**-1,970**	0,235	0,033	0,325	0,090	**-10,849**				
6	-0,445	0,114	0,222	**-1,339**	0,204					
7	**-2,572**	-0,149	0,463	-0,919						
8	-0,993	0,038	0,183							
9	0,426	-0,590								
10	0,000									

We may see that the weight of the errors on incremental payments (the standardized errors) is high in several cells, for example, (2,1), (5,1), (7,1), (2,3), (1,4), (6,1), (5,6), (2,7) and (1,9). Of all these cells, two of them, (2,1) and (5,6), have very high standardized errors. The model has difficulty in explaining the evolution of the second and fifth year of origin but there are several other years with standardized errors around 0.5, that is 50% of the incremental payments.

We may also see in the following Figure 5.6 that the CL is not producing stable estimates of the ultimate costs over the years:

- In most of the origin years the CL increases the ultimate costs significantly over time.
- After a certain year of development, the opposite happens, and the CL starts correcting downwards the original forecast.
- The number of years of development, for these two effects to emerge, is not always the same.
- Finally, the path described above as an inversion for the more recent year considered, the year 9, where we see that the first estimate of the ultimate cost is much higher than the current one. This is an important fact as the year 9 just has two years of development and is far from being closed.

Figure 5.6: Back-Testing CL with Mack Data



Doing the same analysis for the VP we get the following Tables 5.13 and 5.14.

Table 5.13: Data Errors with VP for Mack Data

	1	2	3	4	5	6	7	8	9	10
1	1 960	-457	-1 212	-1 871	-434	1 091	1 135	26	-238	0
2	-2 625	854	-2 335	2 791	1 175	429	-724	160	274	
3	-488	837	-36	-1 270	-175	1 498	-237	-129		
4	1 014	251	-1 644	1 288	-1 138	300	-70			
5	-3 579	2 787	379	2 094	468	-2 149				
6	-1 597	1 146	1 333	-1 590	707					
7	-2 250	46	3 384	-1 180						
8	-2 418	1 008	1 410							
9	700	-700								
10	0									

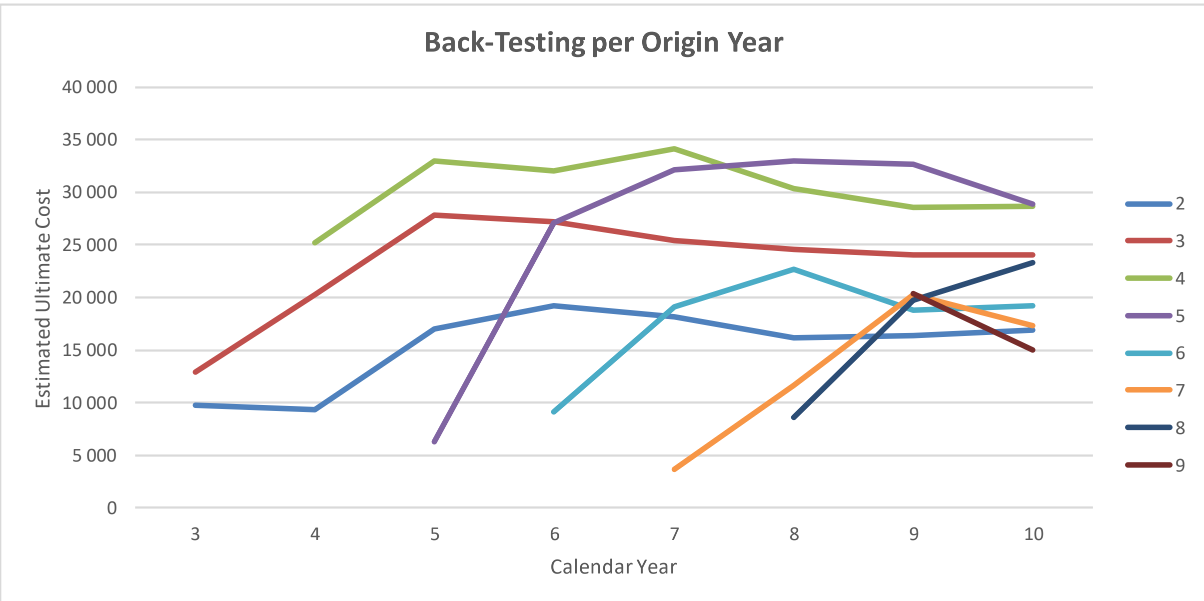
Table 5.14: Data Standardized Errors with VP for Mack Data

	1	2	3	4	5	6	7	8	9	10
1	0,391	-0,140	-0,459	**-2,084**	-0,250	0,413	0,621	0,043	**-4,405**	0,000
2	**-24,768**	0,204	**-2,101**	0,530	0,377	0,236	**7,025**	0,238	0,512	
3	-0,143	0,150	-0,007	-0,560	-0,068	0,431	-0,365	-0,214		
4	0,179	0,043	-0,390	0,234	-0,527	0,113	-0,072			
5	**-3,277**	0,329	0,060	0,331	0,123	**-9,550**				
6	**-1,056**	0,232	0,254	**-1,289**	0,242					
7	**-4,040**	0,013	0,489	-0,862						
8	**-1,790**	0,180	0,229							
9	0,223	-0,309								
10	0,000									

We may see that the weight of the errors on incremental payments is high in the same cells as the CL. However, there are some extra cells in the first development year with high standardized errors, the (6,1) and (8,1).

We may also see in the following Figure 5.7 that the VP, as the CL in Figure 5.6, is not producing stable estimates of the ultimate costs over the years. Also, the problems detected with the CL are also emerging with the VP.

Figure 5.7: Back-Testing VP with Mack Data



We must remember here that both the CL and the VP presented very high prediction errors for this data, see section 5.2.1. Putting together all the analysis done, we may conclude that with the irregular data considered here, Mack (1993a), both the CL and the VP do not fit properly the data. Because of this, there are high errors, unstable estimates of the ultimate costs and high prediction errors. Comparing both methods, the CL and the VP, we know that the former minimizes the square of the errors, Straub (1988). However, in this analysis, Straub (1988) did the calculation of the regression errors and not of the errors presented above (with retrospective calculation).

Mack (1993b) showed that using a regression framework is very useful, as the usual regression analysis instruments become available (and the CL may be seen as a weighted-regression). Mack (1993b) also consider for analysis the regression errors and not the errors calculated retrospectively. In the following tables, we present the regression errors and the standardized regression residual from the CL and the VP. The standardized regression errors are obtained dividing the regression errors obtained as in (3.2) by the observed cumulative payments $C_{i,j}$. With the regression errors the VP has a sum of the square of errors of 204 640 676 and the CL of 258 245 586.

Using CL, the first regression (which errors are presented on column 2 of the Tables 5.15 and 5.16) is the one with higher standardized regression errors, the regression errors divided by the observed payments, see Table 5.16. In this column, five years of origin have errors that represent more than 0.5 of the observed payments.

Table 5.15: Regression Errors with CL for Mack Data

	1	2	3	4	5	6	7	8	9	10
1	0	-6 764	-2 518	-2 057	-293	1 107	1 149	0	-261	0
2	0	3 967	-1 561	3 808	1 285	254	-757	158	261	
3	0	-1 236	-726	-1 490	-177	1 355	-283	-158		
4	0	-5 406	-2 994	1 229	-1 492	2	-110			
5	0	6 290	307	2 043	-20	-2 718				
6	0	1 907	1 238	-1 937	696					
7	0	2 349	4 419	-1 597						
8	0	2 895	1 833							
9	0	-4 002								
10	0									

Table 5.16: Data Standardized Regression Errors with CL for Mack Data

	1	2	3	4	5	6	7	8	9	10
1	0	**-0,8180**	-0,2309	-0,1742	-0,0216	0,0684	0,0638	0,0000	-0,0140	0,0000
2	0	**0,9258**	-0,2893	0,3570	0,0932	0,0163	-0,0489	0,0097	0,0156	
3	0	-0,1374	-0,0523	-0,0923	-0,0094	0,0610	-0,0124	-0,0067		
4	0	-0,4679	-0,1899	0,0578	-0,0637	0,0001	-0,0041			
5	0	**0,6576**	0,0194	0,0922	-0,0008	-0,1038				
6	0	0,2959	0,1058	-0,1497	0,0439					
7	0	**0,5844**	0,4037	-0,1297						
8	0	0,4167	0,1398							
9	0	**-0,7418**								

With the VP, see Table 5.18, the regression of the column 2 is also the one with higher standardized regression errors. In this column, five years of origin have errors that represent more than 0.5 of the observed payments. It is the same conclusion we got from the CL.

However, here the VP, for the same origin year 9, has a much lower residual, when compared with the CL. This happens because the link ratio from year 9, see Table 2.1, is much lower than the historical link ratios and the VP method gives more weight to the years with more payments (see Lemma 5.1.2), which is the case in year 9 for the development year 1. Giving more weight to this link ratio made the VP to have a better fit to this cell.

Table 5.17: Data Regression Errors with VP for Mack Data

	1	2	3	4	5	6	7	8	9	10
1	0	-2 844	-2 067	-1 948	-178	1 292	1 172	19	-242	0
2	0	4 050	-1 327	3 862	1 388	443	-735	174	278	
3	0	1 431	-235	-1 351	-20	1 611	-251	-133		
4	0	-983	-2 363	1 387	-1 285	322	-73			
5	0	7 144	829	2 202	195	-2 363				
6	0	3 090	1 590	-1 820	822					
7	0	2 785	4 639	-1 488						
8	0	3 952	2 212							
9	0	-1 552								
10	0									

Table 5.18: Data Standardized Regression Errors with VP for Mack Data

	1	2	3	4	5	6	7	8	9	10
1	0	**-0,3439**	-0,1895	-0,1650	-0,0132	0,0799	0,0651	0,0010	-0,0129	0,0000
2	0	**0,9452**	-0,2459	0,3621	0,1007	0,0284	-0,0475	0,0108	0,0166	
3	0	0,1592	-0,0169	-0,0837	-0,0011	0,0725	-0,0110	-0,0057		
4	0	-0,0851	-0,1499	0,0652	-0,0549	0,0124	-0,0027			
5	0	**0,7469**	0,0523	0,0993	0,0075	-0,0903				
6	0	0,4795	0,1359	-0,1407	0,0518					
7	0	**0,6928**	0,4238	-0,1208						
8	0	**0,5688**	0,1687							
9	0	-0,2876								

5.7.2 Taylor and Ashe Data

Here we present the same results as in section 5.4.1 but now for Taylor and Ashe (1983) data. When compared with Mack (1993a) data, Taylor and Ashe (1983) data is more regular and fitting the CL and the VP gave lower prediction errors, see Tables 5.4 and 5.5.

Table 5.19: Errors with CL for Taylor and Ashe Data

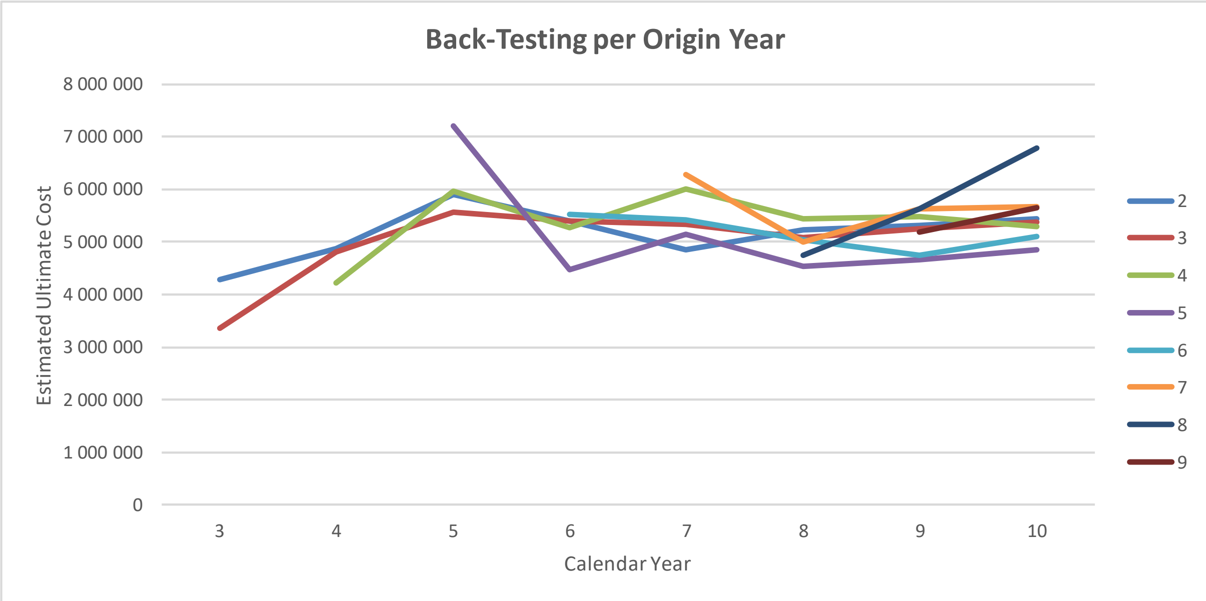
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	87 787	94 323	-93 952	-270 498	109 976	281 827	-122 002	-42 085	-45 377	0
2	-24 007	-52 758	-47 282	133 947	-135 515	-86 478	154 072	12 645	45 377	
3	-81 818	74 483	-45 045	-22 087	175 428	-256 435	126 035	29 439		
4	-56 116	194 885	-180 463	539 286	-294 249	-45 237	-158 105			
5	106 873	-144 369	114 729	-168 712	-14 844	106 323				
6	42 334	55 913	-75 435	-182 016	159 204					
7	48 990	-128 292	109 223	-29 920						
8	-110 168	-108 059	218 227							
9	-13 875	13 875								
10	0									

Table 5.20: Standardized Errors with CL for Taylor and Ashe Data

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	**0,245**	0,123	-0,154	**-0,560**	**0,209**	**0,491**	**-0,834**	**-0,301**	**-0,200**	0,000
2	-0,068	-0,060	-0,051	0,113	**-0,304**	**-0,269**	**0,292**	0,048	0,107	
3	**-0,282**	0,074	-0,049	-0,022	**0,234**	**-1,745**	**0,254**	0,105		
4	-0,181	0,176	**-0,232**	**0,345**	**-1,080**	-0,128	**-0,766**			
5	**0,241**	**-0,208**	0,116	**-0,219**	-0,029	**0,226**				
6	0,107	0,060	-0,089	-0,226	**0,226**					
7	0,111	-0,151	0,097	-0,028						
8	**-0,306**	-0,102	0,151							
9	-0,037	0,014								
10	0,000									

We may see that the weight of the errors on incremental payments is higher in several cells until year of origin 5. The fit is better for more recent years. However, there are only four cells with standardized errors higher than 0.5. These results are better than those obtained for the irregular data set from section 5.4.1. We may also see in the following Figure 5.8 that the CL is producing more stable estimates of the ultimate costs over the years. However, there is no such stability for years of origin 5 and 8 and for the latter the results do not seem to be already stabilized.

Figure 5.8: Back-Testing with CL for Taylor and Ashe Data



Doing the same analysis for the VP we get the following Tables 5.21 and 5.22.

Table 5.21: Errors with VP for Taylor and Ashe Data

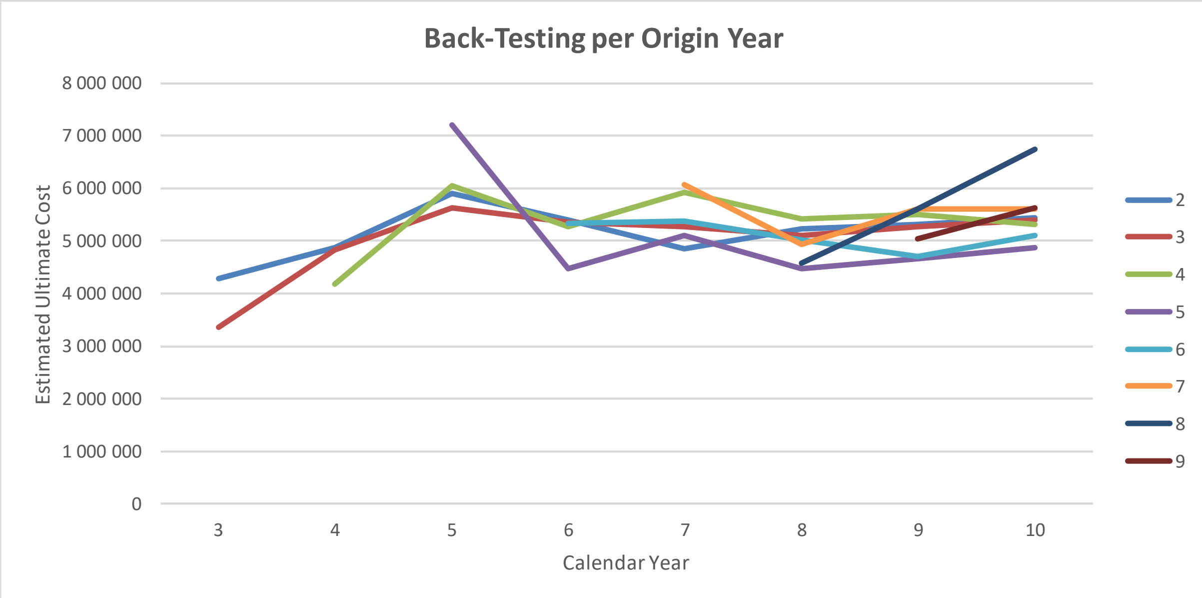
	1	2	3	4	5	6	7	8	9	10
1	80 871	97 257	-98 512	-281 756	123 463	299 084	-124 379	-44 973	-51 056	0
2	-33 639	-48 673	-53 633	118 268	-116 731	-62 444	150 761	8 623	37 468	
3	-91 963	77 053	-52 894	-39 292	193 133	-233 250	122 162	25 051		
4	-66 463	196 557	-189 103	521 359	-277 329	-22 754	-162 268			
5	97 043	-143 661	105 933	-186 092	175	126 601				
6	34 085	61 717	-79 333	-194 525	178 056					
7	42 244	-116 087	111 022	-37 178						
8	-119 709	-96 949	216 658							
9	-22 191	22 191								
10	0									

Table 5.22: Standardized Errors with VP for Taylor and Ashe Data

	1	2	3	4	5	6	7	8	9	10
1	**0,226**	0,127	-0,161	**-0,583**	**0,234**	**0,521**	**-0,850**	**-0,321**	**-0,225**	0,000
2	-0,096	-0,055	-0,057	0,100	**-0,262**	**-0,195**	**0,286**	0,032	0,088	
3	**-0,317**	0,077	-0,057	-0,039	**0,257**	**-1,588**	**0,246**	0,089		
4	**-0,214**	0,177	**-0,244**	**0,334**	**-1,018**	-0,065	**-0,787**			
5	**0,219**	**-0,207**	0,107	**-0,242**	0,000	**0,269**				
6	0,086	0,066	-0,094	**-0,242**	**0,252**					
7	0,096	-0,137	0,098	-0,035						
8	**-0,333**	-0,091	0,150							
9	-0,059	0,022								
10	0,000									

We may see that the weight of the errors on incremental payments is high in the same cells as the CL. However, there are some extra cells with high standardized errors, the (4,1) and (6,4). We may also see in the following Figure 5.9 that the VP, as with the CL, see Figure 5.6, is producing stable estimates of the ultimate costs over the years. Also, the problems detected with the CL are emerging with the VP, mainly the unstable results from origin years 5 and 8 and the current trend for the year 8 to grow.

Figure 5.9: Back-Testing with VP for Taylor and Ashe Data



We must remember here that both the CL and the VP presented lower prediction errors when compared with the Mack data case, see Tables 5.6 and 5.7. Putting together all the analyses done, we may conclude that with the regular data considered here, Taylor and Ashe (1983), both the CL and the VP fit the data much better when compared with the previous case with irregular data. Because of this, there are lower errors, more stable estimates of the ultimate costs and lower prediction errors.

In the following tables we present the regression errors and the standardized regression residual from the CL and the VP. The VP has a sum of the square of errors of 1 871 805 779 252 and the CL of 1 884 835 560 18.

Using CL, the regression of the column 2 is the one with higher standardized regression errors, the regression errors divided by the observed payments, see Table 5.24. In these columns, there are three years of origin has errors that represent more than 0.2 of the observed payments.

Table 5.23: Regression Errors with CL for Taylor and Ashe Data

	1	2	3	4	5	6	7	8	9	10
1	0	-124 319	-230 049	-310 822	141 676	289 341	-140 072	-46 797	-48 851	0
2	0	7 034	10 087	190 688	-137 236	-73 437	172 369	15 771	48 851	
3	0	278 260	-39 563	1 872	188 375	-266 917	139 448	31 025		
4	0	334 648	-284 170	558 357	-380 757	-66 348	-171 745			
5	0	-410 547	142 752	-204 039	1 060	117 362				
6	0	-49 524	-148 859	-192 450	186 882					
7	0	-250 308	168 488	-43 606						
8	0	166 325	381 315							
9	0	48 431								
10	0									

Table 5.24: Standardized Regression Errors with CL for Taylor and Ashe Data

	1	2	3	4	5	6	7	8	9	10
1	0	-0,1105	-0,1326	-0,1401	0,0516	0,0872	-0,0404	-0,0130	-0,0127	0,0000
2	0	0,0057	0,0046	0,0569	-0,0361	-0,0178	0,0371	0,0032	0,0091	
3	0	**0,2153**	-0,0178	0,0006	0,0473	-0,0646	0,0301	0,0063		
4	0	**0,2359**	-0,1295	0,1486	-0,0945	-0,0151	-0,0374			
5	0	**-0,3613**	0,0671	-0,0704	0,0003	0,0303				
6	0	-0,0371	-0,0683	-0,0645	0,0506					
7	0	-0,1943	0,0696	-0,0125						
8	0	0,1170	0,1331							
9	0	0,0355								

Using now the VP, the regression of the column 2 is the one with higher standardized regression errors, the regression errors divided by the observed payments, see Table 5.26. In these column three years of origin has errors that represent more than 0.2 of the observed payments. It is the same conclusion we got with the CL.

Table 5.25: Regression Errors with VP for Taylor and Ashe Data

	1	2	3	4	5	6	7	8	9	10
1	0	-98 275	-231 931	-318 526	157 191	306 754	-143 629	-50 241	-55 052	0
2	0	32 660	8 018	181 054	-113 781	-49 342	167 954	11 152	40 401	
3	0	299 403	-41 726	-7 977	211 003	-241 637	135 019	26 425		
4	0	357 253	-286 544	548 613	-354 475	-40 790	-176 440			
5	0	-378 294	140 850	-213 487	21 328	138 942				
6	0	-20 694	-151 090	-202 131	207 766					
7	0	-218 225	166 332	-54 349						
8	0	192 487	378 937							
9	0	75 846								
10	0									

Table 5.26: Regression Errors with VP for Taylor and Ashe Data

	1	2	3	4	5	6	7	8	9	10
1	0	-0,0874	-0,1337	-0,1436	0,0573	0,0924	-0,0414	-0,0139	-0,0144	0,0000
2	0	0,0264	0,0037	0,0540	-0,0299	-0,0120	0,0361	0,0023	0,0076	
3	0	**0,2317**	-0,0188	-0,0025	0,0529	-0,0585	0,0292	0,0054		
4	0	**0,2518**	-0,1305	0,1460	-0,0880	-0,0093	-0,0385			
5	0	**-0,3329**	0,0662	-0,0737	0,0063	0,0359				
6	0	-0,0155	-0,0693	-0,0677	0,0563					
7	0	-0,1694	0,0687	-0,0156						
8	0	0,1354	0,1323							
9	0	0,0556								

5.7.3 Taylor and McGuire Data

Here we present the same results as in sections 5.4.1 and 5.4.2 but now with data from Taylor and McGuire (2016). This set of data is also regular, but the prediction errors are even lower than the ones obtained for Taylor and Ashe (1983) data, see Table 5.8 and Table 5.9.

Table 5.27: Errors with CL for Taylor and McGuire Data

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	935	1 424	-517	-1 140	54	-698	151	-453	245	0
2	3 864	2 349	-956	-1 484	-1 417	-668	-862	-581	-245	
3	-2 154	8 252	-7 202	-1 389	9	325	1 127	1 034		
4	3 586	-6 631	510	737	278	1 936	-416			
5	4 264	-4 347	992	-685	670	-895				
6	-1 627	-565	231	1 555	406					
7	-4 934	-1 173	3 701	2 407						
8	-2 767	-474	3 241							
9	-1 166	1 166								
10	0									

Table 5.28: Standardized Errors with CL for Taylor and McGuire Data

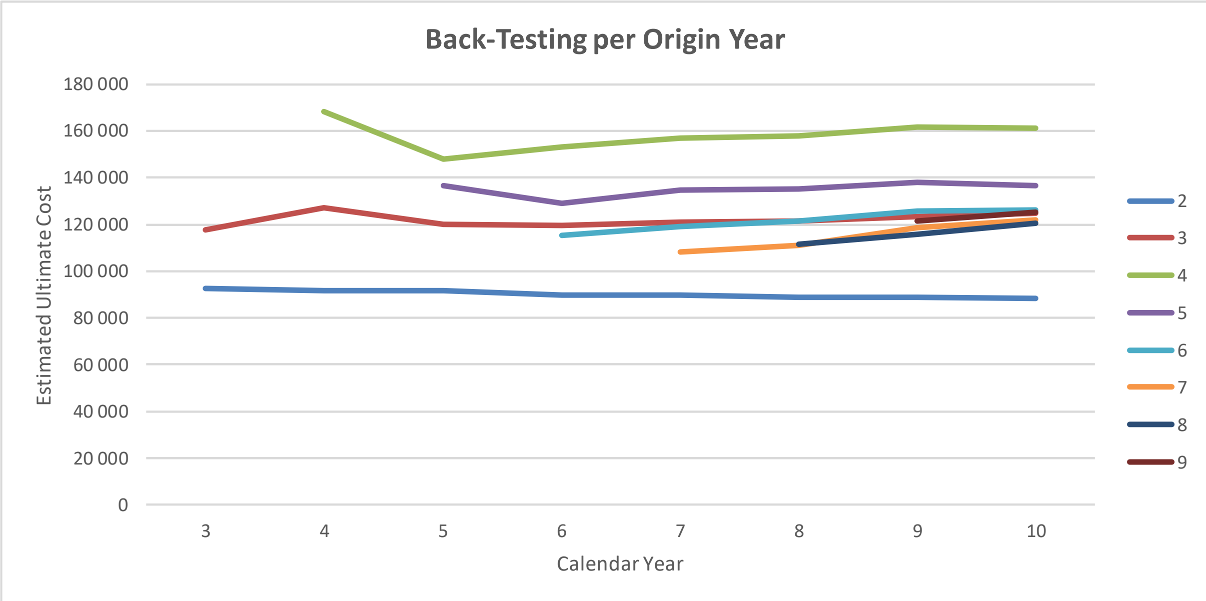
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0,020	0,061	-0,177	**-0,634**	0,027	**-0,580**	0,117	**-0,805**	**0,315**	0,000
2	0,073	0,089	-0,338	**-0,849**	-1,935	**-0,469**	**-2,161**	**-1,081**	**-0,721**	
3	-0,032	0,195	**3,885**	**-0,437**	0,003	0,099	**0,387**	**0,396**		
4	0,038	-0,177	0,069	0,111	0,066	**0,336**	**-0,220**			
5	0,053	-0,131	0,145	-0,158	0,167	**-0,381**				
6	-0,024	-0,017	0,041	**0,252**	0,117					
7	-0,078	-0,036	**0,414**	**0,350**						
8	-0,043	-0,015	**0,385**							
9	-0,017	0,033								
10	0,000									

We may see that the weight of the errors on incremental payments is higher in several cells after development year 2. However, there are only seven cells with standardized errors higher than 0.5. The results are also better than those obtained for the irregular data set from 5.4.1 section.

Comparing with the results from section 5.4.2, we see a good fit in the first two development years, with low standardized errors.

We may also see in the following Figure 5.10, that the CL is producing more stable estimates of the ultimate costs over the years. However, there is a trend to correct ultimate costs projection upwards on more recent years of origin, from 6 to 9.

Figure 5.10: Back-Testing with CL for Taylor and McGuire Data



Doing the same analysis for the VP we get the following tables.

Table 5.29: Errors with VP for Taylor and McGuire Data

	1	2	3	4	5	6	7	8	9	10
1	900	1 764	-484	-1 210	1	-813	145	-560	257	0
2	3 826	2 724	-920	-1 560	-1 474	-795	-870	-698	-232	
3	-2 197	8 787	-7 151	-1 496	-72	146	1 117	868		
4	3 407	-5 999	567	591	167	1 699	-433			
5	4 104	-3 814	1 041	-810	576	-1 096				
6	-1 886	-128	267	1 432	314					
7	-5 239	-777	3 731	2 284						
8	-3 145	-120	3 265							
9	-1 542	1 542								
10	0									

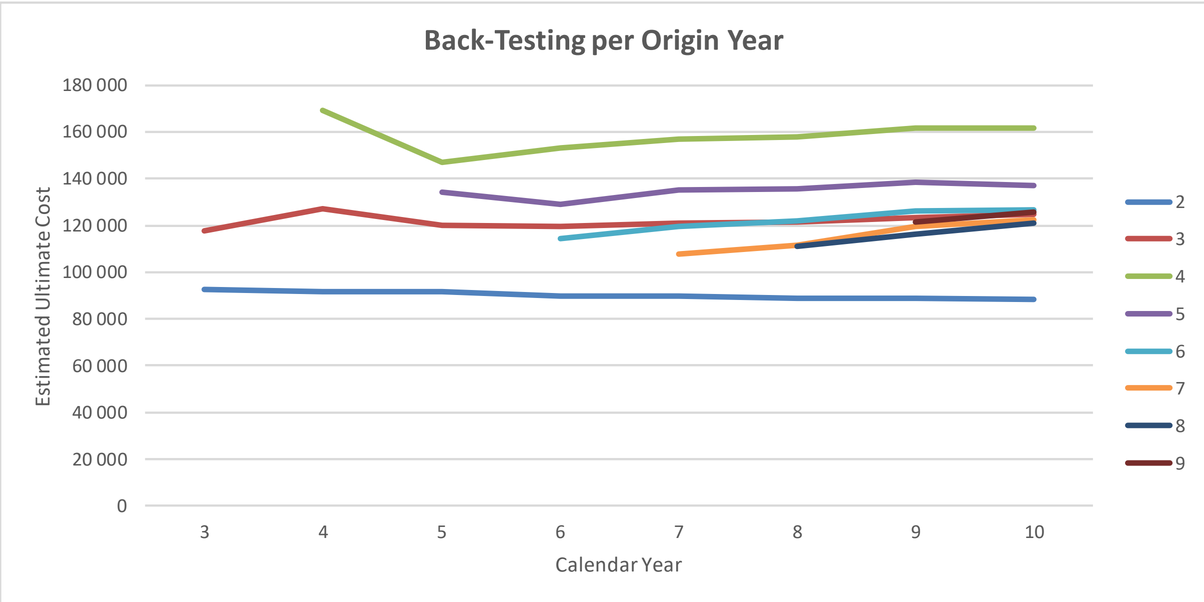
Table 5.30: Standardized Errors with VP for Taylor and McGuire Data

	1	2	3	4	5	6	7	8	9	10
1	0,020	0,076	-0,166	**-0,673**	0,001	**-0,676**	0,111	**-0,995**	**0,331**	0,000
2	0,072	0,103	**-0,325**	**-0,893**	**-2,014**	**-0,558**	**-2,179**	**-1,300**	**-0,683**	
3	-0,033	**0,208**	**3,857**	**-0,471**	-0,024	0,044	**0,384**	**0,332**		
4	0,036	-0,160	0,076	0,089	0,040	**0,295**	**-0,229**			
5	0,051	-0,115	0,152	-0,187	0,144	**-0,467**				
6	-0,027	-0,004	0,047	**0,232**	0,090					
7	-0,083	-0,024	**0,417**	**0,332**						
8	-0,049	-0,004	**0,388**							
9	-0,022	0,044								
10	0,000									

We may see that the weight of the errors on incremental payments is high in the same cells as with the CL. However, there are some extra cells with high standardized errors, the (3,2) and (2,3).

We may also see in the following Figure 5.11 that the VP, as with the CL, see Figure 5.9, is producing stable estimates of the ultimate costs over the years. Also, the problem detected with the CL is also emerging with the VP, mainly a trend to correct ultimate costs projection upwards on more recent years of origin, from 6 to 9.

Figure 5.11: Back-Testing with VP for Taylor and McGuire Data



Both the CL and the VP presented lower prediction errors when compared with the two previous cases, see sections 5.2.1 and 5.2.2. Putting together all the analyses done we may conclude that with the regular data considered in this section, Taylor and McGuire (2016), both the CL and the VP fit better the data when compared with the previous cases with irregular and regular data. Because of this there are lower errors, more stable estimates of the ultimate costs and lower prediction errors.

In the following Tables 5.31 and 5.32 we present the regression errors and the standardized regression residual from the CL and the VP. The VP has a sum of the square of errors of 319 281 386 and the CL of 322 372 956.

Using CL, the regression of the column 2 is the one with higher standardized regression errors, the regression errors divided by the observed payments. None of the years of origin has errors that represent more than 20% of the observed payments.

Table 5.31: Regression Errors with CL for Taylor and McGuire Data

	1	2	3	4	5	6	7	8	9	10
1	0	965	-639	-1 218	35	-717	151	-456	247	0
2	0	453	-1 277	-1 705	-1 517	-728	-887	-591	-247	
3	0	9 309	-7 517	-1 343	76	388	1 159	1 047		
4	0	-8 390	667	844	326	1 975	-422			
5	0	-6 439	997	-723	664	-918				
6	0	233	345	1 637	417					
7	0	1 247	4 016	2 508						
8	0	883	3 408							
9	0	1 738								
10	0									

Table 5.32: Standardized Regression Errors with CL for Taylor and McGuire Data

	1	2	3	4	5	6	7	8	9	10
1	0	0,0140	-0,0089	-0,0165	0,0005	-0,0093	0,0019	-0,0058	0,0031	0,0000
2	0	0,0057	-0,0155	-0,0203	-0,0179	-0,0084	-0,0102	-0,0068	-0,0028	
3	0	0,0849	-0,0697	-0,0121	0,0007	0,0033	0,0096	0,0085		
4	0	-0,0641	0,0048	0,0058	0,0022	0,0127	-0,0027			
5	0	-0,0567	0,0083	-0,0058	0,0052	-0,0070				
6	0	0,0023	0,0032	0,0143	0,0035					
7	0	0,0131	0,0385	0,0226						
8	0	0,0091	0,0324							
9	0	0,0167								

Using now the VP, the regression of the column 2 is again the one with higher standardized regression errors. None of the standardized regression errors represents more than 0.2 of the observed payments. It is the same conclusion we got with the CL.

Table 5.33: Regression Errors with VP for Taylor and McGuire Data

	1	2	3	4	5	6	7	8	9	10
1	0	1 330	-621	-1 304	-26	-840	142	-564	259	0
2	0	877	-1 256	-1 803	-1 587	-865	-897	-712	-234	
3	0	9 847	-7 490	-1 472	-16	203	1 146	880		
4	0	-7 643	700	678	207	1 733	-439			
5	0	-5 795	1 026	-868	562	-1 126				
6	0	783	371	1 508	323					
7	0	1 752	4 041	2 383						
8	0	1 398	3 433							
9	0	2 286								
10	0									

Table 5.34: Regression Errors with VP for Taylor and McGuire Data

	1	2	3	4	5	6	7	8	9	10
1	0	0,0193	-0,0086	-0,0177	-0,0003	-0,0109	0,0018	-0,0072	0,0032	0,0000
2	0	0,0110	-0,0153	-0,0214	-0,0187	-0,0100	-0,0104	-0,0082	-0,0027	
3	0	0,0898	-0,0695	-0,0133	-0,0001	0,0017	0,0095	0,0072		
4	0	-0,0584	0,0051	0,0047	0,0014	0,0112	-0,0028			
5	0	-0,0510	0,0085	-0,0070	0,0044	-0,0086				
6	0	0,0076	0,0034	0,0132	0,0027					
7	0	0,0184	0,0388	0,0214						
8	0	0,0144	0,0326							
9	0	0,0220								

After analysing these three datasets we may conclude that there is a relation between the prediction error and the errors and back-testing results obtained:

- When the prediction error is high, as in the case of irregular data, the errors are high, and the back-testing shows unstable results.

- When the prediction error decreases, as we saw with the two examples of regular data, the same happens with errors and the back-testing shows more stable results.
- As the prediction error of the CL is lower than the one from VP, as in the case with irregular data, its errors and back-testing show better results. However, we cannot say that the method fits properly the data because the errors are high, and the back-testing shows unstable results.
- With regular data the VP shows a low prediction error than the CL. This means lower errors and more stability in the results.
- The VP with regular data appears with more cases with high errors than the CL, even if the overall results are better with the VP. This does not happen with regression errors.
- The errors and back-testing results obtained confirm the conclusions from the prediction error, when the latter is lower the former have better results.
- However, the errors analysis and the back-testing also show some problems with the fitting and that is not observed with the prediction error analysis.

5.8 Summary of the Empirical Findings

We know from Straub (1988), that the classical CL does not minimize the square of the errors. We also saw in the previous sections 5.2 and 5.3 that CL does not provide the best fit, having higher prediction errors in most of the cases.

Straub (1988) showed that the regression through the origin, i.e., the VP, minimizes the square of the errors. This theoretical result has been applied empirically in Section 5.2 by using, Mack (1993a) irregular data and regular data from Taylor and Ashe (1983) and from Taylor and McGuire (2016). Thus, not only the VP has smaller prediction error than the CL method for both datasets with regular data, but simultaneously, we report that the CL requires more reserves.

Even in the case of irregular development of data, see Table 2.1 (Mack, 1993a), the CL method still estimates 19% more of reserves, when compared with VP. This is again because VP minimizes the square of the errors, see Straub (1988). Without doubt, it is an important fact for actuaries[8] (see also our extensive discussion in chapters 1 and 2) and it should not be underestimated. Obviously, actuaries should present to insurers the best estimates and the current Solvency II system promotes a better prediction of the reserves, asking for best estimates without risk margins, see European Union (2015).

Finally, the obtained results based on the random choice of 114 triangles from different companies and for different periods of business using data from the consulting company, Actuarial Group, Lisbon, Portugal, report again a lower prediction error for the VP, and the only cases where that does not happen are those where their triangles have special features, like a lot of zeros on the cells.

5.9 Conclusions

The assessment of financial strength in the insurance industry includes a thorough analysis of outstanding claims reserves, including an assessment of possible variability in the reserves. As we saw in chapters 1 and 2, failure to do so might result in the insolvency or lack of competitiveness of some insurers. Methods of analysis, which help with the reserve estimation as well as provide insight into the variability of those reserves, are according with the Solvency II regulation, see European Union (2015), particularly if they can reduce the prediction errors.

We proposed the stochastic VP methodology using the regression through the origin approach of Murphy (1994), but with heteroscedastic errors instead, to develop it comparably with the Mack (1993a, 1993b, 1994) stochastic distribution-free framework and test it empirically with the CL method. Interestingly, the equation for the loss development factors which is

---

[8] Moreover, we should not forget that when the CL started spreading, back in 70s, there were no microcomputers and any methods had to be easy to apply with a calculator machine. The CL accomplished that task very well and it was a good approximation to the minimum square of the errors (Straub, 1988). Later, with the arrival of the stochastic methods, the CL framework was the simplest approach to start its development. However, after all these years, we have now more experience on the CL application and it is clear to any professional actuary that sometimes the CL estimates have very high prediction errors. The VP aims to solve this problem and to fill this gap in these cases.

formulated on the Mack (1993a, 1994) heteroscedastic errors framework, although, it is assumed that the errors are proportional to the square of payments instead, see (5.1.3), it can also be derived straightforwardly from the Murphy (1994) homoscedastic errors framework.

Obviously, the prediction error is not the only measure to have when the claims reserves are estimated. Additionally, other items should be also addressed, such as the errors, the back-testing and so on and so forth. However, from the datasets analysed, we conclude that when the prediction error is lower, the errors analysis and the back-testing also give better results.

As we saw in chapters 1 and 2, the reserves are crucial for insurer's management and solvency. Due to that it is almost impossible to tolerate a method which might have a higher or even very high prediction error. This means that the method does not follow the insurer experience and, as such, it is not a good predictor of the future. This happens because the triangles data format technique and the link ratios methods assume that the past may be a good predictor for the future.

Finally, three commonly used empirical examples have been applied. We observe that when the data has irregular developments both the CL and VP approaches generate high prediction errors, and thus, they cannot be considered as the best approaches to predict with this class of data. Additionally, we show that the VP, with such a set of irregular data, is not able to outperform CL. On the other hand, however, when more regular data is considered, like in Examples 1 and 2, the prediction error for both methods is improved, and the VP outperforms the CL. In these regular cases, the risk margins of the Vector Projection are also lower comparing with those derived from the CL. Practically, this also implies a lower fair value of reserves for the VP method. The results are also tested and confirmed by using 114 triangles with paid claims and 10 years of information to be comparable directly to the previous cases, where 85% of them appear to give a lower prediction error when the VP method is used. Finally, it should be mentioned that in the present chapter, a direct comparison between CL and VP methods of the link ratio family has been provided, and a Use test is also presented, something not common in the corresponding literature. The only case, as far as we are concerned, where such a comparison is performed is due to Verrall (2000) in the context of the generalized linear models.

As a natural continuation of this chapter, the VP homoscedastic, the multivariate approach and the estimation of several triangles at the same time, will be also considered in the following chapters 6 and 7.

6. Stochastic Univariate and Multivariate Generalized Link Ratios

This chapter makes four major contributions to the claims reserving literature.

First, it develops a framework that is used to introduce two methods: the Generalized Link Ratios (GLR) and the Multivariate Generalized Link Ratios (MGLR). In this regard, both methods are developed with a parameter that allow us to obtain several methods, univariate in the GLR and multivariate in the MGLR. The three methods obtained are the CL, the recently proposed Vector Projection (VP) from Portugal et al. (2017), which is a regression through the origin, and the SA. Moreover, several other methods may be obtained for other parameter values. In the literature, a similar approach was proposed by Murphy (1994) for univariate regressions using recursive formulas with constant variance of the errors (homoscedastic errors). In this case, the prediction errors were calculated with a formula for the first origin year and another formula to the following origin years. In our approach, we are going to generalize the loss development factor in such way to permit the consideration simultaneously of several methods that may have homoscedastic or heteroscedastic errors.[9] Additionally, the prediction errors are calculated, both within a univariate and multivariate regression framework and using matrices that consider information from all the regressions inside the triangle. This approach allows us, simultaneously, to have the loss development factors, the reserves and the prediction errors over all the regressions, without utilizing recursive formulas. Consequently, a GLR method is obtained with homoscedastic or heteroscedastic errors, with method selection based on the lowest prediction error, which also corresponds to a certain level of heteroscedasticity (which may be a homoscedastic method if this level is

---

[9] In Murphy (1994) and Barnett and Zehnwirth (1999), the general regression trough the origin by using assumptions to the errors was developed and they demonstrated that several models might be obtained because of it. Obviously, some of the models are heteroscedastic, such as the CL and the SA methods. Straub (1988) showed that the slope that minimizes the sum of squares of errors is not the CL, but the regression through the origin, i.e., the VP in our case. Thus, Murphy (1994) introduced the VP method, but with homoscedastic errors. Recently, in Portugal et al. (2017), the stochastic VP method with heteroscedastic errors is introduced instead to compete with Mack (1993a, 1993b, 1994) stochastic heteroscedastic CL approach. VP outperforms CL for most datasets.

zero). The output will give us either the VP or CL or SA or other methods depending on the level of heteroscedasticity considered.

Second, it develops a method based on a multivariate regression framework. However, this time with contemporaneous correlations between the equations of the triangle. Thus, the MGLR method is obtained. Furthermore, we obtain again several methods for different values from one parameter: the multivariate VP, the multivariate CL, the multivariate SA and several other multivariate methods. The method selection is based on the lowest prediction error, as in the previous case, but now with contemporaneous correlations between the regressions. This is a very distinct approach from the existing claims reserving literature.

Third, several tests on method's assumptions, from regression techniques, will be performed: heteroscedasticity of the errors, correlations between equations and serial correlations of the errors. We will also see that testing for heteroscedasticity is also an important help for method selection.

Finally, regarding the empirical part, the illustration of our theoretical findings is also benefited by considering[10] 114 triangles from different companies and for different periods of business using data from the consulting company, Actuarial Group, Lisbon Portugal.[11]

The next parts of the chapter are organized as follows. In Section 6.1, the necessary up-to-date review of the multivariate approaches, known already in the reserving literature, is presented. Section 6.2 presents the generalized link ratios. In Section 6.3, the GLR method is developed in the claims reserving context. Thus, a universal formula for the prediction error is developed and with the specification of method assumptions we will be able to apply it to any of the methods considered in this chapter. This will help us develop the MGLR method in Section 6.4. In Section 6.5, we provide several numerical examples obtained for both methods. We also present the replication results of the Mack (1993a, 1993b, 1994)'s method, with the use

---

[10] Practically, we mean that the dataset includes different companies with observations from ten years, see Portugal et al. (2017).

[11] Obviously, it is not possible to disclose any further information about the triangles used and their orientation in the paper.

of two scenarios for the process variance. For this treatment, we use two triangles on cumulative payments, one with irregular and the other with much more regular development of data. A discussion of the empirical findings is provided on Section 6.6 based on 114 triangles from different companies. In section 6.7 some heteroscedasticity tests are applied to the triangles and we show its relationship with method choice. The section 6.8 tests the inexistence of correlations between the equations and the section 6.9 tests the serial correlation of the errors. Finally, Section 6.10 concludes the whole discussion.

6.1 Multivariate Approaches in the Reserving Literature

In the existing reserving literature, several multivariate approaches have been considered to check the existence of structural connections among triangles.

This includes situations, where the development of one triangle might depend upon past information from other triangles (Holmberg, 1994; Halliwell, 1997; Quarg and Mack, 2004; Merz and Wüthrich, 2006) and where joint development is considered with contemporaneous correlations among triangles (Braun, 2004; Pröhl and Schmidt, 2005; Kremer, 2005; Hess et al. 2006; Schmidt, 2006; Merz and Wüthrich, 2007, 2008, 2009; Bardis et al., 2012). Zhang (2010) proposed a general multivariate CL method which does not only specify contemporaneous correlations, but allows structural connections among triangles, simultaneously.
Recently, copulas methodologies have been considered to give another dimension to the standard multivariate reserving approaches, see, for example, Shi (2011) and Shi and Frees (2014).

In what follows in the next sections we present the GLR and the MGLR prediction errors with analytical non-recursive formulas. In the multivariate claims reserving literature there is a concentration on the CL method, see for example Braun (2004), Pröhl and Schmidt (2005), Merz and Wüthrich (2008) and Zhang (2010). As we saw before, see Chapter 2, the same happens with the univariate literature, with several univariate CL. For example, Mack (1993a, 1993b, 1994), Verrall (2000), England and Verrall (2002), Brydon and Verrall (2009) among others.

6.2 Generalized Link Ratios

In insurance practice, it has been confirmed that it is not always possible to apply the same reserving methodology for all triangles involved in the portfolio of activities, see Portugal et al. (2017), and the references therein. The reasoning behind it, is that higher prediction errors will occur, and this is a serious limitation of the existing traditional multivariate techniques.[12] This can be even worse if the method to be applied is known to produce high prediction errors.

The present chapter should not just be considered as an extension of Portugal et al. (2017) approach presented in chapter 5, where the VP is compared with the CL method and it was shown that the VP produces lower prediction errors in most of the cases. Nevertheless, in this study, we allow the best GLR method to be anything among the VP, CL, SA or other methods. To choose a method we will decide based on the prediction error minimization.[13] Then, the multivariate approach will be formulated using contemporaneous correlations among the equations of the GLR method.

As we are going to use regression techniques and its framework, see for example Fomby et al. (1984), we will do same changes in the notation for the cumulative payments. The cumulative payments will be designated by $y_{i,j}$ when used as a dependent variable and by $x_{i,j}$ when used as an independent variable. The dependent variable is a random variable. The independent variable is a non-random variable.

---

[12] However, to be more precise, the existing multivariate approach may still be applied to the estimation of each of the implicit equations from one triangle if we consider contemporaneous correlations between all the equations. This will be presented in detail when the multivariate generalized link ratios method is discussed.

[13] Obviously, the prediction error is not the only measure to have when the claims reserves are estimated, additionally, other items should be equally addressed, such as the errors analysis, the back-testing and so on and so forth. However, in real life applications, it is almost impossible to tolerate a model which might have a high or even very high prediction error (Portugal et al., 2017). As we saw in section 5.4, better results on errors analysis and back-testing are associated with lower prediction errors.

The link ratios between the triangle cells $F_{i,j+1}$, defined in (2.1), are summarized through the loss development factor in one figure. To have the latter, and following Murphy (1994), Barnett and Zehnwirth (1999), and Bardis et al. (2012), we concentrate on the family of link ratios given by a generalized function. In this function, the loss development factor estimation depends on one parameter α. Defining α give us several well-known practical methods. For a generic triangle of claims with T origin years, with cumulative payments on origin year i and development year j given by $y_{i,j}$, we have

$$F_{i,j+1} = \frac{y_{i,j+1}}{y_{i,j}} \text{ and } \hat{b}_j(\alpha) = \frac{\sum_{i=1}^{T-j} y_{i,j+1} x_{i,j}^{1-\alpha}}{\sum_{i=1}^{T-j} x_{i,j}^{2-\alpha}} \tag{6.1}$$

Indeed, when $\alpha = 0$, we get the regression through the origin (or VP), for $\alpha = 1$, we get the CL, and when $\alpha = 2$, we get the SA. These loss development factors can also be seen as the weighted average of the link ratios, being the weights the payments to the power of $2 - \alpha$ (Portugal et al., 2017). The here defined α is equal to the Murphy (1994) δ presented in section 5.1.

$$\hat{b}_j(\alpha) = \frac{\sum_{i=1}^{T-j} y_{i,j}^{2-\alpha} F_{i,j+1}}{\sum_{i=1}^{T-j} y_{i,j}^{2-\alpha}} \tag{6.2}$$

The equation (6.2) might suggest an upper bound for the parameter α. Indeed, if α becomes higher than two, in practice, the results might be considered as problematic and unexplained by actuaries, as very high/low weights are given to the link ratios if the associated cumulative payments used at (6.2) are very different.[14] Bardis et al. (2012) also showed that the value of α should not be negative. Putting these conclusions together we expect $0 \leq \alpha \leq 2$.

Practically, this means that we may have several methods inside this generalized approach that differ from the typical ones, i.e., the VP, CL and SA. Moreover, for each α, we get a different weight for the link ratios, and a different estimator for the loss development factor, b (not necessarily the VP, the CL or the SA). The prediction error is developed from this method using a regression framework. In this regard, we may choose the best parameter α that minimizes the prediction error.

As we saw already in section 3.6, link ratios methods may be seen as a set of regressions, where each loss development factor may be estimated through a regression, see for example

[14] See the case study of 114 triangles (Section 6.6).

Mack (1993a) or Murphy (1994). Using regressions, a triangle may be seen as a set of individual regressions to calculate the loss development factors. However, the estimation of the parameters of each regression and specifically the loss development factor is done without considering the calculations of the parameters from the other regressions. This means that the loss development factors obtained by a link ratios method for one triangle are independent of each other.

Obviously, this helps the calculations, but introduces a very strong assumption on it: any regression from one triangle may be individually estimated without considering the estimation of the other regressions. However, the regressions may be correlated and if that happens the calculation of its parameters depends on those correlations, see for example Fomby et al. (1984).

In this chapter, we also consider the generalized link ratios presented above with contemporaneous correlations between the regressions. Under this framework, we consider a regression method that becomes multivariate due to this feature and using seemingly unrelated regressions (SUR) (Zellner, 1962, 1963; Zellner and Huang, 1962; Srivastava and Giles, 1987). Manipulating the GLR method within the SUR framework, it allows us to have a multivariate method, the MGLR method. We will also select the best method as the one corresponding to the parameter α that produces the lowest prediction error.

Our approach considers the case of homoscedastic as well as heteroscedastic errors. Before we start to present the mathematical framework for our treatment, it should be mentioned that the heteroscedastic feature in claims reserving is crucial and in the traditional link ratios methods there is an implicit assumption on it (Taylor, 2000). Moreover, regression models offer a good opportunity to explore these issues since many years, see for example Taylor (1987):

"*However, it seems that the regression models have not been prevalent in claims analysis leading to loss reserving. The scarcity arises from the suspicion with which many actuaries regard such models … Despite this it appears that regression techniques have a definite place in the actuarial repertoire.*"

6.3 Stochastic Univariate Regression Method

In a regression framework, the loss development factor j is estimated by a linear regression with (or without) intercept between two adjacent columns, $j+1$ and j. In this method, main attention is provided for the case without intercept.

6.3.1 General Univariate Method

Data considered on each of the $k = 1, \dots, T-1$ regressions has $T-k$ elements and the calculations will be provided simultaneously with the $y_{i,j}$ explained by the adjacent triangle column, $x_{i,j-1}$. This means that the payments on column j, $y_{i,j}$ are a function (a regression through the origin) of the payments on column $j-1$, $x_{i,j-1}$. Both variables represent the cumulative payments but $y_{i,j}$ is a random variable and $x_{i,j-1}$ is a non-random variable (because when we want to estimate $y_{i,j}$ we know $x_{i,j-1}$). That is why the notation on cumulative payments, from the last chapters, was changed.

The j loss development factor β_j is the slope of each of the $k = 1, \dots, T-1$ regressions, $\varepsilon_{i,j}$ is the error of each regression j on each observation i, and $y_{i,j}$ is given by

$$y_{i,j} = \beta_j x_{i,j-1} + \varepsilon_{i,j} \quad i = 1, \dots, T-j \ \ and \ j = 2, \dots T \tag{6.3.1}$$

Now, in a matrix form and considering all the equations implicit in the triangle of cumulative payments, our method (6.3.1) will be given by

$$Y = X\beta + \varepsilon \tag{6.3.2}$$

Y is a block-vector with dimension $m \times 1$, where $m = \frac{T\times(T-1)}{2}$, that includes the block-vectors Y_k for $k = 1, \dots, T-1$. Analytically, we have

$$Y = \begin{bmatrix} Y_1 \\ \cdots \\ Y_{T-1} \end{bmatrix}$$

where the generic $Y_k = \begin{bmatrix} y_{1,k+1} \\ \cdots \\ y_{T-k,k+1} \end{bmatrix}$ includes the random variables $y_{i,k+1}$ for $i = 1, \dots, T-k$.

X is defined by a diagonal block matrix with all the $X_{i,j}$ used as explanatory variables and considered as non-random. For each diagonal element X_k, we have a column vector with the number of observations to be equal to $T-k$. The matrix X has dimension $m \times (T-1)$, and it can be represented by

$$X = \begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_{T-1} \end{bmatrix}$$

where each element $X_k = \begin{bmatrix} x_{1,k} \\ \cdots \\ x_{T-k,k} \end{bmatrix}$

We also define the $T-1$ loss development factors, β_j, $j = 1, \dots, T-1$, and the $(T-1) \times 1$ vector of the loss development factors which is given by

$$\beta = \begin{bmatrix} \beta_1 \\ \cdots \\ \beta_{T-1} \end{bmatrix}$$

Indeed, we know from Straub (1988) and Murphy (1994) that the β_j are the loss development factors from a link ratios method.

Finally, the errors vector is a block matrix of size $m \times 1$, and it is given by

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \cdots \\ \varepsilon_{T-1} \end{bmatrix}$$

where $\varepsilon_k = \begin{bmatrix} \varepsilon_{1,k+1} \\ \cdots \\ \varepsilon_{T-k,k+1} \end{bmatrix}$

We define the true unknown future observations of the dependent variables as

$$Y_F = X_F \beta + \varepsilon_F$$

where X_F and ε_F are the future values of X and the future errors, respectively.

Y_F is a block vector with size $m \times 1$ and given by

$$Y_F = \begin{bmatrix} Y^F{}_1 \\ \cdots \\ Y^F{}_{T-1} \end{bmatrix}$$

with each element $Y^F{}_k = \begin{bmatrix} y^F{}_{T-k+1,k+1} \\ \cdots \\ y^F{}_{T,k+1} \end{bmatrix}$ for $k = 1, \dots, T-1$.

X_F is given by the current diagonal of payments and by the estimated payments of the lower triangle. It is block matrix with size $m \times (T-1)$ given by

$$X_F = \begin{bmatrix} X^F{}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X^F{}_{T-1} \end{bmatrix}$$

where each element $X^F{}_k = \begin{bmatrix} x_{T-k+1,k} \\ \cdots \\ x_{T,k} \end{bmatrix}$ for $k = 1, \dots, T-1$.

ε_F is a block vector with size $m \times 1$ and given by

$$\varepsilon_F = \begin{bmatrix} \varepsilon_{F_1} \\ \cdots \\ \varepsilon_{F_{T-1}} \end{bmatrix}$$

with each element $\varepsilon_{F_k} = \begin{bmatrix} \varepsilon_{F_{T-k+1,k+1}} \\ \cdots \\ \varepsilon_{F_{T,k+1}} \end{bmatrix}$ for $k = 1, \dots, T-1$.

The estimated values of the dependent variables are obtained from $\widehat{Y_F} = X_F\hat{\beta}$ due to the assumption to be introduced in (6.3.3.). The X_F matrix has two types of elements:

- The $x_{T-k+1,k}$, which are on the last diagonal of the upper triangle.
- And, the $x_{i>T-k+1,k}$ which are on the lower triangle. They are obtained after estimating all the cells of the lower triangle using $\widehat{Y_F} = X_F\hat{\beta}$.

6.3.2 Assumptions

Having defined the method at section 6.3.1, we present in this section its assumptions.

Proposition 6.3.1 *Considering the method given by (6.3.2), that allows for heteroscedasticity of the errors inside each equation, we assume for our GLR method*

$$\mathbb{E}(\varepsilon|X) = \mathbb{E}(\varepsilon) = 0 \tag{6.3.3}$$

$$\mathbb{E}(\varepsilon\varepsilon') = \sigma^2 W \tag{6.3.4}$$

$$\mathbb{E}(\varepsilon_F {\varepsilon_F}') = \sigma^2 W_F \tag{6.3.5}$$

where W is a $m \times m$-diagonal weighting matrix, which depends on the parameter α. W is given by (6.3.6), where the $diag$ operator transforms one vector on a diagonal matrix. The W diagonal elements are given by the elements of the transformed vectors

$$W = diag(X^{\alpha}) = \begin{bmatrix} diag({X_1}^{\alpha}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & diag({X_{T-1}}^{\alpha}) \end{bmatrix} = \begin{bmatrix} {x_{1,1}}^{\alpha} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{T-1,T-1}}^{\alpha} \end{bmatrix}. \tag{6.3.6}$$

The matrix W_F is the future W and has the same structure as W. However, its elements are the ${X_F}^{\alpha}$ instead of X^{α}.

$$W_F = diag({X_F}^{\alpha}) = \begin{bmatrix} diag\left({X_{F,1}}^{\alpha}\right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & diag\left({X_{F,T-1}}^{\alpha}\right) \end{bmatrix} = \begin{bmatrix} {x_{T,1}}^{\alpha} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{T,T-1}}^{\alpha} \end{bmatrix} \tag{6.3.7}$$

The σ^2 is a diagonal block matrix of size $m \times m$ with $j = 1, \dots, T-1$ blocks

$$\sigma^2 = \begin{bmatrix} {\sigma^2}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {\sigma^2}_{T-1} \end{bmatrix} \tag{6.3.8}$$

where each block ${\sigma^2}_j = diag \begin{bmatrix} {\sigma^2}_{j,k} \\ \cdots \\ {\sigma^2}_{j,T-1} \end{bmatrix}$ *for $k = 1, \dots, T-1$ and with ${\sigma^2}_{j,k} = \cdots = {\sigma^2}_{j,T-1}$.*

Matrix W of size $m \times m$ corresponds to a specific structure of heteroscedasticity through the choice of parameter α. If α is zero, we will get homoscedastic errors.

The way this matrix is defined will provide us with several methods for estimating the loss development factors. Analytically, we get the VP for $\alpha = 0$, the CL for $\alpha = 1$, the SA for $\alpha = 2$, and other methods for different values of α. To have them, we just need to change α to get a different W matrix. We will have homoscedastic errors for the VP and heteroscedastic errors for the CL and the SA.

6.3.3 Estimation

The following Lemmas allow us to get estimators from the parameters presented in (6.3.2) and (6.3.8).

Lemma 6.3.1 *Following, Fomby et al. (1984), we may get the estimation of the* β*, the loss development factors of all the equations.* $\hat{\beta}$ *is obtained using the Aitken generalized least squares and it is the best linear unbiased estimator of* β*.*

$$\hat{\beta} = (X'W^{-1}X)^{-1}XW^{-1}Y \tag{6.3.9}$$

Lemma 6.3.2 *Following Fomby et al. (1984), we may estimate* σ^2*using the equation sum of square of the errors,* SSR_j*, divided by the equation degrees of freedom, the number of observations* T_j *from equation* j *minus the number of parameters from the equation, in this case one.*

$$\hat{\sigma}^2{}_j = \frac{SSR_j}{T_j - 1} \tag{6.3.10}$$

The parameter α from (6.3.6) and (6.3.7) will be estimated as the one that minimizes the prediction error. This α parameter is a method choice parameter and we selected the prediction error as the criterion for method choice. This is due to the following reasons:

- It is an important practical criterion for actuaries, as it summarizes in one figure the error implicit in the reserve's forecasts, from a specific claims reserving method.
- And as we saw already in the numerical results presented in section 5.4, a lower prediction error is usually associated with good indicators from other tools used for method selection, for example, errors analysis and back-testing.

6.3.4 Prediction Error

Regression models, using matrices, allow us to develop very quickly a general non-recursive formula to have the prediction errors (the square root of the mean square error of prediction, very often presented as a percentage of the estimated reserves).

Theorem 6.3.1 *The mean square error of prediction (msep) from the method presented in section 6.3.1 is given by.*

$$msep = \mathbb{E}\left[X_F(\hat{\beta}-\beta)(\hat{\beta}-\beta)'X'_F\right] + \mathbb{E}(\varepsilon_F\varepsilon_F)' \, . \qquad (6.3.11)$$

The first term is the estimation variance and is given by

$$Estimation\ Variance = X_F(X'W^{-1}X)X'W^{-1}\mathbb{E}(\varepsilon\varepsilon')W^{-1}X(X'W^{-1}X)X'_F$$

and the second term is the process variance given by

$$Process\ Variance \ = \mathbb{E}(\varepsilon_F{\varepsilon_F}')$$

All together means that the msep is given by

$$msep = X_F(X'W^{-1}X)^{-1}X'W^{-1}\mathbb{E}(\varepsilon\varepsilon')W^{-1}X\ (X'W^{-1}X)^{-1}{X_F}' + \ \mathbb{E}(\varepsilon_F{\varepsilon_F}') \qquad (6.3.12)$$

Proof. We know that $\widehat{Y_F} = X_F\hat{\beta}$ and $Y_F = X_F\beta + \varepsilon_F$. Using this, we may develop the $msep$ such as

$$\mathbb{E}\big(\widehat{Y_F} - Y_F\big)\big(\widehat{Y_F} - Y_F\big)' = \mathbb{E}(X_F\hat{\beta} - X_F\beta - \varepsilon_F)(X_F\hat{\beta} - X_F\beta - \varepsilon_F)'.$$

This means that the right-hand side of the above equation is rewritten as

$$\mathbb{E}\big(X_F(\hat{\beta}-\beta) - \varepsilon_F\big)\big(X_F(\hat{\beta}-\beta) - \varepsilon_F\big)' = \mathbb{E}\big(X_F(\hat{\beta}-\beta) - \varepsilon_F\big)\left((\hat{\beta}-\beta)'{X_F}' - {\varepsilon_F}'\right).$$

Developing this product, we get that

$$\mathbb{E}\left(X_F(\hat{\beta}-\beta)(\hat{\beta}-\beta)'X_F' + \varepsilon_F\varepsilon_F' - X_F(\hat{\beta}-\beta)\varepsilon_F' - \varepsilon_F(\hat{\beta}-\beta)'X_F'\right).$$

Applying the expected value operator, which is linear, to all parts, the following equation yields?

$$\mathbb{E}\left(X_F(\hat{\beta}-\beta)(\hat{\beta}-\beta)'X_F'\right) + \mathbb{E}(\varepsilon_F\varepsilon_F') - \mathbb{E}\left(X_F(\hat{\beta}-\beta)\right)\mathbb{E}(\varepsilon_F')$$
$$- \ \mathbb{E}(\varepsilon_F)\mathbb{E}\left((\hat{\beta}-\beta)'X_F'\right).$$

Since the expected value of the errors, current or future, is zero, then we get the *msep* as a sum of the estimation variance with the process variance,

$$\mathbb{E}\left(X_F(\hat{\beta}-\beta)(\hat{\beta}-\beta)'X_F'\right)+\mathbb{E}(\varepsilon_F\varepsilon_F').$$

We may develop a little bit further the estimation variance, i.e., the left-hand side term of the last equation, using (6.3.2) and (6.3.3) that allow us to have $\hat{\beta}-\beta=(X'W^{-1}X)^{-1}X'W^{-1}\varepsilon$,

$$\mathbb{E}((X_F(X'W^{-1}X)^{-1}X'W^{-1}\varepsilon)(\varepsilon'W^{-1}X\,(X'W^{-1}X)^{-1}X_F')).$$

Applying the expected value operator and knowing that X is not random, and it is independent of the random errors, we get

$$X_F(X'W^{-1}X)^{-1}X'W^{-1}\mathbb{E}(\varepsilon\varepsilon')W^{-1}X\,(X'W^{-1}X)^{-1}X_F'.$$

Our estimation variance will be dependent on what we assume to be the variance covariance matrix of the current and future errors. The current errors are coming from the estimation variance and the future errors from the process variance

$$\mathbb{E}(\widehat{Y_F}-Y_F)(\widehat{Y_F}-Y_F)'=$$

$$X_F(X'W^{-1}X)^{-1}X'W^{-1}\mathbb{E}(\varepsilon\varepsilon')W^{-1}X\,(X'W^{-1}X)^{-1}X_F'+\mathbb{E}(\varepsilon_F\varepsilon_F').$$

■

Following the results of Theorem 6.3.1, and seeing (6.3.12) we realize that we need to have the following to have the prediction error of this generalized link ratios:

- We need the parameter α to have the W, (6.3.6). Our decision was to choose the α that minimizes the prediction error.
- The W will give the vector of the loss development factors, given by (6.3.9), to have X_F.
- With the first two steps we will have the W_F matrices, see (6.3.7).
- And finally, we need $\mathbb{E}(\varepsilon\varepsilon')$ and $\mathbb{E}(\varepsilon_F\varepsilon_F')$, which implies some assumptions about the method that will use W and W_F, see Proposition (6.3.1).

The following proposition is very useful in what follows. Its proof is omitted as it derives straightforwardly from Theorem 6.3.1.

Proposition 6.3.2 *Following, (6.3.12), assumptions (6.3.4) and (6.3.5) and knowing that* σ^2*and* $X'W^{-1}$ *are diagonal matrices that may commute, the msep is given by*

$$\mathbb{E}\left(\widehat{Y_F} - Y_F\right)\left(\widehat{Y_F} - Y_F\right)' = X_F(X'W^{-1}X)^{-1}\sigma^2 X'_F + \sigma^2 W_F. \qquad (6.3.13)$$

As our regressions estimate the lower triangle cells, the ultimate losses are obtained in the last column and we know already that the *msep* of the estimated reserves equals the *msep* from the ultimate claims, see Lemma 5.1.4. This means that (6.3.13) equals the *msep* of the reserves, (and we know that its root is the prediction error).

6.3.5 Particular Univariate Methods

Thus, the main advantage of our approach is that we choose the alpha, α, which minimizes the prediction error. With α different from 0, 1 and 2, we get other distinct methods. Moreover, the choice of the weights of the link ratios is obtained such as the prediction error is minimized.

In the Proposition 6.3.1, the level of the heteroscedastic errors is given by the matrix W, which depends on X_{ij} that yields from the triangle data, and on the parameter α. Here, we use (6.3.13), i.e., the *msep* minimization, to get α.

Homoscedastic errors arise if $W = I$, the identity matrix. As we saw in section 3.6 and on section 6.1 each development year $j > 1$ may be seen as a dependent variable explained by a non-random independent variable, given by the previous development year $j - 1$. Heteroscedasticity may appear on each regression for reasons such as: an increase/decrease of claims on particular origin years, an increase/decrease of the speed of paying claims on certain origin years for the same development year, the presence of outliers (which should be, if possible, previously removed), a bad specification of the method that may be more severe in certain origin years, for example, we may need other variables to explain method (6.3.1) or to have a different functional form between the dependent and the independent variables.

All the link ratios methods considered here, see next corollaries (6.3.1), (6.3.2), depend on α which represents the level of heteroscedasticity. We want to choose α that minimizes the prediction error. As we saw in section (5.4), the lower the errors and the better the back-testing, the lower is the prediction error.

Particular cases of the method are considered with the next three corollaries. Obviously, the proofs of those corollaries are linked with (6.36), Theorem 6.3.1 and Proposition 6.3.1, and they are omitted.

Corollary 6.3.1 *If* $\alpha = 0$ *, variances are homoscedastic, and we get, see (6.3.4) and (6.3.5)*

$$\mathbb{E}(\varepsilon\varepsilon') = \mathbb{E}(\varepsilon_F{\varepsilon_F}') = \sigma^2 I,$$

Here I *is a diagonal identity matrix with size* $m \times m$*. With* α *=0, the loss development factors are equal to the ones from the VP,* $\hat{\beta}^{VP} = (X'X)^{-1}XY$*, see (6.3.9) with* $W = I$ *. Then, the msep is given by, see (6.3.13)*

$$\mathbb{E}\left(\widehat{Y_F} - Y_F\right)\left(\widehat{Y_F} - Y_F\right)' = X_F(X'X)^{-1}\sigma^2 {X'}_F + \sigma^2 \tag{6.3.14}$$

Corollary 6.3.2 *If* $\alpha = 1$*, variances are heteroscedastic, and we get*

$$\mathbb{E}(\varepsilon\varepsilon') = \sigma^2 W_{CL}$$

and

$$\mathbb{E}(\varepsilon_F{\varepsilon_F}') = \sigma^2 W_{F,CL},$$

with

$$W_{CL} = \begin{bmatrix} x_{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{T-1,T-1} \end{bmatrix} \quad \text{and} \quad W_{F,CL} = \begin{bmatrix} x_{T,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{T,T-1} \end{bmatrix}.$$

With α *=1, the loss development factors are equal to the ones from the CL, see (6.3.6) and (6.3.9). Then, the msep is given by, see (6.3.13),*

$$\mathbb{E}\left(\widehat{Y_F} - Y_F\right)\left(\widehat{Y_F} - Y_F\right)' = X_F(X'{W_{CL}}^{-1}X)^{-1}\sigma^2 {X'}_F + \sigma^2 W_{F,CL}. \tag{6.3.15}$$

Corollary 6.3.3 *If* $\alpha = 2$*, variances are heteroscedastic, and we get*

$$\mathbb{E}(\varepsilon\varepsilon') = \sigma^2 W_{SA},$$

and

$$\mathbb{E}(\varepsilon_F{\varepsilon_F}') = \sigma^2 W_{F,SA},$$

with

$$W_{SA}=\begin{bmatrix} x_{1,1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{T-1,T-1}^2 \end{bmatrix} \quad \textit{and} \quad W_{F,SA}=\begin{bmatrix} x_{T,T}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{m,m}^2 \end{bmatrix}.$$

With α =2, the loss development factors are equal to the ones from the SA, see (6.3.6) and (6.3.9). Then, the msep is given by, see (6.3.13)

$$\mathbb{E}\left(\widehat{Y_F}-Y_F\right)\left(\widehat{Y_F}-Y_F\right)'=X_F\left(X'{W_{SA}}^{-1}X\right)^{-1}\sigma^2X'_F+\sigma^2W_{F,SA}. \tag{6.3.16}$$

Remark 6.3.1: In Portugal et al. (2017), Mack (1993a, 1994)'s framework with heteroscedastic variances was considered. In that context, the variance of the payments was proportional to the weights of the link ratios. Here, we consider a regression method with the homoscedastic and heteroscedastic cases and the VP emerges as the homoscedastic model.

In this GLR method we may see from (6.3.6) that we assume no correlations between the errors of the equations. This means that we may estimate the method by doing univariate independent regressions or using the framework presented in this section.

The reason why we developed the framework above was to have a more flexible and integrated approach and to have the multivariate method, the MGLR. This method will be presented in the following section and will be developed using the GLR framework, but with different assumptions.

6.4 Stochastic Multivariate Generalized Link Ratios

A general multivariate method is presented in this section. Some particular multivariate methods are also identified. The methods are a continuation of the methods presented in section 6.3 but now with a multivariate framework, as summarized in section 6.1.

6.4.1 General Multivariate Method

The method considered here is the same presented in section 6.3.1. However, we will introduce a more complex structure with the errors of all the equations correlated and with the possibility of having a homoscedastic or heteroscedastic structure. Following Srivastava and

Giles (1987): *"we are going to assume that the equations may be linked statistically, even though not structurally"*.

This was called by Zellner (1962) the Seemingly Unrelated Regressions (SUR) method. A SUR method may be needed when the method does not consider all the variables that explain the dependent variable (in our case the insurer´s payments in each development year). If that is the case, then the error of that regression will show the impact of that missing variable. As all the regressions of the method are missing the same variables, it is possible that the equations, even if not structurally linked, have some statistical link.

The MGLR method will become multivariate as a SUR and may also use the homoscedastic or heteroscedastic structure from any α from the GLR, including also the VP, the CL and the SA. The development of the method is easily done with the framework developed in section 6.3. In this MGLR method, we are going to maintain the entire framework presented in section 6.3 but we are going to change the assumptions (6.3.4) and (6.3.5). We are going to assume contemporaneous correlations between the errors of the different equations and we get a multivariate method. The method is still based on (6.3.2) and even (6.3.3) will be similar.

6.4.2 Assumptions

Σ is a block matrix of block-size $T-1 \times \mathrm{T}-1$ that summarizes the variances and the covariance between $j=1,\dots,T-1$ regressions. Expanding each block, we get a matrix of dimension $m \times \mathrm{m}$

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,T-1} \\ \vdots & \ddots & \vdots \\ \Sigma_{T-1,1} & \cdots & \Sigma_{T-1,T-1} \end{bmatrix}. \tag{6.4.1}$$

The generic component of (6.4.1), $\Sigma_{j,j}$ is given by a matrix of size $(T-j)\times(T-j)$

$$\Sigma_{j,j} = \mathrm{s}_{j,j} \begin{bmatrix} {x_{1,j}}^{\alpha} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{T-j,j}}^{\alpha} \end{bmatrix} \tag{6.4.2}$$

The generic component of (6.4.1), $\Sigma_{l,j}$ with $l \neq j$ is given by a matrix of size $(T-j)\times(T-j)$, where I_{T-j} is an identity matrix of size $T - j$.

$$\Sigma_{l,j} = \mathrm{s}_{l,j} I_{T-j} \tag{6.4.3}$$

Σ^F is a block matrix of block-size $T-1\ \times \mathrm{T}-1$ that summarizes the future variances and the covariances between $j = 1, \ldots, T-1$ regressions. Expanding each block, we get a matrix of dimension $m\ \times \mathrm{m}$

$$\Sigma^F = \begin{bmatrix} \Sigma^F_{1,1} & \cdots & \Sigma^F_{1,T-1} \\ \vdots & \ddots & \vdots \\ \Sigma^F_{T-1,1} & \cdots & \Sigma^F_{T-1,T-1} \end{bmatrix}. \tag{6.4.4}$$

The generic component of (6.4.4), $\sum^F_{j,j}$ is given by a matrix of size $(T-j) \times (T-j)$

$$\Sigma^F_{j,j} = \mathrm{s}_{j,j} \begin{bmatrix} {x_{T,j}}^{\alpha} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{T+jj,j}}^{\alpha} \end{bmatrix} \tag{6.4.5a}$$

The generic component of (6.4.4), $\Sigma_{l,j}$ with $l \neq j$ is given by a matrix of size $(T-j) \times (T-j)$

$$\Sigma^F{}_{l,j} = \mathrm{s}_{l,j} I_{T-j} \tag{6.4.5b}$$

Proposition 6.4.1 *Considering a multivariate method that allows for heteroscedasticity of the errors inside each equation and contemporaneous correlations between these equations, we assume for our MGLR method*

$$\mathbb{E}(\varepsilon|X) = \mathbb{E}(\varepsilon) = 0 \tag{6.4.6}$$

$$\mathbb{E}(\varepsilon\varepsilon') = \Sigma \tag{6.4.7}$$

$$\mathbb{E}(\varepsilon_F {\varepsilon_F}') = \Sigma^F \tag{6.4.8}$$

6.4.3 Estimation

The parameters estimation may be obtained by the following Lemma 6.4.1.

Lemma 6.4.1 *Following, Zellner (1962) or Srivastava and Giles (1987), we may get the estimation of the β, that is the estimation of the loss development factors from all the equations. The $\hat{\beta}$ is obtained using the SUR generalized least squares and it is the best linear unbiased estimator of β.*

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X\Sigma^{-1}Y \tag{6.4.9}$$

We also need an expression for the prediction error which is given by the following section.

Lemma 6.4.2 Following Zellner (1962), *the estimators for the parameters of variance and covariance matrix from a multivariate regression are given by*

$$\hat{s}_{j,j} = \frac{1}{T-1}SSR_j \qquad \hat{s}_{i,j} = \frac{1}{T}SSR_j \tag{6.4.10}$$

The SSR_j (the sum of the square of the residuals from equation j) are calculated using for each regression j the Ordinary Least Squares (OLS) sum of the square of the errors (also called residuals). This means that we need to run a second regression using the OLS method, that is (6.4.9) considering Σ given by an identity matrix of equal size.

6.4.4 Prediction Error

The next theorem is similar to Theorem 6.3.1 and gives us a general non-recursive formula to have the prediction error, here for the method (6.3.2).

Theorem 6.4.1 *The mean square error of prediction from the method presented in (6.3.2) is given by.*

$$msep = X_F(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\mathbb{E}(\varepsilon\varepsilon')\Sigma^{-1}X\,(X'\Sigma^{-1}X)^{-1}X_F' + \mathbb{E}(\varepsilon_F\varepsilon_F') \tag{6.4.11}$$

The proof follows directly from Theorem 6.3.1 when (6.4.4), (6.4.5) and (6.4.6) are considered.

Following the results of Theorem 6.4.1, the procedures are like the ones from the univariate method, presented on section 6.3. In the MGLR we need to do the following:

- We must get the $\hat{s}_{i,j}^2$ and $\hat{s}_{j,j}^2$ to estimate the $\sum$ matrix, which implies to have an extra regression, with OLS, to get the sum of the square of the errors.
- Then we need the parameter α to have the Σ, (6.4.2) and the Σ^F (6.4.3) matrices. Our suggestion is to choose the α that minimizes the prediction error.
- We will also have from α the vector of the loss development factors, given by (6.4.9), to have X_F.

- Having $\sum$ and Σ_F we have $\mathbb{E}(\varepsilon\varepsilon')$ and $\mathbb{E}(\varepsilon_F{\varepsilon_F}')$ and we will get the minimum prediction error.

Proposition 6.4.2 *Following (6.3.12) and assumptions from Proposition (6.4.1) the msep is given by*

$$\mathbb{E}\left(\widehat{Y_F} - Y_F\right)\left(\widehat{Y_F} - Y_F\right)' = X_F(X'\Sigma^{-1}X)^{-1}X'_F + \Sigma^F. \qquad (6.4.12)$$

Clearly, the parameters $s_{j,j}$ and $s_{j,j}$ are not known and must be estimated using (6.4.10).

As in the univariate method from section 6.3, we choose the α which minimizes the prediction error. Analytically, we do not get anymore the loss development factors from, the VP for $\alpha = 0$, the CL for $\alpha = 1$ and the SA for $\alpha = 2$. The reason is the consideration of contemporaneous correlations between the regressions that change the loss development factors, indeed, the (6.4.9) equation is not equal to (6.3.9). However, we may say that, when $\alpha = 0$ we get a Multivariate VP, when $\alpha = 1$ we get a Multivariate CL and when $\alpha = 2$ we get a Multivariate SA. The argument for this is the heteroscedasticity level and its relation with these methods.

What defines and differentiates these three methods are the weights given to the link ratios and the weights also define of the heteroscedasticity level. In the VP is zero, $\alpha = 0$, in the CL is one, $\alpha = 1$ and in the SA is two, $\alpha = 2$. We may say that the heteroscedasticity level may be defined by α, the weights of the link ratios, see (6.2), where given by $2 - \alpha$. This means that the homoscedastic case is a particular case of the heteroscedastic methods when the level of heteroscedasticity is zero. These levels of heteroscedasticity are maintained in the multivariate approach.

As with the univariate method, we will get other methods for different α's, as they give other weights to the link ratios (and other levels of heteroscedasticity). As with the univariate method from section 6.3 the optimal α is the one that minimizes the prediction error.

In the Proposition 6.4.1, the level of the heteroscedastic errors and of correlation is given by the matrix Σ. The latter depends on the variance-covariance parameters s_{lj} and s_{jj}, on X_{ij} (that comes from the triangle data) and on the parameter α. Here, we use (6.4.11), i.e., the *msep* minimization, to get α. Homoscedastic errors may also arise also here if $\Sigma_{l,j} = I$. The univariate method is a particular case of the multivariate method when $s_{lj} = 0$. Correlations between regressions appear on the data for several reasons such as an increase/decrease of claims on some development years or an increase/decrease of the speed of paying claims on certain development years.

6.4.5 Particular Multivariate Methods

Particular cases of the method are considered with the next three corollaries. Obviously, the proofs of those corollaries are linked with (6.3.6), Theorem 6.4.1 and Proposition 6.4.1, and they are omitted.

Corollary 6.4.1 *If* $\alpha = 0$, *variances are homoscedastic, and the regressions correlated, we get the Multivariate VP*

$$\mathbb{E}(\varepsilon\varepsilon') = \Sigma_{VP}$$

and

$$\mathbb{E}(\varepsilon_F {\varepsilon_F}') = \Sigma_{VP}^F$$

where Σ_{VP} *and* Σ_{VP}^F *are, respectively,* Σ *and* Σ^F *as defined in (6.4.1) and (6.4.4) with*

$$\Sigma_{j,j} = s_{j,j} \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

$$\Sigma_{j,j}^F = s_{j,j} \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

With α *=0, the loss development factors are equal to the ones from the multivariate VP. Then, the msep is given by*

$$X_F(X'\Sigma^{-1}X)^{-1}X'_F + \Sigma^F$$

$$\mathbb{E}\left(\widehat{Y_F} - Y_F\right)\left(\widehat{Y_F} - Y_F\right)' = X_F(X'{\Sigma_{VP}}^{-1}X)^{-1}{X_F}' + \Sigma_{VP}^F \qquad (6.4.14)$$

Corollary 6.4.2 *If* $\alpha = 1$, *variances are heteroscedastic, and the regressions correlated, we get the Multivariate CL*

$$\mathbb{E}(\varepsilon\varepsilon') = \Sigma_{CL}$$

and

$$\mathbb{E}(\varepsilon_F {\varepsilon_F}') = \Sigma_{CL}^F$$

where Σ_{CL} *and* Σ_{CL}^F *are, respectively,* Σ *and* Σ^F *as defined in (6.4.1) and (6.4.4) with*

$$\Sigma_{j,j} = s_{j,j} \begin{bmatrix} x_{1,j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{T-j,j} \end{bmatrix}$$

$$\Sigma_{j,j}^F = s_{j,j} \begin{bmatrix} x_{T,j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{T+j,j} \end{bmatrix}$$

With α *=1, the loss development factors are equal to the ones from the multivariate CL. Then, the msep is given by*

$$\mathbb{E}\left(\widehat{Y_F} - Y_F\right)\left(\widehat{Y_F} - Y_F\right)' = X_F (X' {\Sigma_{CL}}^{-1} X)^{-1} {X'}_F + \Sigma_{CL}^F \tag{6.4.15}$$

Corollary 6.4.3 *If* $\alpha = 2$, *variances are heteroscedastic, and the regressions correlated, we get the Multivariate SA*

$$\mathbb{E}(\varepsilon\varepsilon') = \Sigma_{SA}$$

and

$$\mathbb{E}(\varepsilon_F {\varepsilon_F}') = \Sigma_{SA}^F$$

where Σ_{SA} *and* Σ_{SA}^F *are, respectively,* Σ *and* Σ^F *as defined in (6.4.1) and (6.4.4) with*

$$\Sigma_{j,j} = s_{j,j} \begin{bmatrix} {x_{1,j}}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{T-j,j}}^2 \end{bmatrix}$$

$$\Sigma_{j,j}^F = s_{j,j} \begin{bmatrix} {x_{T,j}}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{T+j,j}}^2 \end{bmatrix}$$

With α =2, the loss development factors are equal to the ones from the multivariate SA. Then, the msep is given by

$$\mathbb{E}\left(\widehat{Y_F} - Y_F\right)\left(\widehat{Y_F} - Y_F\right)' = X_F(X'\Sigma_{SA}{}^{-1}X)^{-1}X'_F + \Sigma_{SA}^F \tag{6.4.16}$$

6.5 Numerical Results

To have some comparable numerical results we considered two triangles of cumulative payments, used already in section 5.2.1 and 5.2.2.

As we saw in these sections, the payments in the first triangle are irregular and in the second triangle they are regular.

We will see that, in both cases, the conclusions confirm the ones we presented in Portugal et al. (2017). Additionally, a use test was conducted with 114 paid claims triangles and with 10 years of information, as we did in section 5.3.

6.5.1 Irregular Development of Data

In this section we consider the Mack (1993a) irregular data used in section 5.2. The explanation of why is the data considered irregular, see Table 2.1, was already given in section 5.2.

6.5.1.1 Replication of Mack (1993b, 1994) Results

The GLR method for $\alpha = 1$ replicates Mack (1993a, 1993b, 1994)'s results, i.e., the loss development factors, the variance, and the prediction error (see Table 6.1).[15] We are able to obtain Mack's results with a prediction error of 52% when we consider that the process variance equals $\sigma^2 W_{CL}$ instead of $\sigma^2 W_{F,CL}$, see Corollary 6.4.2. In other words, the Mack's

---

[15] This one with a difference of 1%, due to the number of operations performed with the associated rounding.

stochastic distribution-free method is replicated when the prediction error formula assumes that the process variance depends on current weights, that is W_{CL}. As we saw before the weights of the link ratios are related with the heteroscedasticity level, see Corollaries 6.4.1, 6.4.2 and 6.4.3.

When we consider the process variance as given by $\sigma^2 W_{F,CL}$, see Corollary 6.4.2, the Mack method prediction error increases to 99% (see Table 6.2).

Table 6.1: Mack (1993, 1994)'s results with irregular data

Development Year	Loss Development Factors	Variance
2	2,999	27 883,479
3	1,624	1 108,526
4	1,271	691,443
5	1,172	61,230
6	1,113	119,439
7	1,042	40,820
8	1,033	1,343
9	1,017	7,883
10	1,009	1,343

Development Year	Estimated Reserves	Prediction Error
2	154	287%
3	617	138%
4	1 636	177%
5	2 747	135%
6	3 649	80%
7	5 435	146%
8	10 907	115%
9	10 650	150%
10	16 339	129%
Total	52 135	53%

Table 6.2: Replication of Mack (1993, 1994)'s results for irregular data

Future Heteroscedasticity	Our Model	Mack Model
Current	53%	54%
Estimated from Future Payments Zellner (1962) estimator)	99%	

6.5.1.2 Generalized Link Ratios

Here we estimate the parameter α that minimizes the prediction error on the GLR. Numerically, this may be achieved straightforwardly using a simple toolbox like the Excel Solver by defining the prediction error, see (6.3.1.3), as an objective function and the parameter α as the variable to be changed. Thus, as it is shown in Figure 6.1, for $\alpha = 0$, we get a prediction error of 36.1% and we confirm the VP as the ones that minimize the prediction error. This compares with 53% to the CL ($\alpha = 1$). Indeed, it seems that we do not have significant heteroscedastic errors in this triangle. We also report that, as soon we leave the low values of α, the prediction error increases in a non-linear way (see Figure 6.1).

Figure 6.1: Prediction error: Generalized Link Ratios irregular data

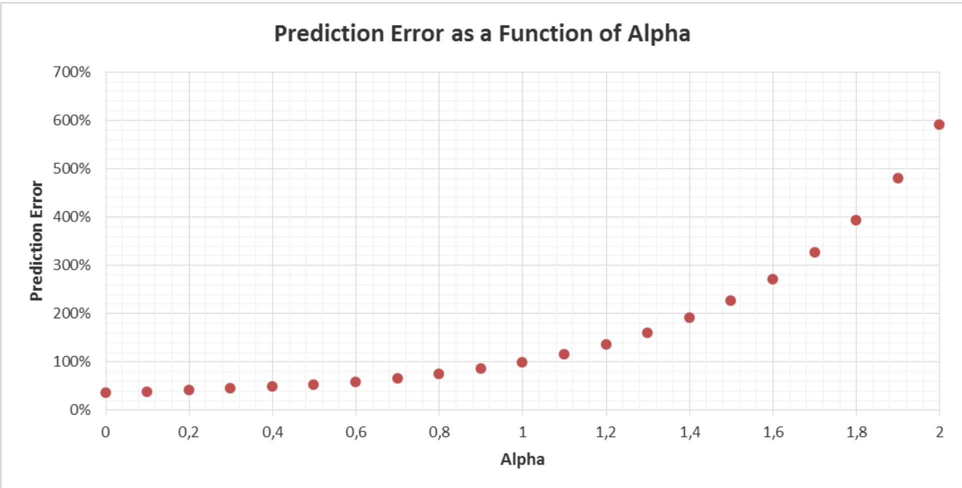


Table 6.3: Generalized Link Ratios α=0 for irregular data

Column	Reserves per Column	Prediction Error	Prediction Error %
2	2 511	3 773	150%
3	5 672	5 359	94%
4	7 501	6 627	88%
5	7 867	7 021	89%
6	7 208	5 801	80%
7	4 283	6 015	140%
8	4 412	4 488	102%
9	2 620	4 421	169%
10	1 698	2 074	122%
Total	**43 772**	**15 811**	**36,1%**

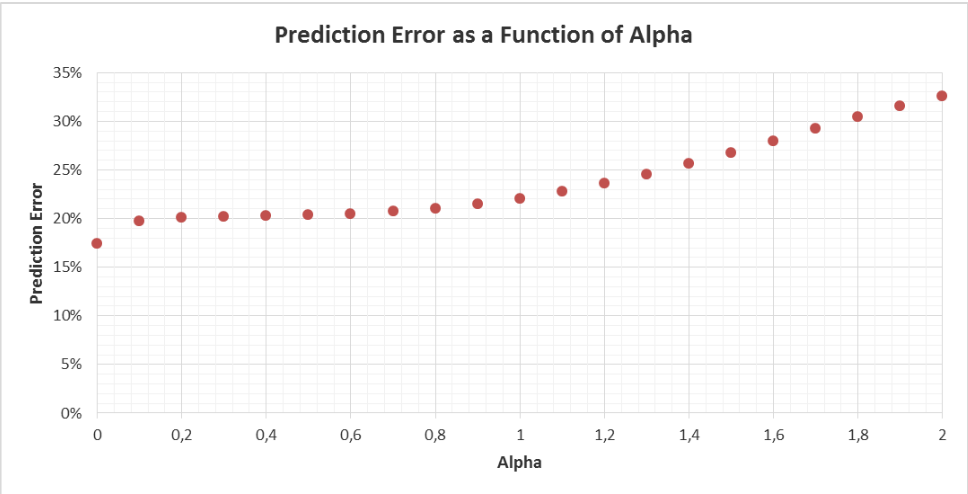
The reserves estimated are of 43 772, and they are reported in Table 6.3, which compare with the CL with 52 135 (see Table 6.1). This result confirms that we may improve the CL with other methods like the VP estimated through the GLR, and here, with homoscedastic errors.

6.5.1.3 Multivariate Generalized Link Ratios

In this section, the same procedure is followed as in the previous method, but now we minimize the prediction error using the MGLR method instead, i.e., by using (6.4.11). The method is multivariate and assumes contemporaneous correlations between the equations that exist inside the triangle. As before, the Excel Solver is used to estimate the parameter α that minimizes the prediction error. For $\alpha = 0$, we get a prediction error of 17.5.%. This is comparable with 22.1%, when we have $\alpha = 1$, i.e., the heteroscedastic structure from the multivariate CL (Figure 6.2).

Once again, we report an improvement over the CL when the VP is used. But here, we may also see that the multivariate method also improves the prediction error in both the VP and the CL.

Figure 6.2: Prediction error: Multivariate Generalized Link Ratios irregular data



The reserves estimated are of 45 638, and they are reported in Table 6.4, which compare with the GLR reserves of 43 772. The results show that we may improve the GLR considering the method as multivariate.

Table 6.4: Multivariate Generalized Link Ratios α=0 irregular data

Development Year	Estimated Reserves	Prediction Error
2	154	126%
3	694	65%
4	1 623	54%
5	2 787	27%
6	3 715	48%
7	5 163	50%
8	10 531	12%
9	9 267	61%
10	11 706	87%
Total	45 638	17,5%

6.5.2 Regular Development of Data

In this section, we consider the set of data from Taylor and Ashe (1983), also used in section 5.2 and presented in Table 5.2.

We start by replicating Mack (1993b, 1994) results.

6.5.2.1 Replication of Mack (1993b, 1994) Results with Regular Data

Again, the GLR method for $\alpha = 1$ replicates Mack (1993b, 1994) results with Taylor and Ashe (1983) data. The replication includes, the loss development factors, the variance, and the prediction error (see Table 6.5.).[16] Additionally, here, we approximate Mack (1993b, 1994) prediction error of 13%. Our prediction error was of 10% (assuming that the future weights matrix, $W_{F,CL}$, equals to the current weights matrix, W_{CL}). This means that the Mack's method is replicated when the prediction error calculation assumes that the process variance is given by $\sigma^2 W_{CL}$. Considering Corollary 6.4.2, which considers that the process variance is given by $\sigma^2 W_{F,CL}$, the prediction error increases to 15% and stays 2 percentage points above Mack's results (see Table 6.6).

[16] This one also with a difference of 3 percentage points, probably due to the number of operations performed with the associated rounding.

Table 6.5: Mack (1993, 1994)'s results for Regular Data

Parameters	
Loss Development Factors	**Variance**
3,491	160280,327
1,747	37736,855
1,457	41965,213
1,174	15182,903
1,104	13731,324
1,086	8185,772
1,054	446,617
1,077	1147,366
1,018	446,617

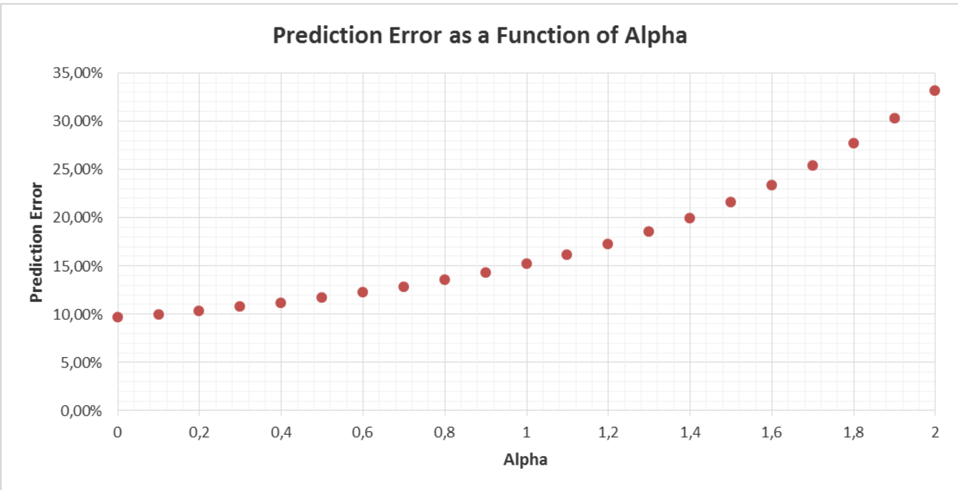
Table 6.6: Replication of Mack (1993, 1994)'s results for Regular Data

Future Heteroscedasticity	Our Model	Mack Model
Current	10%	13%
Estimated from Future Payments Zellner (1962) estimator)	15%	

6.5.2.2 Generalized Link Ratios

Here we estimate the parameter α that minimizes the prediction error on the GLR using again the Excel Solver. For $\alpha = 0$, we get a prediction error of 9.7%. This compares with 15.1% to the CL ($\alpha = 1$). It seems that we do not have heteroscedastic errors (Figure 6.3).

Figure 6.3: Prediction error: Multivariate Generalized Link Ratios regular data



The reserves estimated are of 1 789 811 (see Table 6.7) which compares with the CL estimated reserves of 2 822 035. The results show that we may improve the CL with other methods like the VP.

Table 6.7: Generalized Link Ratios: α=0 for regular data

Row	Reserves per Row	Prediction Error	Prediction Error %
2	94 634	242 020	256%
3	478 103	349 855	73%
4	723 104	478 207	66%
5	1 002 041	570 839	57%
6	1 408 034	660 893	47%
7	2 131 332	837 597	39%
8	3 885 296	700 641	18%
9	4 255 237	715 968	17%
10	4 501 720	571 014	13%
Total	**18 479 500**	**1 789 811**	**9,7%**

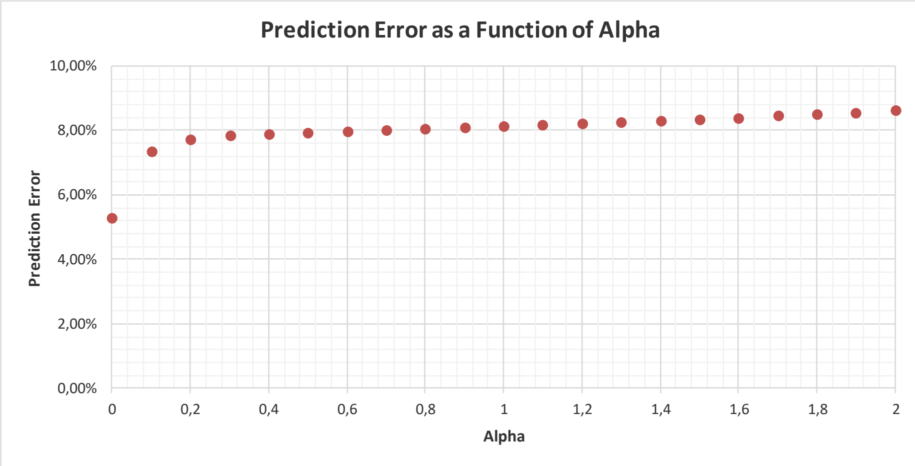
6.5.2.3 Multivariate Generalized Link Ratios

In this section, we follow the same procedure as previously, but now we will minimize the prediction error in the MGLR, i.e., by using (6.4.11). For $\alpha = 0$, we get a prediction error of 5.4%. This compares with 8.1%, when we have $\alpha = 1$, the heteroscedastic structure from the multivariate CL.

Once again, an improvement is observed over the CL when the VP is used (Figure 6.4). We may also see that the VP is the method that minimizes the prediction error of our multivariate method. Finally, the multivariate approach also presents lower prediction errors, both in the CL and VP multivariate methods, when compared with the univariate versions.

The reserves estimated are of 1 065 939, and they are reported in Table 6.8, which compare with the GLR optimum of 1 789 811.

Figure 6.4: Prediction error: Multivariate Generalized Link Ratios regular data



The results show that we may improve the GLR considering the method as multivariate. We also have a significant drop of the reserves.

Table 6.8: Multivariate Generalized Link Ratios: α=0 regular data

Row	Reserves per Row	Prediction Error	Prediction Error %
2	94 634	242 020	256%
3	519 654	324 937	63%
4	765 334	555 500	73%
5	1 127 046	535 655	48%
6	1 563 631	589 963	38%
7	2 323 548	630 746	27%
8	4 106 477	155 152	4%
9	4 502 111	249 467	6%
10	4 664 560	488 305	10%
Total	**19 666 994**	**1 065 939**	**5,4%**

6.6 Use Test

In this subsection, to interpret better the results derived by the previously presented numerical examples and to also provide a “business orientated” analysis, as in Portugal et al. (2017), we select 114 triangles with paid claims and 10 years of information to be comparable directly to the previous cases. Thus, the parameter α that was selected to minimize the prediction error is derived. Note that it was not possible to have a solution for 8 triangles due to the existence of several zeros on the triangle cells thus we decide to exclude them from the analysis.

Observing Table 6.9, we conclude the following for the GLR: 75% of the 106 triangles considered have $\alpha = 0$, which confirms again the preference of VP against other methods (Portugal et al. 2017), however here with homoscedastic errors.

This also means that just 25% of the cases seem to show heteroscedastic errors. However, in total, 89% of the remaining 106 cases have α less or equal to 0.5. What is more, only 9% of the cases are in the CL zone, i.e., $\alpha \in [0.5, 1.5]$, and just 2% are in the SA zone. Interestingly, the only method where parameter α receives an exact number is the VP method, i.e. $\alpha = 0$. Both the CL and SA methods, with $\alpha = 1$ and $\alpha = 2$, respectively, are never confirmed.

The conclusions are similar for the MGLR, but the number of cases with $\alpha = 0$ is reduced to 45% (Table 6.10). However, when we sum up the number of cases including $\alpha = 0.1$, the proportion is quite like the GLR, i.e., 75%.

It seems that the fact that we assume dependencies between the regressions obliges a small increase of the heteroscedasticity.

Table 6.9: Values for parameter α: Generalized Link Ratios

α	**Number of cases**	**% of all cases**
0	80	75%
[0, 0.5]	94	89%
[0, 1]	99	93%
[0, 1.5]	104	98%
[0, 2]	105	99%
[0, 3]	106	100%

Table 6.10: Values for parameter α: Multivariate Generalized Link Ratios

α	**Number of cases**	**% of all cases**
0	45	42%
[0, 0.5]	87	82%
[0, 1]	88	83%
[0, 1.5]	91	86%
[0, 2]	97	92%
[0, 3]	106	100%

6.7 Testing for Heteroscedasticity

As we saw before, in chapters 5 and in current chapter 6, the main methods considered, VP, CL and SA, can be replicated by a regression model. However, in such a model in this chapter, the VP method assumes that there is homoscedasticity (no heteroscedasticity) and the CL and the SA assumes there is heteroscedasticity. In this section, we use a known regression test to verify the existence of heteroscedasticity. Finally, we compare its results with the α obtained when the prediction error is minimized.

The literature presents several heteroscedasticity tests; see for example Fomby et al. (1984) or Hill et al. (2012).

A first test that may be done is the residual plot analysis for all the regressions. We may plot the errors against the cumulative payments and check if there is any visible pattern. For example, if the errors are growing/decreasing with the payments that may be a sign of heteroscedasticity. Otherwise, if there is not any pattern this means that we should have homoscedasticity.

The problem with visual tests is that they are not a statistical test. To do a statistical test we have several alternatives.

The first one is the Lagrange Multiplier Test, also called Breusch-Pagan test, see Hill et al. (2012). It's a large sample test and as the triangles we are using, see Tables (2.1) and (5.19), have $T = 10$, it's not the best approach for analysis when we want to analyse the errors per regression (even knowing that putting all the regressions together we get $m = 45$). Also, very often, insurers work with triangles with $T < 10$. Due to all these reasons, we decide not to apply this test.

The second possibility are the Goldfeld-Quandt test and the Chow test, see Fomby et al. (1984). The tests are designed for two groups of data with possibly different variances. This means that for each regression we need to have two regressions, one for lower variances and another one for higher variances. The consequence of this on claims reserving triangle will be a reduced number of observations on each regression, for instance for $T = 10$ the first regression, which just uses 9 observations from this 10, will be split in two regressions, one with 4 observations and another one with 5 (or maybe one with 3 observations and the other one with 6). The degrees of freedom to estimate the regressions will be low and for this reason we do not apply this test.

The third test, White (1980) test, see Hill et al. (2012), is more feasible to apply to claims reserving triangles: for each equation from the payments triangle, for example, Table 2.1, we just need to use OLS to each regression. For each of these regressions we estimate the errors squared $\hat{\varepsilon}_i^{\,2}$ as a function of the independent variables that explains the errors behaviour, Z_1,$Z_2, \dots$ and its cross-products. For regressions without a constant, as we are considering, Baltagi (2013) suggests to only use $Z_1^{\,2}$.

There will be also an error term v_i which sum of squared of the errors will be minimized to obtain a and b, the parameters from equation (6.6.1). This error has the usual OLS properties, see White (1980) and Hill et al. (2012).

$$\hat{\varepsilon}_i^{\,2} = a + bZ_1^{\,2} + v_i \tag{6.6.1}$$

In claims reserving triangles, when they are considered a regression model, the errors, what we cannot explain on origin year i from a specific column j, may be explained by several factors that we may use as explanatory variables. Indeed, considering Z the level of payments, should reflect several factors that may affect the errors. For example:

- If in one cell the payments are too high, compared with other origin years, the error may be higher, because the method may not be able to capture that feature for that cell. The residual, the difference from the true payments and the forecasted payments will be positive.
- With the same argument, if in one cell the payments are too low, the error may be lower. The residual, the difference from the true payments and the forecasted payments will be negative.

In (6.6.1) the payments will also act like a proxy that may summarize several other factors, like the speed of paying claims or the higher/lower frequency of claims in some years. There is also an advantage in using just one explanatory variable: the degrees of freedom do not decrease so much (something always important in triangles claims reserving due to the low number of observations at some development years).

In applying the White test, we used the information available from the datasets to explain the errors. We did two tests. In the first, we will test the method (6.6.1) for regressions 1 to 7. The regressions 8 to 9 are excluded due to the small number of observations. In a second test, we will apply the White test to all the regressions at the same time (including data on regressions 1 to 9).

The White test is performed as an F-test, see Hill et al. (2012), and test the null hypothesis that the parameter b from the regression is zero against the alternative hypothesis that the parameter b from the regression is different from zero. If we reject the null hypothesis we will have an indication of heteroscedasticity.

We reject the null hypothesis if the statistic χ^2 is higher than the 5% critical value ${\chi^2}_{1-0.05,df}$, where df means the degrees of freedom. The latter are the number of regressions without constant, which is 1 for the first test and 9 to the second test. The statistic χ^2 is the product of the regression number of observations by its R-squared. The latter summarizes how close the data is from the fitted regression.

From the Chi-Square distribution table and for one degree of freedom we get ${\chi^2}_{0.95,1} = 3.84$. For nine degrees of freedom we get ${\chi^2}_{0.95,9} = 16.9$. More details on the statistic may be seen in Hill et al. (2012).

The following analysis is done for the GLS univariate method but using OLS errors. These are necessary to test for heteroscedasticity (the GLS errors already include heteroscedasticity). The GLS is a method that considers heteroscedasticity for $\alpha \neq 0$. The case of $\alpha = 0$, the VP, corresponds to an OLS model.

For the MGLR, another analysis will be done in section 6.8, testing the possibility of having (6.4.3) a diagonal matrix that is $\sum$ showing absence of correlations between the triangle equations.

We will analyse the two datasets considered in the previous numerical analysis.

6.7.1 Irregular Development of Data

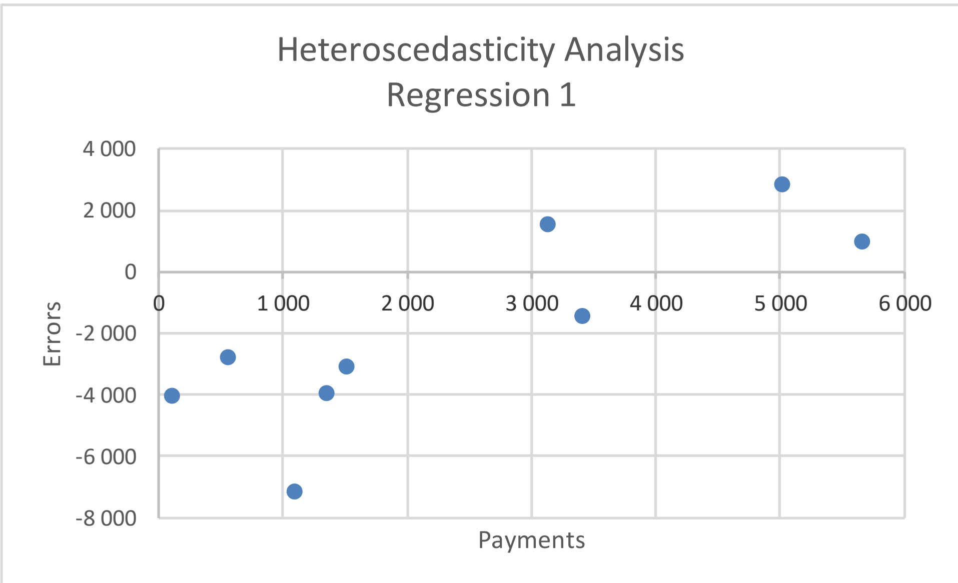
In this section, we consider the same irregular data used on section 5.2 from Mack (1993a, 1993b, 1994). On section 5.2 it was explained why the data is considered irregular.

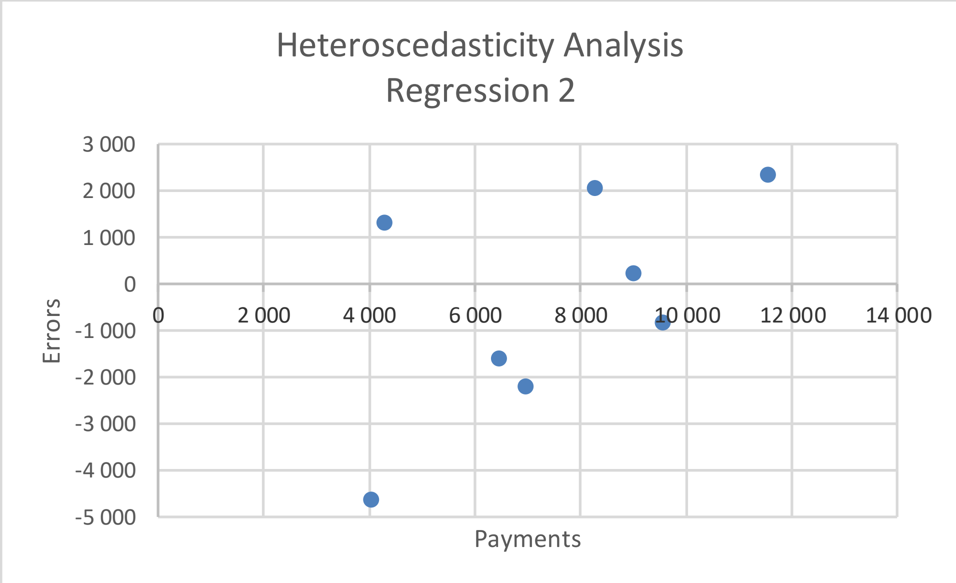
6.7.1.1 Regression Error's Plots

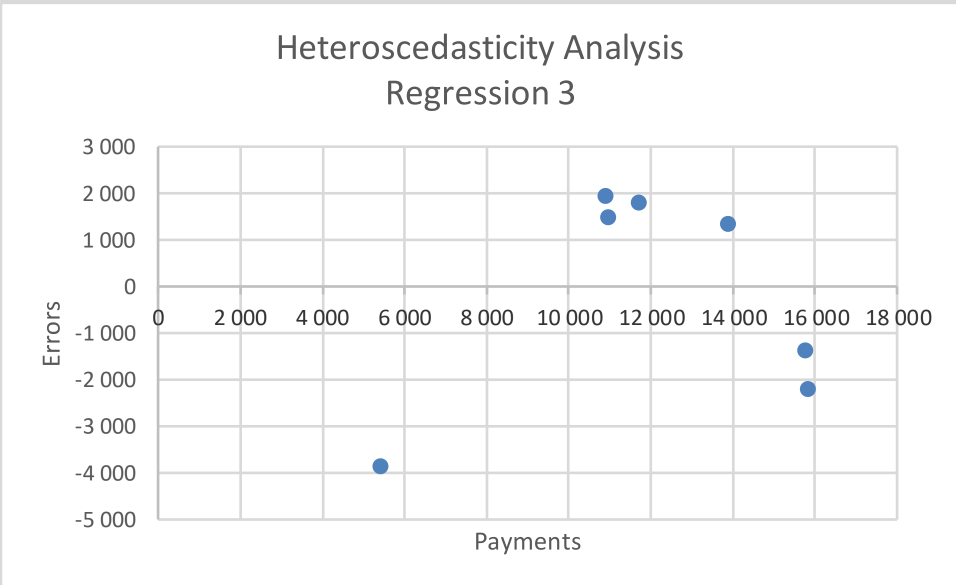
In Figure 6.5 we analyse the first three regressions, $k = 1,2,3$. We may see in regressions $k = 1$ and $k = 2$ plots that the payments increase, and the errors are also increasing.

However, the range of the errors is mostly, between -4000 and 4000 for equation 1, between -3000 and 3000 for equation 2 and between -2000 and 2000 for equation 3. There is just one case, on each equation, outside the respective equation range.

Figure 6.5: Regression Error's Plots for Individual Regressions 1 to 3



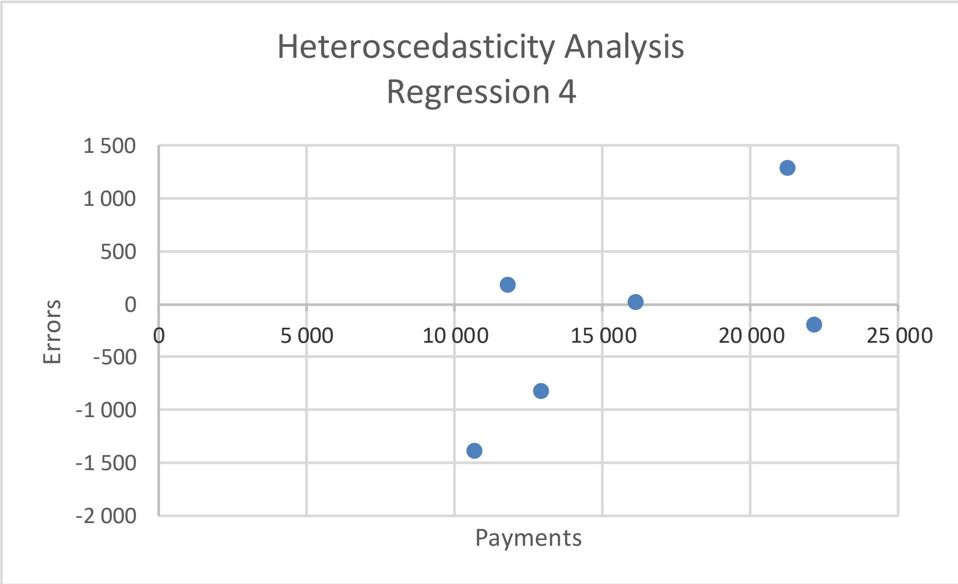


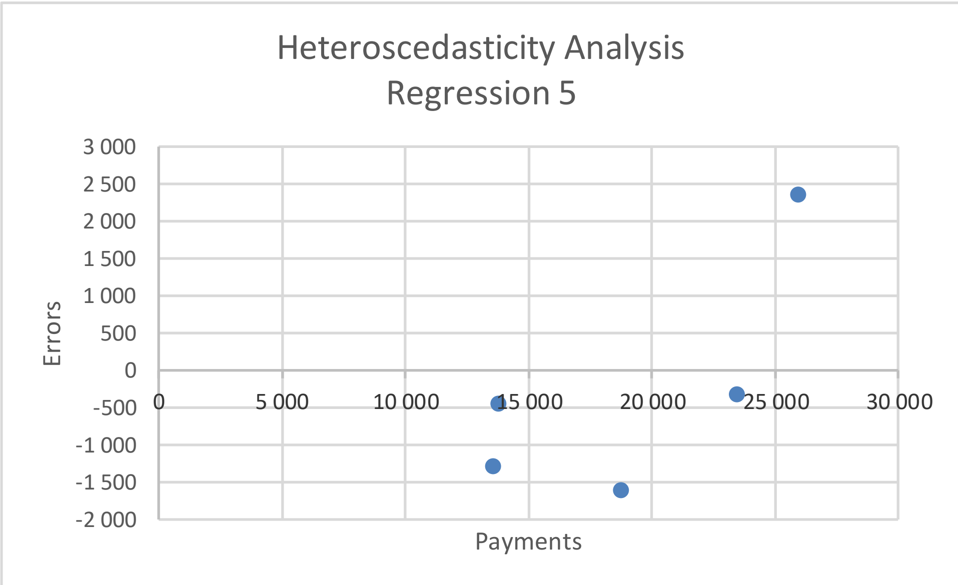


For the remaining regressions $k = 4, \ldots, 9$ the analysis is more difficult to be done, as the number of observations is smaller but the plots for equations 4 and 5 are presented in Figure 6.6.

As some regressions have less observations we did also an analysis with the data from all the regressions in the same plot, Figure 6.7.

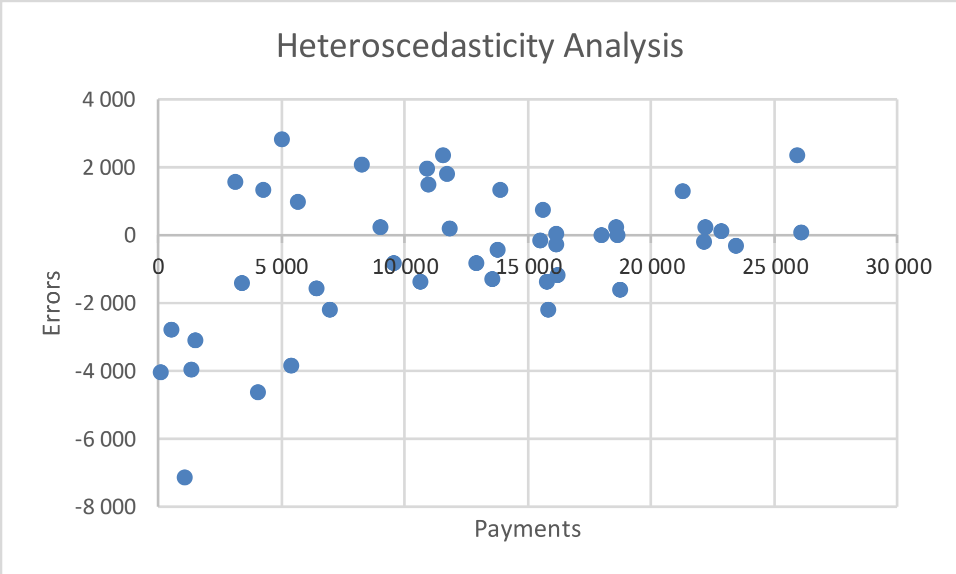
Figure 6.6: Regression Error's Plots or Individual Regressions 4-5





As expected, the previous conclusion of Figure 6.6 is also on the plot of Figure 6.7: the range of the errors is between -4000 and 4000. We also see a different range of values for payments until 5 000.

Figure 6.7: Regression Errors Plot for all the Regressions with Payments



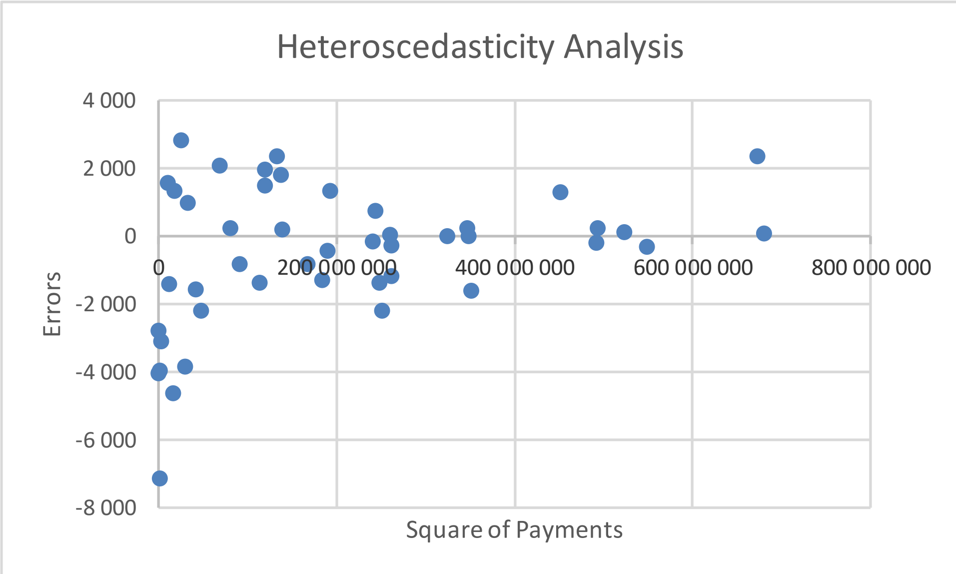
In section 6.5.1 we saw that both the prediction errors from the CL and the homoscedastic VP were high, respectively, 56% and 36%. In section 5.2 we got a prediction error for the heteroscedastic VP of 63%.

As we can see, the homoscedastic VP has a better prediction error than the CL (which is a heteroscedastic method). Also, the homoscedastic VP has a lower prediction error than the heteroscedastic VP. Thus, for this irregular data, when we assume homoscedasticity of the errors, the prediction error is lower. Although, these results are not good enough (36% is still a high prediction error).

However, when we compare the two methods with heteroscedasticity, the CL and the heteroscedastic VP, the CL has a lower prediction error. This happens because some heteroscedasticity was detected on payments up to 5 000 and the level of heteroscedasticity from the CL is higher than the level of heteroscedasticity from the VP.

So far, we assume that the regression errors are comparable with the payments, but other relations may be more appropriate. We show this with another plot where we compare regression errors with the square of the payments.

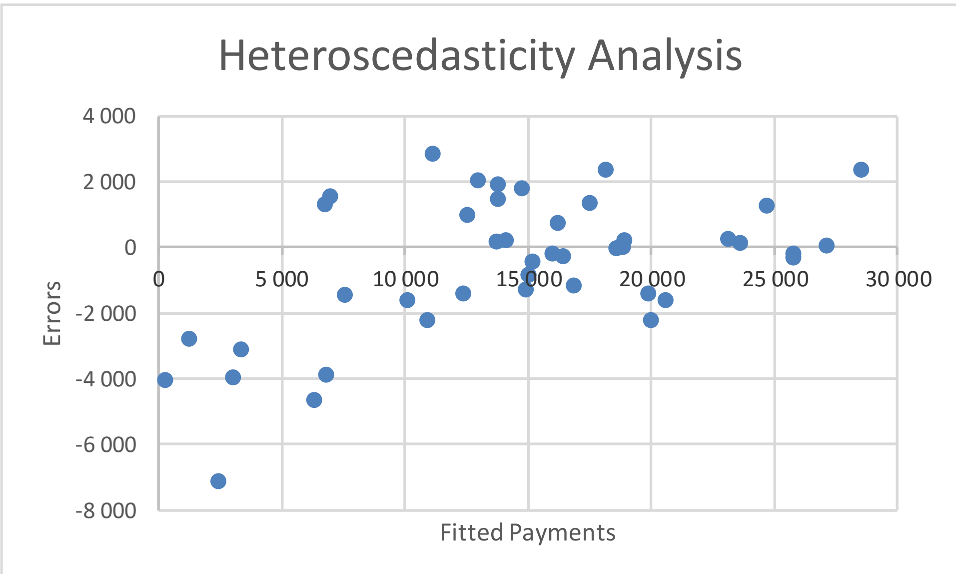
Figure 6.8: Regression Error Plot for all the Regressions with Squared Payments



The change for the square of the payments did not change the conclusions obtained before. Now we may say that, for this triangle, the variance of the errors does not seem to have a strong relation with the payments and the square of the payments.

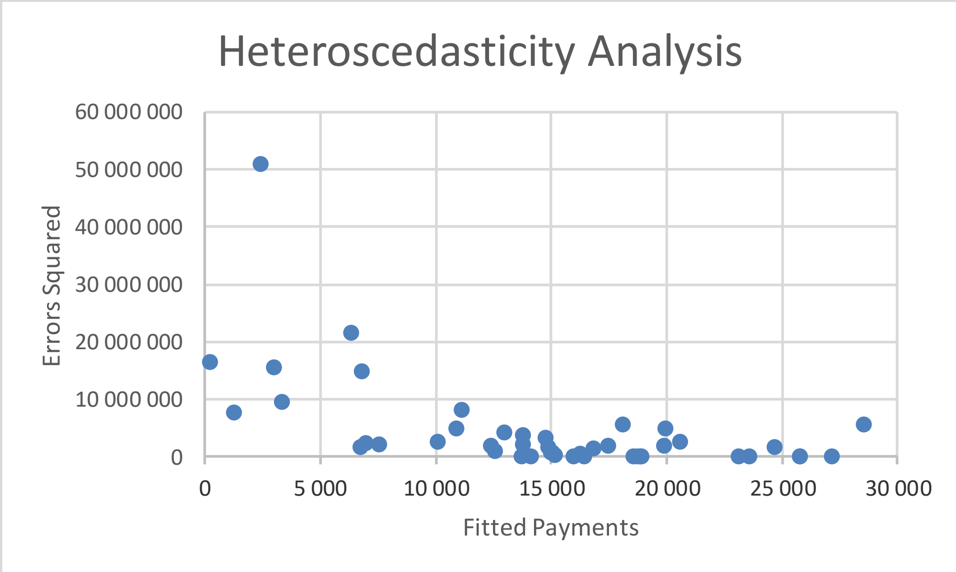
The same analysis was also done for the errors when explained by the fitted payments and we may see that the conclusions are like the ones obtained with the payments.

Figure 6.9: Regression Error Plot for all the Regressions with Fitted Payments



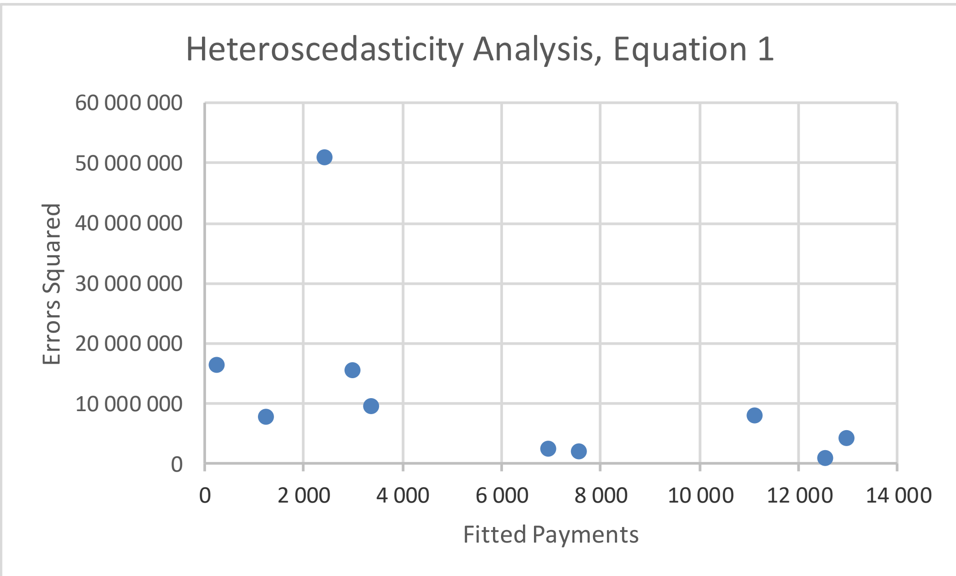
Finally, we analysed the square of the errors when compared with the fitted payments. The results confirm what we saw in previous figures from this section. A constant range of the errors squared in relation to the fitted payments, but with one outlier and four observations above the range.

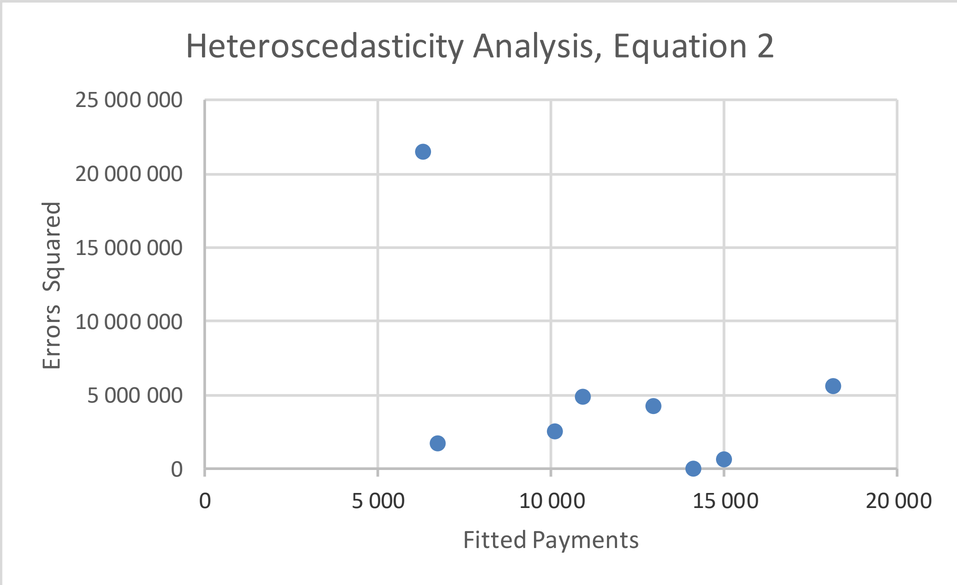
Figure 6.10: Regression Squared Error Plot for all the Regressions with Fitted Payments

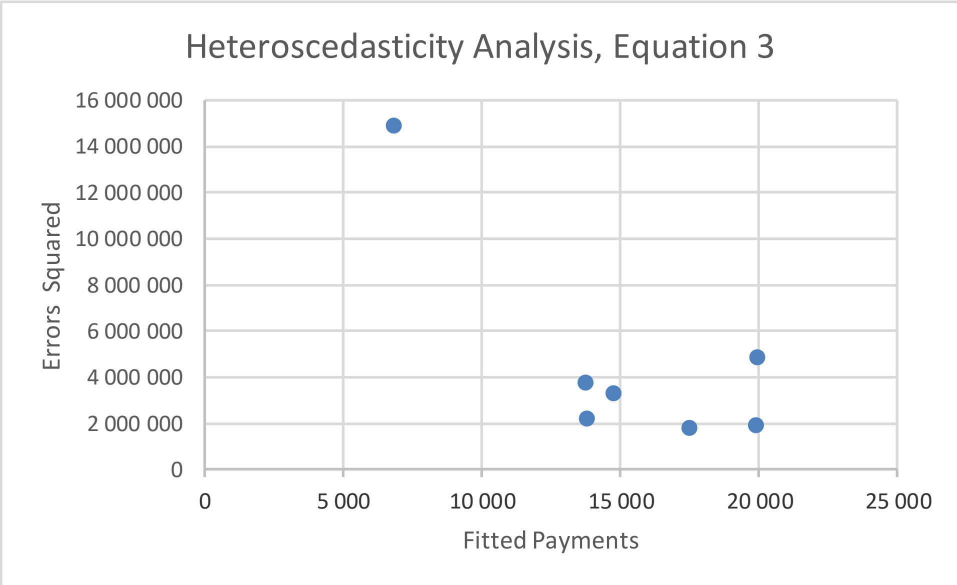


The visual analysis shows some trend to homoscedastic errors with some heteroscedasticity until fitted payments of 7 000. The following Figure 6.11 shows the origin of those deviations with the same plots done for regressions 1, 2 and 3.

Figure 6.11: Squared Error Plot for Regressions 1-3 with Fitted Payments







There is an outlier on regressions 1, 2 and 3, but one outlier does not justify the existence of heteroscedasticity. However, the payments until 5000 from regression 1 show a different range from those after 5000. Visually, it is not possible to conclude if this is evidence of a significant heteroscedasticity.

To completely understand if this fact is enough to classify the regressions as heteroscedastic we will perform, in the following section, the White test. See Hill et al. (2012), for more details on the test.

6.7.1.2 White Test

We start by applying the test for regression 1 to 7. The results are presented in the following table 6.11 and show, for all the regressions, that we should not reject the null hypothesis that the *b* parameter from (6.6.1) is null. This means that we should not reject the hypothesis that the errors are homoscedastic. This conclusion, for this irregular data set, does not reject the homoscedastic VP results and does not confirm the CL heteroscedastic method.

Table 6.11: White 5% Test of Heteroscedasticity for Regression 1 to 7

	Regression						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Chi-Square Statistic	1,9	1,2	3,2	0,1	1,6	2,4	0,1
Critical Value	3,8	3,8	3,8	3,8	3,8	3,8	3,8
Reject Null Hypothesis?	*No*	*No*	*No*	*No*	*No*	*No*	*No*

Then we applied the test for all the regressions, considering all the cells from the upper triangle of the data and calculating the R-squared of all the set of regressions.

Now we have nine regressions, which means nine degrees of freedom.

The critical value is now 16.9. Table 6.12 summarizes the results of the White Test.

Table 6.12: White 5% Test of Heteroscedasticity for all the Regressions

All the Regressions as one Regression	
Chi-Square Statistic	6,9
Critical Value	16,9
Reject Null Hypothesis?	*No*

The results obtained show that we should not reject the null hypothesis that the *b* parameter from (6.6.1) is null in both methods. This means that we should not reject the hypothesis that the errors are homoscedastic. This conclusion, for this irregular data set, does not reject the homoscedastic VP results and does not confirm the CL heteroscedastic method.

6.7.2 Regular Development of Data

In this section, we consider the same regular data used on section 5.2 from Taylor and Ashe (1983).

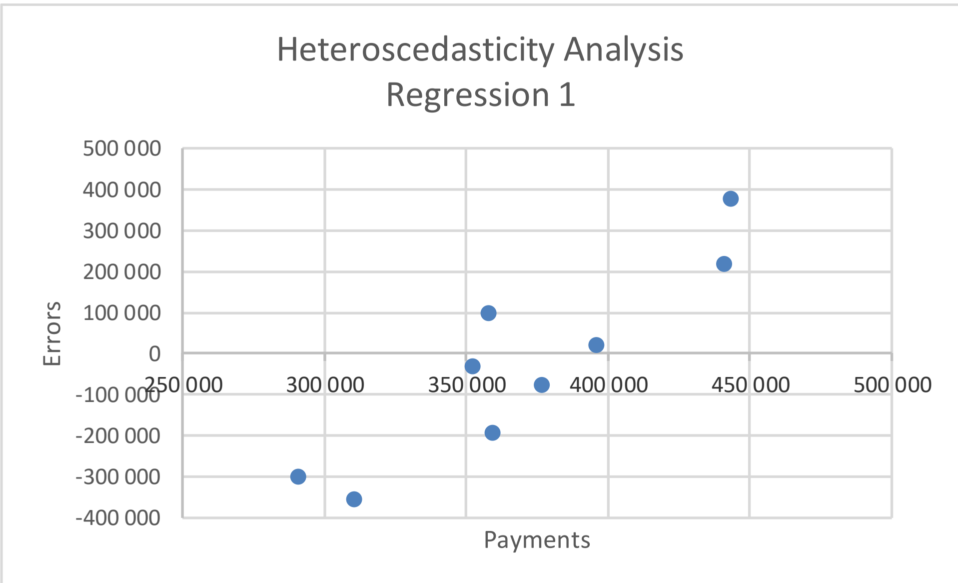
On section 5.2 it was explained why the data is irregular. The analysis is like the one performed in section 6.6.1 but the data is now regular. Regular data was defined in section 5.4.

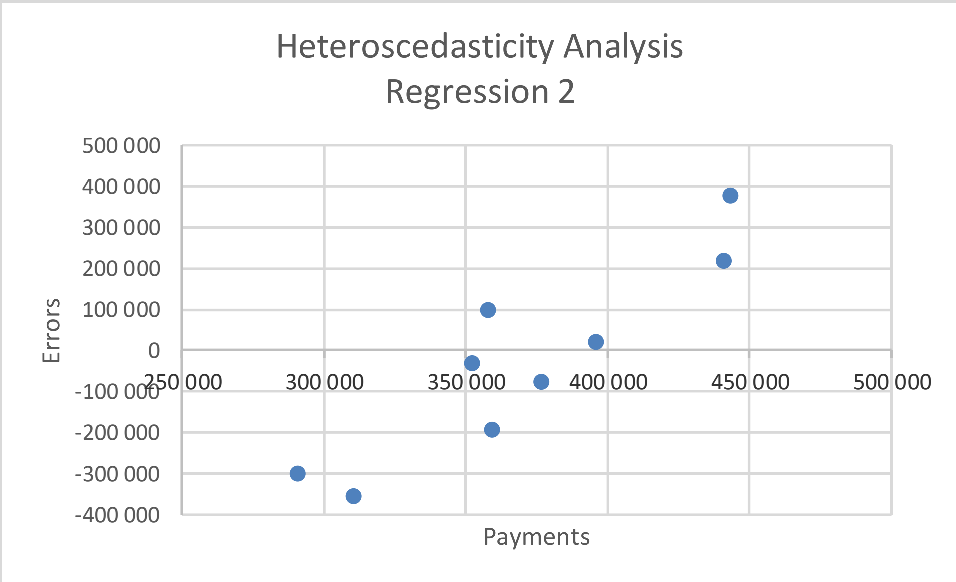
6.7.2.1 Regression Error's Plots

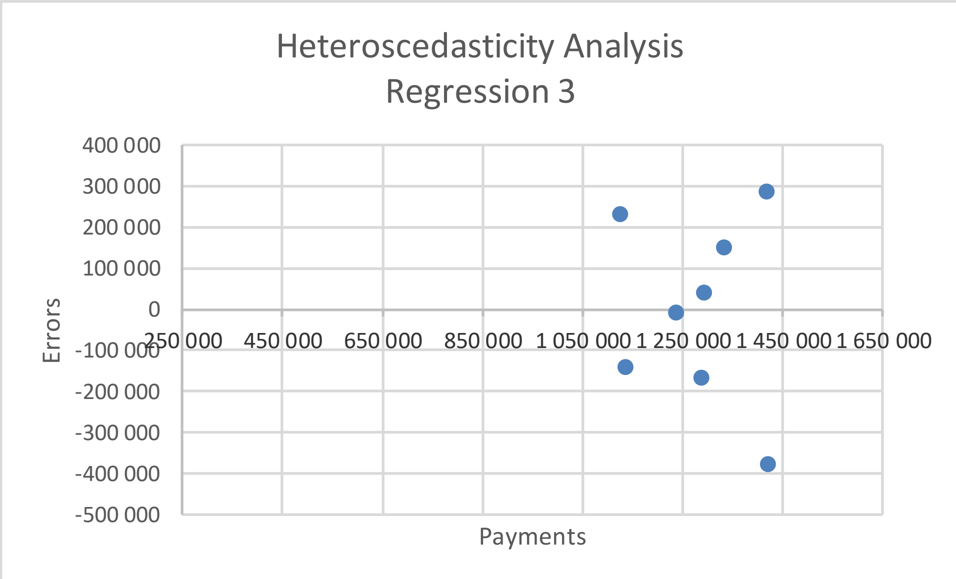
In Figure 6.12 we analyse the first three regressions $k = 1,2,3$. We may see in regressions $k = 1$ and $k = 2$ plots that the payments increase, and the errors are also increasing.

However, the range of the errors is mostly between -400 000 and 400 000 for all three equations and there are no values and outliers outside this range. It is a more regular triangle.

Figure 6.12: Regression Error's Plots for Individual Regressions 1 to 3



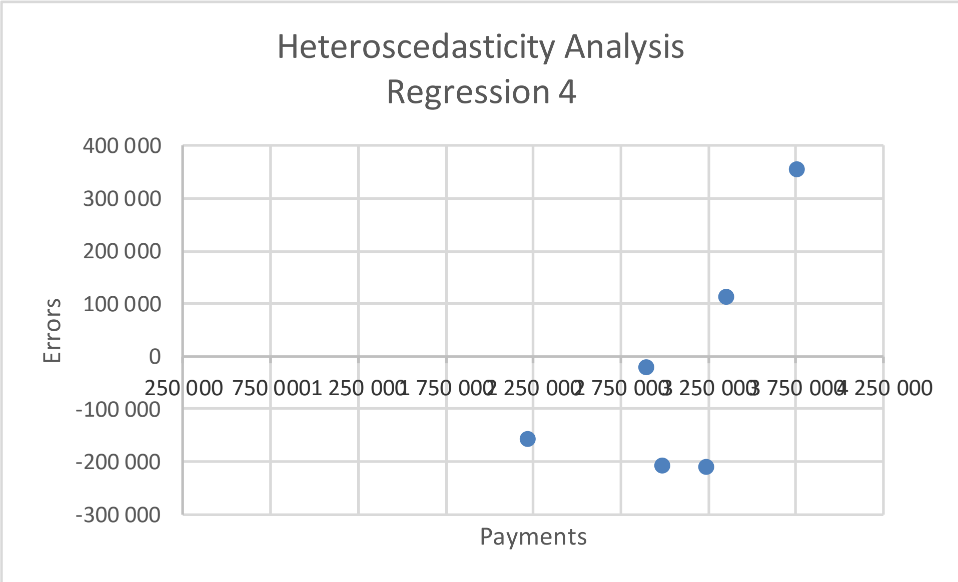


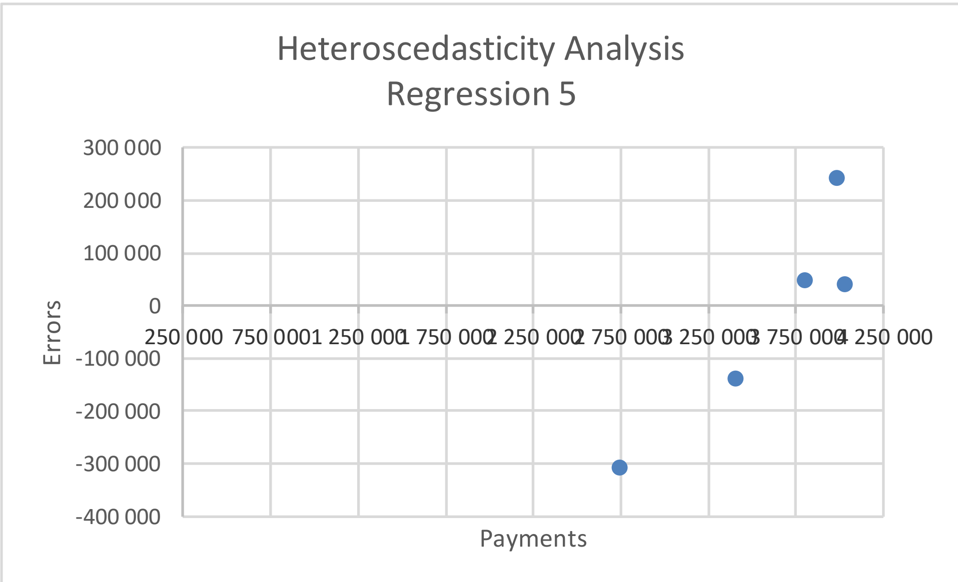


As in section 6.6.1, for the remaining regressions $k = 4, \dots, 9$ the analysis is more difficult to be done, as the number of observations is smaller. Despite this, the plots for equations 4 and 5 are presented on Figure 6.13.

As some regressions have less observations we did also an analysis with the data from all the regressions in the same plot, Figure 6.14.

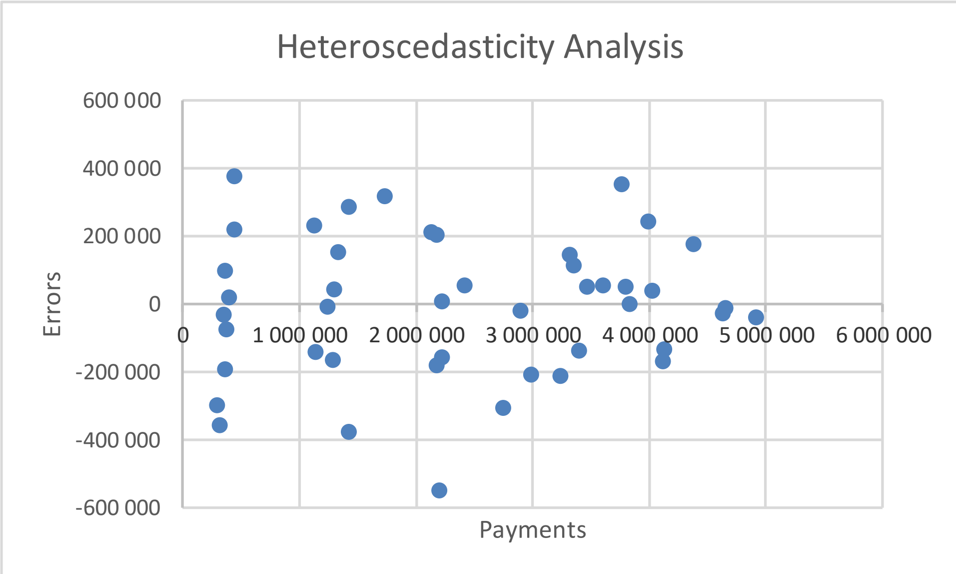
Figure 6.13: Regression Error's Plots or Individual Regressions 4-5





As expected, the previous conclusion of Figure 6.12 is also on the plot of Figure 6.14: the range of the errors is between -400 000 and 400 000. There is just one value outside this range. We also see that the errors are not related with the payments.

Figure 6.14: Regression Errors Plot for all the Regressions with Payments

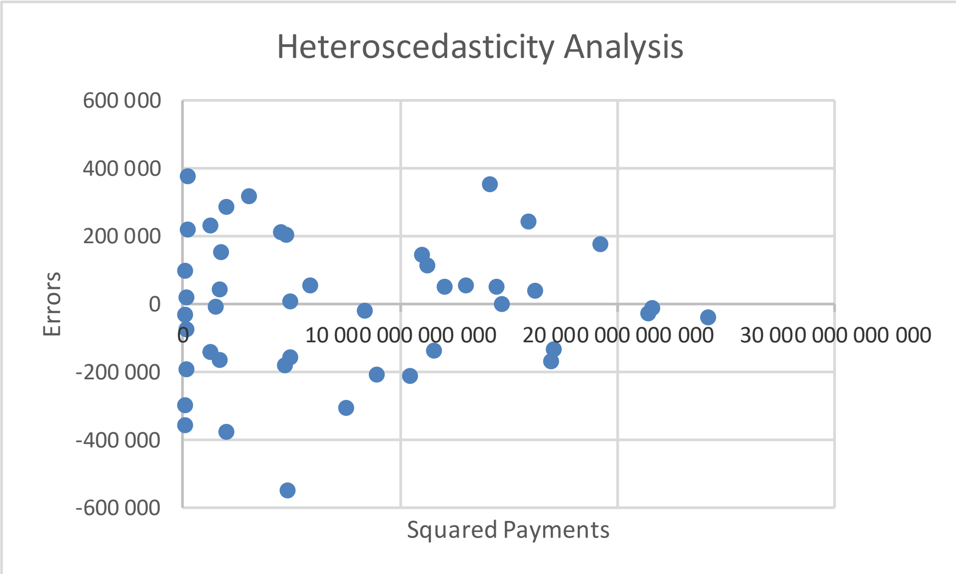


In section 6.5.2, we saw that both the prediction errors from the CL and the homoscedastic VP were low, 15% and 10%, respectively. In section 5.2 we got a prediction error for the heteroscedastic VP of 9%.

As we can see the homoscedastic VP has a better prediction error than the CL (which is a heteroscedastic method). Also, the homoscedastic VP has a similar prediction error to the heteroscedastic VP.

Thus, for this regular data, when we assume homoscedasticity of the errors, the prediction error is lower. So far, we assume that the regression errors are comparable with the payments, but other relations may be more appropriate. As we did for irregular data, we show this with another plot, Figure 6.15, where we compare the regression's errors with the square of the payments.

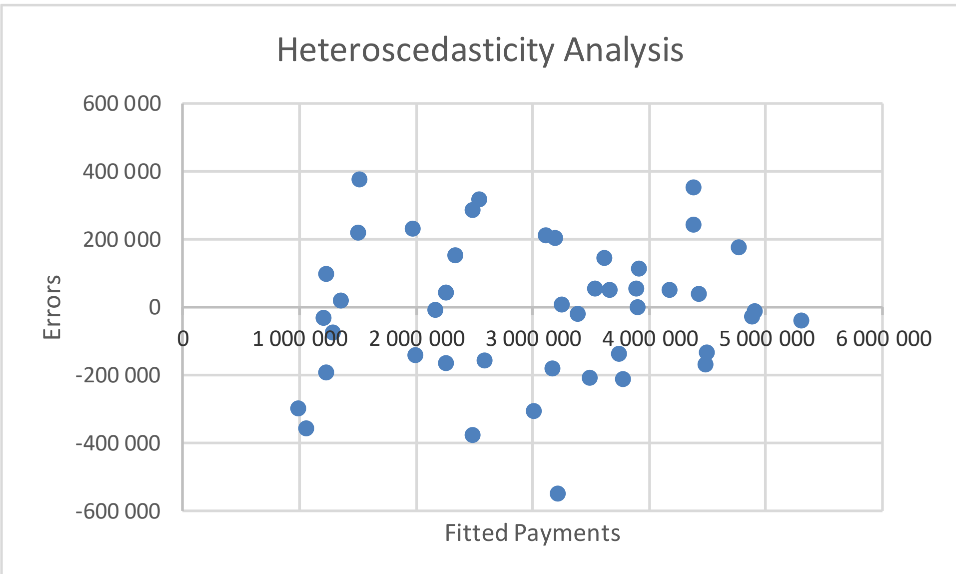
Figure 6.15: Regression Error Plot for all the Regressions with Squared Payments



The change for the square of the payments did not change the conclusions obtained before. Now we may say that, for this triangle, the variance of the errors does not have any relation with the payments and the square of the payments.

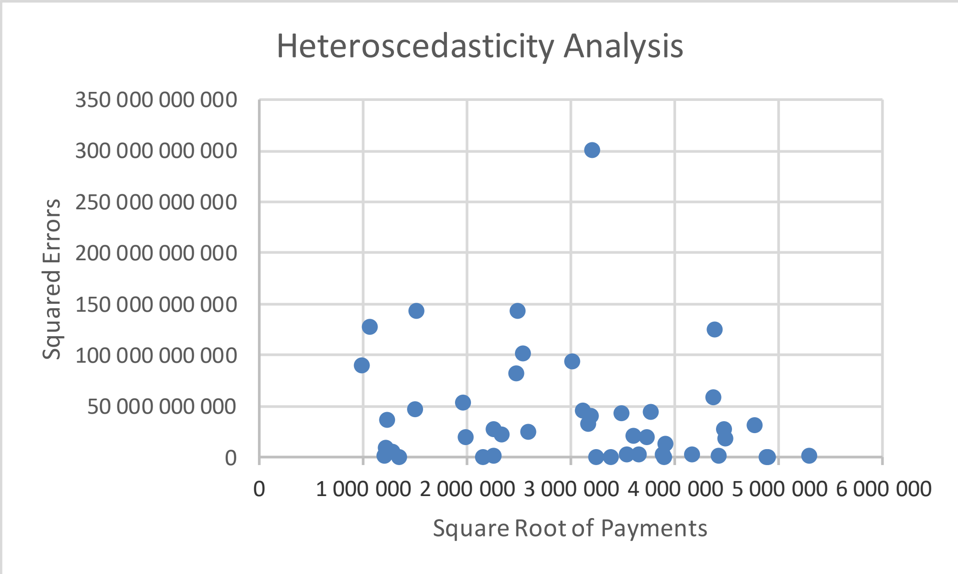
The same analysis was also done for the errors when explained by the fitted payments, Figure 6.16, and we may see that the conclusions are like the ones obtained with the payments, see Figure 6.13. The same already happened before with the irregular data, on section 6.6.1.

Figure 6.16: Regression Error Plot for all the Regressions with Fitted Payments



Finally, we analysed the square of the errors when compared with the fitted payments. The results confirm what we saw in previous figures from this section. A constant range of the errors squared in relation to the fitted payments. There is just one outlier.

Figure 6.17: Regression Squared Error Plot for all the Regressions with Fitted Payments



We may conclude from all these figures that this triangle or regular data shows homoscedastic errors. This means that we should promote the method with such a feature, namely the homoscedastic VP.

6.7.2.2 White Test

As before, we start by applying the test for regression 1 to 7. The results are presented in the following table 6.13 and show, for all the regressions, that we should not reject the null hypothesis that the *b* parameter from (6.6.1) is null. This means that we should not reject the hypothesis that the errors are homoscedastic.

This conclusion, for this regular data set, does not reject the VP homoscedastic results and does not confirm the CL heteroscedastic model.

Table 6.133: White 5% Test of Heteroscedasticity for Regression 1 to 7

	Regression						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Chi-Square Statistic	1,9	1,2	3,2	0,1	1,6	2,4	0,1
Critical Value	3,8	3,8	3,8	3,8	3,8	3,8	3,8
Reject Null Hypothesis?	*No*	*No*	*No*	*No*	*No*	*No*	*No*

Then we applied the White test for all the regressions, considering all the cells from the upper triangle of the data and calculating the R-squared of all the set of regressions. Table 6.14 summarizes the results of the White Test.

Table 6.14: White 5% Test of Heteroscedasticity for all the Regressions

All the Regressions as one Regression	
Chi-Square Statistic	12,1
Critical Value	16,9
Reject Null Hypothesis?	*No*

The results obtained for both methods show, as expected from the plots analysis on section 6.7.2.1, that we should not reject the null hypothesis that the b parameter from (6.6.1) is null. This means that we should not reject the hypothesis that the errors are homoscedastic. This conclusion, for this regular data set, does not reject the homoscedastic VP results and does not confirm the CL heteroscedastic method.

6.8 Test of Correlations between Equations

In this section, we test for the existence of correlations between the equations. That may be done considering as null hypothesis $\sum$ as a diagonal matrix against the alternative hypothesis of one or more off-diagonal elements of $\sum$ are non-zero.

Two tests may be used for this, the LR test (the log-likelihood ratio test) and the Breusch and Pagan test with the LM statistic, see Pesaran (2015). The tests are recommended, see Pesaran (2015), for situations where the number of equations is not large, and the number of observations is large. The tests are asymptotically equivalent.

We concentrated on the Breusch and Pagan test, as it is easier to apply in practice.

The LM statistic is given by

$$LM = m \sum_{j'=2}^{T-1} \sum_{j=1}^{j'-1} \rho_{jj'}{}^2 \tag{6.8.1}$$

Where $\rho_{jj'}$ is the pair-wise correlation coefficient of the OLS errors from regression equations $j = 1, \dots T-1$ and $j' = 1, \dots T-1$. The covariance and the variance of the errors are obtained using (6.4.10).

We will test as null hypothesis $\sum$, as a diagonal matrix, against the alternative hypothesis of one or more off-diagonal elements of $\sum$ are non-zero. If we reject the null hypothesis we will have an indication of correlations between equations.

We reject the null hypothesis if the statistic χ^2 is higher than the 5% critical value ${\chi^2}_{1-0.05,df}$, where df means the degrees of freedom. The degrees of freedom are given by $(T-1)(T-2)/2 = 36$. From the Chi-Square distribution table and for 36 degrees of freedom we get ${\chi^2}_{0.95,36} = 51$. More details on the statistic may be seen in Pesaran (2015).

The results obtained are presented in Table 6.15 and confirm that we should reject the null hypothesis that $\sum$ is a diagonal matrix. This means that we have an indication, in the two triangles tested (with irregular and regular data), that there are correlations between equations inside each of these triangles. That is, we have an indication supporting the MGLR (6.4.1) assumption.

Table 6.15: White 5% Test of Correlations between Regressions

	Irregular Data	**Regular Data**
LM Statistic	262,7	278,2
Critical Value	51,0	51,0
Reject Null Hypothesis	*Yes*	*Yes*

6.9 Testing Serial Correlation of the Errors

Serial correlations are a feature from time-series and the regressions we are considering are based on time-series (by accident year). It arises when we have omitted variables from the method or when the dependent variable is related with the explanatory variable measured on other periods. It may happen, when modelling payments, that we did not consider all the relevant explanatory variables, for instance the change in legislation or the new policy of speeding up the claims payments.

It is not expected, for the same development year, that the errors in one origin year are related with errors in the previous origin year and most of the methods, see for example Mack (1993a, 1993b, 1994), consider that these origin years are not correlated. A reason for this is that what an insurer pays in one year is related with what the same insurer paid in other development years (the correlation studied in the multivariate method and tested on section 6.8). However, it is not expected that the payments in one origin year will influence the payments in the following origin year for the same development year.

As it was assumed in the methods defined in sections 6.3 and 6.4, see (6.6) and (6.7), that the errors inside each equation are not correlated, we decide to apply the Durbin-Watson statistic to the regressions with more than 5 observations and also to the case where all the regressions are estimated as one regression, see Fomby et al. (1984) for more details.

As the method defined in section 6.3.1 does not have an intercept, the Fairbrother table was applied, together with the Savin-White tables due to the existence of regressions with less than 15 observations.

All the tests done to a 5% significance were inconclusive to the existence or not of first order serial correlation. Indeed, this is one of the problems from the Durbin-Watson test: there is a range of the test values where it's not possible to accept or reject the assumption of the errors serial correlation.

Due to this we did a regression, per equation j, of the errors obtained as a function of the errors from the previous origin year, a first order correlation, with parameter ρ and with an error $v_{i,j}$ following the OLS properties.

$$\varepsilon_{i,j} = \rho_j \varepsilon_{i-1,j} + v_{i,j} \quad i = 2, \dots, T-j+1 \text{ and } j = 2, \dots T \tag{6.4.17}$$

The conclusion for all the regressions, for the regular and irregular data, was that the parameter ρ_j was not statistically significant for a 5% T-test.

It's difficult to imagine a reason to have higher order correlations (the payment of one origin year to depend on the payments of two or three origin years ago) and due to that those cases were not tested.

The test of correlation is important. If there is serial correlation of the errors, of first order, the $\widehat{Y_F} \neq X_F\hat{\beta}$ because $\mathbb{E}(\varepsilon|X) = \mathbb{E}(\varepsilon) \neq 0$. The true relation will depend on the structure of the serial correlation. If we consider, as structure of the serial correlation, the one on (6.4.17), we will have,

$$\mathbb{E}(\varepsilon) = \rho\, \varepsilon^*$$

where ε^* is the ε vector with a lag of one origin year. This means that if the correlation coefficient of the errors, $\rho,$ is high as well as the errors, the estimated payments may be significantly different from the ones forecasted by $\widehat{Y_F} = X_F\hat{\beta}$.

Clearly, the methods developed and analysed, GLR and MGLR, assumes that there is no serial correlation of the errors.

6.10 Conclusions

In the insurance industry, the analysis of outstanding claims reserves plays a core role in the assessment of financial strength and solvency of a company. Failure to do so might result in the insolvency or lack of competitiveness of some insurers, see chapters 1 and 2. Furthermore, the flexibility to control the prediction errors level is very much appreciated.

The framework presented is very flexible and allows us to have several methods with the same general formulas, including the one for the prediction error. We may generate new methods and replicate other known methods. Also, we may have the prediction error method with the lowest prediction error for a set of LRT stochastic methods.

In this chapter, the main conclusions can be summarized as follows.

First, heteroscedasticity is an important feature of insurers' triangles to be analysed and may arise in some circumstances. There is a clear connection between its level, and the method selected from the LRT methods, when the prediction error is minimised.

Second, the GLR method replicates the loss development factors, the variance parameter, and the prediction error of the Mack (1993a, 1993b, 1994)'s method, when the structure of the future heteroscedasticity is assumed to be equal to the current one. Otherwise, the method shows that the prediction error increases significantly. This means that the traditional stochastic CL may have even higher prediction errors. Moreover, a generalized approach to the link ratios using the heteroscedasticity level as parameter optimized allows us to have a solution that is not necessarily the one from the traditional LRT methods presented on chapter 3.1. Other weights for the link ratios may arise that are not necessarily the ones from the VP, CL or SA methods. The criterion used in our method was the minimization of the prediction error in a univariate and a multivariate framework. The results obtained show a lower prediction error when homoscedastic errors or a low level of heteroscedasticity are considered. The formula yields for the prediction error is computationally easy, it can be applied using the Excel Solver, and allows the implementation of the GLR method when the lowest prediction error is required.

Third, implementing correlation of the errors between the triangle equations, a MGLR method, reduces even further the prediction error. It also brings the possibility of considering several methods with heteroscedastic errors, including the structure of well-known methods from the LRT techniques. Using the same examples as in Portugal et al. (2017), we confirm that lower prediction errors are obtained when the VP is used instead of the typical CL. Additionally and as in the univariate methods, it is also shown that between the latter and the former, there are several other methods obtained if other values of the heteroscedasticity parameter are considered.

Fourth, the survey conducted with 114 triangles seems to show, as more frequent, the inexistence of heteroscedasticity in the GLR. Some cases arise with a low level of heteroscedasticity, and the possibility to have the CL and SA methods as the best ones is very low. When we use the MGLR the conclusions are similar, but we notice a small increase of the heteroscedasticity level.

Fifth, the heteroscedasticity is a feature assumed from claims reserving methods like the CL and the SA but most of the triangles show indications of homoscedasticity. The heteroscedasticity may be easily tested using the White test. If the test does not reject the null hypothesis we have an indication to select methods with homoscedastic errors, like the VP and to be careful in selecting methods with heteroscedastic errors, like the CL and the SA.

The existence or not of heteroscedasticity may, in some cases, be detected just with a plot of the errors vs the payments (or the square of the payments).

Sixth, the correlation between equations, inside each triangle, may be easily tested. The triangles used in this chapter show statistical evidence of correlation between the equations of each triangle. This is an important indication towards the use of multivariate techniques.

Finally, the LRT methods seem to rely on the determination of an appropriate level of heteroscedasticity in the methods (permitting also the possibility of homoscedastic errors). However, it is also shown that the consideration of a multivariate method with contemporaneous correlations in the errors may improve the results of the method, while it also defines the correct level of heteroscedasticity. Regarding the multivariate case, the MGLR method, the method changes the level of heteroscedasticity obtained in the GLR method. This appears to be very interesting, since it may be an indication that in the univariate context, the level of heteroscedasticity may be hidden because dependencies are not considered in the method.

7. Stochastic Portfolio Claims Reserving

The framework presented in the last chapter may be extended to the situation where we have more than one triangle to be estimated at the same time. In the literature, as we saw, for example in Merz and Wüthrich (2007), this is called a multivariate approach. However, this just happens due to the consideration of correlations between the triangles. Indeed, in such cases that we see in the literature, there is the absence of correlations between the regressions inside each triangle. Since we are considering, in one of the methods of this chapter, correlations between each triangle regressions and simultaneously correlations between the triangles, it is more appropriate to call these methods something different than multivariate.

A first possibility is to call it multivariate claims reserving with panel data, sometimes also called longitudinal data, see for example Frees (2010). Panel data, see Fomby et al. (1984) or Hill et al. (2012), consists of N cross-sectional units (for example people, countries, firms, lines of business) who are observed over T years when those cross-sectional units are the same in all the points in time. There are several types of panel data, see for example Hill et al. (2012):

- "Long and narrow", where $T > N$.
- "Wide and short", where N is much larger than T.
- "Long and wide", where both T and N are large.

Insurer's triangles, when considered correlated, are the same over each period of the T years considered. It is not common to have a large T, for example 20 years, although is not impossible with lines of business like credit insurance, general liability and inward reinsurance (see section 2.1). Also, it is not common to have N much larger than T. Putting all this together, insurer's triangles seem to be more of the type "Short and Narrow", with a short T and a low number of N triangles, which is not matching the types presented by Hill et al. (2012).

Due to this, a second possibility was considered and we decided to call it portfolio claims reserving (univariate or multivariate). Indeed, the methods to be presented in the following sections are to be applied to a portfolio of claim's triangles (without or with correlation effects).

Working with a portfolio of triangles shows the importance of having data split, by lines of business and covers, to get homogeneous triangles. Just in particular cases may be indifferent to have either aggregated data or data split in several triangles. The following Lemma 7.0 shows that.

Lemma 7.0 *The reserves estimated when using an aggregated triangle are equal to the ones obtained when considering that triangle split in two or more triangles if the individual triangles ultimate factors (see 2.3) are equal to the one from the aggregated data triangle. If that is not the case, different levels of reserves will arise - a similar Lemma with incremental data may be seen at Radtke et al. (2010).*

Proof. We know that to have equivalence between the triangles aggregate reserves and the sum of the individual triangles reserves we need to have (the bold identifies the aggregated triangle variables and parameters and the $\hat{f}_{j,t}$ are the ultimate factors from the triangle $t = 1, \dots, N$).

$$\hat{\boldsymbol{C}}_{i,T}.\hat{\boldsymbol{f}}_j = \hat{C}_{i,t-i+1,1}.\hat{f}_{j,1} + \dots + \hat{C}_{i,t-i+1,N}.\hat{f}_{j,N}$$

This means that

$$\hat{\boldsymbol{C}}_{i,T} = \hat{C}_{i,t-i+1,1}.\frac{\hat{f}_{j,1}}{\hat{\boldsymbol{f}}_j} + \dots + \hat{C}_{i,t-i+1,N}.\frac{\hat{f}_{j,N}}{\hat{\boldsymbol{f}}_j}.$$

If the ratios of the individual triangles ultimate factors to the aggregated triangle ultimate factor are equal to one, we get the requested equivalence. In that case we may say that the speed of arrival to the ultimate cost is the same in all the triangles, the individual ones and the aggregate one. This means that all the triangles behave in the same way and we do not lose anything in aggregating the triangles. However, this situation of having the same ratios above is probably rare, even if we considered those ratios as similar.

Finally, having a portfolio means that we may have correlations between the portfolio triangles. For example:

- The increase of payments on some triangles may be a consequence of measures taken for all the lines of business, which may originate some anticipation of payments, in the same direction, in several triangles.

- Increase on claim's payments in one triangle may have a correlation with different triangles, if the driver of the payment is the same, e.g., courts decisions or inflation (even if with a different impact on each triangle).
- Some triangles may show payments adverse movements when others show the opposite (or even not show anything).

Due to all these reasons, aggregating triangles may not be a good procedure. Also, not considering the correlation between them implies losing information about claims patterns. This means, that it may be important to have a portfolio analysis, with the same triangles estimated at the same time. The next sections develop two different methods for that.

Baltagi (2013) identifies six good reasons and four problems in using panel data (our portfolio data). The good reasons identified are the following:

- Controlling for heterogeneity. Claims features, see section 2.1, suggests that triangles are heterogeneous. This may happen, for example, if we do not split motor insurance data between bodily injury claims and material damage claims.
- We will have more information, more variability, less collinearity among variables, more degrees of freedom and more efficiency. Multicollinearity is not an issue with triangle's regressions when we just have one explanatory variable (as in the regression methods of chapters 6 and 7), but the degrees of freedom is an important aspect of claims reserving. Some of the columns of the triangles have a low number of observations. With more data, from several triangles, we have more information, for example about certain development years where the link ratios are not so much stable.
- It is easier to understand the dynamics of the variables under study. With more than one triangle we may understand if the claims paid are related or not with claims paid in other triangles.
- We may identify what is not clear on pure time-series data or pure cross-section data. Indeed, correlations between several triangles will allow understanding the true overall reserves and not just the reserves from one triangle (the sum of the reserves estimated from several triangles is not, necessarily, the true overall reserve).
- It is possible to construct and test more complex behaviour. For example, material damages claims evolution when bodily injured claims payments increase.

- Bias coming from aggregated data is eliminated. We saw already the disadvantages of aggregated data with Lemma 7.0.

But there also some problems with this portfolio approach. Baltagi (2013) identifies the following problems, but we will see that most of them do not exist in insurer's triangles:

- Design and data collection problems. Indeed, to apply portfolio analysis we need to have triangles for all the lines of business and with the same number of years of origin and development. This usually happens with insurers.
- Distortions and measurement errors. This is a disadvantage from panel data as they rely on the information of several individuals surveyed. However, in insurance, data is not collected by survey and insurers are used to have data in triangle format. Also, the European Union (2015) oblige insurers to have quality on the data.
- Selectivity problems arise in panel data due to the absence of response in some surveys. This problem does not exist with insurers, as all the information is registered in the information system. Even if the RBNR's are not there, see chapter 2, they will be estimated inside IBNR and IBNER estimation. Eventually, we may have some triangles without payments, for example, in the first development year. These triangles may require a different approach from the other triangles (to avoid the zeros on the triangle first column we may aggregate the first two development years) and we may avoid this problem.
- Short time-series dimension. Indeed, the same may happen with insurers if the triangles just have a few years of information, in which case the T is much lower than the N. In such a case, the first year of origin may not be closed, which brings other type of problems, with the need to have a tail factor (see section 2.4).

There are several types of panel data models, see for example (Frees 2010):

- Fixed effects model, which consider the slope of the regression as the same in all the subjects of the regression (the subjects would correspond to the triangles). These models consider an extra parameter to attend to the differences over the subjects: a variable constant (one-way fixed effects model and two-way fixed effects model) or a second variable constant, to attend differences over time (two-way fixed effects model). It is difficult (or impossible) to justify the same slope in all the insurer's triangles for the same development year.

- Variable coefficients model, which allows the slope of the regression to vary over the subjects (the triangles in our case), a typical situation from the insurer's triangles.
- Serial correlation models, which consider correlations over time between the errors. We saw already in section 6.9 that this is not an issue in insurer's triangles (or seem not to be an issue).
- And random effects models, where the subjects of the regression (the triangles), are randomly selected from a population. Insurer´s triangles are not selected randomly.

Clearly the panel data model of interest to the extension of the GLR and MGLR methods is the variable coefficients model. This is the one we will develop here.

Firstly, in section 7.1 we develop the Portfolio Generalized Link Ratios (PGLR), which uses the GLR presented on chapter 6 in the context of a portfolio of triangles. Secondly, in section 7.2 we develop the Portfolio Multivariate Generalized Link Ratios (PMGLR), which uses the MGLR presented on chapter 6, also in the context of a portfolio of triangles.
Some numerical examples are presented in section 7.3 and some tests are done in section 7.4. The final section 7.5, presents the conclusions that we got from this chapter.

7.1 Portfolio Generalized Link Ratios

We have now data from $t = 1, \ldots, N$ triangles. On each of these triangles we will have $k = 1, \ldots, T-1$ equations (regressions) and on each of these equations we have $T-k$ observations (years of origin).

The estimations for the *t* triangles will be provided simultaneously with the $y_{t,i,j}$ explained by the adjacent triangle column $x_{t,i,j-1}$. This means that the claim's payments on column j from triangle t, $y_{t,i,j}$ are a function (a regression through the origin) of the claim's payments on column $j-1$ from triangle t, $x_{t,i,j-1}$. Both variables represent the cumulative payments, but $y_{t,i,j}$ is a random variable and $x_{t,i,j-1}$ is a non-random variable. We follow the same notation from chapter 6.

The following model is based on GLR method presented in section 6.3 but considering $t = 1, \ldots, N$ triangles at the same time with k equations on each triangle.

We define $\beta_{t,j}$ as the slope (loss development factor) of the j equation from the triangle t. Also, each $\varepsilon_{t,i,j}$ is the error from year of origin i, development year j and triangle t. For, $t = 1, \ldots, N$, $i = 1, \ldots, T-j+1$ and $j = 2, \ldots T$, the cumulative payments dependent variable $y_{t,i,j}$ is given by

$$y_{t,i,j} = \beta_{t,j} x_{t,i,j-1} + \varepsilon_{t,i,j} \tag{7.1.1}$$

7.1.1 Portfolio Univariate Model

In a matrix format and considering all the triangles $t = 1, \ldots, N$ and all the $T-1$ equations implicit in each triangle of cumulative payments, the model is similar from the one in equation (6.2). The difference is the size of each of its components, as now we have N triangles. To differentiate the two models, here we consider the same notation as before but now in bold.

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \tag{7.1.2}$$

$\boldsymbol{Y}$ is the block-vector that includes the previous Y from the GLR method of section 6.3.1 but now for N triangles. $\boldsymbol{Y}$ as dimension $(m \times N) \times 1$. The Y from the GLR method is now $\boldsymbol{Y}_t$ and $t = 1, \ldots, N$

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \ldots \\ \boldsymbol{Y}_N \end{bmatrix},$$

The $\boldsymbol{Y}_t = \begin{bmatrix} Y_{t,1} \\ \ldots \\ Y_{t,T-1} \end{bmatrix}$ represents the dependent variables of the set of $T-1$ equations (see 7.1.1) on the triangle t, where the generic equation $Y_{t,k} = \begin{bmatrix} y_{t,1,k+1} \\ \ldots \\ y_{t,T-k,k+1} \end{bmatrix}$ includes the random variables $y_{t,i,k+1}$ for $t = 1, \ldots, N$, $k = 1, \ldots, T-1$ and $i = 1, \ldots, T-k$.

$\boldsymbol{X}$ is defined by a diagonal block matrix which includes the previous X from the GLR method of section 6.3.1 but now for N triangles. $\boldsymbol{X}$ has dimension $(m \times N) \times [N \times (T-1)]$, and it can be represented by

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{X}_N \end{bmatrix}$$

$\boldsymbol{X}_t = \begin{bmatrix} X_{t,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_{t,T-1} \end{bmatrix}$, where each generic element $X_{t,k} = \begin{bmatrix} x_{t,1,k} \\ \cdots \\ x_{t,T-k,k} \end{bmatrix}$ belongs to equation k and includes the non-random variables $x_{t,i,k}$ for $t = 1, \dots, N$, $k = 1, \dots, T-1$ and $i = 1, \dots, T-k$.

$\boldsymbol{\beta}$ is defined by a block-vector that includes the previous β from the GLR method of section 6.3.1 but now for N triangles. $\boldsymbol{\beta}$ has dimension $[N \times (T-1)] \times 1$, and it can be represented by

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \cdots \\ \boldsymbol{\beta}_N \end{bmatrix}$$

$\boldsymbol{\beta}_t = \begin{bmatrix} \beta_{t,1} \\ \cdots \\ \beta_{t,T-1} \end{bmatrix}$ where the generic $\beta_{t,k}$ is the non-random parameter that represents the slope (loss development factor) from triangle $t = 1, \dots, N$ and equation $k = 1, \dots, T-1$.

$\boldsymbol{\varepsilon}$ is the block-vector that includes the previous ε from the GLR method of section 6.3.1, but now for N triangles. $\boldsymbol{\varepsilon}$ as dimension $(m \times N) \times 1$. The ε from the GLR method is now $\boldsymbol{\varepsilon}_t$ and $t = 1, \dots, N$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \cdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix},$$

The $\boldsymbol{\varepsilon}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_{t,1} \\ \cdots \\ \boldsymbol{\varepsilon}_{t,T-1} \end{bmatrix}$ where the generic $\boldsymbol{\varepsilon}_{t,k} = \begin{bmatrix} \varepsilon_{t,1,k+1} \\ \cdots \\ \varepsilon_{t,T-k,k+1} \end{bmatrix}$ includes the random variables $\boldsymbol{\varepsilon}_{t,i,k+1}$ for $t = 1, \dots, N$, $k = 1, \dots, T-1$ and $i = 1, \dots, T-k$.

We define the true unknown future observations of the dependent variables as $\boldsymbol{Y_F} = \boldsymbol{X_F}\boldsymbol{\beta} + \boldsymbol{\varepsilon_F}$, where $\boldsymbol{X_F}$ and $\boldsymbol{\varepsilon_F}$ are, respectively, the future values of $\boldsymbol{X}$ and the future errors.

$\boldsymbol{Y_F}$ is a block-vector with size $(N \times m) \times 1$ given by

$$\boldsymbol{Y_F} = \begin{bmatrix} \boldsymbol{Y^F}_1 \\ \ldots \\ \boldsymbol{Y^F}_N \end{bmatrix},$$

with each element $\boldsymbol{Y^F}_t = \begin{bmatrix} \boldsymbol{Y^F}_{t,1} \\ \ldots \\ \boldsymbol{Y^F}_{t,T-1} \end{bmatrix}$ and the generic $\boldsymbol{Y^F}_{t,k} = \begin{bmatrix} y^F{}_{t,T-k+1,k+1} \\ \ldots \\ y^F{}_{t,T,k+1} \end{bmatrix}$ for $t = 1, \ldots, N$ and $k = 1, \ldots, T-1$.

$\boldsymbol{X_F}$ is given by the current diagonal of payments from each triangle and by the estimated payments of the lower triangle from each triangle. It is a block matrix given by

$$\boldsymbol{X_F} = \begin{bmatrix} \boldsymbol{X^F}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{X^F}_N \end{bmatrix}$$

where each element $\boldsymbol{X^F}_k = \begin{bmatrix} X^F{}_{t,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X^F{}_{t,T-1} \end{bmatrix}$ and $X^F{}_{t,k} = \begin{bmatrix} x_{t,T-k+1,k} \\ \ldots \\ x_{t,T,k} \end{bmatrix}$ for $t = 1, \ldots, N$ and $k = 1, \ldots, T-1$.

$\boldsymbol{\varepsilon_F}$ is a block-vector with size $(N \times m) \times 1$ given by

$$\boldsymbol{\varepsilon_F} = \begin{bmatrix} \boldsymbol{\varepsilon_F}_1 \\ \ldots \\ \boldsymbol{\varepsilon_F}_N \end{bmatrix},$$

with each element $\boldsymbol{\varepsilon_F}_t = \begin{bmatrix} \boldsymbol{\varepsilon^F}_{t,1} \\ \ldots \\ \boldsymbol{\varepsilon^F}_{t,T-1} \end{bmatrix}$ and the generic $\boldsymbol{\varepsilon^F}_{t,k} = \begin{bmatrix} \varepsilon^F{}_{t,T-k+1,k+1} \\ \ldots \\ \varepsilon^F{}_{t,T,k+1} \end{bmatrix}$ for $t = 1, \ldots, N$ and $k = 1, \ldots, T-1$.

The estimated values of the dependent variables are obtained from $\widehat{\boldsymbol{Y_F}} = \boldsymbol{X_F}\widehat{\boldsymbol{\beta}}$.

The $\boldsymbol{X_F}$ matrix has two types of elements (in all the *t* triangles):

- The $x_{t,T-k+1,k}$, which are on the last diagonal of the upper triangle.
- And the $x_{t,i>T-k+1,k}$ which are on the lower triangle. They are obtained starting with the $x_{t,T-k+1,k}$ (from the last diagonal) multiplied by the estimated loss development factors.

This will give us another diagonal and we will repeat the procedure to get the remaining ones.

7.1.2 Assumptions

Having defined the method given by equation 7.1.2, we present in this section its assumptions.

Proposition 7.1.1 *Considering the method given by (7.1.2), that allows for heteroscedasticity of the errors inside each equation we assume for our PGLR method*

$$\mathbb{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \mathbb{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0} \tag{7.1.3}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\sigma}^2\boldsymbol{W} = \boldsymbol{\Psi} \tag{7.1.4}$$

$$\mathbb{E}(\boldsymbol{\varepsilon_F}\boldsymbol{\varepsilon_F}') = \boldsymbol{\sigma}^2\boldsymbol{W_F} = \boldsymbol{\Psi_F} \tag{7.1.5}$$

Where $\boldsymbol{0}$ *is a vector of zeros of* size $(N \times m) \times 1$, $\boldsymbol{W}$ *is a* $(N \times m) \times (N \times m)$ *diagonal weighting matrix, which depends on the parameter* α *on each non-zero cell.* $\boldsymbol{W}$ *is given by equation (7.1.6) where the* $diag$ *operator transforms one vector on a diagonal matrix. The* $\boldsymbol{W}$ *diagonal elements are given by the elements of the transformed vectors*

$$\boldsymbol{W} = diag(\boldsymbol{X}^{\alpha}) = \begin{bmatrix} diag({\boldsymbol{X}_1}^{\alpha}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & diag({\boldsymbol{X}_N}^{\alpha}) \end{bmatrix} = \begin{bmatrix} {x_{1,1,1}}^{\alpha} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{N,T-1,T-1}}^{\alpha} \end{bmatrix} \tag{7.1.6}$$

The matrix $\boldsymbol{W_F}$ *is the future* $\boldsymbol{W}$ *and has the same structure as* $\boldsymbol{W}$. *However, its elements are the* ${\boldsymbol{X_F}}^{\alpha}$ *instead of* $\boldsymbol{X}^{\alpha}$. $\boldsymbol{W}$ *corresponds to a specific structure of heteroscedasticity through the choice of parameter α, and* $\boldsymbol{W_F}$ *has the same structure as* $\boldsymbol{W}$ *but is based on the predicted payments.*

$$\boldsymbol{W_F} = diag\ (\mathbf{X}) = \begin{bmatrix} diag\left({\boldsymbol{X}_{\boldsymbol{F}_1}}^{\alpha}\right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & diag\left({\boldsymbol{X}_{\boldsymbol{F}_N}}^{\alpha}\right) \end{bmatrix} = \begin{bmatrix} {x_{1,T,1}}^{\alpha} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{N,T,T-1}}^{\alpha} \end{bmatrix} \tag{7.1.7}$$

The $\boldsymbol{\sigma}^2$ *is diagonal block matrix with* N *blocks and of size* $(N \times m) \times (N \times m)$, *when expanded.*

$$\boldsymbol{\sigma^2} = \begin{bmatrix} \boldsymbol{\sigma^2}_{\mathbf{1,1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\sigma^2}_{\boldsymbol{N,N}} \end{bmatrix} \quad (7.1.8)$$

where each block $\boldsymbol{\sigma^2}_{\boldsymbol{t,t}} = diag \begin{bmatrix} \sigma^2{}_{t,k} \\ \dots \\ \sigma^2{}_{t,T-1} \end{bmatrix}$ *for* $t = 1, \dots, N$ *and* $k = 1, \dots, T-1$.

Seeing (7.1.7) with (7.1.8) we understand that the method will be homoscedastic in each triangle when $\alpha = 0$. Otherwise, it will be heteroscedastic.

However, being homoscedastic at each triangle level doesn't imply that the method is homoscedastic when all the triangles are considered and the estimated variances between triangles are different (see 7.1.8).

7.1.3 Estimation

The following two Lemmas allow us to have estimators for $\boldsymbol{\beta}$ and for $\boldsymbol{\sigma^2}$.

Lemma 7.1 *Following, Fomby et al. (1984), we may get the estimation of the* $\boldsymbol{\beta}$*, the loss development factors vector of all the equations from all the triangles. The* $\widehat{\boldsymbol{\beta}}$ *is obtained using the Aitken generalized least squares method with* $\boldsymbol{\Psi}$ *as weights matrix and it is the best linear unbiased estimator of* $\boldsymbol{\beta}$.

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'\Psi^{-1}X})^{-1}\boldsymbol{X\Psi^{-1}Y} \quad (7.1.9)$$

The parameter α from (7.1.6) and (7.1.7) will be estimated as the one that minimizes the prediction error. This α parameter is a method choice parameter and as in chapter 6 we selected the model with the prediction error minimization. The reasons were already given in 6.1.3.

Lemma 7.1.2 *Following Srivastava and Giles (1987) we may estimate* $\hat{\sigma}^2{}_{t,k}$ *using the equation k (from triangle t) sum of square of the errors,* $SSR_{t,k}$*, divided by this equation degrees of freedom, the number of observations* $T_{t,k}$ *from this equation minus the number of parameters from the equation, in this case one.*

$$\hat{\sigma}^2{}_{t,k} = \frac{SSR_{t,k}}{T_{t,k}-1} \quad (7.1.10)$$

7.1.4 Prediction Error

We also need an expression for the prediction error which is given by the following theorem.

Theorem 7.1.1 *Knowing that the prediction error (i.e., the root of the mean square error) is given by the root of the expected value of* $\widehat{\boldsymbol{Y}_{\boldsymbol{F}}} - \boldsymbol{Y}_{\boldsymbol{F}}$ *and its transpose,* $E\left(\widehat{\boldsymbol{Y}_{\boldsymbol{F}}} - \boldsymbol{Y}_{\boldsymbol{F}}\right)\left(\widehat{\boldsymbol{Y}_{\boldsymbol{F}}} - \boldsymbol{Y}_{\boldsymbol{F}}\right)'$*, we get it as the root of the sum of the estimation variance with the prediction variance.*

$$Mean\ Square\ Error = \mathbb{E}\left[\boldsymbol{X}_{\boldsymbol{F}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{X}'_{\boldsymbol{F}}\right] + \mathbb{E}(\boldsymbol{\varepsilon}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_{\boldsymbol{F}})' \qquad (7.1.11)$$

The estimation of variance is given by

$$Est.Variance. = \boldsymbol{X}_{\boldsymbol{F}}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})\boldsymbol{X}'_{\boldsymbol{F}}$$

and the process variance comes as

$$Process\ Var. = \mathbb{E}(\boldsymbol{\varepsilon}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_{\boldsymbol{F}}')\ .$$

All together means that the msep is given by

$$\boldsymbol{X}_{\boldsymbol{F}}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})\boldsymbol{X}'_{\boldsymbol{F}} + \mathbb{E}(\boldsymbol{\varepsilon}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_{\boldsymbol{F}}') \qquad (7.1.12)$$

Proof. This can be done following the same steps of the GLR and MGLR case.

Proposition 7.1.2 *Following (7.1.12), assumptions (7.1.4) and (7.1.5) the msep is given by*

$$msep = \boldsymbol{X}_{\boldsymbol{F}}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})\boldsymbol{X}'_{\boldsymbol{F}} + \boldsymbol{\Psi}_{\boldsymbol{F}} \qquad (7.1.13)$$

7.1.5 Particular Portfolio Univariate Methods

As in the case of the GLR and the MGLR, from sections 6.3 and 6.4, and following the results of Theorem 7.1, we need to know which weighting matrices we are using, in the method of this section 7.1 this means $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}_{\boldsymbol{F}}$. For the latter, we need the parameter α and we may obtain it, as before, searching for the α that minimizes the prediction error presented in (7.1.13).

The parameter α also corresponds to a specific structure of heteroscedasticity. If α is zero, we will get homoscedastic errors inside each triangle. This means, as in the GLR and MGLR methods, that the way $\boldsymbol{\Psi}$ is defined will provide us with several claims reserving methods for estimating the loss development factors.

Analytically, we get the VP for $\alpha = 0$, the CL for $\alpha = 1$, the SA for $\alpha = 2$, and other methods for different values of α. To have them, we just need to change α to get a different $\boldsymbol{\Psi}$ matrix. For the VP, we will have homoscedastic errors, for the CL and the SA we will have heteroscedastic errors.

Thus, the main advantage of this approach is that we choose α that minimizes the prediction error for N triangles at the same time. With α different from 0, 1 and 2, we get other distinct methods: the optimal choice of the weights of the link ratios is obtained as the prediction error is minimized.

All the link ratios methods considered here, see next corollaries (7.1.1), (7.1.2) and (7.1.3), depend on α which represents the level of heteroscedasticity and we want to choose α that minimizes the prediction error. As we saw in section 5.4, the lower is the prediction error, the better the errors analysis and the back-testing results.

Particular cases of the method are considered with the next three corollaries. Obviously, the proofs of those corollaries are linked with, Theorem 6.3.1, equation (7.1.13), Theorem 7.1.1 and Proposition 7.1.2. Due to this they are omitted.

Corollary 7.1.1 *If* $\alpha = 0$, *the triangle's variances are homoscedastic, and we get, see (7.4.4) and (7.1.5)*

$$\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\sigma}^2 \boldsymbol{I}_{(N\times m)\times(N\times m)} = \boldsymbol{\Psi}_{\boldsymbol{VP}}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_{\boldsymbol{F}}') = \boldsymbol{\sigma}^2 \boldsymbol{I}_{(N\times m)\times(N\times m)} = \boldsymbol{\Psi}_{\boldsymbol{VP,F}}$$

Here $\boldsymbol{I}_{(N\times m)\times(N\times m)}$ *is a diagonal identity matrix with size* $(N \times m) \times (N \times m)$. *With* α *=0, the loss development factors are the ones from the VP applied with a portfolio context, the Portfolio Vector Projection (PVP), see (7.1.9), where* $\boldsymbol{\Psi}_{\boldsymbol{VP}}$ *is* $\boldsymbol{\Psi}$ *with* $\boldsymbol{W}$=$\boldsymbol{I}_{(N\times m)\times(N\times m)}$

$$\hat{\boldsymbol{\beta}}^{PVP} = \left(\boldsymbol{X}'\boldsymbol{\Psi}_{\boldsymbol{VP}}{}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{\Psi}_{\boldsymbol{VP}}{}^{-1}\boldsymbol{Y}$$

Then, the msep is given by, see (7.1.13), where $\boldsymbol{\Psi_{VP,F}}$ *is* $\boldsymbol{\Psi}$ *with* $\boldsymbol{W_F}$=$\boldsymbol{I}_{(N\times m)\times(N\times m)}$

$$\mathbb{E}\left(\widehat{\boldsymbol{Y_F}}-\boldsymbol{Y_F}\right)\left(\widehat{\boldsymbol{Y_F}}-\mathrm{Y_F}\right)'=\boldsymbol{X_F}(\boldsymbol{X}'\boldsymbol{\Psi_{VP}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{\Psi_{VP}}^{-1}\boldsymbol{X}\right)\boldsymbol{X}'_{\boldsymbol{F}}+\boldsymbol{\Psi_{VP,F}} \quad (7.1.14)$$

Corollary 7.1.2 *If* $\alpha=1$, *the triangle's variances are heteroscedastic, and we get*

$$\mathbb{E}(\boldsymbol{\varepsilon\varepsilon}')=\boldsymbol{\sigma}^2\boldsymbol{W}_{CL}=\boldsymbol{\Psi}_{CL}$$

$$\mathbb{E}(\boldsymbol{\varepsilon_F\varepsilon_F}')=\boldsymbol{\sigma}^2\boldsymbol{W}_{CL,F}=\boldsymbol{\Psi}_{CL,F}$$

with

$$\boldsymbol{W}_{CL}=\begin{bmatrix}x_{1,1,1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & x_{N,T-1,T-1}\end{bmatrix} \quad and \quad \boldsymbol{W}_{CL,F}=\begin{bmatrix}x_{1,T,1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & x_{N,T,T-1}\end{bmatrix}.$$

With α *=1, the loss development factors are the ones from the CL applied in a portfolio context, the Portfolio Chain-Ladder (PCL), see (7.1.9), where* $\boldsymbol{\Psi_{CL}}$ *is* $\boldsymbol{\Psi}$ *with* **W**=$\boldsymbol{W}_{CL}$.

$$\widehat{\boldsymbol{\beta}}^{PCL}=\left(\boldsymbol{X}'\boldsymbol{\Psi_{CL}}^{-1}\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{X}\boldsymbol{\Psi_{CL}}^{-\mathbf{1}}\boldsymbol{Y}$$

Then, the msep is given by, see (7.1.13), where $\boldsymbol{\Psi_{CL,F}}$ *is* $\boldsymbol{\Psi}$ *with* $\boldsymbol{W_F}$=$\boldsymbol{W}_{CL,F}$

$$\mathbb{E}\left(\widehat{\boldsymbol{Y_F}}-\boldsymbol{Y_F}\right)\left(\widehat{\boldsymbol{Y_F}}-\boldsymbol{Y_F}\right)'=\boldsymbol{X_F}(\boldsymbol{X}'\boldsymbol{\Psi}_{CL}{}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{\Psi}_{CL}{}^{-1}\boldsymbol{X}\right)\boldsymbol{X}'_{\boldsymbol{F}}+\boldsymbol{\Psi}_{CL,F} \quad (7.1.15)$$

Corollary 7.1.3 *If* $\alpha=2$, *the triangle's variances are heteroscedastic, and we get*

$$\mathbb{E}(\boldsymbol{\varepsilon\varepsilon}')=\boldsymbol{\sigma^2}\boldsymbol{W}_{SA}=\boldsymbol{\Psi}_{SA}$$

$$\mathbb{E}(\boldsymbol{\varepsilon_F\varepsilon_F}')=\boldsymbol{\sigma}^2\boldsymbol{W}_{SA,F}=\boldsymbol{\Psi}_{SA,F}$$

with

$$\boldsymbol{W}_{SA}=\begin{bmatrix}x_{1,1,1}{}^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & x_{N,T-1,T-1}{}^2\end{bmatrix} \quad and \quad \boldsymbol{W}_{SA,F}=\begin{bmatrix}x_{1,T,1}{}^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & x_{N,T,T-1}{}^2\end{bmatrix}$$

With α *=2, the loss development factors are the ones from the SA, see (7.1.3), applied in a portfolio context, the Portfolio Simple Average (PSA), see (7.1.9), where* $\boldsymbol{\Psi_{SA}}$ *is* $\boldsymbol{\Psi}$ *with* **W**=$\boldsymbol{W}_{SA}$.

$$\widehat{\boldsymbol{\beta}}^{PSA} = \left(\boldsymbol{X}'\boldsymbol{\Psi_{SA}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{\Psi_{SA}}^{-1}\boldsymbol{Y}$$

. Then, the msep is given by, see (7.1.13)

$$\mathbb{E}\left(\widehat{\boldsymbol{Y}_F} - \boldsymbol{Y}_F\right)\left(\widehat{\boldsymbol{Y}_F} - \boldsymbol{Y}_F\right)' = \boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{\Psi}_{SA}{}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{\Psi}_{SA}{}^{-1}\boldsymbol{X}\right)\boldsymbol{X}'_F + \boldsymbol{\Psi}_{SA,F} \qquad (7.1.16)$$

Following the results of Theorem 7.2.1, the procedures are like the ones from the univariate GLR method, presented on section 6.3. In the PGLR we need:

- The $\hat{\sigma}^2{}_{t,j,k}$ using OLS, see Lemma (7.1.2.).
- The parameter α to have the $\boldsymbol{W}$, (7.1.6) and the $\boldsymbol{W}_F$ (7.1.7) matrices. Our decision was to choose the α that minimizes the prediction error.
- The latter will give the vector of the loss development factors, given by (7.1.9), to have $\boldsymbol{X}_F$.
- And finally, we need $\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$ and $\mathbb{E}(\boldsymbol{\varepsilon_F}\boldsymbol{\varepsilon_F}')$, which implies some assumptions about the method.

7.2 Portfolio Multivariate Generalized Link Ratios

In this section, we develop the section 7.1 method to the case where there are contemporaneous correlations between equations inside the same triangle and between triangles. It is a development similar to the one done on the previous chapter section 6.4 but now including correlations between triangles.

7.2.1 Portfolio Multivariate Method

The method considered here is the same presented in section 7.1.1. However, we will change the assumptions introducing a more complex structure for the errors: the SUR method.

The method will become multivariate as a SUR and may also use the heteroscedastic structure from the GLR, including also the VP, the CL and the SA. In this PMGLR method, we are going to maintain all the framework presented in section 7.1 but we are going to change the assumptions (7.1.4) and (7.1.5).

We are going to assume contemporaneous correlations between the errors of the different equations and between the triangles. To do that, we get a portfolio multivariate method. The method is still based on (7.1.2) and even (7.1.9) will be similar.

7.2.2 Assumptions

$\mathbf{\Sigma}$ is a block matrix of block-size $[N \times (T-1)] \times [N \times (T-1)]$ that summarizes the variances and the covariances between, $k = 1, \dots, T-1$ regressions in each of the $t = 1, \dots, N$ triangles and also between each triangle, for observations in the same origin year. Expanding each block, we get a matrix of dimension $(N \times m) \times (N \times m)$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{1,1,1} & \cdots & \mathbf{\Sigma}_{N,1,T-1} \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{1,T-1,1} & \cdots & \mathbf{\Sigma}_{N,T-1,T-1} \end{bmatrix} \tag{7.2.1}$$

The generic component of (7.2.1), $\mathbf{\Sigma}_{t,k,k}$ is given by a matrix of size $[N \times (T-k)] \times [N \times (T-k)]$ and by $\mathrm{s}_{t,k,k}$, the variance parameter from triangle t and regression k.

$$\mathbf{\Sigma}_{t,k,k} = \mathrm{s}_{t,k,k} \begin{bmatrix} {x_{1,1,k}}^{\alpha} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{N,T-k,k}}^{\alpha} \end{bmatrix} \tag{7.2.2}$$

The generic component of (7.2.1), $\mathbf{\Sigma}_{t,k,j}$ with $k \neq j$ is given by a matrix of size $[N \times (T-k)] \times [N \times (T-k)]$ and by $\mathrm{s}_{t,k,j}$, the covariance parameter for triangle t between regression k and j, where $k \neq j$.

$$\mathbf{\Sigma}_{t,k,j} = \mathrm{s}_{t,k,j} \boldsymbol{I}_{N \times (T-k)} \tag{7.2.3}$$

$\mathbf{\Sigma}^F$ is a block matrix of block-size $[N \times (T-1)] \times [N \times (T-1)]$ that summarizes the future variances and the covariances between $k = 1, \dots, T-1$ regressions. Expanding each block, we get a matrix of dimension $(N \times m) \times (N \times m)$

$$\mathbf{\Sigma}^F = \begin{bmatrix} \mathbf{\Sigma}^F_{1,1,1} & \cdots & \mathbf{\Sigma}^F_{N,1,T-1} \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}^F_{1,T-1,1} & \cdots & \mathbf{\Sigma}^F_{N,T-1,T-1} \end{bmatrix} \tag{7.2.4}$$

The generic component of (7.2.4), $\mathbf{\Sigma}^F_{t,k,k}$, is given by a matrix of size $[N \times (T-k)] \times [N \times (T-k)]$

$$\mathbf{\Sigma}^F_{t,k,k} = \mathrm{s}_{t,k,k} \begin{bmatrix} {x_{1,T,k}}^{\alpha} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {x_{N,T+k,k}}^{\alpha} \end{bmatrix} \tag{7.2.5}$$

The generic component of (7.2.4), $\mathbf{\Sigma}^F_{t,k,j}$ with $k \neq j$, is given by a matrix of size $[N \times (T - k)] \times [N \times (T - k)]$

$$\mathbf{\Sigma}^F{}_{t,k,j} = \mathrm{s}_{t,k,j} \boldsymbol{I}_{N\times(T-k)} \tag{7.2.6}$$

Proposition 7.2.1 *Considering a multivariate method that allows for heteroscedasticity of the errors inside each equation and contemporaneous correlations between these equations, we assume for our PMGLR method*

$$\mathbb{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \mathbb{E}(\boldsymbol{\varepsilon}) = 0 \tag{7.2.7}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{\Sigma} \tag{7.2.8}$$

$$\mathbb{E}(\boldsymbol{\varepsilon_F}\boldsymbol{\varepsilon_F}') = \mathbf{\Sigma}_F \tag{7.2.9}$$

Our assumptions for the PMGLR are going to reflect the existence of correlations between each of the regressions and also between triangles. We now have as new assumptions for errors the following weighting matrix given by $\mathbf{\Sigma}$.

Inside this method we may also have three specific cases for correlations:

- We may assume correlations between triangles and correlations between each regression equation, which is the method presented with Proposition 7.2.1.
- We may assume correlations between triangles and no correlations between each regression equation. In this case we will have,

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{1,1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{\Sigma}_{N,T-1,T-1} \end{bmatrix}$$

  and

$$\mathbf{\Sigma}^F = \begin{bmatrix} \mathbf{\Sigma}^F_{1,1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{\Sigma}^F_{N,T-1,T-1} \end{bmatrix}$$

- And we may assume correlations between each regression equation and no correlations between triangles.

7.2.3 Estimation

The parameters estimation may be obtained by the following Lemma 7.2.1.

Lemma 7.2.1 [Srivastava and Giles (1987)] *We may get the estimation of the $\boldsymbol{\beta}$, that is the estimation of the loss development factors from all the equations. The $\widehat{\boldsymbol{\beta}}$ is obtained using the SUR generalized least squares for panel data with heteroscedasticity and contemporaneous correlations and is the best linear unbiased estimator of $\boldsymbol{\beta}$.*

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \tag{7.2.10}$$

We also need an expression for the prediction error which is given by the following section.

Clearly, the parameters $s_{j,j}$ and $s_{j,j}$ are not known and must be estimated. Thus, with the following Lemma, the estimators $\hat{s}_{j,j}$ and $\hat{s}_{l,j}$ are provided.

Lemma 7.2.2 [Srivastava and Giles, 1987] *Estimators for the parameters of variance and covariance matrix from a multivariate regression with panel data are given by*

$$\hat{s}_{t,k,k} = \frac{1}{T-1} SSR_{t,k} \qquad \hat{s}_{t,k,j} = \frac{1}{T} SSR_{t,k} \tag{7.2.11}$$

$SSR_{t,k}$ are to be calculated using for each equation t the regression k Ordinary Least Squares (OLS) sum of the square of the errors.

7.2.4 Prediction Error

The following theorem follows from Theorem 7.2.1 and gives us a general non-recursive formula to have the prediction error.

Theorem 7.2.1 *The mean square error of prediction from the method presented in (7.1.2) is given by.*

$$msep = \boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}_F{}' + \mathbb{E}(\boldsymbol{\varepsilon}_F\boldsymbol{\varepsilon}_F{}') \tag{7.2.11}$$

The proof follows directly from Theorem 6.3.1 when (7.2.7), (7.2.8) and (7.2.9) are considered.

Following the results of Theorem 7.2.1, the procedures are like the ones from the univariate method, presented on section 7.1. In the PMGLR we need:

- We must get the $\hat{s}_{t,j,j}$ and $\hat{s}_{t,l,j}$ to estimate the $\mathbf{\Sigma}$ matrix, which implies to have a first regression, with OLS, to get the sum of the square of the errors.
- We need the parameter α to have $\mathbf{\Sigma}$, (7.2.1) and the Σ_F (7.2.4) matrices. Our suggestion is to choose α that minimizes the prediction error.
- This will also give us the vector of the loss development factors, given by (7.2.10) and with that we will have $\boldsymbol{X_F}$.
- Having $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_F$ we have $\mathbb{E}(\boldsymbol{\varepsilon\varepsilon}')$ and $\mathbb{E}(\boldsymbol{\varepsilon_F\varepsilon_F}')$ and we calculate the prediction error.

Proposition 7.2.2 *Following (7.2.12) and assumptions from Proposition (7.2.1) the msep is given by*

$$\mathbb{E}\left(\widehat{\boldsymbol{Y}_F}-\boldsymbol{Y}_F\right)\left(\widehat{\boldsymbol{Y}_F}-\boldsymbol{Y}_F\right)' = \boldsymbol{X}_F(\boldsymbol{X}'\mathbf{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'_F+\mathbf{\Sigma}_F. \tag{7.2.12}$$

7.2.5 Particular Portfolio Multivariate Methods

As in the univariate method from section 7.1 we choose α that minimizes the prediction error. Analytically, we do not get anymore the loss development factors from, the VP for $\alpha = 0$, the CL for $\alpha = 1$ and SA for $\alpha = 2$. The reason is the consideration of contemporaneous correlations between the regressions that change the loss development factors, see (7.2.9) which is different from (7.1.9).

However, we may say that, when $\alpha = 0$ we get a Portfolio Multivariate VP, when $\alpha = 1$ we get a Portfolio Multivariate CL and when $\alpha = 2$ we get a Portfolio Multivariate SA. The argument for this is the heteroscedasticity level. What defines and differentiates these three methods is the weights given to the link ratios and the former define of the heteroscedasticity level. In the VP is zero, $\alpha = 0$, in the CL is one, $\alpha = 1$ and in the SA is two, $\alpha = 2$. As in chapter 6, we may say that the heteroscedasticity level may be defined by α. These levels of heteroscedasticity are maintained in the multivariate approach.

As with the univariate portfolio data method, we will get other methods that give other weights to the link ratios through the α selection. As with the univariate method from section 7.1, the optimal α is the one that minimizes the prediction error.

In the Proposition 7.2.1, the level of the heteroscedastic errors and of correlation is given by the matrix $\boldsymbol{\Sigma}$, which depends, on $s_{t,lj}, s_{t,jj}$, $\boldsymbol{X}_{ij}$ (that is given from the triangle data), and on the parameter α. Here, we use (7.2.12), i.e., the *msep* minimization, to get α. Homoscedastic errors in the portfolio multivariate method, PMGLR, may also arise if $\boldsymbol{\Sigma}_{t,l,j} = \boldsymbol{I}_{N\times(T-j)}$. The univariate method, PGLR, is a particular case of the multivariate method, PMGLR, when $s_{t,lj} = 0$ when $l \neq j$.

Correlations between regressions appear on the data for several reasons such as an increase/decrease of claims on some development years or an increase/decrease of the speed of paying claims on certain development years.

Corollary 7.2.1 *If* $\alpha = 0,$ *the variances are homoscedastic and the regressions correlated. We get*

$$\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma}_{\boldsymbol{VP}}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}_F\boldsymbol{\varepsilon}_F{}') = \boldsymbol{\Sigma}_{\boldsymbol{VP},\boldsymbol{F}}$$

$\boldsymbol{\Sigma}_{\boldsymbol{VP}}$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{VP},\boldsymbol{F}}$ *are the* $\boldsymbol{\Sigma}$ *defined in (7.2.1) with the following relations*

$$\boldsymbol{\Sigma}_{t,j,j} = \mathrm{s}_{t,j,j}\boldsymbol{I}_{N\times(T-j)}$$

$$\boldsymbol{\Sigma}_{t,l,j} = \mathrm{s}_{t,l,j}\boldsymbol{I}_{N\times(T-j)}$$

Here $\boldsymbol{I}_{N\times(T-j)}$ *is a diagonal identity matrix with size* $N \times (T-j)$. *With* α *=0, the loss development factors are the ones from the VP within a portfolio multivariate context (PMVP). Also, the* $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{VP}}$.

$$\widehat{\boldsymbol{\beta}}_{PMVP} = \left(\boldsymbol{X}'\boldsymbol{\Sigma}_{\boldsymbol{VP}}{}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{\Sigma}_{\boldsymbol{VP}}{}^{-1}\boldsymbol{Y}$$

Then, the msep is given by

$$\mathbb{E}\left(\widehat{\boldsymbol{Y}_F} - \boldsymbol{Y}_F\right)\left(\widehat{\boldsymbol{Y}_F} - \boldsymbol{Y}_F\right)' = \boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{\Sigma}_{\boldsymbol{VP}}{}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'_F + \boldsymbol{\Sigma}_{\boldsymbol{VP},\boldsymbol{F}}. \qquad (7.2.13)$$

Corollary 7.2.2 *If* $\alpha = 1,$ *the variances are heteroscedastic and the regressions correlated. We get*

$$\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma}_{\boldsymbol{CL}}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}_F\boldsymbol{\varepsilon}_F{}') = \boldsymbol{\Sigma}_{\boldsymbol{CL,F}}$$

The $\boldsymbol{\Sigma}_{\boldsymbol{CL}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{CL,F}}$ are respectively the $\boldsymbol{\Sigma}$ and the $\boldsymbol{\Sigma}_{\boldsymbol{F}}$ defined in (7.2.1) and (7.2.4) with

$$\boldsymbol{\Sigma}_{t,j,j} = s_{t,j,j}\begin{bmatrix} x_{1,1,j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{t,T-j,j} \end{bmatrix}$$

$$\boldsymbol{\Sigma}^F_{t,j,j} = s_{t,j,j}\begin{bmatrix} x_{1,T,j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{t,T+j,j} \end{bmatrix}$$

With α =1, the loss development factors are the ones from the CL within a portfolio multivariate context (PMCL).

$$\widehat{\boldsymbol{\beta}}_{PMCL} = \left(\boldsymbol{X}'\boldsymbol{\Sigma}_{\boldsymbol{CL}}{}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{\Sigma}_{\boldsymbol{CL}}{}^{-1}\boldsymbol{Y}$$

Then, the msep is given by

$$\mathbb{E}\left(\widehat{\boldsymbol{Y}_F} - \boldsymbol{Y}_F\right)\left(\widehat{\boldsymbol{Y}_F} - \boldsymbol{Y}_F\right)' = \boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{\Sigma}_{\boldsymbol{CL}}{}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'_F + \boldsymbol{\Sigma}_{\boldsymbol{CL,F}} \qquad (7.2.14)$$

Corollary 7.2.3 *If $\alpha = 2,$ the variances are heteroscedastic and the regressions correlated. We get*

$$\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma}_{\boldsymbol{SA}}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}_F\boldsymbol{\varepsilon}_F{}') = \boldsymbol{\Sigma}_{\boldsymbol{SA,F}}$$

The $\boldsymbol{\Sigma}_{\boldsymbol{SA}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{SA,F}}$ are respectively the $\boldsymbol{\Sigma}$ and the $\boldsymbol{\Sigma}_{\boldsymbol{F}}$ defined in (7.2.1) and (7.2.4) with

$$\boldsymbol{\Sigma}_{t,j,j} = s_{t,j,j}\begin{bmatrix} x_{1,1,j}{}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{t,T-j,j}{}^2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}^F_{t,j,j} = s_{t,j,j}\begin{bmatrix} x_{1,T,j}{}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{1,T+j,j}{}^2 \end{bmatrix}$$

With α =2, the loss development factors are the ones from the SA within a portfolio multivariate context (PMSA).

$$\widehat{\boldsymbol{\beta}}_{PMSA} = \left(\boldsymbol{X}'\boldsymbol{\Sigma_{SA}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{\Sigma_{SA}}^{-1}\boldsymbol{Y}$$

Then, the msep is given by

$$\mathbb{E}\left(\widehat{\boldsymbol{Y}_F} - \boldsymbol{Y}_F\right)\left(\widehat{\boldsymbol{Y}_F} - \boldsymbol{Y}_F\right)' = \boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{\Sigma_{SA}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'_F + \boldsymbol{\Sigma_{SA,F}} \qquad (7.2.15)$$

7.3 Numerical Results

We consider for the numerical results the three paid claims triangles used before, three of them on chapter 5 and two on the chapter 6. We call them triangle 1, triangle 2 and triangle 3:

- Triangle 1, Mack (1993a).
- Triangle 2, Taylor and Ashe (1983).
- And Triangle 3, Taylor and McGuire (2016).

The results obtained, once again, confirm the VP as the method that minimizes the prediction error.

We present results for the PGLR and the PMGLR methods estimating simultaneously the above three triangles. For the PMGLR we give two results. One result with correlations between all the equations, from each triangle and between the triangles, and another one just with correlations between the triangles.

We also compare these results with the ones obtained from an aggregated triangle. Such one will result from the sum of the three triangles and it is a good illustration of a practical problem that all the actuaries face: sometimes it is not possible to split the data, from a certain line of business, in more than one triangle or if it is, then it is not possible to trust the split due to the lack of quality on data.

7.3.1 Portfolio Generalized Link Ratios

We obtained the results presented in the Table 7.1 for the portfolio of three triangles. The α that minimizes the prediction error was zero, confirming once again the VP as the best solution, according with this criterion.

The prediction error obtained was of 8.9% and the total reserves estimated was of 18 896 187, see Table 7.1.

Table 7.1: Portfolio Generalized Link Ratios Results

Column	Reserves per Column	Prediction Error	Prediction Error %
2	869 981	240 331	28%
3	1 952 143	347 595	18%
4	3 435 107	475 663	14%
5	2 419 986	567 808	23%
6	2 021 461	658 827	33%
7	2 319 736	837 251	36%
8	1 847 867	698 526	38%
9	3 141 606	594 305	19%
10	888 300	321 286	36%
Total	**18 896 187**	**1 675 306**	**8,9%**

Had we considered just one triangle that corresponds to the sum of the three triangles, the results would have been the ones presented in Table 7.2.

Table 7.2: Generalized Link Ratios Aggregated Triangle Results

Column	Reserves per Column	Prediction Error	Prediction Error %
2	865 608	230 665	27%
3	1 937 534	334 062	17%
4	3 397 751	459 846	14%
5	2 406 814	553 000	23%
6	2 016 933	657 000	33%
7	2 309 872	837 757	36%
8	1 844 373	700 222	38%
9	3 138 605	712 769	23%
10	886 667	571 490	64%
Total	**18 804 158**	**1 772 148**	**9,4%**

When compared with Table 7.1 results, the aggregated triangles results are similar: the prediction error increases to 9.4% and the reserves decrease to 18 804 158. The differences are small but we must be aware that the triangle 2 has 98% of the level of the total reserves.

As expected, the reserves obtained in Table 7.1 correspond to the reserves sum from the three triangles when the GLR method is applied to each triangle, see Table 7.3.

Table 7.3: Portfolio Generalized Link Ratios Results - Totals from the 3 Triangles

Column	Reserves per Column	Prediction Error	Prediction Error %
2	869 981	244 074	28%
3	1 952 143	360 140	18%
4	3 435 107	491 128	14%
5	2 419 986	585 065	24%
6	2 021 461	668 160	33%
7	2 319 736	846 658	36%
8	1 847 867	706 287	38%
9	3 141 606	600 946	19%
10	888 300	324 446	37%
Total	**18 896 187**	**1 707 793**	**9,0%**

The only thing that changes is the prediction error: now, instead of three prediction errors, we have one prediction error.

This same level of total reserves happens because we did not consider any correlations between triangles (neither between equations regressions). We just used portfolio data to estimate all the regressions and triangles at the same time.

The prediction error obtained is a weighted average from the prediction errors of the three triangles. The weights are the reserves estimated.

The individual results from the three triangles are presented in Table 7.4. Here we may see that the 9.0% prediction error obtained is a weighted average from those in the individual triangles:

$$9.0\% = \frac{43\,772 \times 35.8\% + \ 18\,479\,500 \times 9.1\% + 372\,915 \times 4.6\%}{18\,896\,187}$$

Also, the sum of the prediction error from all the triangles (in monetary units), see Table 7.4, is equal to the same indicator obtained from Table 7.1:

$$1\,707\,793 = 15\,651 + 1\,675\,147 + 16\,995$$

Table 7.4: Portfolio Generalized Link Ratios Results from the 3 Triangles

Triangle 1			
Column	**Reserves per Column**	**Prediction Error**	**Prediction Error %**
2	2 511	3 773	150%
3	5 672	5 359	94%
4	7 501	6 627	88%
5	7 867	7 021	89%
6	7 208	5 801	80%
7	4 283	6 015	140%
8	4 412	4 488	102%
9	2 620	3 957	151%
10	1 698	1 775	105%
Total	**43 772**	**15 651**	**35,8%**
Triangle 2			
Column	**Reserves per Column**	**Prediction Error**	**Prediction Error %**
2	831 767	240 301	29%
3	1 901 782	347 477	18%
4	3 373 994	475 532	14%
5	2 363 113	567 670	24%
6	1 970 809	658 792	33%
7	2 276 227	837 223	37%
8	1 806 584	698 504	39%
9	3 102 951	594 286	19%
10	852 273	321 278	38%
Total	**18 479 500**	**1 675 147**	**9,1%**
Triangle 3			
Column	**Reserves per Column**	**Prediction Error**	**Prediction Error %**
2	35 703	0	0%
3	44 689	7 304	16%
4	53 611	8 969	17%
5	49 007	10 373	21%
6	43 445	3 567	8%
7	39 225	3 420	9%
8	36 871	3 295	9%
9	36 035	2 703	7%
10	34 329	1 393	4%
Total	**372 915**	**16 995**	**4,6%**

7.3.2 Portfolio Multivariate Generalized Link Ratios

We start the PMGLR method by having the results from Proposition 7.2.1, that is, with correlations between triangles and between equations inside each triangle.

For this method, the PMGLR, we also got, as the lowest prediction error, α = 0. The prediction error of 2.7% represents an important decrease in respect of the PGLR result (8.9% and also with α = 0). The reserves increase to 23 619 959 (they were 18 896 187 with PGLR). See Table 7.5 for the PMGLR (and Table 7.1 for the PGLR case).

Table 7.5: Portfolio Multivariate Generalized Link Ratios Results

Column	Reserves per Column	Prediction Error	Prediction Error %
2	806 936	240 633	30%
3	1 992 818	347 132	17%
4	2 452 163	588 011	24%
5	3 162 513	521 512	16%
6	3 675 198	557 398	15%
7	3 838 679	695 253	18%
8	1 886 760	240 575	13%
9	2 338 183	151 203	6%
10	3 463 709	109 640	3%
Total	**23 616 959**	**630 782**	**2,7%**

Using an aggregate triangle would decrease the reserves to 19 889 001, but with an important increase of the prediction error to 5.3%, see Table 7.6.

Table 7.6: Multivariate Generalized Link Ratios Aggregated Triangle Results

Column	Reserves per Column	Prediction Error	Prediction Error %
2	858 647	230 665	27%
3	1 971 159	327 504	17%
4	3 421 429	557 641	16%
5	2 426 784	529 668	22%
6	2 098 779	583 189	28%
7	2 765 878	624 811	23%
8	1 898 505	155 946	8%
9	3 542 188	225 593	6%
10	905 634	488 026	54%
Total	**19 889 001**	**1 052 298**	**5,3%**

The reason for the increase of the reserves level between the PGLR and the PMGLR lies in the change of the loss development factors, $\boldsymbol{\beta}$, mainly the ones from the triangle 2, as the latter weights 98% on the total reserves.

Several loss development factors increase and some of them decrease but the increase of 5% from the one correspondent to $j = 9$ has 5% impact on all the ultimate factors from all the origin years and justifies the increase of reserves of around 25%.

The change of the loss development factors is a consequence of the change of the weights matrix as the latter is now considering the contemporaneous correlations between the triangles. The changes of these factors are presented on the Table 7.7.

Table 7.7: Changes in Loss Development Factors with PMGLR

Triangle	**Loss Development Factors**	
Number	Number	Variation
1	1	99,1%
1	2	12,8%
1	3	17,8%
1	4	-8,1%
1	5	-38,6%
1	6	-19,2%
1	7	-6,1%
1	8	6,1%
1	9	0,1%
2	1	-4,4%
2	2	0,9%
2	3	-9,1%
2	4	6,0%
2	5	7,8%
2	6	4,7%
2	7	-0,3%
2	8	-2,3%
2	9	5,0%
3	1	-19,8%
3	2	19,0%
3	3	-2,9%
3	4	-0,9%
3	5	1,7%
3	6	0,8%
3	7	0,0%
3	8	-0,2%
3	9	0,2%

Finally, we present another result for a variant of the PMGLR: we assume that there are correlations between triangles but that there are no correlations between each triangle equations. This should correspond to the PGLR method. However, the results will be different as the PGLR and the PMGLR methods estimate the correlations between triangles in a different way, see Lemmas 7.1.2 and 7.2.2. Despite this we may compare the PMGLR results with this calculation without correlations between equations of each triangle.

We can see in Table 7.8 that the prediction error increases from 2.7% (see Table 7.5) to 6.4%. We conclude from these figures that, with these triangles, the correlations between equations are more important than the correlations between triangles. Also, the level of estimated reserves drops from 23 616 959 to 19 114 443.

Table 7.8: PMGLR with no correlations between Equations of each Triangle

Column	Reserves per Column	Prediction Error	Prediction Error %
2	890 992	240 214	27%
3	1 935 591	324 597	17%
4	3 492 070	515 633	15%
5	2 367 356	549 370	23%
6	1 907 664	640 951	34%
7	2 191 227	662 738	30%
8	1 839 764	187 908	10%
9	3 292 299	61 799	2%
10	1 197 480	128 837	11%
Total	**19 114 443**	**1 225 642**	**6,4%**

7.4 Test of Pooled Data

If we have several triangles, a question that may arise is the possibility of aggregating all the triangles in just one. This could save some working time and, in some circumstances, where we do not have credible information (not enough claims), may be a way of handling this problem of insufficient claims in some triangles. Aggregating triangles means having more claims and more credible information but also means mixing covers, which may bring heterogeneity and unstable projections.

Baltagi (2013), presents a test for aggregation, also called pool-ability test. This tests the hypothesis that the slopes of the equations between different entities are the same. Applied to the insurer's triangles this means that we test the hypothesis of the loss development factors for the same development year being equal between the triangles analysed. If the slopes are equal, we may aggregate the triangles.

To do the test we need to compare two methods. One method, the restricted method, assumes that, for the same development year, all the loss development factors from the different triangle are equal. This means that the restricted method is given by (7.4.1)

$$\boldsymbol{Y} = \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}, \tag{7.4.1}$$

$$\widetilde{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{X}_1 \\ \dots \\ \boldsymbol{X}_N \end{bmatrix}$$

$$\widetilde{\boldsymbol{\beta}} = \begin{bmatrix} \tilde{\beta}_1 \\ \dots \\ \tilde{\beta}_k \end{bmatrix}$$

The unrestricted method will be the method summarized with (7.1.2).

The test has one version for OLS methods and a second version to GLS methods. The latter is the former applied to a transformed method where each variable was multiplied by $\boldsymbol{W}^{-1/2}$. After this transformation the method satisfies the assumptions from the traditional OLS method, Baltagi (2013), and we may apply the OLS test.

As we do the test for the case of α = 0, that is the VP that got the lowest prediction errors, we do not need to change the variables and we may apply the OLS test directly. We saw in Corollaries 6.3.1 and 7.1.1 that the homoscedastic VP is an OLS method. Had we applied the CL or the SA, would require the use of the test for GLS methods. As we saw in Corollaries 6.3.2, 6.3.3, 7.1.2 and 7.1.3, the CL and the SA are GLS methods.

The null hypothesis is $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$ and it may be shown that under this hypothesis, the SSR is the pooled OLS SSR, see Baltagi (2013). The test statistic is given by

$$\tilde{F} = \frac{\dfrac{(RSSR - URSS)}{tr(\widetilde{M}) - tr(M)}}{\dfrac{URSS}{tr(M)}} \sim F[(N-1)(T-1), N]$$

Where RSSR is the restricted method errors sum of squares and URSS is the unrestricted method errors sum of squares. M and $\widetilde{M}$ are idempotent matrices, respectively, for the unrestricted and the restricted methods, given by

$$\boldsymbol{M} = \boldsymbol{I}_{NT} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

$$\widetilde{\boldsymbol{M}} = \boldsymbol{I}_{NT} - \widetilde{\boldsymbol{X}}\left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\right)^{-1}\widetilde{\boldsymbol{X}}'$$

$\boldsymbol{I}_{NT}$ is the identity matrix of size $N \times T$. The test is F distributed with $[(N-1)(T-1), N]$ degrees of freedom. Working with 5% of significance level and with the triangles used in section 7.3 we get $F_{5\%,(18,3)} = 8.675$. We also obtain as test statistic $\tilde{F} = 0.015$.

This means that we do not reject the null hypothesis that, for the same development years, the loss development factors are the same in the three triangles. The reason for this is the enormous weight of the triangle 2 on the future payments, 98%.

Table 7.9: Testing Aggregation

RSSR	1 972 740 041 457
tr($\widetilde{M}$)	126
URSS	1 872 256 901 279
tr(M)	27
F	**0,015**
Significance	5%
N	3
T	10
Critical Value	**8,675**
Accept H0	Yes

We saw in section 7.3, that using aggregated data increases the prediction error by just 0.5%. This small deterioration of the prediction error is the price to be paid for the aggregation, we have loss development factors specific from each triangle and correlations between the triangles that are not used anymore.

7.5 Conclusions

From these methods with portfolio data, in this case with three triangles estimated at the same time, we got several conclusions:

- The use of a portfolio of triangles confirms the use of the VP as the solution that minimizes the prediction error. This happens both in the univariate (PGLR) and in the multivariate case (PMGLR).

- When we use the PMGLR, the prediction error decreases when compared with the PGLR. It seems that it worth to work with more information to predict the reserves.
- However, using such information also produces an increase of the level of reserves due to the correlations between triangles.
- As expected the level of reserves is not the same as the one that arises when we have all the triangles aggregated. The aggregation of the triangles in just one triangle, gives us a lower level of estimated reserves but the prediction errors are higher. This is a good example of the danger of not using homogeneous triangles in claims reserving, even if the prediction error is low.
- A pooled data test may help us decide if we shall or not aggregate triangles. If the loss development factors are not statistically different that may be an indication to work with aggregated that if the covers are not significantly different.

8. General Conclusions

The CL was very important to the insurance sector to understand the risk and the uncertainty in the claims process. It gave more importance to actuaries as being the best professionals to use this method with care and professionalism. It allowed also the development of other methods for claims reserving.

However, the method does not minimize the prediction error. The latter is not the only factor to consider on method selection but when is minimized is associated with better results on other common analysis, for example: lower errors, lower volatility of the errors and better performance when back-testing is performed. We also understood that the CL, very often, has very high prediction errors when applied to the triangles. Despite all this most of the actuaries use the CL.

The literature produced several methods on claims reserving but several methods are also CL based or mixtures with the CL. Some of them are also replications of the CL.

It was shown in this work that the homoscedastic and the heteroscedastic VP's, both regression through the origin, produces better results than the CL when the prediction error is analysed. This is a very important issue to insurers to manage the companies as it is now compulsory to them to have better estimates of the reserves, due to the implementation of the Solvency II systems in Europe (and of the risk-based systems already in force in other countries).

Considering claims reserving in a regression context opened the door to have other methods and to replicate known methods like the CL and the SA. The minimization of the prediction error, of these methods, shows to us, once again, that the VP is the best one to achieve this objective. It also highlighted that the stochastic CL was presented in the literature considering has process error the past errors obtained by a regression. A proper approach, with the future errors, show to us that the CL prediction errors are usually higher. But more than this, we could see that despite the objection of some actuaries in using regression analysis, the stochastic CL and the SA are also a regression, but with an approximation to the minimization of the square of the errors, due to the inexistence of significant heteroscedastic errors in the triangles.

It was also shown that with a simple test we may confirm or not the existence of heteroscedasticity, the main assumption from CL and SA when presented as regression methods. In the triangles considered in this work the tests exclude the possibility of heteroscedasticity.

Although methods as the CL and the SA may be important if the regressions exhibit heteroscedasticity. The CL assumes a lower heteroscedasticity than the SA.

We could also develop a multivariate regression in the sense that we estimate all the triangle regressions at the same time and considering contemporaneous correlations between all the regressions. This is an important practice issue, due to the relation between the cumulative payments in one column with the payments we do in the following columns. Having this extra information, allowed us to have the previous methods on a multivariate framework, reducing even further the prediction errors. The VP on a multivariate framework was once again confirmed as the one that minimizes the prediction error.

The tests performed to test the non-existence of correlations between the equations confirmed that we should reject this hypothesis, giving an indication of the need to consider multivariate regressions.

We also developed a general method for a portfolio of triangles. The previous framework was extended, and the results obtained confirmed once again the VP has the best method to minimize the prediction error. The portfolio method was presented with the previous univariate and multivariate framework and confirmed several things.

- We may have an improvement of the results if we consider several triangles at the same time.
- We may see the importance of having homogeneous triangles when we compared their results with the aggregate triangle result, where we put in one triangle all the other triangles: despite the lower level or reserves from the aggregate triangle the prediction error is higher when we compare it with the portfolio of triangles univariate or multivariate analysis.
- The correlations between triangles are important and even more when mixed with correlations between equations inside each regression.

Claims reserving techniques using regression models seem to confirm four things, which are important for modelling:

- Heteroscedasticity may exist in some triangles but does not seem to be an issue in most of the insurer's triangles. If it exists, is on irregular data triangles but in this case the method with heteroscedasticity may present a high prediction error.
- Serial correlation does not seem to be important in claims reserving, confirming the independence of origin years.
- Issues like correlations between equations and triangles seem to be important for claims reserving.
- Probably the most important issue in claims reserving is the method error, not considered in the prediction error calculations. Several variables, important to explain insurer's payments, are omitted from the methods. However, regression methods may be useful to fill this gap as it is very easy, on the methods that we developed, to have other explanatory variables, and not just the payments from the previous development year.

All these together seem to confirm the need for insurers and actuaries to switch from the traditional CL that has a fantastic history of implementation worldwide, to the VP, promising the best estimates that we need to manage and to know the health of an insurance company. The VP implementation done so far promises that.

Regression models are a useful tool to implement the VP in all the variants we saw in this thesis: univariate, multivariate, univariate portfolio or multivariate portfolio. They will also be useful in developing other claims reserving techniques: Bayesian regression and quantile regression may also play a role here, respectively, with LR methods and LRT when the link ratio is an ordinal statistic.

Bibliography

A.D.W. (1974), *Review of Claims Provisions for Non-Life Insurance Business*, Symposium Proceedings No. 3, Transactions of the Faculty of Actuaries, Vol. 34, No. 246

Acted (2016), *General Insurance Manual for 2016 Examinations on the Institute of Actuaries*, Acted

Alai, D., Merz, M. and Wüthrich (2009), *Mean Square Error of Prediction in the BF Claims Reserving Method*, Annals of Actuarial Science, Institute of Actuaries, Vol. 4, Part 1, pp 7-31

Alai, D., Merz, M. and Wüthrich (2010), *Prediction Uncertainty in the BF Claims Reserving Method*, Annals of Actuarial Science, Institute of Actuaries, Vol. 5, Part 1, pp 7-17

Antonio, K. and Plat, R. (2012), *Micro–level Stochastic Loss Reserving for General Insurance*, Leuven Catholic University, Faculty of Business and Economics

Ashe, F. (1986), *An Essay at Measuring the Variance of Estimates of Outstanding Claims Payments,* Astin Bulletin, April, pp 99-113

Baltagi, B. (2013), *Econometrics Analysis of Panel Data*, 5th Edition Wiley

Barnett, G. and Zehnwirth, B.(2000) *Best Estimates for Reserves* Casualty Actuarial Society, November 2000

Bardis, M., Majidi, A. and Murphy, D. (2012), *A family of chain-ladder factor models for selected link ratios*, Variance 6-2, pp 143 -160.

Beard, R. E. (1969), *Technical Reserves in Non-Life Insurance with Particular Reference to Motor Insurance*, Astin Bulletin Vol. 5, Part II

Beard, R. E. (1974), *Claims Provisions for Non-Life Insurance Business,* Institute of Financial Mathematics and its Applications

Benedikt, V. (1969), *Estimating Incurred Claims*, Astin Bulletin Vol.5, Part II

Benjamin, S. and Eagles, L. (1986), *Reserves in the Lloyd's and the London Market*, Journal of the Institute of Actuaries

Benktander, G. (1976), *An Approach to Credibility in Calculating IBNR for Casualty Excess Reinsurance,* The Actuarial Review, April 1976, Vol. 312-7.

Booth, P, Chadburn, R., Haberman, S., James, D., Khorasanee, Z., Plumb, R. and Rickayzen, B. (2005), *Modern Actuarial Theory and Practice*, Ed. Chapman & Hall/CRC second edition, ISBN 1-58488-368-5

Bowers, L. Gerber, H., Hickman, J., Jones, A and Nesbitt, C. (1986), *Actuarial Mathematics*, Ed. Society of Actuaries, ISBN 10: 0938959107

Braun, C. (2004), *The Prediction Error of the Chain-Ladder Method Applied to Run-Off Triangles*, Astin Bulletin, 34, pp 399-423

Brydon, D. and Verrall, R. 2009. *Calendar Year Effects, Claims Inflation and the Chain-Ladder Technique*, A.A.S. 4-II, pp 287-301

Bornhuetter, R. and Ferguson, R. (1972), *The Actuary and IBNR* Proceedings of the Casualty Actuarial Society, LIX, pp 181-195

Brehm, P. (2002), *Correlation and the aggregation of unpaid loss distributions*, CAS Forum Fall 2002, pp 1-23

Breusch, T. and Pagan, A. (1980), *Lagrange Multiplier Test and its Applications*, Review of Economic Studies, 47, pp 239-253

Brookshear, J. (2013), *Computer Science an Overview*, Ed. Addison-Wesley, ISBN: 978-0-13-256903-3

Brosius, Eric, *Loss Development Using Credibility*, Casualty Actuarial Society Part 7 Exam Study Kit, 1992.

Brown, R. (1993), *Introduction to Ratemaking and Loss Reserving for Property and Casualty Insurance*, Ed. Actex, ISBN 0-936031-11-5

Buhlmann, H., Schnieper, R. and Straub. E. (1980), *Claims reserves in Casualty Insurance based on a Probability Model*, Ed. Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker

Buhlmann, H. and Gisler, A. (2005), *A Course in Credibility Theory and its Applications*, Universitext, ISBN 3-540-25753-5

Christofides, S. (1997), *Regression Models based on Log-Incremental Payments*, Institute of Actuaries Claims Reserving Manual Vol 2

CAS (1996), *Foundations of Casualty Actuarial Science*, Third Edition, ISBN 0-9624762-0-X

CAS (2017), Accessed by the author on the month of July 2015 http://www.casact.org/admissions/syllabus/index.cfm?fa=5syllabi&parentID=163, [online]

Davis, H. (1941), *Analysis of Economic Time Series*, Bloomington

Denuit, M. and Charpentier, A. (2005), *Mathématiques de L'Assurance Non-Vie*, Vol. 2., Economica

De Vylder (1978), *Estimation of IBNR Claims by Least Squares*, Mitteilungen der Vereinigung Schwezerischer Versicherungsmathematiker, pp 249-254

De Vylder (1982), *Estimation of IBNR Claims by Credibility Theory*, Insurance Mathematics and Economics, 1, pp 35-40

De Jong, P. and Zehnwirth, B. (1983a), *Credibility Theory and the Kalman Filter*, Insurance Mathematics and Economics 2, pp 281-286

De Jong, P. and Zehnwirth, B. (1983b), *Claims Reserving, State Space Models and the Kalman Filter*, Journal of the Institute of Actuaries 110, pp 157-182

Efron, B. (1979), *Bootstrap Methods: Another Look at the Jackknife*, Annals of Statistics. 7-1

Efron, B. and Tibshirani, R. (1993), *An Introduction to Bootstrap*, Ed. Chapman and Hall, ISBN, 978-0-412-04231-7

England, P. and Verrall, R. (1998), *Standard Errors of Prediction in Claims Reserving: A Comparison of Methods*, Astin Colloquium, Glasgow

England, P. and R. Verrall (1999), *Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving*, Insurance Mathematics and Economics 25, pp. 281-293

England, P. (2002), *Addendum to "Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving," Insurance Mathematics and Economics* 31, 2002, pp. 461-466

England, P. and Verrall, R. (2002), Stochastic claims reserving in general insurance, *British Actuarial Journal* 8-3, pp 443-544

England, P., Verrall, R. and Wüthrich, M. (2012), *Bayesian Over dispersed Poisson Model and the BF Claim Reserving Method*, Annals of Actuarial Science, 6-2, pp. 258-283

European Community (1999) *Solvency Margin Review*, DIV 9049 (06/99)

European Union (2015), Delegated Regulation 2015/35 of 10 October 2014

European Union (2009), Directive 2009/138/CE of 25th November 2009

Fallquist, R. and Jones, B. (1987), *Loss Reserving in the Microcomputer Environment*, Astin Colloquium Speakers Corner, Tel Aviv

Fisher, W. and Lange, J. (1973), *Loss Reserve Testing: A Report Year Approach* Proceedings of the Casualty Actuarial Society, 60, pp 189-207

Fomby, T., Hill, R. and Johnson, S. (1984), *Advanced Econometric Methods*, 2nd Edition, Springer, New York.

Frees, E. (2010), *Regression Modelling with Actuarial and Financial Applications*, Ed. Cambridge University Press, ISBN 978-0-521-13596-2

Gentle, J. 2007. *Matrix Algebra - Theory, Computations, and Applications in Statistics*, Springer, New York.

Gisler, A. and Wüthrich, M. (2008), *Credibility for the Chain-Ladder Reserving Method*, Ed. Astin Bulletin, 38-2, pp 565-600.

Goovaerts, M., Kaas, R, Heerwarden, A. and Bauwelinckx (1990), *Effective Actuarial Methods*, Ed. North Holland Insurance Series 3, ISBN 0 444 88399 1.

Haastrup, S. & Arjas, E. (1996), *Claims Reserving in Continuous Time, a Nonparametric Bayesian approach*. ASTIN Bulletin, 26-2, pp 139-164

Hachemeister, C. (1975), *Credibility for Regression Models with Applications to Trend, in Credibility Theory with Applications,* Academic Press.

Hachemeister, C. and Stanard, J. (1975), *IBNR Claim Count Estimation with Static Lag Functions,* Astin Colloquium, Portimão, Portugal

Hachemeister, C.A. (1980), *A Stochastic Model for Loss Reserving*, Transactions of the 21st International Congress of Actuaries, pp 185-194

Halliwell, L. (2007), *Chain-Ladder Bias: Reasoning and Meaning*, Variance Journal, Vol. 1-2, pp 214-247

Harrington, S. and Niehaus, G. (1999), *Risk Management and Insurance*, Ed. McGraw-Hill International, ISBN 0-256-21018-7

Hess, K., Schmidt, K. and Zocher, M. (2006), *Multivariate Loss Prediction in the Multivariate Additive Model*, Insurance Mathematics and Economics 39, pp 185-191

Hindley, D. (2018), *Claims Reserving in General Insurance*, Cambridge University Press, ISBN 978 1 107 07693 8

Hilbe, J. (2011), *Negative Binomial Regression*, Cambridge University Press, 2nd Edition, ISBN 978 052 11 981 58

Hill, R, Griffiths, W. and Lim, G. (2012), *Principle of Econometrics*, Wiley, 4th Edition, ISBN 978 0-470-873 724

Hossack, I., Pollard, J. and Zehnwirth, B. (1993), *Introductory Statistics with Applicants in General Insurance*, Ed. Cambridge University Press, ISBN 0 521 28957 2

Halliwell, L. (1997), *Conjoint Prediction of Paid and Incurred Losses*, Casualty Actuarial Society Forum Summer (1), pp 241-379

Halliwell, L. (2007), *Chain-Ladder Bias: Its Reason and Meaning*, Casualty Actuarial Society Volume 01/Issue 02 (1), pp 214-247

Hess, K.T., Schmidt, K.D. and Zocher, M. (2006), Multivariate loss prediction in the multivariate additive model, *Insurance Mathematics and Economics* 39-2, pp 185 – 191.

Holmberg, R. (1994), *Correlation and Measurement of Loss Reserve Variability*, Casualty Actuarial Society Forum Spring, pp 247-277

Hovinen, E. (1981), *Additive and Continuous IBNR*, Astin Colloquium, Loen, Norway

Hurlimann (2006), *Approximate Bounds for the IBNR Reserves based on the Bivariate Chain-Ladder*, Belgian Actuarial Bulletin 5, pp 46-51

International Actuarial Association (2017), *Survey on Claims Reserving Methods*.

Institute of Actuaries (1989), *Claims Reserving Manual*, Vols. 1 and 2

Institute of Actuaries (1997), *Claims Reserving Manual*, Second Edition, Vols. 1 and 2

Jewell, W.S. (1989). *Predicting IBNR events and delays I: continuous time*. ASTIN Bulletin, 19, pp 25-55

Jewell, W.S. (1990). *Predicting IBNR events and delays II: discrete time*. ASTIN Bulletin, 20, pp 93-111

Johnson, P. and Hey, G. (1972) *Statistical Review of Motor Insurance Portfolio*, Astin Bulletin Vol VI, Part III

Kamreiter, H. and Straub, E. (1973), *On the calculation of IBNR reserves II*, Ed. Swiss Actuarial Journal, pp177-190.

Kirschner, G, Kerly, C and Isaacs, B. (2002) *Two Approaches to Calculating Correlated Reserve Indications across Lines of Business*, CAS Forum Fall 2002, pp 211-245

Kremer, E. (1982), *IBNR Claims and the Two-Way Model of Anova*, Scandinavian Actuarial Journal, pp 47-55

Kremer, E. (1984), *A Class of Autoregressive Models for Predicting the Final Claims Amount*, Insurance Mathematics and Economics, 3, 1984, pp 111-119

Kremer, E. (1985), *Einführung in die Versicherungsmathematik*, Vandenhoeck & Ruprecht, Göttingen & Zürich.

Kremer, W. (2005), *The correlated chain-ladder method for reserving in case of correlated claims developments*. Blätter *DGVFM* 27, pp 315 – 322.

Klugman, S. (1991), *Bayesian Statistics in Actuarial Science with Emphasis on Credibility*, Ed. Kluwer, ISBN 0-7923-9212-4

Kuang, D., Nielsen, B. and Nielsen, J.P. 2009. *Chain-Ladder as Maximum Likelihood Revisited*, Annals of Actuarial Science, 4-1, pp 105-121.

Kupper, J. (1967), *The Recent Developments of Risk Theory and its Applications*, Astin Bulletin, Vol IV, Part II

Larsen, C.R. (2007), *An Individual Claims Reserving Model*, ASTIN Bulletin, 37, pp 113-132

Lee, P. (1997), *Bayesian Statistics – An Introduction*, Ed. Arnold, ISBN 0 340 67785 6

Lindley, D. and Smith, A. (1972), *Bayes Estimates for the Linear Model*, Journal of the Royal Statistical Society, Series B, 35-1, pp 67-75

Lemaire, J., Mélard, G. and Vandermeulen, E. (1981), *Claims Reserves an Autoregressive Model*, Ed. Université Libre de Brusselles (unpublished)

Lowe, J. (1994), *A Practical Guide to Measuring Reserve Variability Using: Bootstrapping, Operational Time and a Distribution-Free Approach*, 1994 General Insurance Convention

Mack, T. (1990), *Improved Estimation of IBNR Claims by Credibility Theory*, Ed. Insurance Mathematics and Economics, 9, pp 51-57

Mack, T. (1991), *A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves*, Astin Bulletin, 2, pp 93-109

Mack, T. (1993a), *Distribution-Free Calculation of the Standard Error of the Chain Ladder Method Reserves Estimates*, Astin Bulletin, 23-2, pp 213-225

Mack, T. (1993b), *Measuring the Variability of Chain Ladder Reserve Estimates*, Casualty Actuarial Society meeting May 1993

Mack, T. (1994), *Which Stochastic Model is Underlying the Chain Ladder Method*, Ed. Insurance Mathematics and Economics 15, 2/3, pp 133-138

Mack, T. and Venter, G. (1999), *A Comparison of Stochastic Models that Reproduce the Chain-Ladder Reserve Estimates*, Tokyo Astin Colloquium Proceedings pp 263-273

Mack, T. (1999). *The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor*, ASTIN Bulletin, 29-2, pp 361-366

Mack, T. (2000), *Credible Reserves: The Benktander Method*, Astin Bulletin 30-2, pp 333-347

Mack, T. (2006), *Parameter Estimation for Bornhuetter/Ferguson*, Casualty Actuarial Society Forum, Fall 2006

Mack, T. (2008), *The Prediction Error of Bornhuetter/Ferguson*, Astin Bulletin, 38-1, pp 87-103

Marcuson, T. (2013), *Discussions on Triangle Free Reserving*, British Actuarial Journal ISSN 1357-3217

Masterson, N. (1962), *Problems in Motor Insurance - Claim Reserves*, Astin Bulletin, 2-1

Miranda, M, Nielsen, J. and Verrall, R. 2012. *Double Chain Ladder*, ASTIN Bulletin, 42-1, pp 59 - 76

Merz, M. and Wüthrich, M. (2006), *A credibility approach to the Munich chain-ladder reserving method*, Blätter *DGVFM* 27, pp 619-628

Merz, M. and Wüthrich, M. (2007), *Prediction error of the chain-ladder reserving method applied to correlated run-off triangles*, Annals of Actuarial Science 2-1, pp 25-50.

Merz, M. and Wüthrich, M. (2007a), *Prediction Error with Multivariate Additive Loss Reserving Method for Dependent Lines of Business*, CAS, Vol 3, Issue 1, pp 131-151

Merz, M. and Wüthrich, M. (2007b), *Prediction Error of the Multivariate Chain-Ladder Method*, Insurance Mathematics and Economics 42-1, pp 378-388

Merz, M. and Wüthrich, M. (2007c), *Prediction Error of the Multivariate Chain-Ladder Reserving Method*, North American Actuarial Journal, 12, pp 175-197.

Merz, M. and Wüthrich, M. (2008), *Prediction error of the multivariate chain-ladder reserving method*, North American Actuarial Journal 12-2, pp 175 – 197.

Merz, M. and Wüthrich, M. (2009), *Prediction error with multivariate additive loss reserving method for dependent lines of business*, Variance 3-1, pp 131 – 151.

Meyers, G. and Shi, P. (2011), *The Retrospective Testing of Stochastic Loss Reserving Methods*, CAS Forum

Mildenhall, S. (2006), *A Multivariate Bayesian Claim Count Development Model with Closed Form Posterior and Predictive Distributions*, CAS Forum Winter pp 451-493

Moitinho de Almeida, J.C (1971), *The Insurance Contract on Portuguese Law and Compared Law*, Livraria Sá da Costa

Murphy, D. (1994), *Unbiased Loss Development Factors*, Casualty Actuarial Society Proceedings

McCullagh, P. e Nelder, J. (1989), *Generalized Linear Models*, Ed. Chapman and Hall, ISBN 0 412 31760 5

Norberg, R. (1986), "*A Contribution to Modelling of IBNR Claims*" Scandinavian Actuarial, Journal, pp 155-203

Norberg, R. (1993). *Prediction of Outstanding Liabilities in Non-Life Insuranc*e. ASTIN Bulletin, 23-1, pp 95-115

O'Hagan, A. (1994), *Kendall's Advanced Theory of Statistics*, Edward Arnold, ISBN 0 340 52922 9

Oxford Dictionary (2014), Accessed by the author on the month of June 2014 *http://www.oxforddictionaries.com/definition/english/insurance?q=insurance*, [online] Parodi, P. (2014), *Triangle-free Reserving: A Non-Traditional Framework for Estimating Reserves and Reserve*, British Actuarial Journal, 19-1, pp 219-233

Pesaran, M. (2015), *Time Series and Panel Data Econometrics*, Oxford, ISBN 978-0-19875998-0

Pinheiro, P., Andrade e Silva, J. and Centeno, Lurdes (2003), *Bootstrap Methodology in Claims Reserving*, Ed. The Journal of Risk and Insurance, Vol 70, 4, pp 701-714

Portugal, L. (2007), *Gestão de Seguros Não-Vida*, Ed. IFA

Portugal, L. (2009) *Stochastic Claims Reserving with Negative Numbers*, MSc Thesis on Applied Mathematics at Heriot Watt

Portugal, L., Pantelous A.A., Assa, H. (2017) *Claims Reserving with a Stochastic Vector Projection*, North American Actuarial Journal, November 2017 https://www.tandfonline.com/doi/abs/10.1080/10920277.2017.1353429.

Portugal, L., Pantelous A.A., Assa, H. (2018) *Claims Reserving with a Stochastic Vector Projection*, North American Actuarial Journal, 2018 Vol 22-1, pp 22-39.

Prohl, C. and Schmidt, K. (2005), *Multivariate Chain-Ladder*, Dresdner Schriften zur Versicherungsmathematik

Pooser, D.M. and Walker P.L. (2015). *Own Risk and Solvency Assessment: Origins and Implications for Enterprise Risk Management*, Journal of Insurance Regulation, 34-9, pp 1-19.

Quarg, G. and Mack, T. (2004a), *Munich Chain-Ladder: A Reserving Method that Reduces the Gao between IBNR Projections Based on Paid Losses and IBNR Projections Based on Incurred Losses*, Variance, Casualty Actuarial Society, Vol. 2, 2, pp 266-299

Quarg, G., Mack, T. (2004b), *Munich chain ladder*, Blätter DGVFM 26, pp 597-630

Radtke, M, Schmidt, K. and Schnaus, A. (2010), *Handbook of Loss Reserving*, EEA Series, Ed. Springer.

Reid, D. H. (1987), *Claims Reserving in General Insurance*, Journal of the Institute of Actuaries, Vol. 105, Part III, pp 211-296

Rejda, G. (2005), *Principles of Risk Management and Insurance*, Addison-Wesley, ISBN 0-321-24846-5

Renshaw, A., (1989), *Chain-Ladder and Interactive Modelling*, Journal of the Institute of Actuaries, 116, pp 559-587

Renshaw, A. and Verrall, R. (1994), *A Stochastic Model Underlying the Chain-Ladder Technique*, Astin Colloquium Cannes

Renshaw, A. and Verrall, R. (1998), *A Stochastic Model Underlying the Chain-Ladder Technique*, British Actuarial Journal 4, IV

Rosenlund, S. (2012), *Bootstrapping Individual Claim Histories*, ASTIN Bulletin, Volume 42-1, pp 291 – 324

Saluz, A., Gisler, A., Wüthrich, M. (2011), *Development Pattern and Prediction Error for the Stochastic BF Claims Reserving Method,* Astin Bulletin, 41-2, pp 279-317

Searle, S. (1987) *Linear Models for Unbalanced Data*, Ed. John Wiley and Sons, ISBN 0-471-84096-3.

Shi, P. (2014), *A copula regression for modelling multivariate loss triangles and quantifying reserving variability*, ASTIN Bulletin 44-1, pp 85 – 102.

Shi, P. and Frees, E. W. (2011), *Dependent loss reserving using copulas*, ASTIN Bulletin 41-3, pp 449 - 486.

Schmidt, K. (2006a), *Methods and Models of Loss Reserving Based on Run-Off Triangles: A Unifying Survey,* CAS Forum Fall, pp 269-317

Schmidt, K. (2006b), *Optimal and Additive Loss Reserving for Dependent Lines of Business,* CAS Forum Fall, pp 319-351

Schmidt, K. and Zocher, M. (2008), *BF as a General Principle of Loss Reserving,* Astin Colloquium, Manchester

Schnieper, R. (1989), *A Pragmatic IBNR Method*, Astin Colloquium New York

Shi, P. and Hartman, B. (2014), *Credibility in Loss Reserving,* CAS Forum-Summer Vol. 2

Simon, L-J (1957) *Discussions*, Proceedings of the Casualty Actuarial Society, 44

Srivastava, V. and Giles, D. (1987), *Seemingly Unrelated Regression Equation Models,* Dekker.

Straub, E. (1971), *On the Calculation of IBNR Reserves*, Boleslaw Monic Fund, NGR Publication, Amsterdam pp 123-131

Straub, E. (1988), *Non-Life Insurance Mathematics*, Springer-Verlag, ISBN: 978-3-642-05741-0

Tarbell, T. (1934), *Incurred But Not Reported Claim Reserves*, Proceedings of the Casualty Actuarial Society, Vol. XX 275-280

Taylor, G. (1977), *Separation of Inflation and Other Effects from the Distribution of Non-Life Insurance Claims Delays*, Astin Bulletin, 9, pp 217-230

Taylor, G. (1978), *Regression Models in Claims Analysis I: Theory,* Casualty Actuarial Society, May 1987

Taylor, G. and Ashe, F. (1983), *Second Moments of Estimates of Outstanding Claims*, Journal of Econometrics, 23, pp 37-61

Taylor, G. (1986), *Claims Reserving in Non-Life Insurance*, North Holland, ISBN 0444878467.

Taylor, G. (1987), *Regression Models in Claims Analysis I: Theory*, Casualty Actuarial Society, pp 353-384.

Taylor. G. (1988), *Regression Models in Claims Analysis, Theory*, Proceedings of the Casualty Actuarial Society, 74, pp 354-383

Taylor, G. (2000), *Loss Reserving*, Ed. Kluwer Academic Publishers, ISBN 0-7923-8502-0

Taylor, G. (2000). *Loss Reserving: An Actuarial Perspective,* Huebner International Series on Risk, Insurance and Economic Security, Vol. 21, Springer New York, USA.

Taylor. G. (2002), Written Discussion of the Paper *"Stochastic Claims Reserving in General Insurance" by England, P. and Verrall, R.*, British Actuarial Journal, 8, Part III, pp 540-542

Taylor, G. and McGuire, G. (2007), *Synchronous Bootstrap to Account for Dependencies between Lines of Business in the Estimation of Loss Reserve Prediction Error*, North American Actuarial Journal, 11-2, pp 70-88

Taylor, G., McGuire, G., and Sullivan, J. (2008) *Individual Claim Loss Reserving Conditioned by Case Estimates*, Annals of Actuarial Science, 3, pp 215-256

Taylor, G. and McGuire, G. (2016), *Stochastic loss reserving using generalized linear models,* Casualty Actuarial Society

Turkman, M. (2000), *Modelos Lineares Generalizados da Teoria à Prática*, Ed. Sociedade Portuguesa de Estatística

Vaughan, E. and Vaughan, T. (1995), *Essentials of Insurance: A Risk Management Perspective*, Wiley ISBN 0-471-10758-1

Verrall, R. (1988), *Bayes Linear Model and the Claims Run-Off Triangle*, Actuarial Research Report, 7, City University of London

Verrall, R. (1989), *A State Space Representation of the Chain Ladder Model*, Journal of the Institute of Actuaries, 116, pp 589-609

Verrall, R. (1990), *Bayes and Empirical Bayes Estimation of the Chain Ladder Model*, Astin Bulletin, 20, pp 217-243

Verrall, R. (1991a), *On the Estimation of Reserves from Loglinear Models*, Insurance Mathematics and Economics, 10, pp 75-80

Verrall, R. (1991b), *Chain Ladder and Maximum Likelihood*, Journal of the Institute of Actuaries, 118, pp 489-499

Verrall, R. (1993) and Li, Z., *Negative Incremental Claims: Chain Ladder and Linear Models,* Journal of the Institute of Actuaries, 120, pp 171-183

Verrall, R. (1995), *Claims Reserving and Generalized Additive Models,* Astin Colloquium Leuven

Verrall, R. (2000), *An Investigation into Stochastic Claims Reserving Models and the Chain-Ladder Technique,* Insurance Mathematics and Economics, 26, pp 91-99

Verrall, R. (2001), *A Bayesian Generalized Linear Model for the BF Method of Claims Reserving,* Actuarial Research Paper, 139, City University of London

Verrall, R. (2004), *A Bayesian Generalized Linear Model for the BF Method of Claims Reserving,* North American Actuarial Journal, 8-3, pp 67-89

Verrall R., Brydon, D. (2009), *Calendar Year Effects, Claims Inflation and the Chain-ladder Technique, Annals of Actuarial Science*, 4-2, pp 287-30

Verbeek, H.G., (1972), *An Approach to the Analysis of Claims Experience in Motor Liability Excess Loss Reinsurance*, Astin Bulletin Vol. VI, Part III

White, H. (1980), *A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity*, Econometrica, 48, pp 1805-1813

Wright, T. (1990), *A Stochastic Method for Claims Reserving in General Insurance*, Journal of the Institute of Actuaries, 117, pp 677-731

Wright, T. (1992), *Stochastic Claims Reserving when Past Claim Numbers are Known,* Proceedings of the Casualty Actuarial Society, 151, pp 255-361

Wüthrich, M. and Merz, M. (2008), *Stochastic Claims Reserving Methods in Insurance,* Ed. Wiley, ISBN 978-0-470-72346-3

Wüthrich, M. and Merz, M. 2014. *Modified Munich Chain Ladder Method*, Swiss Finance Institute Research Paper nº 14-65

Wright, T.S. 1990. *A Stochastic Method for Claims Reserving in General Insurance*, Journal of the Institute of Actuaries, 117, pp 677 – 731

Zellner, A. (1962), *An Efficient method of estimating seemingly unrelated regressions equations and tests for aggregation bias*, Journal of the American Statistical Association 57, pp 348-368

Zellner, A., Huang, D.S. (1962), *Further properties of efficient estimators for seemingly unrelated regression equations*, International Economic Review 3-3, pp 300 – 313

Zellner, A. (1963), *Estimators for seemingly unrelated regressions equations: some exact finite sample results*, Journal of the American Statistical Association 58, pp 977-992

Zhang, Y. (2010), *A General Multivariate Chain-Ladder,* Insurance Mathematics and Economics. 46, pp 588-599

Zehnwirth, B. (1985), *Interactive Claims Reserving Forecasting System,* Insureware

Zehnwirth, B. (1989), *The Chain-Ladder Technique - A Stochastic Model,* Institute of Actuaries Claims Reserving Manual, vol. 2

Zweifel, P. and Eisen, R. (2012), *Insurance Economics,* Ed. Springer, ISBN 978-3-642-20547-7