Universiti Teknologi MARA Cawangan Kedah Kampus Sungai Petani

2021 ICMS INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

Multivariate Evolutionary and Tweedie GLM Methods for Estimating Estimation of Motor Vehicle Insurance Claims Reserves

Adhitya Ronnie Ronnie Effendie

Universitas Gadjah Mada

4 - 5 AUGUST 2021

CONTENTS

Introduction

An insurance policy is a contract agreement between the policyholder and the insurance company. For the contract agreement to run, policyholders need to pay premiums to insurance companies. On the other hand, the insurance company must underwrite the risk if the policyholder does submission of claims. It is necessary to estimate the reserves of claims for the company insurance accurately to prepare several funds for settlement of claim.

Generalised Linear Model (GLM) can be used to estimate the claim values in a univariate form which **only consists of 1 LoB (Line of Bussiness). In practice**, almost every **insurance company has various types of LoB** which depends on one another. Therefore, the GLM can be expanded to a **multivariate GLM** which can be used **to estimate the claim data with more than one LoB**. The researcher also wants to compare between an estimated reserve calculations of **Swiss Re Group's** claims using the **Multivariate Evolutionary GLM Adaptive Simple Method** and **GLM with the Tweedie Family Distribution** Approach to find a more accurate method of finding claim reserves for each line of Swiss Re Group's business data.

2. Methods

1. Multivariate Evolutionary GLM Framework
2. Adaptive Estimation

2.1 MULTIVARIATE EVOLUTIONARY GLM FRAMEWORK

2.1.1. Specifications
2.1.2. State Space Matrix Representation

2.1.1 Specifications

- **Loss reserving data is represented by using the Loss Triangle**
- **Each loss triangle represents a corresponding line of business (LoB).**
- **Each triangle contains a set of incremental claims** $Y_{i,j}^{(n)}$ $\forall i, j, n \in \mathbb{N}$ **where** $i = (1, I), j = (1, J)$ **and** $n = (1, N)$ **each symbol represents an accident year, development year and LoB.**
- **Claim symbolized by** $Y_{i,j}^{(n)}$ **are included in the calendar year** $t = i + j - 1$ **with** $t \in \mathbb{N}$**,** $t = (1, T)$**.**
- **The multivariate evolutionary GLM framework consists of 2 components, namely: an observation component and a state component.**

2.1.1.1 Observation component

This observation component determines the relationship between observations and random factors. Observation components are determined by means and variances as follows:

$$E\left[Y_{i,j}^{(n)}\right] = \mu_{i,j}^{(n)} \qquad (2.1) \qquad\qquad Var\left[Y_{i,j}^{(n)}\right] = \phi^{(n)}.V\left(\mu_{i,j}^{(n)}\right) \qquad \textbf{(2.2)}$$

where $\mu_{i,j}^{(n)}$ is the mean parameter, $\phi^{(n)}$ is the dispersion parameter, and $V(.)$ is the corresponding variance function. For the Tweedie family distribution, apply:

$$V\left(\mu_{i,j}^{(n)}\right) = \left(\mu_{i,j}^{(n)}\right)^{p} \qquad \textbf{(2.3)}$$

where *p* is the power parameter that maps to the Tweedie family distribution. Therefore, this distribution is useful in reserve modeling because it allows uncertainty in the model via *p* estimation.

The mean structure relates observations to a set of explanatory factors.

The evolutionary reserve uses a modified Hoerl curve that allows the use of the calendar year effect (systematic change in claims activity over time).

Using the log link function on the Hoerl Curve, the mean structure can be written as follows:

$$\boldsymbol{log\left(\mu_{i,j}^{(n)}\right) = a_i^{(n)} + r_i^{(n)}.log(j) + s_i^{(n)}.j + h_t^{(n)},} \qquad \textbf{(2.4)}$$

where $a_i^{(n)}$ **is accident year factor,** $r_i^{(n)}$ **and** $s_i^{(n)}$ **are factors of Hoerl curve that determine the pattern of development of the *i*-th accident year, and** $h_t^{(n)}$ **is a factor of the calendar year.**

One special case of the multivariate evolutionary GLM framework is the multivariate Gaussian model which allows calendar year dependencies between LoBs and the observation component:

$$Y_{i,j}^{(n)} = a_i^{(n)} + r_i^{(n)}.\log(j) + s_i^{(n)}.j + h_t^{(n)} + \varsigma_{i,j}^{(n)}$$

$$\varsigma_{i,j}^{(n)} \sim Normal(0, \sigma_{\varsigma^{(n)}}^2) \qquad \textbf{(2.5)}$$

2.1.1.2 State component

This state component determines the recursive evolution of the random factor.

The observation component discussed earlier is a standard structure for GLM in reserves that uses the same fixed parameter values used to capture the development year effect (eg $r_i^{(n)} = r^{(n)}, s_i^{(n)} = s^{(n)}$in terms of the Hoerl curve) in all accident years.

The factors in the evolutionary multivariate GLM framework, namely $a_i^{(n)}$, $r_i^{(n)}$, $s_i^{(n)}$, and $h_t^{(n)}$are random and continue to evolve over time. As a result, each accident year has its own pattern of development. The evolution of each accident year can be determined using a random process.

The evolution of $a_i^{(n)}$, $r_i^{(n)}$ and $s_i^{(n)}$ is as follows:

$$a_i^{(n)} = a_{i-1}^{(n)} + {}_a\epsilon_i^{(n)} \qquad {}_a\epsilon_i^{(n)} \sim Normal\left(0, \sigma^2_{{}_a\epsilon^{(n)}}\right), \quad (2.6)$$

$$r_i^{(n)} = r_{i-1}^{(n)} + {}_r\epsilon_i^{(n)} \qquad {}_r\epsilon_i^{(n)} \sim Normal\left(0, \sigma^2_{{}_r\epsilon^{(n)}}\right), \quad (2.7)$$

$$s_i^{(n)} = s_{i-1}^{(n)} + {}_s\epsilon_i^{(n)} \qquad {}_s\epsilon_i^{(n)} \sim Normal\left(0, \sigma^2_{{}_s\epsilon^{(n)}}\right), \quad (2.8)$$

where $\sigma^2_{{}_a\epsilon_i^{(n)}}, \sigma^2_{{}_r\epsilon_i^{(n)}}, \sigma^2_{{}_s\epsilon_i^{(n)}}$ is the variance of the respective errors ${}_a\epsilon_i^{(n)}$, ${}_r\epsilon_i^{(n)}$, ${}_s\epsilon_i^{(n)}$ in evolution that must be estimated.

The evolution of the calendar factor $h_t^{(n)}$ is determined by a random process that has been modified to involve interdependencies between calendar years via a common shock approach :

$$\boxed{h_t^{(n)} = h_{t-1}^{(n)} + {}_h\epsilon_t^{(n)} + \lambda^{(n)} \cdot {}_h\tilde{\epsilon}_t} \qquad {}_h\epsilon_t^{(n)} \sim Normal\left(0, \sigma^2_{{}_h\epsilon^{(n)}}\right),$$

$${}_h\tilde{\epsilon}_t \sim Normal\left(0, \sigma^2_{{}_h\tilde{\epsilon}}\right)$$

There are 2 sources of disturbance in this evolution, namely line specific disturbance ${}_h\epsilon_t^{(n)}$ and common shock disturbance ${}_h\tilde{\epsilon}_t$.

The evolution of the random factor in 1 row of the run-off triangle can be described as shown in Figure 2.1 on the side with the assumption that the evolution occurs in the AY dimension, which means that the random factor develops and its estimate is updated when it moves to the next row of the run-off triangle.

The black arrow illustrates evolution, while the red arrow illustrates a one-to-one mapping with AY and DY. So that the claim data in the first accident year will be the beginning of the adaptive estimation process which will be discussed further in 2.2.

The random factor $a_i^{(n)}, r_i^{(n)}, s_i^{(n)}$ will develop from one accident year to the next accident year. Meanwhile, the calendar factor $h_t^{(n)}$ grows diagonally from one calendar year to the next

The calendar factors $h_1^{(n)}, h_2^{(n)}, \dots, h_t^{(n)}$ are mapped one by one to the calendar factors in the next row $h_2^{(n)}, \dots, h_t^{(n)}$ and so for the next line. This is because claims on the same diagonal in the run-off triangle have the same calendar year effect.

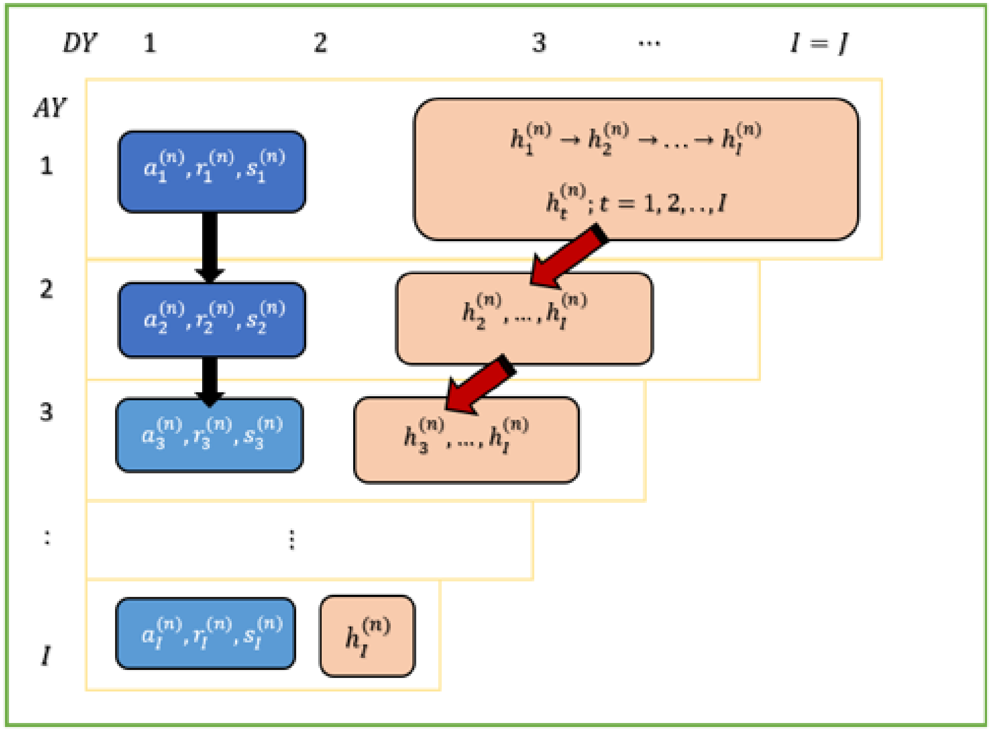


Figure 2.1 Evolution of the random factors in each row of the run-off triangle

Figure 2.2 on the side is a conclusion from the material in section 2.1.1 which is the structure of the framework for the multivariate evolutionary GLM, namely:

- ✓**The parameters that are the basis of the calculation, are constant and become the basis for determining the evolution of the random factor as the observation progresses.**
- ✓**The random factor develops over time and is an explanatory factor from observation.**
- ✓**The observation is in the form of incremental claim data**

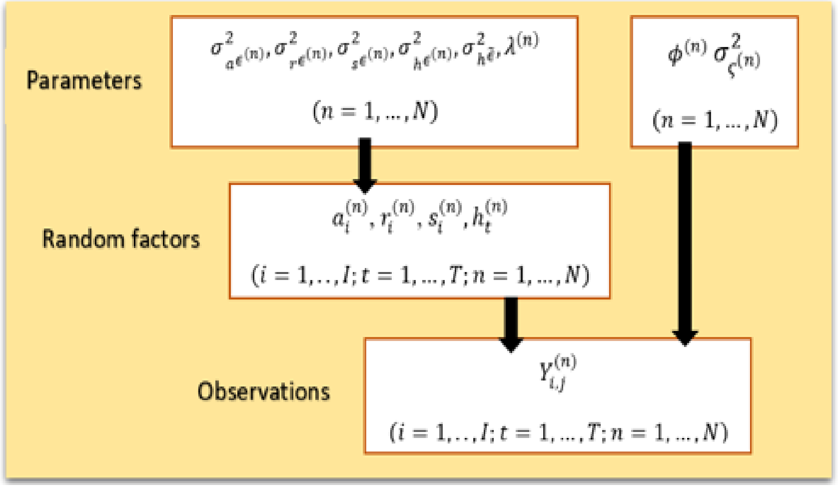


Figure 2.2 Conclusion of framework structure for multivariate evolutionary GLM

2.1.2. State Space Matrix Representation

Representation of the relationship between observations, random factors and the development of random factors over time.

2.1.2.1 observation component

In this framework, claim reserving data can be thought of as a multivariate time series process. Every time there are new observations, each accident year will be the basis for calculating this process. The observation vector in the accident year *i* or Y_i is the vector of all $N.(I - i + 1)$ data claims in the same accident year in the run-off triangle.

$$Y_i = \left.\begin{pmatrix} Y_{i,1}^{(1)} \\ \vdots \\ Y_{i,I-i+1}^{(1)} \\ Y_{i,1}^{(2)} \\ \vdots \\ Y_{i,I-i+1}^{(N)} \end{pmatrix}\right\} (I - i + 1)\ rows$$

with the mean and dispersion :

$$E[Y_i] = \mu_i = \begin{pmatrix} \mu_{i,1}^{(1)} \\ \vdots \\ \mu_{i,I-i+1}^{(1)} \\ \mu_{i,1}^{(2)} \\ \vdots \\ \mu_{i,I-i+1}^{(N)} \end{pmatrix} \qquad \phi = \begin{pmatrix} \left.\begin{matrix} \phi^{(1)} \\ \vdots \\ \phi^{(1)} \end{matrix}\right\} (I-i+1)\, baris \\ \phi^{(2)} \\ \vdots \\ \phi^{(N)} \end{pmatrix} \qquad (2.11)$$

The mean structure is calculated using a linear predictor with a log-link as below

$$log(\mu_i) = A_i . \gamma_i + E_i . \psi_I$$

with the mean structure in equation (2.4) it can be obtained :

$$\gamma_i^{(n)} = \begin{pmatrix} a_i^{(n)} \\ r_i^{(n)} \\ s_i^{(n)} \end{pmatrix}, \qquad \gamma_i = \begin{pmatrix} \gamma_i^{(1)} \\ \vdots \\ \gamma_i^{(N)} \end{pmatrix} \qquad (2.13)$$

$$A_i^{(n)} = \begin{pmatrix} 1 & \log(1) & 1 \\ 1 & \log(2) & 2 \\ \vdots & \vdots & \vdots \\ 1 & \log(I-i+1) & I-i+1 \end{pmatrix}, \qquad (2.14)$$

$$E_i^{(n)} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \Big\} (I-i+1)\, baris$$

$$(i-1) kolom \qquad (I-i-1)\; kolom$$

$$\psi_I^{(n)} = \begin{pmatrix} h_1^{(n)} \\ \vdots \\ h_I^{(n)} \end{pmatrix}, \qquad \psi_I = \begin{pmatrix} \psi_I^{(1)} \\ \vdots \\ \psi_I^{(N)} \end{pmatrix} \qquad A_i = \begin{pmatrix} A_i^{(1)} & 0 & \dots & 0 \\ 0 & A_i^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_i^{(N)} \end{pmatrix}$$

$$E_i = \begin{pmatrix} E_i^{(1)} & 0 & \dots & 0 \\ 0 & E_i^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_i^{(N)} \end{pmatrix} \qquad (2.15)$$

For the special case of the multivariate Gaussian model, the matrix form of the observation equation is

$$Y_i = A_i . \gamma_i + E_i . \psi_i + \varsigma_i, \qquad \varsigma_i \sim Normal(0, H_i)$$

where

$$\varsigma_i = \begin{pmatrix} \varsigma_{i,1}^{(1)} \\ \vdots \\ \varsigma_{i,I-i+1}^{(1)} \\ \varsigma_{i,1}^{(2)} \\ \vdots \\ \varsigma_{i,I-i+1}^{(N)} \end{pmatrix} \qquad (2.17)$$

$$\mathrm{H}_i^{(n)} = \begin{pmatrix} \sigma^2_{\varsigma^{(n)}} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varsigma^{(n)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{\varsigma^{(n)}} \end{pmatrix} \qquad \mathrm{H}_i = \begin{pmatrix} H_i^{(1)} & 0 & \cdots & 0 \\ 0 & H_i^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_i^{(N)} \end{pmatrix}$$

$(I - i - 1)\ kolom$

2.1.2.2 STATE component

The random evolution of γ_i can be represented in the form of a matrix as

$$\gamma_i = \gamma_{i-1} + {}_{\gamma}\epsilon_i, \qquad {}_{\gamma}\epsilon_i \sim Normal\,(0, Q_{\gamma^\epsilon})$$

where

$${}_{\gamma}\epsilon_i^{(n)} = \begin{pmatrix} {}_{a}\epsilon_i^{(n)} \\ {}_{r}\epsilon_i^{(n)} \\ {}_{s}\epsilon_i^{(n)} \end{pmatrix}, \qquad {}_{\gamma}\epsilon_i = \begin{pmatrix} {}_{\gamma}\epsilon_i^{(1)} \\ \vdots \\ {}_{\gamma}\epsilon_i^{(N)} \end{pmatrix} \qquad Q_{\gamma^\epsilon}^{(n)} = \begin{pmatrix} \sigma^2_{a^{\epsilon(n)}} & 0 & 0 \\ 0 & \sigma^2_{r^{\epsilon(n)}} & 0 \\ 0 & 0 & \sigma^2_{s^{\epsilon(n)}} \end{pmatrix} \qquad Q_{\gamma^\epsilon} = \begin{pmatrix} Q_{\gamma^\epsilon}^{(1)} & 0 & \cdots & 0 \\ 0 & Q_{\gamma^\epsilon}^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q_{\gamma^\epsilon}^{(N)} \end{pmatrix}$$

The evolution of the calendar year factors is as follows

$$\psi_t = R_{t-1}.\psi_{t-1} + S_{t-1}.\,{}_{h}\epsilon_t, \qquad {}_{h}\epsilon_t \sim Normal(0, Q_{h^\epsilon})$$

where

$$\psi_t^{(n)} = \begin{pmatrix} h_1^{(n)} \\ \vdots \\ h_t^{(n)} \end{pmatrix}, \quad \psi_t = \begin{pmatrix} \psi_t^{(1)} \\ \vdots \\ \psi_t^{(N)} \end{pmatrix}, \quad {}_{h}\epsilon_t = \begin{pmatrix} {}_{h}\epsilon_t^{(1)} \\ \vdots \\ {}_{h}\epsilon_t^{(N)} \\ {}_{h}\tilde{\epsilon}_t \end{pmatrix}$$

$$R_{t-1}^{(n)} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}}_{(t-1)\ columns} \left.\vphantom{\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}}\right\} t\ rows \qquad R_{t-1} = \begin{pmatrix} R_{t-1}^{(1)} & 0 & \cdots & 0 \\ 0 & R_{t-1}^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{t-1}^{(N)} \end{pmatrix}$$

$$Q_{h^{\epsilon}} = \begin{pmatrix} \sigma^2_{h^{\epsilon(1)}} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \sigma^2_{h^{\epsilon(N)}} & 0 \\ 0 & \cdots & \cdots & \sigma^2_{h^{\epsilon(1)}} \end{pmatrix}$$

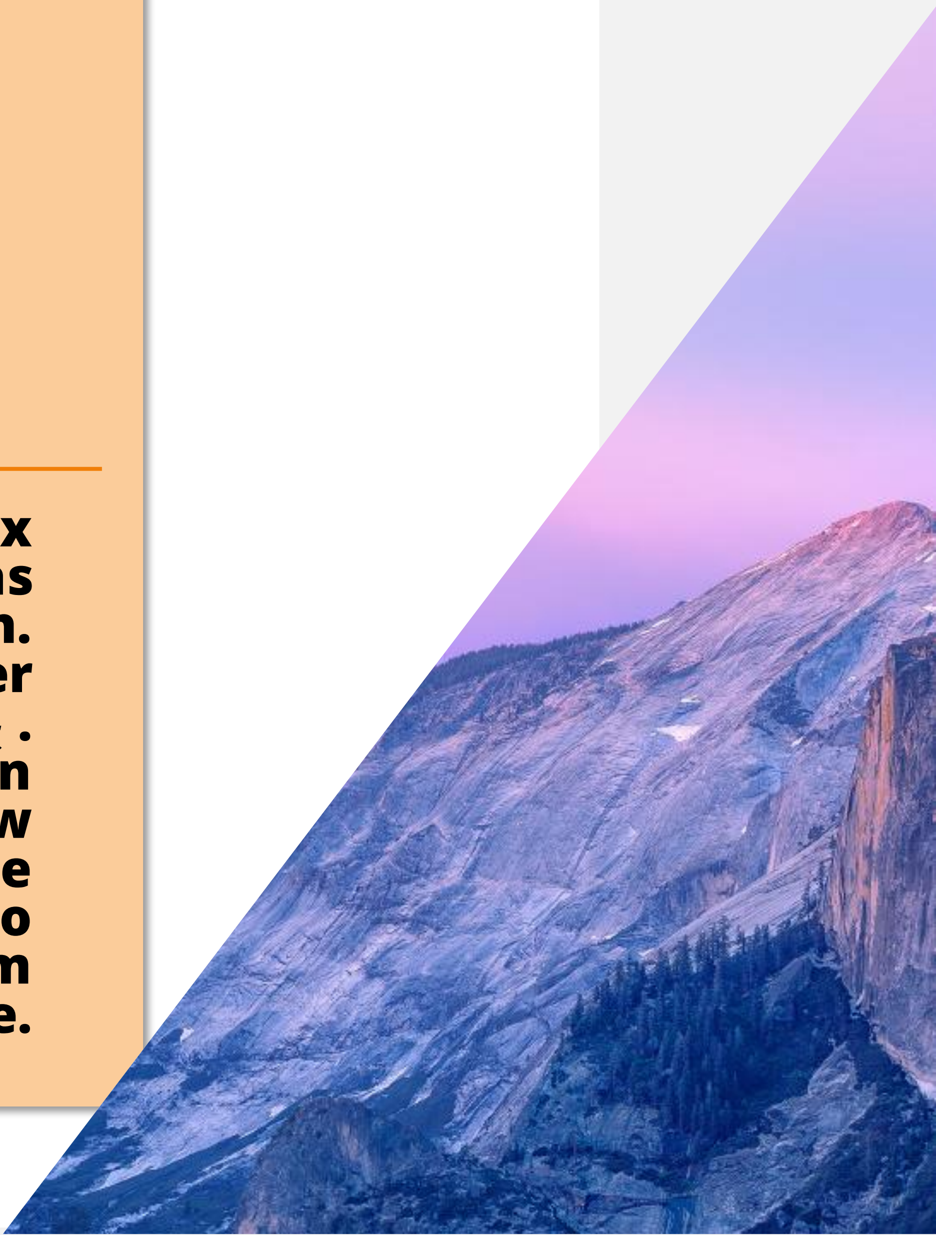
$$\Lambda^{(n)} = \left.\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \lambda^{(n)} \end{pmatrix}\right\} \text{t rows}$$

$$S^{(n)}_{t-1} = \left.\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}\right\} \text{t rows}$$

$$S_{t-1} = \begin{pmatrix} S^{(1)}_{t-1} & 0 & \cdots & 0 & \Lambda^{(1)} \\ 0 & S^{(2)}_{t-1} & \cdots & 0 & \Lambda^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & S^{(N)}_{t-1} & \Lambda^{(N)} \end{pmatrix}$$

2.2 ADAPTIVE ESTIMATION

Representation of space matrix circumstances from this framework used as developtment from approximation of estimation. That random factor including year non-calender factor y_i (i = 1, …, I) and year calender factor ψ_i . Adaptive Estimation uses a Recursive Bayesian Structure which will give more weight for new data that used where it can gave more response to prediction model where this is more actual to reality and change could be seen gradually from time to time.



2.2.1 Particle Filtering with Parameter Learning Approach

In Adaptive estimation from evolutionary models, Random factor estimated recursively continuosly. While the Parameter, could be estimated with 2 methods, using off-line and on-line Estimation method. off-line Estimation method is parameter estimation that using all observation result that obtained in last process of time series. while on-line Estimation method, combine parameter estimation to Adaptive estimation from Random factor.

Phase 1 : Initial

In i = 1, for m = 1, . . . ,M, for sample (also called as particle) from parameter Θ{m;1} from previous density. Superscript 1 is time filter (accident year) index and superscript m is sample index.

$$\gamma_1^{\{m\}} \sim f_{\gamma 1}(\gamma_1;\Theta^{\{m;1\}}). \qquad (2.27)$$

count importance weights in time 1

$$\omega_1^{\{m\}} = f_{Y1|\gamma 1;\psi I,\Theta}(y_1|\gamma_1^{\{m\}};\psi_I^{\{m;1\}},\Theta^{\{m;1\}}). \qquad (2.28)$$

Phase 2

Count Random factor that projected forward (look-ahead) and parameter in time i recursively using

i – 1. Calender time factor and parameter in time i projected using

$$\widetilde{\Theta}^{\{m;i\}} = \xi \cdot \Theta^{\{m;i-1\}} + (1-\xi) \cdot \frac{1}{M} \cdot \sum_{r=1}^{M} \Theta^{\{r;i-1\}}, \qquad (2.29)$$

$$\tilde{\psi}_I^{\{m;i\}} = \xi \cdot \psi_I^{\{m;i-1\}} + (1-\xi) \cdot \frac{1}{M} \cdot \sum_{r=1}^{M} \psi_I^{\{r;i-1\}}, \qquad (2.30)$$

sample mth from non-year calender factor in time i is determined with

$$\tilde{\gamma}_i^{\{m\}} = E\left[\gamma_i \mid \gamma_{i-1}^{\{m\}}, \Theta^{\{m;i-1\}}\right]. \qquad (2.31)$$

Phase 3

determined the projection and normalization of importance weigths in time i.

$$\widetilde{\omega}_i^{\{m\}} = \omega_{i-1}^{\{m\}} \cdot f_{Y_i \mid y_i, \psi_I; \Theta}\left(\boldsymbol{y}_i \mid \tilde{\gamma}_i^{(m)}, \tilde{\psi}_I^{\{m;i\}}; \widetilde{\Theta}^{\{m;i\}}\right), \qquad (2.32)$$

Normalization of importance weights

$$W_i^{\{m\}} = \frac{\widetilde{\omega}_i^{\{m\}}}{\sum_{m=1}^{M} \widetilde{\omega}_i^{\{m\}}} \qquad (2.33)$$

Phase 4 : *Re-sample*

Resample M samples (***particles***) from $\left\{\gamma_{i-1}^{\{m\}}, \tilde{\psi}_I^{\{m;i\}}; \widetilde{\Theta}^{(m;i)}\right\}_{m=1}^{M}$ with the probability $\left\{W_i^{\{m\}}\right\}_{m=1}^{M}$

Samples from Random factor and parameter that gives better probability than recent observation Yi that received in time *i* will be sampled with better probability.

Phase 5

Parameter and filtered year calender factor sampled using

$$\Theta^{\{m;i\}} \sim \text{Normal}\left(\widetilde{\Theta}^{\{m;i\}}, (1-\xi^2)\Sigma_{\Theta^{(i-1)}}\right), \qquad \textbf{(2.34)}$$

$$\psi_I^{\{m;i\}} \sim \text{Normal}\left(\tilde{\psi}_I^{\{m;i\}}, (1-\xi^2)\Sigma_{\psi_l^{(i-1)}}\right) \qquad \textbf{(2.35)}$$

And also from sample from filtered value from non-calender factor in time *i* is same as the distribution

$$\gamma_i^{(m)} \sim f_{\gamma_i|\gamma_{i-1}}\left(\gamma_i \mid \gamma_{i-1}^{(m)}; \Theta^{(m;i)}\right) \qquad \textbf{(2.36)}$$

Phase 6

Count importance weights in filtered samples Importance weights counted and updated with

$$\omega_i^{\{m\}} = \frac{f_{Y_i|\gamma_i,\psi_I;\Theta}\left(y_i|\gamma_i^{\{m\}},\psi_I^{\{m;i\}};\Theta^{\{m;i\}}\right)}{f_{Y_i|\gamma_i,\psi_I;\Theta}\left(y_i|\widetilde{\gamma}_i^{\{m\}},\widetilde{\psi}_I^{\{m;i\}};\widetilde{\Theta}^{(m;i\}}\right)} \qquad \textbf{(2.37)}$$

Phase 7

repeat Phase 2–6 till i = I.

2.2.2 Dual Kalman Filtering Approach for Gaussian Model

Dual Kalman filter is the best solution we found for Gaussian Model. In fact, this method is has been used for solving order estimation from dynamic factors as good as static factors or parameter from Gaussian model in variety work field, including civil engineering and vehicle system. Dual Kalman filter involving 2 filters that worked on parallel, one in calender time factor, and else in non-calender time factor. Using Dual Kalman filter in Gaussian Case, executed like this.

<u>Phase 1</u> : Initial

In $i = 1$, determine early estimation from average and variance from year calender factor

$$\tilde{\psi}_I^{\{1\}} = E\left[\psi_I^{\{1\}}\right]$$
$${}_h\widetilde{\boldsymbol{P}}_1 = \mathrm{Cov}\left[\psi_I^{\{1\}}\right] \qquad (2.38)$$

These could be obtained by simulating N sample drom $\psi_I^{\{1\}}$. Average and variance could be estimated with counting average of sample and covariance matrix from this sample.

$$\tilde{\gamma}_1 = E[\gamma_1]$$
$${}_\gamma\widetilde{\boldsymbol{P}}_1 = \mathrm{Cov}[\gamma_1] \qquad (2.39)$$

That could be obtained from *preliminary analysis GLM* with constant factor

Phase 2 : Filter or update calender year factor

Count Kalman result ${}_h\boldsymbol{G}_i$ for year calender factor

$${}_h\boldsymbol{G}_i = {}_h\widetilde{\boldsymbol{P}}_i \cdot \boldsymbol{E}_i' \cdot \left(\boldsymbol{E}_i \cdot {}_h\widetilde{\boldsymbol{P}}_i \cdot \boldsymbol{E}_i' + \boldsymbol{H}_i\right)^{-1} \qquad (2.40)$$

Update estimation from calender factor, including weighted average $\hat{\psi}_I^{(i]}$ and covariance matrix ${}_h\widehat{\boldsymbol{P}}_i$ in early observation $\boldsymbol{Y}_i$ in time i

$$\hat{\psi}_I^{(i]} = E\left[\psi_I^{\{i\}} \mid Y_i\right] = \tilde{\psi}_I^{\{i\}} + {}_h G_i \cdot \left(Y_i - A_i \cdot \hat{\gamma}_{i-1} - E_i \cdot \tilde{\psi}_I^{\{i]}\right), \qquad (2.41)$$

$${}_h\widehat{\boldsymbol{P}}_i = \mathrm{Cov}\left[\psi_l^{\{i\}} \mid \boldsymbol{Y}_i\right] = {}_h\widetilde{\boldsymbol{P}}_i - {}_h\boldsymbol{G}_i \cdot \boldsymbol{E}_i \cdot {}_h\widetilde{\boldsymbol{P}}_i \qquad (2.42)$$

This can be treated as *posterior mean* and variance in *conventional Bayesian setting*

Phase 3 : Filter or Update non-calender year factor

Count Kalman Result ${}_\gamma\boldsymbol{G}_i$ for non-year calender factor

$${}_\gamma\boldsymbol{G}_i = {}_\gamma\widetilde{\boldsymbol{P}}_i \cdot \boldsymbol{A}_i' \cdot \left(\boldsymbol{A}_i \cdot {}_\gamma\widetilde{\boldsymbol{P}}_i \cdot \boldsymbol{A}_i' + \boldsymbol{H}_i\right)^{-1} \qquad (2.43)$$

Update estimation from calender factor, including weighted average ${}_\gamma\widetilde{\boldsymbol{P}}_i$ and covariance matrix ${}_\gamma\widehat{\boldsymbol{P}}_i$ in early observation $\boldsymbol{Y}_i$ in time i

$$\hat{\gamma}_i = E[\gamma_i \mid \boldsymbol{Y}_i] = \tilde{\gamma}_i + {}_\gamma\boldsymbol{G}_i \cdot \left(\boldsymbol{Y}_i - \boldsymbol{A}_i \cdot \tilde{\gamma}_i - \boldsymbol{E}_i \cdot \hat{\boldsymbol{\psi}}_I^{(i\}}\right), \qquad (2.44)$$

$${}_\gamma\widehat{\boldsymbol{P}}_i = \mathrm{Cov}[\gamma_i \mid \boldsymbol{Y}_i] = {}_\gamma\widetilde{\boldsymbol{P}}_i - {}_\gamma\boldsymbol{G}_i \cdot \boldsymbol{A}_i \cdot {}_\gamma\widetilde{\boldsymbol{P}}_i \qquad (2.45)$$

This could be treated as *posterior mean* and variance in *conventional Bayesian setting*

Phase 4 : **Predict year calender factor (*time update*)**

project year calender factor for next period

$$\tilde{\psi}_I^{\{i+1\}} = E\left[\psi_I^{\{i+1\}} \mid \boldsymbol{Y}_i\right] = \hat{\psi}_I^{\{i\}} \qquad (2.46)$$

And project covariance error from that factor

$$_h\widetilde{\boldsymbol{P}}_{i+1} = \mathrm{Cov}\left[\psi_I^{\{i+1\}} \mid \boldsymbol{Y}_i\right] = {}_h\widehat{\boldsymbol{P}}_i + Q_{n_I\epsilon} \qquad (2.47)$$

Where $Q_{n_I\epsilon}$ is artificial dynamic that featured to the covariance specification. These things could be done for showing level from uncertainty related with estimation from calender factor. Estimation that will be obtained in this Phase could be treated as previous average and variance conventional Bayesian setting.

Phase 5 : **Predict non-calender year factor (‘*time update*’)**

Project non-year calender factor for next period

$$\tilde{\gamma}_{i+1} = E[\gamma_{i+1} \mid Y_i] = \hat{\gamma}_i \qquad (2.48)$$

And project covariance error from that factor

$$\widetilde{\boldsymbol{P}}_{i+1} = \mathrm{Cov}[\boldsymbol{\gamma}_{i+1} \mid \boldsymbol{Y}_i] = \widehat{\boldsymbol{P}}_i + \boldsymbol{Q}_{\gamma^\epsilon} \qquad (2.49)$$

Phase 6 : **Repeat** Phase 2-5 till $i = I$

3. Data Analysis and Results

Data Overview

The dataset used in this research are **claim** values **and premium** values for **line of business liability reinsurance and motor reinsurance** services from **Swiss Re Group**, which is a Comprise P&C Reinsurance and Corporate solution.

The dataset has **no missing value**.

The range time that we used is **January – October 2018**.

SOFTWARE

Google Colab — with Python language

R-Gui 4.0.3 version — with library ChainLadder, plyr, readxl, lubridate, dplyr, magic, MASS, matlib, tweedie, and statmod

Excel

Cumulative RUN OFF TRIANGLE

LIABILITY REINSURANCE

	Premium	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18	Oct-18
01/18	2668.12	199.753	432.583	619.326	839.025	1083.06	1157.37	1191.24	1215.8	1274.03	1309.06
02/18	2579.27	215.194	572.446	825.973	1077.15	1210.44	1301.42	1353.45	1405.11	1421.28	
03/18	2412.58	49.5659	371.774	560.001	711.634	901.323	985.286	1040.27	1059.39		
04/18	2003.16	80.7781	411.143	539.873	783.886	942.896	1089.33	1089.04			
05/18	1727.32	68.09	328.383	547.178	681.812	780.503	883.676				
06/18	1329.15	70.596	335.527	520.456	650.11	716.347					
07/18	1212.89	68.514	256.022	441.441	637.458						
08/18	1112.92	66.9074	215.111	345.75							
09/18	1156.29	47.532	192.312								
10/18	1500.18	48.5309									

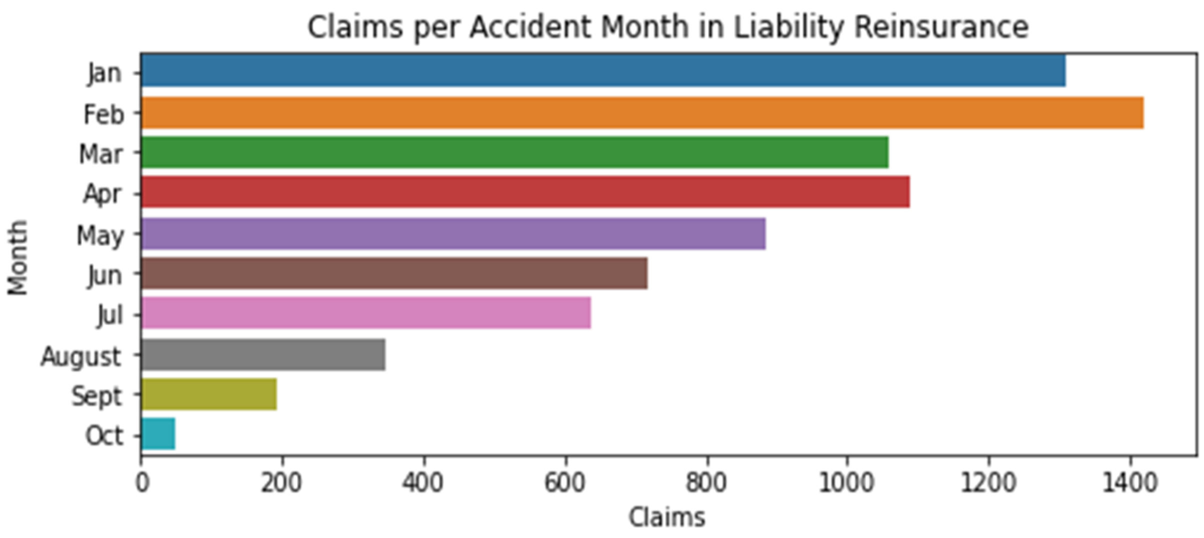
CUMULATIVE RUN OFF TRIANGLE

MOTOR REINSURANCE

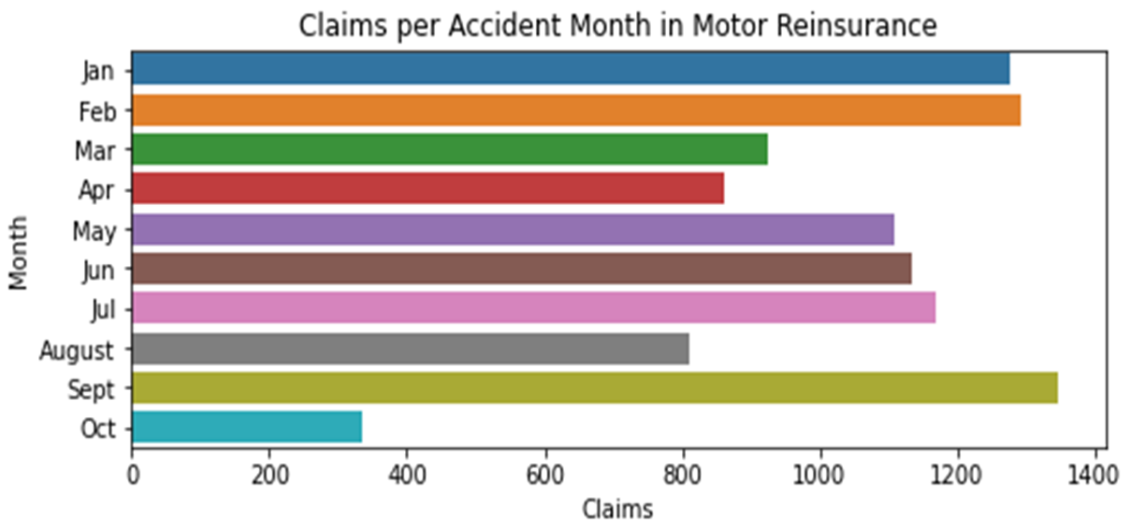
	Premium	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18	Oct-18
01/18	1555.11	680.333	1061.07	1180.27	1221.81	1247.98	1259.09	1264.65	1257.15	1259.06	1268.84
02/18	1573.92	615.401	1061.63	1177.17	1275.66	1280.48	1288.96	1289.54	1291.86	1293.93	
03/18	1299.77	222.421	787.973	871.215	888.448	919.959	922.329	923.492	923.552		
04/18	1153.15	38.8192	656.959	795.225	838.976	851.555	860.039	855.358			
05/18	1343.87	168.378	890.638	1057.06	1080.98	1094.22	1108.35				
06/18	1355.12	299.019	942.295	1070.57	1105.54	1133.51					
07/18	1395.34	280.799	937.08	1117.4	1170.66						
08/18	1115.95	147.219	670.267	810.73							
09/18	1902.29	334.382	1348.15								
10/18	2446.88	335.299									

EXPLORATORY DATA ANALYSIS

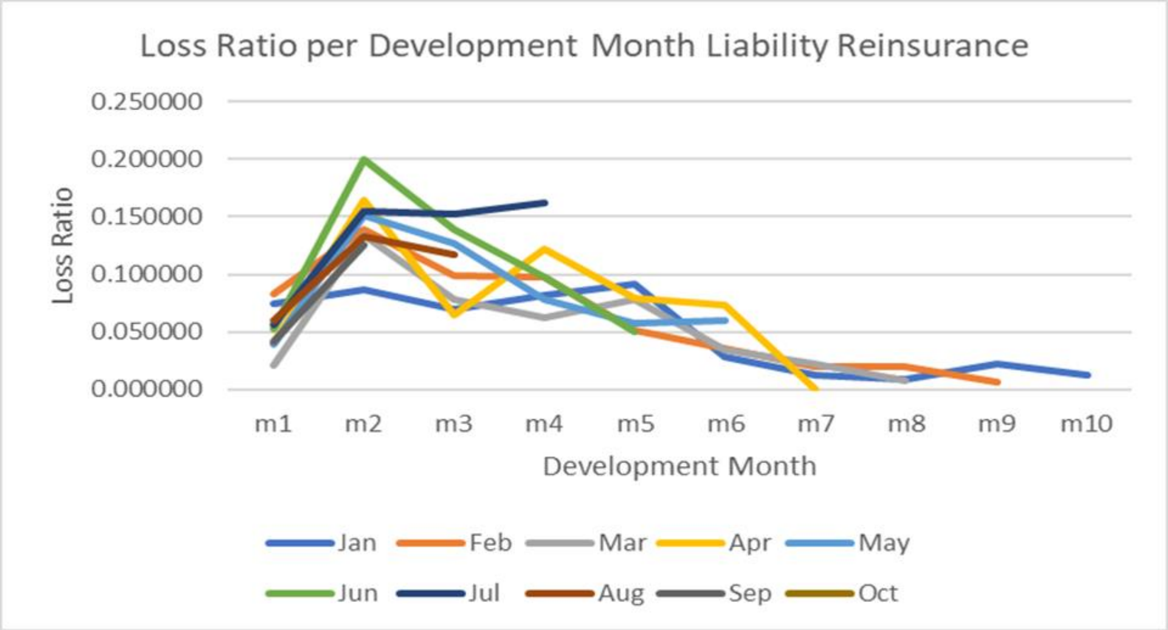
(a) Liability Reinsurance



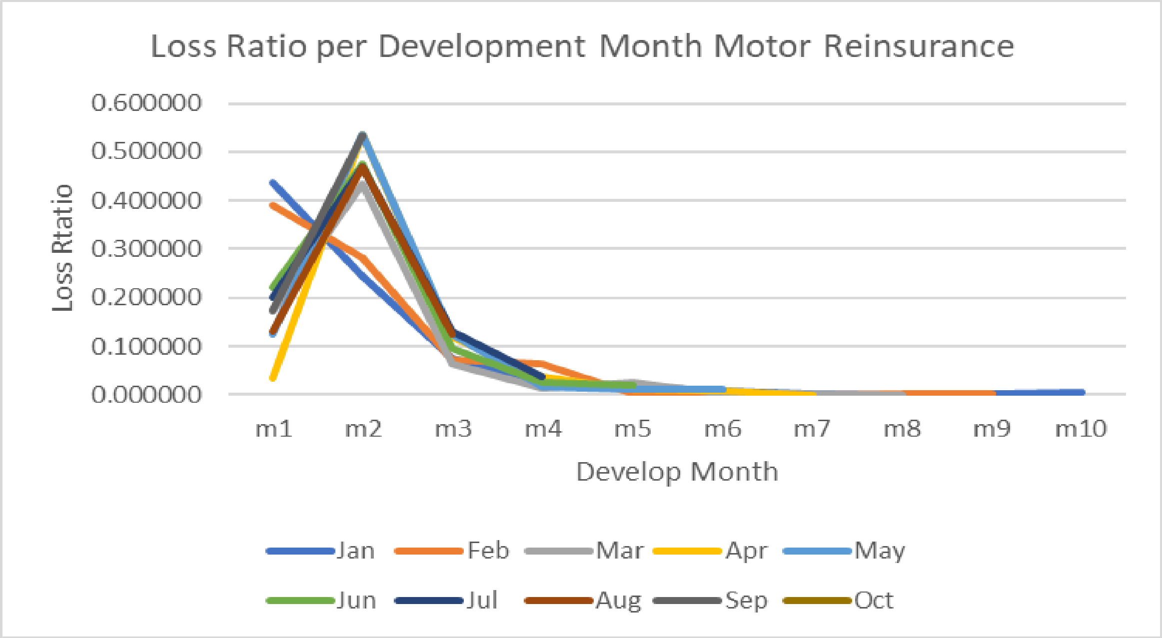
(b) Motor Reinsurance



LOSS RATIO PER DEVELOPMENT MONTH

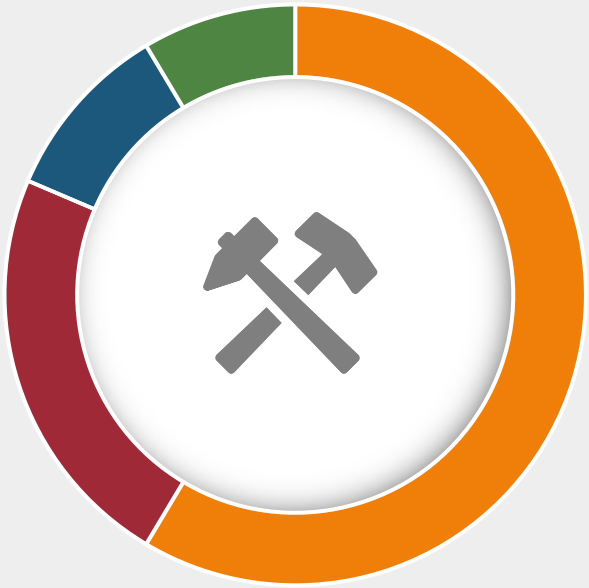


(a)



(b)

A Multivariate Evolutionary GLM Framework with Adaptive Estimation



This research revolves more around the application of GLM with a different Tweedie distribution approach with a different power parameter p so as the model output might give a more flexible dispersion model. In addition, 50.000 samples are used for every time step, filter initiation is also used with static GLM estimation with a mean structure such as:

$$a_i^{(n)} + r_i^{(n)} \log j + s_i^{(n)} j + b_{i,1}^{(n)} \mathbb{1}_{\{j=1\}} + b_{i,2}^{(n)} \mathbb{1}_{\{j=2\}} + h_t^{(n)}$$

On the other hand, changes in claim pattern per month are also monitored.

FILTER VALUES FROM RANDOM FACTORS

FOR LIABILITY REINSURANCE

n	i	$a_i^{(n)}$	$r_i^{(n)}$	$s_i^{(n)}$	$b_{i,1}^{(n)}$	$b_{i,2}^{(n)}$
1	1	-2.087543252	1.309957	-0.55933	0.524395	-3.54663
	2	-2.013931471	1.326508	-0.57422	0.370883	-3.38264
	3	-2.124276212	1.386341	-0.60775	-0.23523	-3.18238
	4	-2.059553345	1.428034	-0.57666	-0.39653	-3.25512
	5	-2.015587325	1.401532	-0.56355	-0.53013	-3.09946
	6	-1.958315133	1.426051	-0.57668	-0.36193	-2.98265
	7	-1.831956881	1.465512	-0.52227	-0.39831	-2.85423
	8	-1.858938768	1.471362	-0.58565	-0.32022	-2.86975
	9	-1.848340336	1.464519	-0.61427	-0.48828	-2.8444
	10	-1.82580777	1.430496	-0.62445	-0.76258	-2.87436

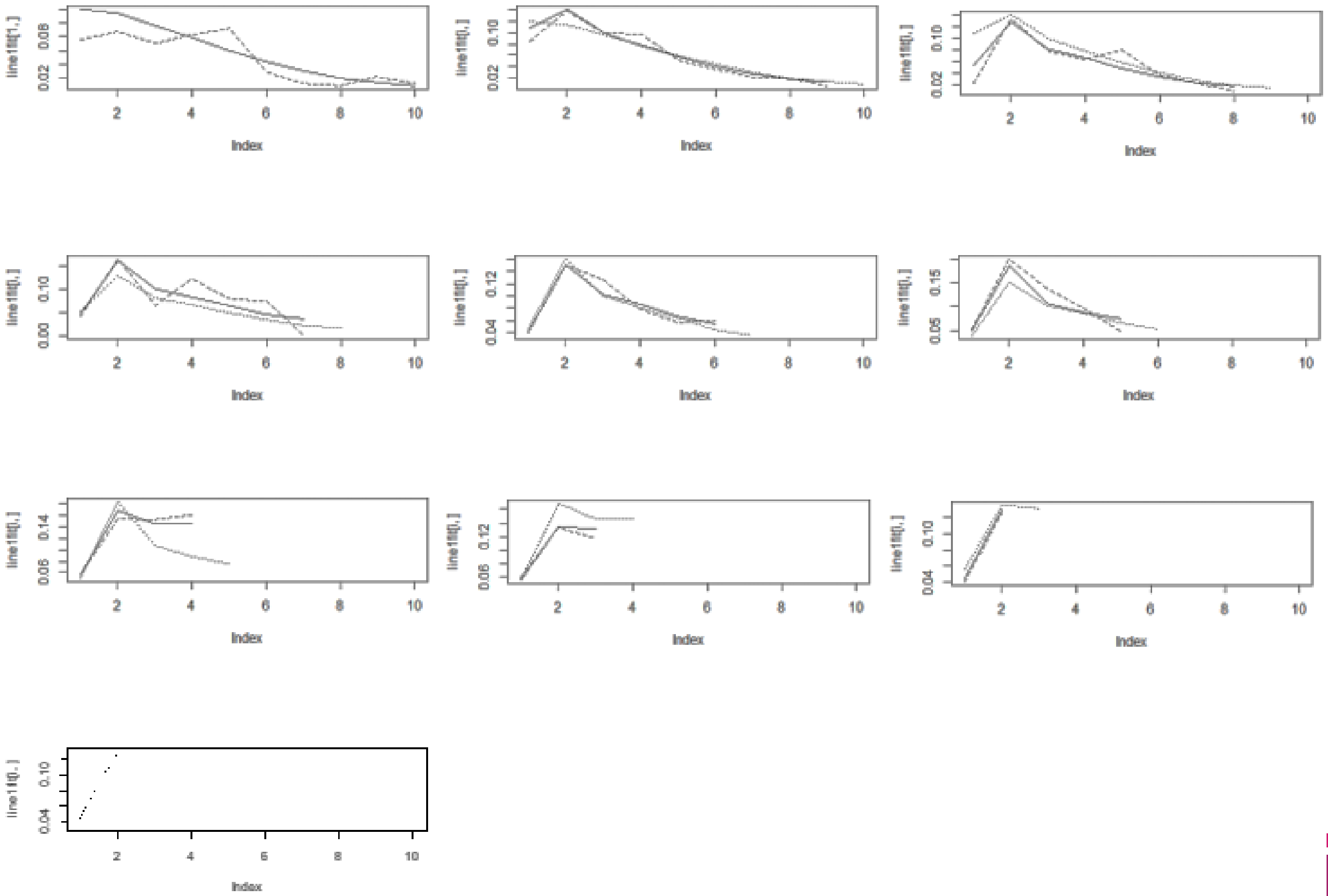
FILTER VALUES FROM RANDOM FACTORS

FOR MOTOR REINSURANCE

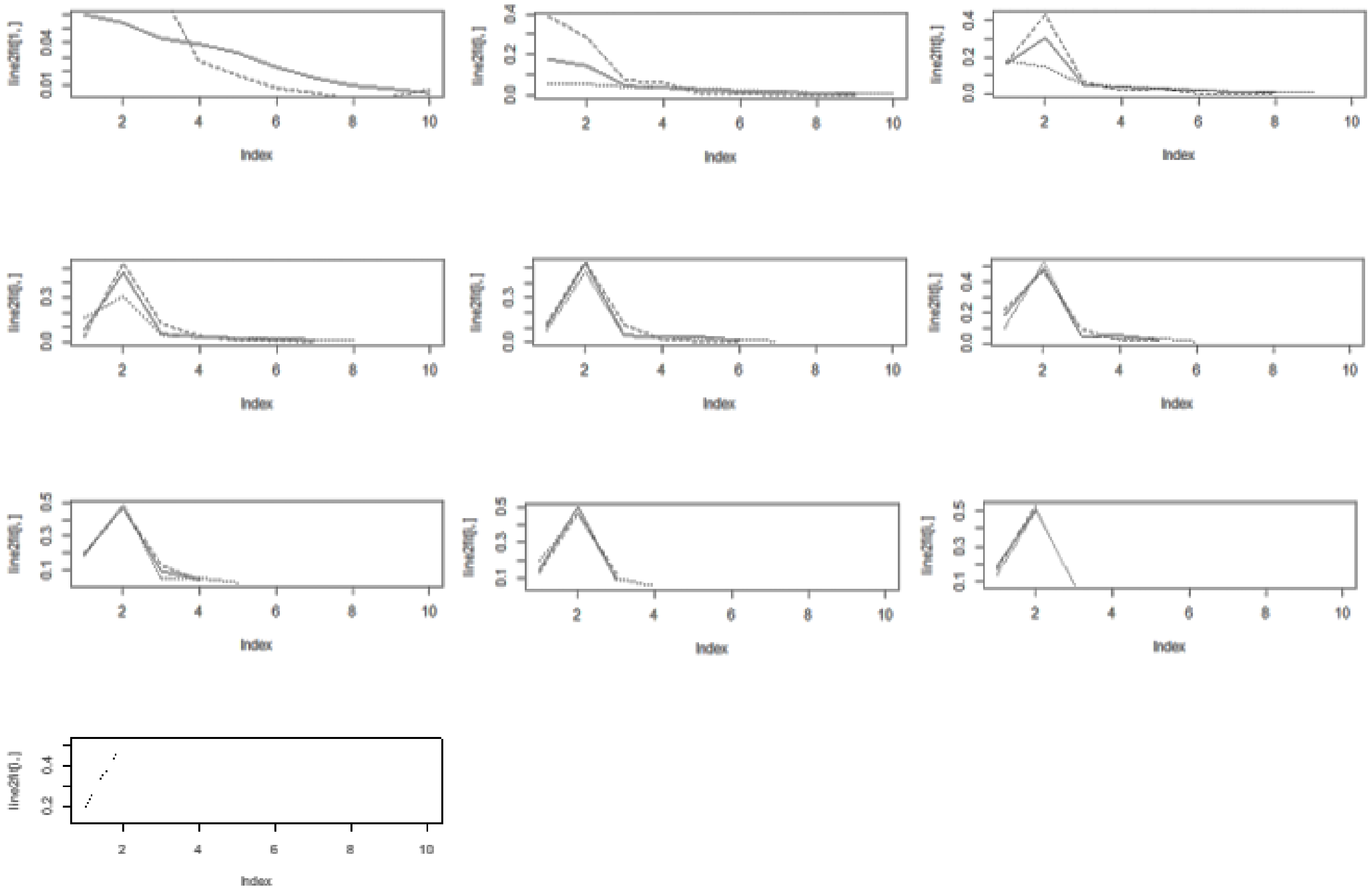
n	i	$a_i^{(n)}$	$r_i^{(n)}$	$s_i^{(n)}$	$b_{i,1}^{(n)}$	$b_{i,2}^{(n)}$
2	1	2.160242943	-0.71604	1.463018	0.455415	0
	2	2.047210591	-0.77495	2.157653	1.189497	0.276643
	3	2.049707678	-0.85201	1.837196	1.790101	0.386594
	4	2.304503967	-0.8926	1.153137	2.313895	0.486357
	5	2.25675042	-0.90987	1.364345	2.30707	0.378763
	6	2.223863193	-0.91475	1.79176	2.231462	0.417468
	7	2.192819508	-0.85437	1.739	2.023258	0.330494
	8	2.363158001	-0.81734	1.494443	1.528353	0.265106
	9	2.481085559	-0.80687	1.323609	1.791329	0.648373
	10	2.464978385	-0.79961	1.383014	1.840059	0.275103

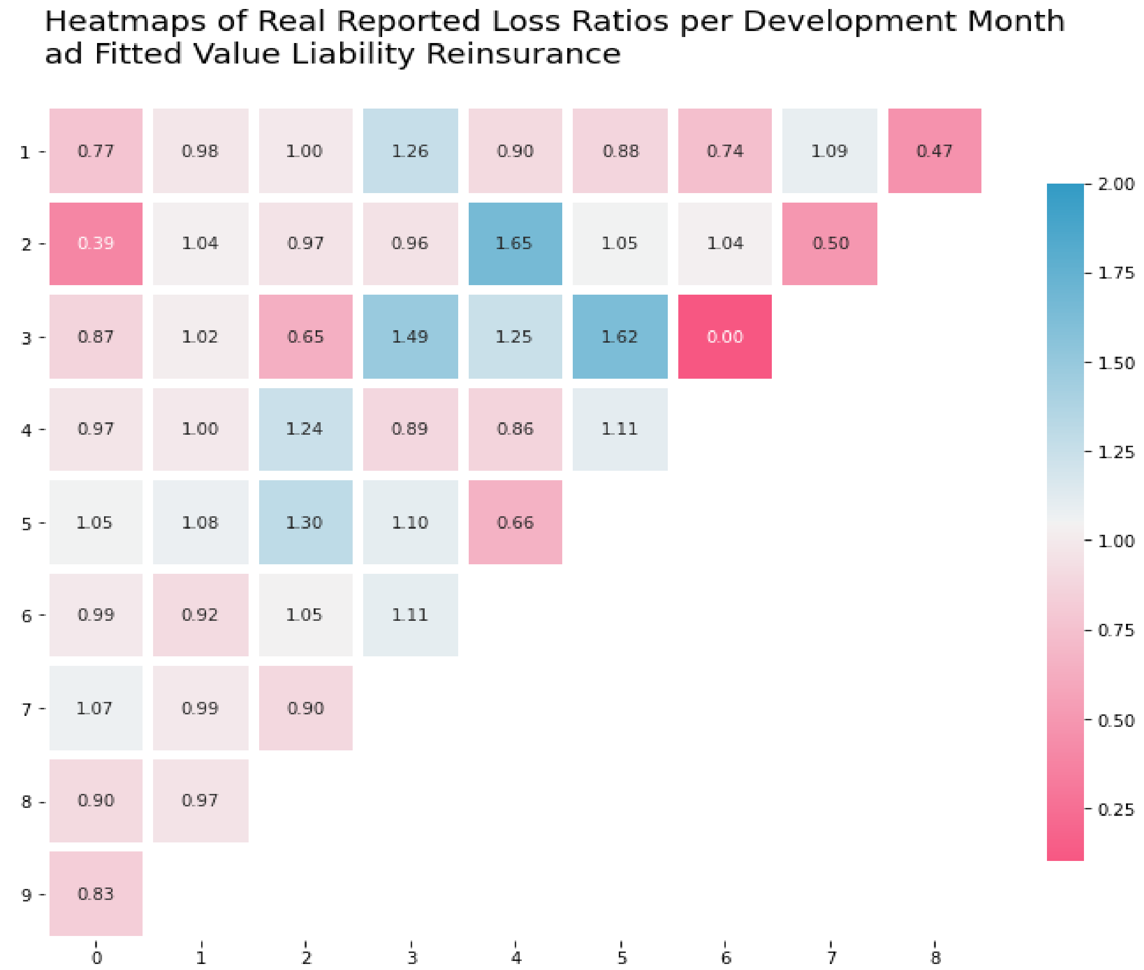
Reveal Loss Ratios Patterns with Development Month every Accident Month

In liability reinsurance

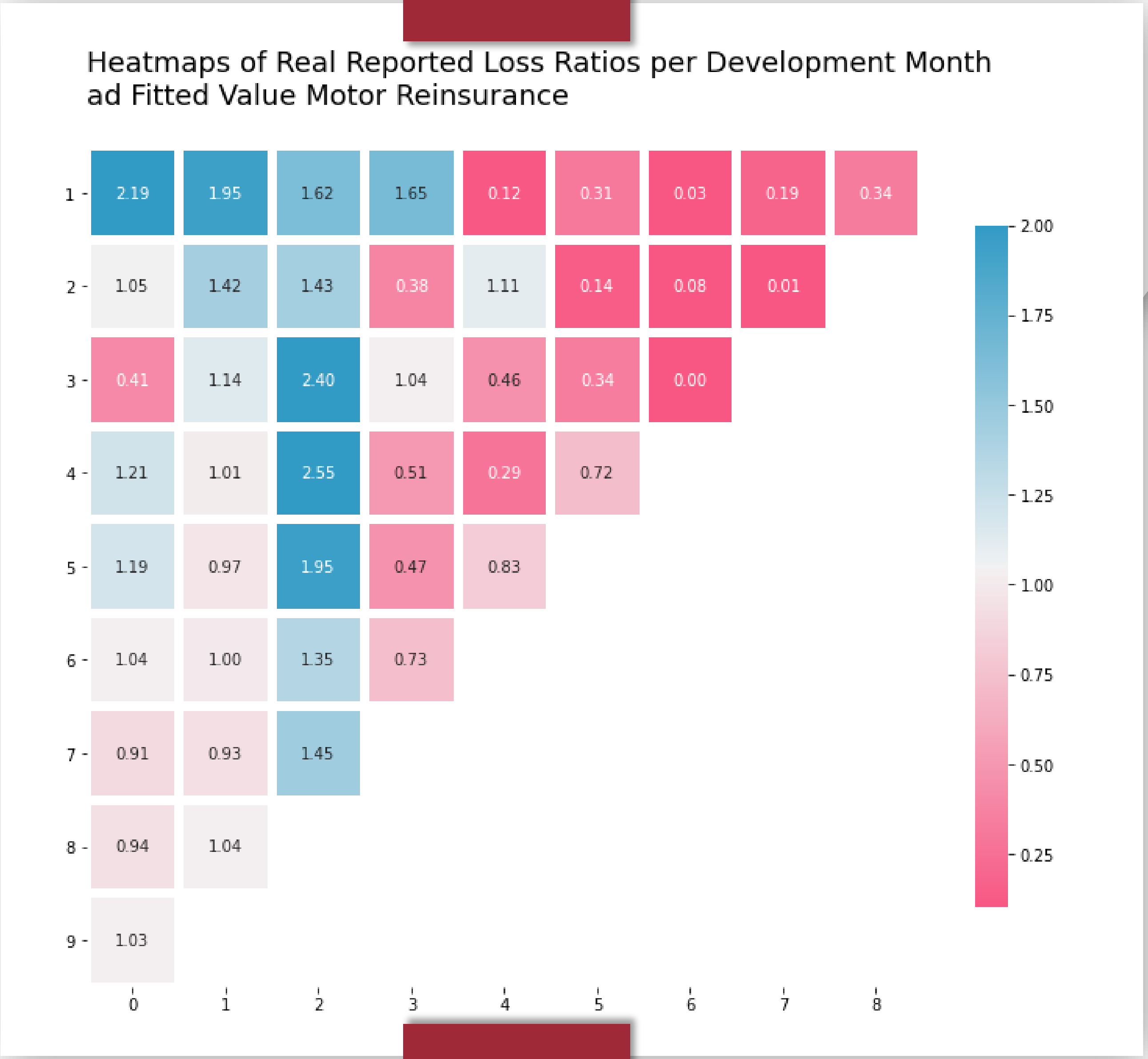


In motor reinsurance





Heatmaps of Real Reported Loss Ratios per Development Month and Fitted Value Liability Reinsurance



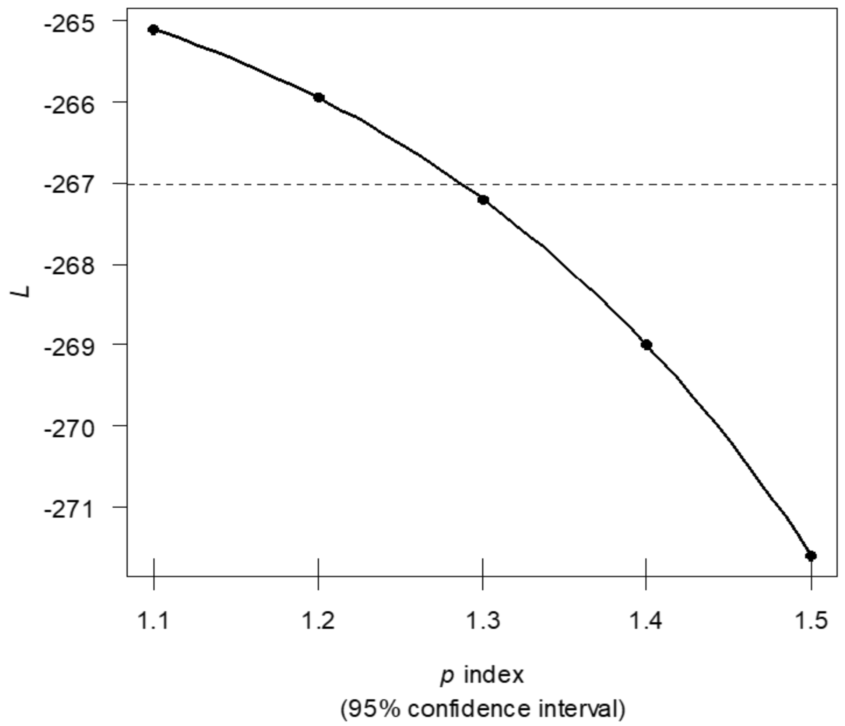
Heatmaps of Real Reported Loss Ratios per Development Month and Fitted Value Motor Reinsurance

Generalised Linear Model Simple method with the Tweedie Family Distribution Approach

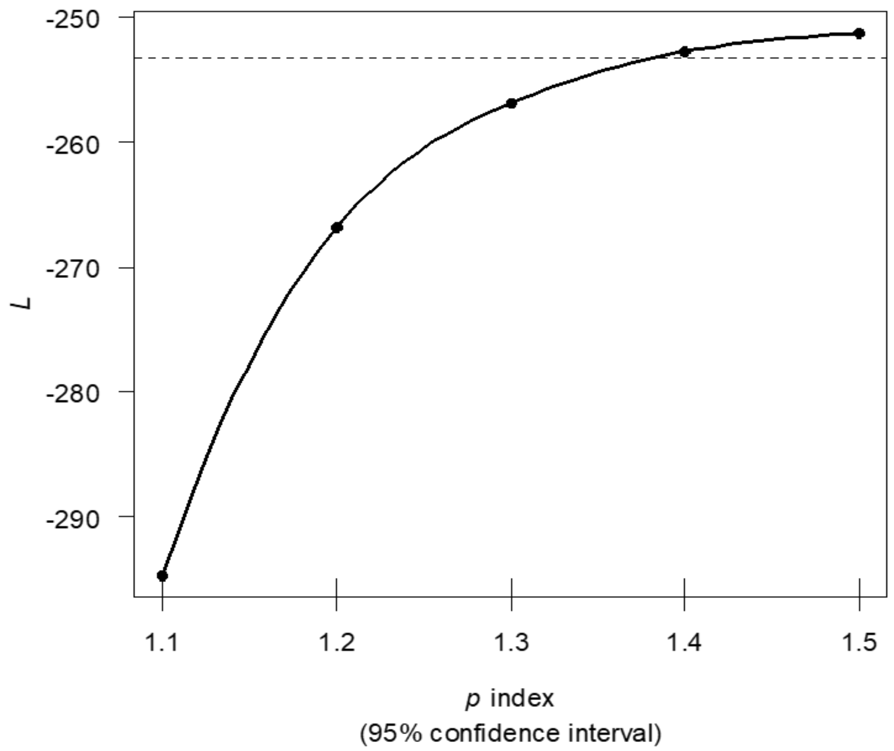
A Generalised Linear Model Simple Method with the Tweedie Family Distribution Approach with parameter p is used to maximize a function. To validate the Tweedie p parameter, we can make the plot as the likelihood profile at defined p values (by using argument p check) for our dataset. In order to make our model able to use parameter p which can maximize a function, we set the argument to p.optim=TRUE.

The Likelihood Profile at Defined p Values in :

(a) Liability Reinsurance



(b) Motor Reinsurance



Predicted Claims per Month (decrement)

for Liability Reinsurance

	Premium	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18	Oct-18
01/18	2668.12	125.311	325.999	228.857	222.594	169.240	108.756	36.653	32.391	35.226	35.000
02/18	2579.27	138.764	360.998	253.427	246.492	187.409	120.433	40.588	35.868	39.008	
03/18	2412.58	106.142	276.130	193.848	188.543	143.351	92.120	31.046	27.436		
04/18	2003.16	111.512	290.100	203.655	198.082	150.603	96.780	32.616			
05/18	1727.32	93.667	243.676	171.065	166.383	126.502	81.293				
06/18	1329.15	83.355	216.851	152.233	148.067	112.576					
07/18	1212.89	88.611	230.523	161.831	157.403						
08/18	1112.92	63.898	166.231	116.697							
09/18	1156.29	53.437	139.017								
10/18	1500.18	49.000									

Predicted Claims per Month (decrement)

for Motor Reinsurance

	Premium	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18	Oct-18
01/18	2668.12	344.577	693.435	158.704	51.355	22.855	10.107	23.317	0.725	1.982	10.000
02/18	2579.27	350.985	706.330	161.655	52.310	23.280	10.295	23.751	0.739	2.018	
03/18	2412.58	245.112	493.269	112.893	36.531	16.258	71.895	16.586	0.516		
04/18	2003.16	229.009	460.862	105.476	34.131	15.189	67.171	15.497			
05/18	1727.32	295.019	593.703	135.879	43.969	19.568	86.533				
06/18	1329.15	306.455	616.718	141.146	45.673	20.326					
07/18	1212.89	323.905	651.833	149.183	48.274						
08/18	1112.92	232.932	468.758	107.283							
09/18	1156.29	444.816	895.157								
10/18	1500.18	335.000									

Root Mean Square ERROR COMPARISON

Line of Business	A Multivariate Evolutionary GLM Framework with Adaptive Estimation	A GLM Simple Method with the Tweedie Family Distribution Approach
Liability Reinsurance	846.352257754092	36.56627
Motor Reinsurance	658.413498294015	95.23479
Average	752.382878027046	65.90053

Claim Reserves (Decrement) Estimation

for Liability Reinsurance

	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18	Oct-18	Total
01/18											0
02/18										35	35
03/18									33	33	66
04/18								27	30	30	87
05/18							27	25	28	27	107
06/18						77	25	23	26	25	176
07/18					107	71	23	21	24	23	269
08/18				130	98	65	21	19	22	21	376
09/18			123	119	90	60	19	18	20	20	469
10/18		159	113	110	83	55	18	16	18	18	590

Claim Reserves (Decrement) Estimation

for Motor Reinsurance

	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18	Oct-18	Total
01/18											0
02/18										3	3
03/18									1	3	4
04/18								25	1	4	30
05/18							9	1	1	4	15
06/18						10	9	1	1	4	25
07/18					22	10	9	1	1	4	47
08/18				53	23	10	10	1	1	4	102
09/18			163	55	24	11	10	1	1	4	269
10/18		741	169	57	24	11	11	1	1	5	1020

CONCLUSIONS

According to this research, we can summarize **filter particles** that we **GLM Generalised Linear Model Framework with Adaptive Estimation are sufficient** in **detecting loss ratios patterns** by development month. However, if we look at the RMSE generated for each line of business, **the RMSE generated by the GLM Simple Method with the Tweedie Family Distribution Approach is smaller** than the RMSE generated by the Multivariate Evolutionary GLM Framework with Adaptive Estimation so we can conclude that the **GLM Simple Method with the Tweedie Family Distribution Approach is more suitable** for us to use in calculating Swiss Re Group's claim reserves, especially for motor reinsurance and liability reinsurance.

THANK YOU