

Reserving

Portfolio Claims Reserving with Univariate and Multivariate Generalized Link Ratios

Luis Portugal^a

Keywords: Reserving, Multivariate regression, Pooled data and portfolio data, Homoscedastic and heteroscedastic errors, Seemingly unrelated regression, Prediction errors

Variance

Vol. 16, Issue 2, 2023

Correlations between claims in insurers' lines of business may change results obtained on claims reserving. Here, we develop univariate and multivariate generalized link ratios for use with portfolio data, that is, estimating several triangles at the same time with correspondent correlations. Two options are presented, with the univariate case just considering correlations between triangles and not inside each triangle, and the multivariate case considering correlations between triangles and inside each triangle. Due to this extension, we call these methods as from portfolio data, because they consider more than one triangle at the same time. Existence/no-existence of correlations inside each triangle is called a multivariate/univariate approach with portfolio data. We also show that, from these methods, we may extract several others with different ways of estimating loss development factors. It is the case of heteroscedastic methods as chain ladder and simple average and homoscedastic ones as vector projection. Other heteroscedastic methods are also extracted with different levels of heteroscedasticity as the ones obtained for chain ladder and simple average. Some numerical examples, with standard data, are presented, and they confirm the use of vector projection as the solution that minimizes the prediction error, both for univariate and multivariate cases, and when using portfolio data.

Address for Correspondence: luis.portugal@actuarial.pt

1. INTRODUCTION

In this paper, we extend our previous article on univariate and multivariate claims reserving with generalized link ratios, Portugal, Pantelous, and Verrall (2021), to the case where we want to estimate reserves with several triangles at the same time. We call it portfolio claims reserving. In the literature (see, for example, Merz and Wüthrich 2007), this is called a multivariate approach. It happens due to the consideration of correlations between the triangles. Correlations between regressions appear in the data for several reasons, such as an increase/decrease of claims on some development years or an increase/decrease of the speed of payment of claims on certain development years.

We extend our previous article because the latter considered correlations inside triangles but not correlations between lines of business. This means we assume that claim's triangles are independent of each other. That is what an insurer does in practice, with the use of these triangles. However, there are situations where correlations exist between triangles.

For example, if comprehensive policies' claim payments increase/decrease, third-party liability claims may increase/decrease because they may share a common driver, for example, the country's economic activity may increase or decrease claims frequency. Another example comes from claims development: if we increase the speed of payment of claims in one line of business, that may also impact other

^a I started working in insurance in 1984 at the Portuguese Association of Insurers. Two years later I became Technical Director. In 1990 I moved to Grupo VITALCIO Portugal as CEO where I worked until 1995. I was President of the Portuguese Institute of Actuaries and in 1996 I founded ACTUARIAL Group, a set of companies specialized in actuarial services, software, outsourcing and training for insurers and pension funds. I am the Partner and Managing Director in Portugal, Dubai International Financial Centre, and Hong Kong.

For twenty years I was also a Lecturer of University of Barcelona and ASEA University of Bucharest, where I taught subjects as actuarial science and insurance companies management. I hold a PhD in Applied Mathematics applied to Actuarial Science from the University of Liverpool and several MSc's (Actuarial Science, Economics and Quantitative Finance). I also have graduate studies in Probabilities and Statistics, and I did the Portuguese exams on Non-Life Actuarial Science.

I wrote a book about Non-Life Insurance Management and have been publishing several articles at scientific journals and in the press.

lines of business, because the same claims department may manage both.

Correlations measure all these effects and are important in reserving because they allow us to extend our knowledge of total reserves. It is worth mentioning that IFRS 17, the insurance international financial reporting standard that became effective on 1 January 2023 (see IASB 2017), incentivizes insurers to consider correlations because they demand an economic balance sheet, considering all risks and diversification effects. As we will see, in the numerical examples provided, considering correlations may reduce the reserves level overall.

In our case, we are going to have correlations between each triangle regression and between the triangles estimated at the same time. This means that we may have correlations inside one triangle (Portugal, Pantelous, and Verrall 2021), giving a multivariate method, and when we do that with several triangles at the same time, we have multivariate methods with portfolio data. This is equivalent to what happens in econometrics, where they call the latter panel data; see, for example, Fomby, Johnson, and Hill (1984). Panel data always have correlations between different sets of data, as happens with insurers' triangles, which, by definition, must have correlations when estimated together. This means that we save the multivariate feature for the existence of correlations inside each triangle.

Multivariate papers in the literature (see Section 2) address this problem but with two limitations:

- only considering the chain-ladder method or the additive method, and for all the triangles, increasing the possibility of model error;
- not considering correlations inside each triangle, which may be important in practice due to the existence of correlations in triangle data.

Finally, considering all lines of business at the same time will allow us to have a prediction error for all portfolios instead of one prediction error per line of business. This is important because, when we work more than one triangle at the same time, we apply the same method to all triangles. In this paper we do that, but we do not restrict ourselves to one method. We will use the method from the link ratios family that minimizes the portfolio prediction error.

This study contributes to the reserving literature in the following *three* distinctive ways. Firstly, we develop the foundations of the *univariate* and *multivariate* generalized link ratios methods in the context of portfolio data with more than one triangle. Current literature usually restricts the approach to chain-ladder and does not consider correlations inside each triangle. Secondly, an analytical formula is presented for the prediction error, which is general to any generalized link ratio method and not restricted to chain-ladder. Finally, to demonstrate clearly the advantages of our approach, this paper contains an empirical investigation using *real* data. In this context, a comparison with Zhang (2010) is also presented.

The remainder of the paper is organized as follows. In Section 2, a brief overview of other trends in the existing literature on claims reserving with more than one triangle

is presented. Section 3 presents the portfolio generalized link ratios framework, which is developed in Section 4 with the univariate portfolio model and in Section 5 for the multivariate case. In Section 6, numerical examples are provided for both methods, and with the Zhang (2010) method. Finally, Section 7 contains the main conclusions.

2. MODELS IN THE LITERATURE WITH MORE THAN ONE TRIANGLE

The literature considers several models where authors calculate claims reserves using several triangles at the same time, known as multivariate, due to correlations between triangles. For example, Holmberg (1994), Halliwell (1997), Brehm (2002), Kirschner, Kerley, and Isaacs (2002), Quarg and Mack (2004), Merz and Wüthrich (2007) and Taylor and McGuire (2007) include cases where the development of one triangle might depend upon past information from other triangles. Braun (2004), Kremer (2005), Prohl and Schmidt (2005), Hess, Schmidt, and Zocher (2006), Schmidt (2006), and Merz and Wüthrich (2007, 2008) consider joint development with contemporaneous correlations among triangles. Finally, Zhang (2010) proposes a general multivariate chain-ladder method with contemporaneous correlations and structural connections among the triangles. Merz, Wüthrich, and Hashorva (2012) presents a multivariate approach using the lognormal distribution and also give a closed formula for claims uncertainty. As an alternative to correlations, some copula-based methodologies have been proposed (Shi and Frees 2011; Shi 2014).

Recently, Avanzi et al. (2020) have developed multivariate models with several triangles in the context of generalized linear models, and Winarta, Novita, and Nurrohman (2021) extends multivariate models to credibility theory.

However, these papers do not include correlations within *each* triangle, and restrict their use to one method (chain-ladder, additive, and some parametric approaches with probability distributions) and they usually do not present an analytical formula for the prediction error. Two examples where prediction errors are presented may be seen in Merz and Wüthrich (2007, 2008).

The present paper aims to overcome all these issues in estimating reserves for portfolio data (which means considering the data's correlations) but using generalized link ratios within a non-parametric context and presenting a general formula for prediction error, which may be applied to any generalized link ratio method, when considering correlations inside each triangle or not.

3. PORTFOLIO GENERALIZED LINK RATIOS FORMULATION

We extend here, to several triangles, the framework presented in Portugal, Pantelous, and Verrall (2021) for one triangle. Now we have data from $t = 1, \dots, N$ triangles. For each of these triangles, we have $k = 1, \dots, T - 1$ equations (regressions) and for each of these equations we have $T - k$ observations (years of origin).

Considering now $t = 1, \dots, N$ triangles with k equations for each triangle, 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 in column j from triangle i , $y_{t,i,j}$, are a function (a regression through the origin) of the claim's payments in 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 define $\beta_{t,j}$ as the slope (loss development factor) of j equation from triangle t . Also, each $\varepsilon_{t,i,j}$ is the error from year of origin i , development year j , and triangle t . For $t = 1, \dots, N$, $i = 1, \dots, T-j+1$, and $j = 2, \dots, T$, the cumulative payments dependent variable, $y_{t,i,j}$, is given by

$$y_{t,i,j} = \beta_{t,j} x_{t,i,j-1} + \varepsilon_{t,i,j}.$$

In matrix format we will have

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \quad (3.3)$$

where $\mathbf{Y}$ is the block-vector that includes the $\mathbf{Y}$ from each triangle case but now for N triangles. $\mathbf{Y}$ has dimensions $(m \times N) \times 1$, with generic $\mathbf{Y}_t$ for triangle t and $t = 1, \dots, N$,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \dots \\ \mathbf{Y}_N \end{bmatrix}.$$

$\mathbf{Y}_t = \begin{bmatrix} Y_{t,1} \\ \dots \\ Y_{t,T-1} \end{bmatrix}$ represents the dependent variables of the set of $T-1$ equations for the triangle t , where the generic equation $Y_{t,k} = \begin{bmatrix} y_{t,1,k+1} \\ \dots \\ y_{t,T-k,k+1} \end{bmatrix}$ includes the random variables $y_{t,i,k+1}$ for $t = 1, \dots, N$, $k = 1, \dots, T-1$, and $i = 1, \dots, T-k$.

$\mathbf{X}$ is defined by a diagonal block matrix. $\mathbf{X}$ has dimensions $(m \times N) \times [N \times (T-1)]$ and can be represented by

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{X}_N \end{bmatrix}.$$

$\mathbf{X}_t = \begin{bmatrix} X_{t,1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_{t,T-1} \end{bmatrix}$, where each generic element $X_{t,k} = \begin{bmatrix} x_{t,1,k} \\ \dots \\ x_{t,T-k,k} \end{bmatrix}$ 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 $i = 1, \dots, T-k$.

β is defined by a block-vector that includes the previous β from the one-triangle model, but now for N triangles. β has dimensions $[N \times (T-1)] \times 1$ and can be represented by

$$\beta = \begin{bmatrix} \beta_1 \\ \dots \\ \beta_N \end{bmatrix}.$$

$\beta_t = \begin{bmatrix} \beta_{t,1} \\ \dots \\ \beta_{t,T-1} \end{bmatrix}$, 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 $k = 1, \dots, T-1$.

ε is the block-vector that includes the ε from the one-triangle case, but now for N triangles. ε has dimensions $(m \times N) \times 1$. The ε from the one-triangle case is now ε_t and $t = 1, \dots, N$,

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_N \end{bmatrix},$$

$\varepsilon_t = \begin{bmatrix} \varepsilon_{t,1} \\ \dots \\ \varepsilon_{t,T-1} \end{bmatrix}$, where the generic $\varepsilon_{t,k} = \begin{bmatrix} \varepsilon_{t,1,k+1} \\ \dots \\ \varepsilon_{t,T-k,k+1} \end{bmatrix}$ includes the random variables $\varepsilon_{t,i,k+1}$ 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

$$\mathbf{Y}_F = \mathbf{X}_F \beta + \varepsilon_F$$

where $\mathbf{X}_F$ and ε_F are, respectively, the future values of $\mathbf{X}$ and the future errors. $\mathbf{Y}_F$ is a block-vector with size $(N \times m) \times 1$, given by

$$\mathbf{Y}_F = \begin{bmatrix} \mathbf{Y}_F^1 \\ \dots \\ \mathbf{Y}_F^N \end{bmatrix},$$

with each element $\mathbf{Y}_t^F = \begin{bmatrix} Y_{t,1}^F \\ \dots \\ Y_{t,T-1}^F \end{bmatrix}$ and the generic $\mathbf{Y}_{t,k}^F = \begin{bmatrix} y_{t,T-k+1,k+1}^F \\ \dots \\ y_{t,T,k+1}^F \end{bmatrix}$ for $t = 1, \dots, N$ and $k = 1, \dots, T-1$.

$\mathbf{X}_F$ 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

$$\mathbf{X}_F = \begin{bmatrix} \mathbf{X}_1^F & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{X}_N^F \end{bmatrix} = \begin{bmatrix} X_{1,1}^F & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_{N,T-1}^F \end{bmatrix}$$

where each element $\mathbf{X}_k^F = \begin{bmatrix} X_{t,1}^F & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_{t,T-1}^F \end{bmatrix}$ and $X_{t,k}^F = \begin{bmatrix} x_{t,T-k+1,k}^F \\ \dots \\ x_{t,T,k}^F \end{bmatrix}$ for $t = 1, \dots, N$ and $k = 1, \dots, T-1$.

ε_F is a block-vector with size $(N \times m) \times 1$ given by $\varepsilon_F = \begin{bmatrix} \varepsilon_{F1} \\ \dots \\ \varepsilon_{FN} \end{bmatrix}$,

with each element $\varepsilon_{Ft} = \begin{bmatrix} \varepsilon_{t,1}^F \\ \dots \\ \varepsilon_{t,T-1}^F \end{bmatrix}$ and the generic $\varepsilon_{t,k}^F = \begin{bmatrix} \varepsilon_{t,T-k+1,k+1}^F \\ \dots \\ \varepsilon_{t,T,k+1}^F \end{bmatrix}$ for $t = 1, \dots, N$ and $k = 1, \dots, T-1$.

The estimated values of the dependent variables are obtained from $\widehat{\mathbf{Y}}_F = \mathbf{X}_F \widehat{\beta}$.

4. PORTFOLIO UNIVARIATE GENERALIZED LINK RATIOS

4.1. ASSUMPTIONS

Having defined the framework of these methods, we present in this section the Portfolio Univariate Generalized Link Ratios (PUGLR) assumptions with the following proposition.

Proposition 4.1 Considering equation (3.3) we assume for PUGLR

$$\mathbb{E}(\varepsilon|\mathbf{X}) = \mathbb{E}(\varepsilon) = \mathbf{0} \quad (4.1.1)$$

$$\mathbb{E}(\varepsilon\varepsilon') = \sigma^2 \mathbf{W} = \mathbf{\Psi} \quad (4.1.2)$$

$$\mathbb{E}(\varepsilon_F \varepsilon_F') = \sigma^2 \mathbf{W}_F = \mathbf{\Psi}_F \quad (4.1.3)$$

where $\mathbf{0}$ is a vector of zeros of size $(N \times m) \times 1$, and $\mathbf{W}$ is an $(N \times m) \times (N \times m)$ diagonal weighting matrix, which depends on the parameter α in each non-zero cell. This parameter is related to the heteroscedasticity level in the triangles and will be crucial to identifying which link ratio method minimizes the prediction error; for more details, see Portugal, Pantelous, and Verrall (2021).

$\mathbf{W}$ is given by equation (4.1.4), where the *diag* operator transforms a vector into a diagonal matrix. $\mathbf{W}$'s diagonal elements are given by the elements of the transformed vector:

$$\begin{aligned} \mathbf{W} &= \text{diag}(\mathbf{X}^\alpha) \\ &= \begin{bmatrix} \text{diag}(\mathbf{X}_1^\alpha) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \text{diag}(\mathbf{X}_N^\alpha) \end{bmatrix} \\ &= \begin{bmatrix} x_{1,1,1}^\alpha & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{N,T-1,T-1}^\alpha \end{bmatrix}. \end{aligned} \quad (4.1.4)$$

The matrix $\mathbf{W}_F$ is the future $\mathbf{W}$ and has the same structure as $\mathbf{W}$. However, its elements are the $\mathbf{X}_F^\alpha$ instead of the $\mathbf{X}^\alpha$. $\mathbf{W}$ corresponds to a specific structure of heteroscedasticity through the choice of parameter α , and $\mathbf{W}_F$ has the same structure as $\mathbf{W}$ but is based on the predicted payments.

$$\begin{aligned} \mathbf{W}_F &= \text{diag}(\mathbf{X}) \\ &= \begin{bmatrix} \text{diag}(\mathbf{X}_{F1}^\alpha) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \text{diag}(\mathbf{X}_{FN}^\alpha) \end{bmatrix} \\ &= \begin{bmatrix} x_{1,T,1}^\alpha & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{N,T,T-1}^\alpha \end{bmatrix} \end{aligned} \quad (4.1.5)$$

σ^2 is diagonal block matrix with N blocks and of size $(N \times m) \times (N \times m)$, when expanded:

$$\sigma^2 = \begin{bmatrix} \sigma_{1,1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{N,N}^2 \end{bmatrix} \quad (4.1.6)$$

where each block $\sigma_{t,t}^2 = \text{diag} \begin{bmatrix} \sigma_{t,k}^2 \\ \vdots \\ \sigma_{t,T-1}^2 \end{bmatrix}$ for $t = 1, \dots, N$

and $k = 1, \dots, T-1$.

Seeing equation (4.1.5), we can understand that the method will be homoscedastic in each triangle when $\alpha = 0$. Otherwise, it will be heteroscedastic.

4.2 PARAMETERS' ESTIMATION

The following two lemmas allow us to have estimators for β and σ^2 .

Lemma 4.2.1 Following Fomby, Johnson, and Hill (1984), we can obtain the estimation of β , the loss development factors vector of all the equations from all the triangles. $\hat{\beta}$ is obtained using the Aitken generalized least squares method with $\mathbf{\Psi}$ as the weights matrix and is the best linear unbiased estimator of β ,

$$\hat{\beta} = (\mathbf{X}'\mathbf{\Psi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Psi}^{-1}\mathbf{Y} \quad (4.2.1)$$

The parameter α from equations (4.1.4) and (4.1.5) will be estimated as the value that minimizes the prediction error. This α parameter is a method choice parameter and, as in Portugal, Pantelous, and Verrall (2021), we select the model with the lowest prediction error.

Lemma 4.2.2 Following Srivastava and Giles (1987), we can estimate $\hat{\sigma}_{t,k}^2$ using equation k (from triangle t)'s sum of the square of the errors, $SSR_{t,k}$, divided by this equation's degrees of freedom, which is the number of observations $T_{t,k}$ for this equation, minus the number of parameters in the equation, in this case one:

$$\hat{\sigma}_{t,k}^2 = \frac{SSR_{t,k}}{T_{t,k} - 1} \quad (4.2.2)$$

4.3 PREDICTION ERROR

We also need an expression for the prediction error, which is given by the following theorem.

Theorem 4.3.1 Knowing that the prediction error (i.e., the root of the mean square error) is given by the root of the expected value of $\widehat{\mathbf{Y}}_F - \mathbf{Y}_F$ and its transpose, $E(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)'$, we get the mean square error prediction (MSEP) from the following expression:

$$\mathbb{E} \left[\mathbf{X}_F'(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \mathbf{X}_F' \right] + \mathbb{E}(\varepsilon_F \varepsilon_F') \quad (4.3.1)$$

The estimation of the variance is given by $\mathbf{X}_F'(\mathbf{X}'\mathbf{\Psi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Psi}^{-1}\mathbb{E}(\varepsilon\varepsilon')\mathbf{X}(\mathbf{X}'\mathbf{\Psi}^{-1}\mathbf{X})^{-1}\mathbf{X}_F'$

and the process variance comes from $\mathbb{E}(\varepsilon_F \varepsilon_F')$.

Altogether, this means that the MSEP will be obtained from

$$\mathbf{X}_F'(\mathbf{X}'\mathbf{\Psi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Psi}^{-1}\mathbb{E}(\varepsilon\varepsilon')\mathbf{X}(\mathbf{X}'\mathbf{\Psi}^{-1}\mathbf{X})^{-1}\mathbf{X}_F' + \mathbb{E}(\varepsilon_F \varepsilon_F') \quad (4.3.2)$$

Proof. This can be obtained by following the same steps as are presented in Portugal, Pantelous, and Verrall (2021) for the single-triangle case.

Proposition 4.3.1 Following expression (4.3.2), and assumptions (4.1.2) and (4.1.3), the MSEP is obtained from

$$\mathbf{X}_F'(\mathbf{X}'\mathbf{\Psi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{\Psi}^{-1}\mathbf{X})^{-1}\mathbf{X}_F' + \mathbf{\Psi}_F \quad (4.3.3)$$

As with univariate generalized link ratios and multivariate generalized link ratios, see Portugal, Pantelous, and Verrall (2021), and following the results of Theorem 4.3.1, we need to know which weighting matrices we are using, namely $\mathbf{\Psi}$ and $\mathbf{\Psi}_F$. For the latter, we need the parameter α and we can

obtain it by searching for the α that minimizes the prediction error presented in expression (4.3.3).

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 that the way Ψ is defined will provide us with several claims reserving methods for estimating the loss development factors.

Analytically, we get several portfolio data methods: Vector Projection (VP), see Portugal, Pantelous, and Assa (2017) for $\alpha = 0$, chain-ladder (CL) for $\alpha = 1$, simple average (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 Ψ 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 the α that minimizes the prediction error for N triangles at the same time. With α different from 0, 1, or 2, we would get other methods: the optimal choice for the weights of the link ratios is obtained as the prediction error is minimized.

All the link ratios methods considered here (see the next section for special cases) depend on α , which represents the level of heteroscedasticity, and we want to choose the α that minimizes the prediction error. The lower is the prediction error, the better are the errors analysis and back-testing results; see Portugal, Pantelous, and Assa (2017).

4.4. SPECIAL CASES

Special cases of the method are considered with the next three corollaries. Obviously, the proofs of these corollaries are linked to Theorem 4.3.1 and Proposition 4.3.1. Thus, they are omitted.

Corollary 4.4.1 If $\alpha = 0$, the triangle's variances are homoscedastic and, looking at expressions (4.1.2) and (4.1.3), we get

$$\begin{aligned}\mathbb{E}(\varepsilon\varepsilon') &= \sigma^2 \mathbf{I}_{(N \times m) \times (N \times m)} = \Psi_{VP} \\ \mathbb{E}(\varepsilon_F \varepsilon_F') &= \sigma^2 \mathbf{I}_{(N \times m) \times (N \times m)} = \Psi_{VP,F}\end{aligned}$$

Here, $\mathbf{I}_{(N \times m) \times (N \times m)}$ is a diagonal identity matrix with size $(N \times m) \times (N \times m)$. With $\alpha = 0$, the loss development factors are the ones from the VP applied with a portfolio context, that is, the portfolio vector projection (PVP) (see expression (4.2.1)), where Ψ_{VP} is Ψ with $W = \mathbf{I}_{(N \times m) \times (N \times m)}$. Then, the MSEF is obtained from

$$\begin{aligned}\mathbb{E}(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)' (\widehat{\mathbf{Y}}_F - \mathbf{Y}_F) \\ = \mathbf{X}_F' (\mathbf{X}' \Psi_{VP}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} (\mathbf{X}' \Psi_{VP}^{-1} \mathbf{X}) \mathbf{X}'_F \\ + \Psi_{VP,F}\end{aligned} \quad (4.4.1)$$

(where $\Psi_{VP,F}$ is Ψ with $\mathbf{W}_F = \mathbf{I}_{(N \times m) \times (N \times m)}$).

Corollary 4.4.2 If $\alpha = 1$, the triangle's variances are heteroscedastic, and we get

$$\begin{aligned}\mathbb{E}(\varepsilon\varepsilon') &= \sigma^2 \mathbf{W}_{CL} = \Psi_{CL} \\ \mathbb{E}(\varepsilon_F \varepsilon_F') &= \sigma^2 \mathbf{W}_{CL,F} = \Psi_{CL,F}\end{aligned}$$

with

$$\mathbf{W}_{CL} = \begin{bmatrix} x_{1,1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{N,T-1,T-1} \end{bmatrix} \quad \text{and} \quad \mathbf{W}_{CL,F} = \begin{bmatrix} x_{1,T,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{N,T,T-1} \end{bmatrix}.$$

With $\alpha = 1$, the loss development factors are the ones from the CL applied in a portfolio context, that is the portfolio chain-ladder (PCL), (see (4.2.1)), where Ψ_{CL} is Ψ with $W = \mathbf{W}_{CL}$. Then, the MSEF is obtained from

$$\begin{aligned}\mathbb{E}(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)' (\widehat{\mathbf{Y}}_F - \mathbf{Y}_F) \\ = \mathbf{X}_F' (\mathbf{X}' \Psi_{CL}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} (\mathbf{X}' \Psi_{CL}^{-1} \mathbf{X}) \mathbf{X}'_F + \Psi_{CL,F}\end{aligned} \quad (4.4.2)$$

(where $\Psi_{CL,F}$ is Ψ with $\mathbf{W}_F = \mathbf{W}_{CL,F}$).

Corollary 4.4.3 If $\alpha = 2$, the triangle's variances are heteroscedastic, and we get

$$\begin{aligned}\mathbb{E}(\varepsilon\varepsilon') &= \sigma^2 \mathbf{W}_{SA} = \Psi_{SA} \\ \mathbb{E}(\varepsilon_F \varepsilon_F') &= \sigma^2 \mathbf{W}_{SA,F} = \Psi_{SA,F}\end{aligned}$$

with

$$\mathbf{W}_{SA} = \begin{bmatrix} x_{1,1,1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{N,T-1,T-1}^2 \end{bmatrix} \quad \text{and} \quad \mathbf{W}_{SA,F} = \begin{bmatrix} x_{1,T,1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{N,T,T-1}^2 \end{bmatrix}$$

With $\alpha = 2$, the loss development factors are the ones from the SA applied in a portfolio context, i.e., the portfolio simple average (PSA), see (4.2.1), where Ψ_{SA} is Ψ with $W = \mathbf{W}_{SA}$.

Then, the MSEF is obtained from

$$\begin{aligned}\mathbb{E}(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)' (\widehat{\mathbf{Y}}_F - \mathbf{Y}_F) \\ = \mathbf{X}_F' (\mathbf{X}' \Psi_{SA}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} (\mathbf{X}' \Psi_{SA}^{-1} \mathbf{X}) \mathbf{X}'_F + \Psi_{SA,F}\end{aligned} \quad (4.4.3)$$

(where $\Psi_{SA,F}$ is Ψ with $\mathbf{W}_F = \mathbf{W}_{SA,F}$).

5. PORTFOLIO MULTIVARIATE GENERALIZED LINK RATIOS

In this section, we develop the Section 4 method for the case where there are contemporaneous correlations between equations inside the same triangle and between triangles. The method considered here is the same as that presented in Section 4. However, we will change the assumptions, introducing a more complex structure for the errors: seemingly unrelated regression (SUR) (Srivastava and Giles 1987). Our method will become multivariate as an SUR and may also use the heteroscedastic structure from the generalized link ratios, also including VP, CL, and SA. In this portfolio multivariate generalized link ratios (PMGLR) method, we are going to maintain the entire framework presented in Section 4 but change assumptions (4.1.2) and (4.1.3).

5.1. ASSUMPTIONS

We are going to assume contemporaneous correlations between the errors of the different equations and between

the triangles. To do that, we obtain a portfolio multivariate method.

Σ is a block matrix of block-size $[N \times (T-1)] \times [N \times (T-1)]$ that summarizes the variances and covariances between the $k = 1, \dots, T-1$ regressions in each of the $t = 1, \dots, N$ triangles, and between each triangle, for observations in the same origin year. Expanding each block, we get a matrix with dimensions $(N \times m) \times (N \times m)$,

$$\Sigma = \begin{bmatrix} \Sigma_{1,1,1} & \cdots & \Sigma_{N,1,T-1} \\ \vdots & \ddots & \vdots \\ \Sigma_{1,T-1,1} & \cdots & \Sigma_{N,T-1,T-1} \end{bmatrix} \quad (5.1.1)$$

The generic component of (5.1.1), $\Sigma_{t,k,k}$ is given by a matrix of size $[N \times (T-k)] \times [N \times (T-k)]$ and by $s_{t,k,k}$, the variance parameter from triangle t and regression k :

$$\Sigma_{t,k,k} = s_{t,k,k} \begin{bmatrix} x_{1,1,k}^\alpha & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{N,T-k,k}^\alpha \end{bmatrix} /$$

The generic component of (5.1.1), $\Sigma_{t,k,j}$ with $k \neq j$, is given by a matrix of size $[N \times (T-k)] \times [N \times (T-j)]$ and by $s_{t,k,j}$, the covariance parameter for triangle t between regressions k and j , where $k \neq j$:

$$\Sigma_{t,k,j} = s_{t,k,j} \mathbf{I}_{N \times (T-k)}. \quad (5.1.3)$$

Σ^F is a block matrix of block-size $[N \times (T-1)] \times [N \times (T-1)]$ that summarizes the future variances and covariances between the $k = 1, \dots, T-1$ regressions. Expanding each block, we get a matrix of dimensions $(N \times m) \times (N \times m)$,

$$\Sigma^F = \begin{bmatrix} \Sigma_{1,1,1}^F & \cdots & \Sigma_{N,1,T-1}^F \\ \vdots & \ddots & \vdots \\ \Sigma_{1,T-1,1}^F & \cdots & \Sigma_{N,T-1,T-1}^F \end{bmatrix}. \quad (5.1.4)$$

The generic component of (5.1.4), $\Sigma_{t,k,k}^F$ is given by a matrix of size $[N \times (T-k)] \times [N \times (T-k)]$,

$$\Sigma_{t,k,k}^F = s_{t,k,k} \begin{bmatrix} x_{1,T,k}^\alpha & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{N,T+k,k}^\alpha \end{bmatrix}. \quad (5.1.5)$$

The generic component of (5.1.4), $\Sigma_{t,k,j}^F$ with $k \neq j$, is given by a matrix of size $[N \times (T-k)] \times [N \times (T-j)]$,

$$\Sigma_{t,k,j}^F = s_{t,k,j} \mathbf{I}_{N \times (T-k)}. \quad (5.1.6)$$

Proposition 5.1 Considering (3.3) we assume, for the PMGLR method,

$$\mathbb{E}(\varepsilon|\mathbf{X}) = \mathbb{E}(\varepsilon) = 0 \quad (5.1.7)$$

$$\mathbb{E}(\varepsilon\varepsilon') = \Sigma \quad (5.1.8)$$

$$\mathbb{E}(\varepsilon_F\varepsilon_F') = \Sigma_F \quad (5.1.9)$$

5.2. PARAMETER ESTIMATION

The parameter estimation can be obtained from the following Lemma 5.2.1.

Lemma 5.2.1 (Srivastava and Giles 1987) We can obtain the estimation of β , that is the estimation of the loss development factors from all the equations. $\hat{\beta}$ is obtained using the SUR generalized least squares for panel data with heteroscedasticity and contemporaneous correlations, and is the best linear unbiased estimator of β ,

$$\hat{\beta} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}. \quad (5.2.1)$$

We also need an expression for the prediction error, which will be given in the following paragraphs. Clearly, the parameters $s_{j,j}$ and $s_{i,j}$ are not known and must be estimated. Thus, with the following Lemma, the estimators $\hat{s}_{j,j}$ and $\hat{s}_{i,j}$ are provided.

Lemma 5.2.2 (Srivastava and Giles 1987) Estimators for the parameters of the variance and covariance matrix, from a multivariate regression with panel data, are given by

$$\hat{s}_{t,k,k} = \frac{1}{T-1} SSR_{t,k} \quad \hat{s}_{t,k,j} = \frac{1}{T} SSR_{t,k} \quad (5.2.2)$$

The $SSR_{t,k}$ are to be calculated using, for each equation t , the regression k ordinary least squares (OLS) sum of the square of the errors.

5.3 PREDICTION ERROR

The following theorem follows from Theorem 4.3.1 and gives us a general analytical formula for obtaining the prediction error.

Theorem 5.3.1 The MSEF for the method presented in (3.3) and Proposition 5.1 is obtained from

$$\mathbf{X}_F(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbb{E}(\varepsilon\varepsilon')\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}_F' + \mathbb{E}(\varepsilon_F\varepsilon_F') \quad (5.3.1)$$

The proof follows directly from Theorem 4.3.1 when (5.1.7), (5.1.8), and (5.1.9) are considered.

Following on from the results of Theorem 5.3.1, the procedures are like those of the univariate method, presented in the previous section. In the PMGLR, we need to obtain the following:

- The $\hat{s}_{t,j,j}$ and $\hat{s}_{t,i,j}$ to estimate the Σ matrix, which implies the need to have a first regression, with OLS, to get the sum of the square of the errors.
- The parameter α so as to have the Σ (5.11) and Σ_F (5.14) matrices. Our suggestion, as in Portugal, Pantelous, and Verrall (2021), is to choose the α that minimizes the prediction error.

This will also give us the vector of the loss development factors, given by (5.2.1) and with that we will have $\mathbf{X}_F$. Having Σ and Σ_F , we then have $\mathbb{E}(\varepsilon\varepsilon')$ and $\mathbb{E}(\varepsilon_F\varepsilon_F')$ and we can calculate the prediction error.

Proposition 5.3.1 Using (5.3.1) and the assumptions from Proposition (5.1), the MSEF is obtained from

$$\mathbb{E}(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)' = \mathbf{X}_F(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}_F' + \Sigma_F. \quad (5.3.2)$$

5.4. SPECIAL CASES

As in the univariate method in Section 4, we choose the α that minimizes the prediction error. Analytically, we no longer obtain the loss development factors from VP, for $\alpha = 0$, 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 (4.2.1), which is different from expression (5.2.1). However, we can 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. This is due to the heteroscedasticity level. What defines and differentiates these three methods is the weights given to the

link ratios, which defines the heteroscedasticity level. In VP, it is zero, with $\alpha = 0$, in CL it is one, with $\alpha = 1$, and in SA it is two, with $\alpha = 2$.

As with the univariate portfolio data method, we can obtain other methods that give other weights to the link ratios, through the selection of α . As with the univariate method from Section 4, the optimal α is the one that minimizes the prediction error.

Corollary 5.4.1 If $\alpha = 0$, the variances are homoscedastic, and the regressions correlated. We get

$$\begin{aligned}\mathbb{E}(\varepsilon\varepsilon') &= \Sigma_{VP} \\ \mathbb{E}(\varepsilon_F\varepsilon_F') &= \Sigma_{VP,F}\end{aligned}$$

Σ_{VP} and $\Sigma_{VP,F}$ are the Σ defined in expressions (5.1.1) and (5.1.4), with the following relations,

$$\begin{aligned}\Sigma_{t,j,j} &= s_{t,j,j} \mathbf{I}_{N \times (T-j)} \\ \Sigma_{t,l,j} &= s_{t,l,j} \mathbf{I}_{N \times (T-j)}\end{aligned}$$

Here, $\mathbf{I}_{N \times (T-j)}$ is a diagonal identity matrix with size $N \times (T-j)$. With $\alpha = 0$, the loss development factors are the ones from the VP within a portfolio multivariate context (PMVP). Also, $\Sigma = \Sigma_{VP}$.

Then, the MSEF comes from

$$\begin{aligned}\mathbb{E}(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)' \\ = \mathbf{X}_F(\mathbf{X}'\Sigma_{VP}^{-1}\mathbf{X})^{-1}\mathbf{X}_F' \\ + \Sigma_{VP,F}.\end{aligned}\quad (5.4.1)$$

Corollary 5.4.2 If $\alpha = 1$, the variances are heteroscedastic, and the regressions correlated. We get

$$\begin{aligned}\mathbb{E}(\varepsilon\varepsilon') &= \Sigma_{CL} \\ \mathbb{E}(\varepsilon_F\varepsilon_F') &= \Sigma_{CL,F}\end{aligned}$$

The Σ_{CL} and $\Sigma_{CL,F}$ are respectively the Σ and Σ_F defined in expression (5.1.1) and (5.1.4) with

$$\begin{aligned}\Sigma_{t,j,j} &= s_{t,j,j} \begin{bmatrix} x_{1,1,j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{t,T-j,j} \end{bmatrix} \\ \Sigma_{t,j,j}^F &= s_{t,j,j} \begin{bmatrix} x_{1,T,j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{t,T+j,j} \end{bmatrix}\end{aligned}$$

With $\alpha = 1$, the loss development factors are the ones from the CL within a portfolio multivariate context (PMCL).

Then, the MSEF comes from

$$\begin{aligned}\mathbb{E}(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)' \\ = \mathbf{X}_F(\mathbf{X}'\Sigma_{CL}^{-1}\mathbf{X})^{-1}\mathbf{X}_F' + \Sigma_{CL,F}\end{aligned}\quad (5.4.2)$$

Corollary 5.4.3 If $\alpha = 2$, the variances are heteroscedastic, and the regressions correlated. We get

$$\begin{aligned}\mathbb{E}(\varepsilon\varepsilon') &= \Sigma_{SA} \\ \mathbb{E}(\varepsilon_F\varepsilon_F') &= \Sigma_{SA,F}\end{aligned}$$

Σ_{SA} and $\Sigma_{SA,F}$ are respectively the Σ and Σ_F defined in expressions (5.1.1) and (5.1.4) with

$$\Sigma_{t,j,j} = s_{t,j,j} \begin{bmatrix} x_{1,1,j}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{t,T-j,j}^2 \end{bmatrix}$$

$$\Sigma_{t,j,j}^F = s_{t,j,j} \begin{bmatrix} x_{1,T,j}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{1,T+j,j}^2 \end{bmatrix}$$

With $\alpha = 2$, the loss development factors are the ones from the SA within a portfolio multivariate context (PMSA).

Then, the MSEF comes from

$$\begin{aligned}\mathbb{E}(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)(\widehat{\mathbf{Y}}_F - \mathbf{Y}_F)' \\ = \mathbf{X}_F(\mathbf{X}'\Sigma_{SA}^{-1}\mathbf{X})^{-1}\mathbf{X}_F' + \Sigma_{SA,F}\end{aligned}\quad (5.4.3)$$

6. APPLICATION TO STANDARD DATA

We consider for the numerical results three paid claims triangles from the literature. We call them triangle 1, triangle 2, and triangle 3:

- triangle 1, Mack (1993);
- triangle 2, Taylor and Ashe (1983);
- triangle 3, Taylor and McGuire (2016).

The results obtained, once again, see Portugal et al. (2017; Portugal, Pantelous, and Verrall 2021), confirm VP as the method that minimizes the prediction error. We present results for the PUGLR (see Section 4) and PMGLR (see Section 5). We also compare these results with those obtained from an aggregated triangle, those we get if we apply generalized link ratios methods separately to each triangle, and those obtained using the multivariate chain-ladder from Zhang (2010).

6.1. PORTFOLIO UNIVARIATE GENERALIZED LINK RATIOS

The results obtained are presented in Table 6.1. The α that minimized the prediction error was zero, confirming once again that VP was the best solution, according to this criterion. The prediction error obtained was 8.9% and the total reserves estimated 18 896 187.

Had we considered just one triangle that corresponded to the sum of the three triangles, the results would have been those presented in Table 6.2.

When compared with the Table 6.1 results, the aggregated triangle results are similar: the prediction error increases to 9.4% and the reserves decrease to 18 804 158. The difference seems small, but we must be aware that triangle 2 has far greater reserves than the other two.

As expected, the reserves obtained in Table 6.1 correspond to the sum of the reserves from the three triangles when the generalized link ratios method is applied to each triangle; see Table 6.3. The prediction error increases.

This same level of total reserves was obtained because we did not consider any correlations between triangles (nor between equation 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 of the prediction errors of the three triangles. The weights are the estimated reserves.

The individual results from the three triangles are presented in Table 6.4. Here, we can see that the 9.0% predic-

Table 6.1. Portfolio univariate 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%

Table 6.2. 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%

Table 6.3. Portfolio generalized link ratios results - Totals from the three 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%

tion error obtained is a weighted average of those for 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 errors from all the triangles (in monetary units) (see [Table 6.4](#)) is equal to the same indicator obtained from [Table 6.1](#):

$$1\,707\,793 = 15\,651 + 1\,675\,147 + 16\,995$$

6.2. PORTFOLIO MULTIVARIATE GENERALIZED LINK RATIOS

For the PMGLR, we also found that, to obtain the lowest prediction error, $\alpha = 0$. The prediction error of 2.7% represents an important decrease relative to the PUGLR result

Table 6.4. Portfolio generalized link ratios results from the three 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%

(8.9% with $\alpha = 0$). The reserves increase to 23 619 959 (they were 18 896 187 with PUGLR). See [Table 6.5](#) for the PMGLR results and [Table 6.1](#) for PUGLR.

Using an aggregate triangle would decrease the reserves to 19 889 001, but with an important increase in the prediction error to 5.3%, is shown in [Table 6.6](#).

The reason for the increase in the reserves level between the PUGLR and the PMGLR lies in the change of the loss development factors, β , mainly those for triangle 2.

Several loss development factors increase and some of them decrease, but the increase of 5% from that corresponding to $j = 9$ has a 5% impact on all the ultimate factors from all the origin years, and justifies the increase in the reserves of around 25%.

The change in the loss development factors is a consequence of the change in the weights matrix, as the latter is now considering the contemporaneous correlations between the triangles. The changes in these factors are presented in [Table 6.7](#).

Table 6.5. 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%

Table 6.6. 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%

Now we compare our results with those obtained when applying Zhang (2010)'s multivariate chain-ladder.

As we can see, the reserves decrease significantly from 23 616 959 to 18 974 544. However, this is due to a lack of fit between the chain-ladder and this data. Indeed, the prediction error increases from 2.7% to 14.3%.

Finally, we present another result for a variant of the PMGLR: we assume that there are correlations between triangles but no correlations between each triangle's equations. This should correspond to the PUGLR method. However, the results will be different from those of the PUGLR. The reason is that the PMGLR methods estimate the correlations between triangles in a different way; see Lemmas 4.2.1 and 5.2.1. Despite this, we may compare the PMGLR results with this calculation. We can see in [Table 6.9](#) that the prediction error increases from 2.7% (see [Table 6.5](#)) 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.

Comparing now with Zhang (2010), with correlations inside each triangle, to obtain comparable results with the ones presented above, we can see that the reserves difference is much smaller, from 19 114 443 to 18 974 545,

which is reflected by the prediction errors being close, 6.4% against 8.0%.

7. CONCLUSIONS

From these methods, using portfolio data, we obtained several conclusions in respect of the theory and from the numerical examples given.

Firstly, it is straightforward to move from one triangle to several triangles, when regression techniques are considered.

Secondly, prediction error formulas come very easily from well-known regression models, using the same approach as for the one-triangle case, adapted to several triangles.

Thirdly, considering several triangles allow us to estimate generalized link ratios adapted to our data. Using prediction error minimization helps us to have more accuracy in the reserves calculation for all triangles. Otherwise, important errors might arise due to the method leverage effect when it is applied to several triangles without accuracy.

Fourthly, the use of a portfolio of triangles confirms the use of VP as the solution that minimizes the prediction error.

Table 6.7. Changes in loss development factors with PMGLR

Triangle	Loss Development Factors		
	Number	Number	Variation
1	1	1	99,1%
1	1	2	12,8%
1	1	3	17,8%
1	1	4	-8,1%
1	1	5	-38,6%
1	1	6	-19,2%
1	1	7	-6,1%
1	1	8	6,1%
1	1	9	0,1%
2	2	1	-4,4%
2	2	2	0,9%
2	2	3	-9,1%
2	2	4	6,0%
2	2	5	7,8%
2	2	6	4,7%
2	2	7	-0,3%
2	2	8	-2,3%
2	2	9	5,0%
3	3	1	-19,8%
3	3	2	19,0%
3	3	3	-2,9%
3	3	4	-0,9%
3	3	5	1,7%
3	3	6	0,8%
3	3	7	0,0%
3	3	8	-0,2%
3	3	9	0,2%

Table 6.8. Zhang model results

Column	Reserves per Column	Prediction Error	Prediction Error %
2	886 549	222 622	25%
3	1 947 950	594 171	31%
4	3 380 539	1 015 477	30%
5	2 503 361	1 293 832	52%
6	2 150 766	976 630	45%
7	2 304 673	1 142 572	50%
8	1 824 189	938 833	51%
9	3 086 871	840 964	27%
10	889 646	611 639	69%
Total	18 974 544	2 707 360	14.3%

ror. This happens in both the univariate (PUGLR) and multivariate (PMGLR) case.

Fifthly, the use of data shows that differences may arise between the models considered. When we use the PMGLR,

the prediction error decreases when compared to the PUGLR. It seems that it is worth working with more information to predict the reserves. However, using such information also produces an increase in the level of reserves,

Table 6.9. PMGLR with no correlations between the 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%

Table 6.10. Zhang model results with correlations inside each triangle

Column	Reserves per Column	Prediction Error	Prediction Error %
2	886 549	222 622	25%
3	1 947 950	319 591	16%
4	3 380 540	600 569	18%
5	2 503 361	552 862	22%
6	2 150 766	694 711	32%
7	2 304 673	682 973	30%
8	1 824 189	212 400	12%
9	3 086 871	473 315	15%
10	889 646	494 062	56%
Total	18 974 545	1 510 083	8.0%

due to the correlations between triangles. This shows that the latter have a role to play in claims reserving.

Sixthly, as expected, the level of reserves is not the same as 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 us-

ing homogeneous triangles in claims reserving, even if the prediction error is low.

Submitted: October 06, 2022 EDT. Accepted: March 14, 2023 EDT. Published: June 22, 2023 EDT.

REFERENCES

- Avanzi, Benjamin, Greg Taylor, Phuong Anh Vu, and Bernard Wong. 2020. "A Multivariate Evolutionary Generalised Linear Model Framework with Adaptive Estimation for Claims Reserving." *Insurance: Mathematics and Economics* 93 (July):50–71. <https://doi.org/10.1016/j.insmatheco.2020.04.007>.
- Braun, Christian. 2004. "The Prediction Error of the Chain Ladder Method Applied to Correlated Run-off Triangles." *ASTIN Bulletin* 34 (2): 399–423. <https://doi.org/10.2143/ast.34.2.505150>.
- Brehm, P. 2002. "Correlation and Aggregation of Unpaid Loss Distributions." *CAS Forum Fall*, 1–24.
- Fomby, Thomas B., Stanley R. Johnson, and R. Carter Hill. 1984. *Advanced Econometric Methods*. New York: Springer New York. <https://doi.org/10.1007/978-1-4419-8746-4>.
- Halliwell, L. 1997. "Conjoint Prediction of Paid and Incurred Losses." *CAS Forum*, 241–379.
- Hess, Klaus T., Klaus D. Schmidt, and Mathias Zocher. 2006. "Multivariate Loss Prediction in the Multivariate Additive Model." *Insurance: Mathematics and Economics* 39 (2): 185–91. <https://doi.org/10.1016/j.insmatheco.2006.02.004>.
- Holmberg, R. 1994. "Correlation and Measurement of Loss Reserve Variability." *CAS Forum Spring*, 247–77.
- IASB. 2017. *IFRS 17 Insurance Contracts*. International Accounting Standards Board.
- Kirschner, G., C. Kerley, and B. Isaacs. 2002. "Two Approaches to Calculating Correlated Reserves Indications across Multiple Lines of Business." *CAS Forum Fall*, 211–46.
- Kremer, W. 2005. "The Correlated Chain-Ladder Method for Reserving in Case of Correlated Claims Developments." *Blatter DGVFM* 27 (2): 315–22.
- Mack, Thomas. 1993. "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates." *ASTIN Bulletin* 23 (2): 213–25. <https://doi.org/10.2143/ast.23.2.2005092>.
- Merz, Michael, and Mario V. Wüthrich. 2007. "Prediction Error with Multivariate Additive Loss Reserving Method for Dependent Lines of Business." *Variance* 3 (1): 131–51.
- . 2008. "Prediction Error of the Multivariate Chain Ladder Reserving Method." *North American Actuarial Journal* 12 (2): 175–97. <https://doi.org/10.1080/10920277.2008.10597509>.
- Merz, Michael, Mario V. Wüthrich, and Enkelejd Hashorva. 2012. "Dependence Modelling in Multivariate Claims Run-off Triangles." *Annals of Actuarial Science* 7 (1): 3–25. <https://doi.org/10.1017/s1748499512000140>.
- Portugal, Luís, Athanasios A. Pantelous, and Hirbod Assa. 2017. "Claims Reserving with a Stochastic Vector Projection." *North American Actuarial Journal* 22 (1): 22–39. <https://doi.org/10.1080/10920277.2017.1353429>.
- Portugal, Luís, Athanasios A. Pantelous, and Richard Verrall. 2021. "Claims Reserving with Univariate and Multivariate Generalized Link Ratios." *Insurance: Mathematics and Economics* 97 (March):57–67. <https://doi.org/10.1016/j.insmatheco.2020.11.011>.
- Prohl, C., and K. Schmidt. 2005. *Multivariate Chain-Ladder*. Report Lehrstuhl für Versicherungsmathematik, Technische Universität Dresden.
- Quarg, Gerhard, and Thomas Mack. 2004. "Munich Chain Ladder." *Blätter Der DGVFM* 26 (4): 597–630. <https://doi.org/10.1007/bf02808969>.
- Schmidt, K. 2006. "Optimal and Additive Loss Reserving for Dependent Lines of Business." *CAS Forum*, 319–51.
- Shi, Peng. 2014. "A Copula Regression for Modelling Multivariate Loss Triangles and Quantifying Reserving Variability." *ASTIN Bulletin* 44 (1): 85–102. <https://doi.org/10.1017/asb.2013.23>.
- Shi, Peng, and E.W. Frees. 2011. "Dependent Loss Reserving Using Copulas." *ASTIN Bulletin* 41 (3): 449–86.
- Srivastava, V., and D. Giles. 1987. *Seemingly Unrelated Regression Equation Models: Estimation and Inference*. Vol. 80. Boca Raton, FL: CRC Press.
- Taylor, Greg, and F.R. Ashe. 1983. "Second Moments of Estimates of Outstanding Claims." *Journal of Econometrics* 23 (1): 37–61. [https://doi.org/10.1016/0304-4076\(83\)90074-x](https://doi.org/10.1016/0304-4076(83)90074-x).

Taylor, Greg, and G. McGuire. 2016. *Stochastic Loss Reserving Using Generalized Linear Models*. Casualty Actuarial Society (CAS) Monograph Series 3. Virginia, USA.

Taylor, Greg, and Gráinne McGuire. 2007. "A Synchronous Bootstrap to Account for Dependencies Between Lines of Business in the Estimation of Loss Reserve Prediction Error." *North American Actuarial Journal* 11 (3): 70–88. <https://doi.org/10.1080/10920277.2007.10597467>.

Winarta, V T, M Novita, and S Nurrohmah. 2021. "Multivariate Bühlmann-Straub Credibility Model for Claim Reserving." *Journal of Physics: Conference Series* 1725 (1): 012026. <https://doi.org/10.1088/1742-6596/1725/1/012026>.

Zhang, Yanwei. 2010. "A General Multivariate Chain Ladder Model." *Insurance: Mathematics and Economics* 46 (3): 588–99. <https://doi.org/10.1016/j.insmatheco.2010.03.002>.

SUPPLEMENTARY MATERIALS

PDF Supplement

Download: <https://variancejournal.org/article/77867-portfolio-claims-reserving-with-univariate-and-multivariate-generalized-link-ratios/attachment/163368.pdf>
