APLICACIÓN DE MODELOS LINEALES GENERALIZADOS A LA ESTIMACIÓN DE RESERVAS: CASO DE ESTUDIO

Prieto, Augusto D. 888.287

Pérez, Tomás A. 891.843

Franco, Patricio 880.851

Valicenti, Fernando 891.634

Universidad de Buenos Aires, Facultad de ciencias económicas

Computación científica actuarial

Docente a cargo: Del Rosso, Rodrigo

Julio, 2020

Resumen

En el presente trabajo se pondrá a prueba el modelo de Mack y distintos modelos lineales generalizados, o simplemente GLM, en el contexto de estimación de provisiones de seguros con datos representados en forma triangular. Para ello se estimarán las reservas a través del método de Chain Ladder y, luego, se estimarán las correspondientes al método de GLM. Se compararán los resultados obtenidos, validándolos con el método de retención de datos –Holdout method- introducido por Highleyman (1962). El objetivo será, dilucidar si existen estimaciones de reservas más precisas que el método de Chain Ladder, sin perder la robustez que lo caracteriza. El software que se utilizará es R 3.6.2 (R Core Team), y se hará principal uso del paquete *ChainLadder* cuya autoría pertenece a Carrato et. al. Se utilizará también, aunque en menor medida, el paquete *tidyverse* de Wickham et. al. El código completo se encuentra en el siguiente repositorio, junto al archivo de los datos necesarios para su reproducción:

https://github.com/grupo10-glm/codigo-reservas

Palabras clave: Chain Ladder, IBNR, Holdout method, R, modelos lineales generalizados.

Marco teórico

Chain Ladder

El método de Chain Ladder es uno de los métodos más antiguos, pero a su vez, de los más utilizados en la actualidad por todo el mundo debido a la robustez del mismo. Introducido inicialmente por Harnek (1966), según lo citado en Taylor (1986), como un modelo determinístico, debido a que no propone la inclusión de un componente de azar o de aleatoriedad. Otra particularidad de este método radica en la utilización datos que son presentados en forma triangular. Es decir, todos los datos individuales se agrupan o combinan por período de accidente y luego, se computa el desarrollo de los mismos a través de los períodos subsiguientes.

Gráficamente, el panel de datos triangular, toma la forma de la Figura 1.

Año de accidente (i)	Año de desarrollo (k)				
	1	2	3	4	5
1	$L_{1,1}$	$L_{1,2}$	$L_{1,3}$	$L_{1,4}$	$L_{1,5}$
2	$L_{2,1}$	$L_{2,2}$	$L_{2,3}$	$L_{2,4}$	$L_{2,5}$
3	$L_{3,1}$	$L_{3,2}$	$L_{3,3}$	$L_{3,4}$	$L_{3,5}$
4	$L_{4,1}$	$L_{4,2}$	$L_{4,3}$	$L_{4,4}$	$L_{4,5}$
5	$L_{5,1}$	$L_{5,2}$	$L_{5,3}$	$L_{5,4}$	$L_{5,5}$

Figura 1. Datos de siniestros agrupados en forma triangular.

Donde $L_{i,k}$ representa los siniestros incurridos acumulados para el año del accidente i y el año de desarrollo del siniestro k. La representación de los datos es triangular debido a que no se conocen los pagos acumulados para los años $i + k \leq K + 1$, i.e, la parte sombreada del panel de datos, donde K es el máximo año de desarrollo de los siniestros.

En 1993, Thomas Mack, introduce la teoría de un modelo estocástico subyacente que replica en forma exacta el método propuesto y, a su vez, permite el cálculo de estimaciones de error estándar. Para ello, se proponen tres supuestos:

Dado $(\Omega, \mathcal{F}, \mathbb{P})$, un espacio probabilístico, y sea $\{L_{i,k}\}$ una secuencia de variables aleatorias asociadas al mismo, entonces:

1. $\mathbb{E}\left(L_{i,k} \mid L_{i,0}, L_{i,1}, \dots, L_{i,k-1}\right) = L_{i,k-1}. f_k$
2. $\mathbb{V}\left(L_{i,k} \mid L_{i,0}, L_{i,1}, \dots, L_{i,k-1}\right) = L_{i,k-1}. \sigma_k^2$
3. $(L_{i,0}, L_{i,1}, \dots, L_{i,K})\ y\ (L_{j,0}, L_{j,1}, \dots, L_{j,K})$ son independientes para todo $i \neq j$.

Donde f_k es el factor de desarrollo de los siniestros desde el año de desarrollo $k-1$ hasta k, y donde σ es un parámetro.

Las consecuencias de estos supuestos son dos, de acuerdo a Brown & Lennox (2018): en primer lugar, el mismo factor de desarrollo se aplica a distintos años de accidente y, en segundo lugar, el valor particular depende únicamente del valor del año previo de desarrollo.

Como resultado del supuesto 2, se obtiene el siguiente estimador de f_j (Ver Anexo I), llamado factor de desarrollo por promedios ponderados:

$$\hat{f}_j = \frac{\sum_{i=1}^{K-j+1} L_{i,j}}{\sum_{i=1}^{K-j+1} L_{i,j-1}}$$

Luego, distintos autores flexibilizaron este supuesto, obteniéndose como resultados estimadores que tienen relación con medias aritméticas, medias armónicas, etc.

Finalmente, el modelo estocástico que representa los valores siniestrales es el siguiente:

$$L_{i,j} = r_j \cdot L_{i,K} + \varepsilon_{i,j}, j = 0, \dots, K-1$$

Donde $r_j = \frac{1}{f_{j,K}}$ y $\varepsilon_{i,j}$ es un término de error. Como consecuencia de este modelo específico, se pueden calcular las estimaciones de las reservas restando las estimaciones de las comúnmente llamadas “ultimate incurred losses” con respecto a los siniestros incurridos acumulados hasta el momento de estimación, pero más importante (ya que el método propuesto en 1966 produce los mimos resultados), se pueden estimar la varianza y los intervalos de confianza con respecto a las estimaciones anteriormente mencionadas (Mack, 1994).

Las reservas, entonces, se obtienen de la siguiente manera

$$\widehat{R_\iota} = \widehat{L_{\iota,K}} - L_{i,K-i+1}$$

La varianza queda definida por

$$\left(s.e(\widehat{L_{\iota,K}})\right)^2 = \widehat{L_{\iota,K}}^2 . \sum_{j=K+1-i}^{K-1} \frac{\widehat{\sigma_j^2}}{\hat{f}_j^{\,2}} . \left(\frac{1}{\widehat{L_{\iota,J}}} + \frac{1}{\sum_{n=1}^{K-j} L_{n,j}}\right)$$

Donde $\widehat{\sigma_j^2} = \frac{1}{K-j-1} \sum_{n=1}^{K-j} L_{n,j} \left(\frac{L_{n,j+1}}{L_{n,j}} - \hat{f}_j\right)^2$, $1 \leq j \leq K-2$, es un estimador insesgado de σ_j^2 .

Dado que $s.e\left(\widehat{R_\iota}\right) = s.e(\widehat{L_{\iota,K}})$, y suponiendo que haya un volumen de datos de siniestros lo suficientemente grande, por el Teorema Central del Límite, el intervalo con 95% de confianza queda definido por

$$\left(\widehat{R_\iota} - 2\, s.e\left(\widehat{R_\iota}\right);\ \widehat{R_\iota} + 2\, s.e(\widehat{R_\iota})\right)$$

Finalmente, el error cuadrático medio en el contexto de cálculos de reservas será

$$MSE\left(\widehat{L_{\iota,K}}\right) = \mathbb{E}\left(\left(L_{i,K} - \widehat{L_{\iota,K}}\right)^2 \mid D\right)$$

Donde $D = \left\{L_{i,k} \middle| i + k \ \leq K + 1 \right\}$ es el conjunto de todos los datos observados. En otras palabras, todos los datos que conforman el panel triangular.

Tras operar algebraicamente

$$MSE\left(\widehat{L_{\iota,K}}\right) = \mathbb{V}\left(L_{i,K} \mid D\right) + \left(\mathbb{E}\left(L_{i,K} \mid D\right) - \widehat{L_{\iota,K}}\ \right)^2$$

Es decir que el error cuadrático medio se puede representar como la suma de error aleatorio futuro y el error en la estimación.

Modelos Lineales Generalizados

Los modelos lineales generalizados, desarrollados inicialmente por Nelder & Wedderburn (1972), permiten unificar distintos modelos paramétricos que cumplen con determinadas características en un mismo marco teórico dentro del cual, por ejemplo, son casos particulares: el modelo de regresión lineal, regresión logística, Poisson, etc.

En este apartado se utilizarán matrices y vectores específicos. La notación, a posteriori, es la siguiente:

$y = \begin{matrix} y_1 \\ \vdots \\ y_n \end{matrix}$ es el vector de observaciones de $\{Y_i, i = 1, 2, \dots, n\}$, i.e, lo que llamaremos nuestra variable dependiente.

$X = \begin{bmatrix} 1 & \cdots & x_{1,k} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{n,k} \end{bmatrix}$ es la matriz de observaciones de las variables explicativas, donde x_i^T es la i-ésima fila de la matriz X.

$\beta = \begin{matrix} \beta_0 \\ \vdots \\ \beta_n \end{matrix}$ es el vector de parámetros asociados a las variables explicativas.

Según lo explicitado en Dobson (2008) y en Frees (2009), un modelo lineal generalizado cuenta con tres características principales:

1. Componente aleatoria: Esta está relacionada a la distribución de $\{Y_i, i = 1, 2, \dots, n\}$. El modelo presupone que $\{Y_i, i = 1, 2, \dots, n\}$ es un conjunto de variables aleatorias independientes que pertenecen a la familia exponencial de distribuciones que cumplen con los siguientes criterios:
    a. La distribución de cada Y_i está expresada en forma canónica, en otras palabras,

    $f(y_i, \theta_i) = \exp\left[\frac{y_i.\theta_i - b_i(\theta_i)}{\phi} + S_i(y_i, \phi)\right]$

b. La distribución de las Y_i son iguales. Por lo tanto, los subíndices para b y S no son necesarios.

2. Componente sistemática: Se puede expresar la esperanza de la variable dependiente como combinación lineal de las variables explicativas, i.e, $\eta_i = x_i^T.\beta$
3. Función de enlace: Existe una función $g(.)$, llamada función de enlace, que relaciona la esperanza de la variable respuesta con la componente sistemática, es decir que, dado $\mathbb{E}[Y_i] = \mu_i$, entonces

   $g(\mu_i) = x_i^T.\beta = \eta_i.$

Una distribución perteneciente a la familia exponencial que es de especial interés en el ámbito actuarial es la llamada distribución de Tweedie (1984): Esta distribución es una suma de N variables aleatorias, i.e, $S_N := \sum_{i=1}^{N} Z_i$ donde Z_i son variables aleatorias independientes e idénticamente distribuidas como Gamma (α, γ), y donde N es una variable aleatoria que está distribuida como Poisson (λ) y es independiente de Z_i .

La densidad de probabilidades de la distribución de Tweedie, en el contexto de GLM, se representa de la siguiente manera:

$$f_s(z) = \exp\left[\frac{-1}{\phi} \cdot \left(\frac{\mu^{2-p}}{2-p} + \frac{z}{(p-1)\mu^{p-1}}\right) + S(z,\phi)\right]$$

Se deduce que $\mathbb{P}(S_N = 0) \neq 0$ (ver Anexo II). Lo cual es de importancia para representar las pérdidas por siniestros acumulados. Además, esta distribución tiene la particularidad de poder expresar la varianza como potencia de la esperanza, es decir

$$\mathbb{V}(S_N) = \phi \,.\, \mu^p$$

Se pueden caracterizar, entonces, algunas de las distribuciones más conocidas de la familia exponencial según el parámetro p:

- Distribución Normal, $p = 0$
- Poisson, $p = 1$
- Poisson-Gamma, $1 < p < 2$
- Gamma, $p = 2$
- Normal Inversa, $p = 3$

Hasta aquí, los modelos lineales generalizados propuestos tienen una relación determinística entre variable respuesta y variables explicativas. Para una correcta representación del modelo de reservas, se utilizarán los llamados Modelos Lineales Generalizados Mixtos, que agrega un componente de error aleatorio, i.e,

$$\eta_i = x_i^T.\beta + Z_i.b_i + \varepsilon_i$$

Donde Z es la matriz de diseño para los efectos aleatorios, b es el vector de efectos aleatorios y ε es un error aleatorio.

Relación Chain Ladder – Modelos Lineales Generalizados

En el apartado que corresponde a Chain Ladder, se representó el método como un modelo estocástico sin suponer una distribución particular. Esta breve sección tiene como objetivo destacar que el mismo método puede ser representado en un marco de trabajo de modelos lineales generalizados mixtos, es decir, con una distribución supuesta. No se explicarán los modelos en detalle, ya que no es el objetivo de este trabajo, pero existen distribuciones para la cual se obtienen los mismos resultados para estimaciones puntuales que los arrojados por el método de Mack. Entre ellos, se pueden encontrar, el modelo con distribución de Poisson (con sobre-dispersión), con distribución Binomial Negativa, y con una aproximación Normal a la distribución Binomial Negativa. La diferencia entre estos y el propuesto por Mack radica en los supuestos utilizados, lo cual tiene un impacto directo en la varianza y los intervalos de confianza, como así se indica en Verral (2000), según lo citado en Gould (2008).

Análisis de resultados

Preparación de datos

En primer lugar, los datos refieren al desarrollo de siniestros del ramo automotores y fueron obtenidos de la Superintendencia de Seguros de la Nación (2018). Por normativa de la SSN, para coberturas distintas a Responsabilidad Civil, se requieren los últimos cinco períodos anuales de ocurrencia. Para el caso, y con el objetivo de comparar los datos reales de siniestralidad con las reservas estimadas, se utilizó la información que data del año 2007 hasta el año 2011 inclusive para los siniestros incurridos consignados en "daño total". Como resultado, se consigue una matriz $\mathbb{R}^{5x5}$ con los siniestros incurridos desarrollados en su totalidad, es decir, se cuenta con $L_{i,5}$ para todos los años de accidente $i = \{2007; 2008; 2009; 2010; 2011\}$ (Ver Tabla 1).

Año de ocurrencia	Año de desarrollo				
	1	2	3	4	5
2007	148813736.31	176967324.56	182322295.92	184720805.41	186305862.85
2008	214207675.40	243243446.33	245747844.78	246895033.93	250562475.09
2009	277701790.07	306262288.25	309431145.89	312180891.74	317792087.17
2010	379729035.32	418197058.72	425606329.84	427164647.12	429384416.69
2011	535269655.12	596996474.56	597771588.26	606862601.61	616582625.77

Tabla 1. Panel de datos con información completa de los siniestros incurridos acumulados desde el año 2007 hasta el año 2011, inclusive.

Luego, se separaron los datos de manera tal que la fecha de estimación de reservas refiera a la finalización del año 2011. En otras palabras, se obtiene un panel de datos triangular con los siniestros incurridos acumulados $L_{i,k}$ para $i + k \leq 6$ (Ver Tabla 2).

Año de ocurrencia	Año de desarrollo				
	1	2	3	4	5
2007	148813736.31	176967324.56	182322295.92	184720805.41	186305862.85
2008	214207675.40	243243446.33	245747844.78	246895033.93	
2009	277701790.07	306262288.25	309431145.89		
2010	379729035.32	418197058.72			
2011	535269655.12				

Tabla 2. Panel de datos triangular con información de los siniestros incurridos acumulados desde el año 2007 hasta el año 2011, inclusive.

Aplicación del método de Chain Ladder (Mack)

Tras aplicar el algoritmo, los resultados de las estimaciones puntuales de las reservas de IBNR y sus respectivos errores estándar son los que se observan en la Tabla 3.

Chain Ladder (Mack)		
Año de ocurrencia	IBNR	S.E
2008	2118564	1110334
2009	5240182	2425175
2010	13538073	5257910
2011	84594756	18165905
Total	105491575	20636534

Tabla 3. Resultados del método Chain Ladder según el modelo de Mack. Incluye las reservas de IBNR y el error estándar (S.E).

Aplicación de GLM

Para el caso de los Modelos Lineales Generalizados (mixtos), en el contexto de cálculo de reservas con triángulos, la variable dependiente es el valor de los siniestros incurridos acumulados y las variables explicativas son dos: el año de ocurrencia del accidente y el año de desarrollo.

Distribución de Poisson (con sobre-dispersión)

Como fue anteriormente descrito, si se presupone una distribución de Poisson (con sobre-dispersión) y, además, se utiliza la función de enlace logarítmica, se pueden alcanzar las mismas estimaciones puntuales que con el método de Mack. Esto no quiere decir que son el mismo modelo, de hecho, se corrobora en la Tabla 4 que el error estándar del modelo de Poisson difiere del modelo provisto por Mack.

GLM – Poisson (con sobre-dispersión)

Año de ocurrencia	IBNR	S.E
2008	2118564	3041521
2009	5240182	4673844
2010	13538073	7452503
2011	84594756	17841865
Total	105491575	25063287

Tabla 4. Resultados de GLM con distribución de Poisson (con sobre-dispersión) y función de enlace logarítmica. Incluye las reservas de IBNR y el error estándar (S.E).

Distribución Poisson-Gamma

Este modelo es el que refiere a la distribución de Tweedie con $1 < p < 2$. En este caso, el algoritmo estima el parámetro p a partir de los datos. Al igual que en el modelo de Poisson con sobre-dispersión, también se utilizó una función de enlace logarítmica. Luego, los resultados son los que se observan en la Tabla 5.

GLM – Poisson-Gamma

Año de ocurrencia	IBNR	S.E
2008	2112922	2500953
2009	5229850	3807633
2010	13529557	6068675
2011	84655040	14455772
Total	105527369	21963400

Tabla 5. Resultados de GLM con distribución compuesta Poisson-Gamma y función de enlace logarítmica. Incluye las reservas de IBNR y el error estándar (S.E).

Distribución Normal

En este caso, la distribución que se presupone es Normal y, también, función de enlace logarítmica. En la Tabla 6 se observan las estimaciones.

GLM – Normal		
Año de ocurrencia	IBNR	S.E
2008	2266907	9817338
2009	5378323	14379597
2010	13138109	19860850
2011	79536605	27496065
Total	100319944	62959542

Tabla 6. Resultados de GLM con distribución Normal y función de enlace logarítmica. Incluye las reservas de IBNR y el error estándar (S.E).

Validación con método de retención o Holdout

El método de validación por retención, también llamado Out of Sample Validation o, simplemente Holdout, consiste en realizar una partición del conjunto de datos en dos subconjuntos: uno que servirá para ajustar el modelo (datos de entrenamiento o training set) y otro para poner a prueba dicho modelo (datos de prueba o testing set). Por las características particulares que tiene la información en el contexto de cálculo de reservas con triángulos, se utilizan los datos del triángulo para modelar y, luego, la resta entre los siniestros incurridos acumulados con desarrollo completo y los siniestros incurridos hasta el año de estimación de reservas, i.e, $R_i = L_{i,K} - L_{i,K-i+1}$, constituyen los datos para evaluar los modelos obtenidos. Finalmente, la raíz del error cuadrático medio, medida calculada en el conjunto de datos de evaluación o, para simplificar notación, sólo RMSE, queda definido por

$$RMSE := \sqrt{\frac{\sum_{j \in i}\left(R_j - \widehat{R}_j\right)^2}{K}}$$

Donde $\widehat{R}_j$ es la estimación puntual de la reserva e i toma los valores que corresponden a los años de ocurrencia.

Los resultados con respecto al RMSE correspondiente a cada modelo ajustado se detallan en la Tabla 7.

Raíz del error cuadrático medio

Año de ocurrencia	Mack	Poisson	Poisson-Gamma	Normal
2008	692679	692679	695202	626338
2009	1395646	1395646	1400267	1333867
2010	1051272	1051272	1047463	872402
2011	1467659	1467659	1494619	794415
Total	4607256	4607256	4637551	3627022

Tabla 7. RMSE del modelo Mack y los GLM: Poisson, Poisson-Gamma y Normal.

Como se puede observar, el modelo lineal generalizado que presupone distribución normal con enlace logarítmico tuvo una predicción más acertada que el modelo de Mack. Aun así, el modelo de Chain Ladder sin supuesto de distribución tuvo mejor rendimiento en comparación al modelo de Poisson-Gamma y, teniendo en cuenta los errores estándar (Ver Tabla 3 y Tabla 4), al modelo de Poisson con sobre-dispersión.

Conclusiones

Cada uno de los modelos presentados en este trabajo tiene una serie de ventajas y desventajas que devienen en la elección de uno sobre otro sólo de forma subjetiva. Para este caso de estudio particular:

- El modelo de Mack tiene un desvío estándar menor con respecto a los modelos lineales generalizados.
- El modelo GLM con distribución Normal y enlace logarítmico tiene mejor predicción.

Como se introdujo en el marco teórico, el modelo de Mack es un modelo robusto, esto se debe al bajo desvío estándar con respecto a las estimaciones de las reservas. Este atributo se vio reflejado en este caso específico y, en consecuencia, responde a por qué es tan elegido por los organismos reguladores nacionales e internacionales. En principio, los modelos lineales generalizados también gozan de cierta robustez, pero ha quedado demostrado en este trabajo que, en un contexto de datos agregados por triángulo, es posible que no sea así.

Luego, a diferencia de los modelos lineales generalizados, el modelo de Mack admite valores siniestrales incrementales negativos. En el marco de GLM, es importante elegir una función de enlace que tenga en cuenta esta cuestión como, por ejemplo, la función de enlace identidad.

Finalmente, aunque las estimaciones de los modelos lineales generalizados son más flexibles, es posible determinar reservas con más precisión que en modelos de Chain Ladder, según la validación provista.

Futuras investigaciones y limitaciones

Debido a las características de la información que ofrecen los organismos regulatorios, esto es, en forma agregada y en triángulos, los modelos que pueden ajustarse son escasos:

- En el caso de reproducir una distribución con "Bootstrapping", no se puede validar el modelo con un subconjunto de información real.
- En la elección de modelos no paramétricos como, por ejemplo, bosques aleatorios o redes neuronales, no se presupone una distribución y, por lo tanto, se necesita una gran cantidad de observaciones (muchas más que en modelos paramétricos) para realizar estimaciones con cierto grado de precisión (James et. al., 2013). En consecuencia, es necesario obtener información individual de los siniestros.

Quedará pendiente para futuras investigaciones la aplicación de estos modelos con información que cumpla con los requisitos anteriormente mencionados. También es de interés la aplicación de modelos lineales generalizados en combinación con el método de frecuencia y severidad, pretendiéndose mejorar el rendimiento de las estimaciones puntuales de reservas.

Referencias

Brown, R.L. and Lennox, W.S. (2018). Supplement to chapter 3 of Introduction to Ratemaking and Loss Reserving for Property and Casualty Insurance. *ACTEX learning.*

Carrato et. al. (2020). ChainLadder. R package version 0.2.11. Recuperado de: https://cran.rproject.org/package=ChainLadder

Dobson, Annette J. (2002). An introduction to generalized linear models *Chapman & Hall/CRC*

Frees, E. (2009). Regression Modeling with Actuarial and Financial Applications (International Series on Actuarial Science). *Cambridge: Cambridge University Press.*

James et. Al. (2013). An introduction to statistical learning: with applications in R. *New York: Springer*.

Gould, I. L. (2008). Stochastic chain-ladder models in nonlife *insurance (Master's thesis, The University of Bergen)*.

Harnek, R. F. (1966). Formula Loss Reserves. *Insurance Accounting and Statistical Association Proceedings.*

Highleyman, W. (1962*)*. The Design and Analysis of Pattern Recognition Experiments. *Bell System Technical Journal.*

Mack, T. (1993). Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *Astin Bulletin Vol: 23*.

Mack, T. (1994). Measuring the Variability of Chain Ladder Reserve Estimates. *CAS Forum 1 (Spring): 101-82.*

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370-384.

Superintendecia de Seguros de la Nación (2018). Expediente: EX-2018-28639504-APN-GA#SSN. Desarrollo de Siniestros del Ramo Automotores - Ejercicio 2016/2017. Recuperado de: https://www.argentina.gob.ar/superintendencia-de-seguros/estadisticas/desarrollo-de-siniestros-del-ramo-automotores

Taylor, G. C. (1986). Claim Reserving in Non-Life Insurance.

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. *In Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. (Eds. J. K. Ghosh and J. Roy), pp. 579-604. Calcutta: Indian Statistical Institute.*

Verral, R. J. An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics 26, 91-99. 2000.*

Wickham et. al. (2018). Tidyverse. R package version 1.3.0. Recuperado de: https://cran.rproject.org/package=tidyverse

Anexo I

Dado que $\hat{f}_j = \frac{L_{i,j}}{L_{i,j-1}}$ es un estimador insesgado de f_j, entonces se puede representar el mismo de la siguiente manera:

$$\hat{f}_j = \sum_{i=1}^{K-j+1} w_{i,j} \cdot \frac{L_{i,j}}{L_{i,j-1}}$$

Donde los $w_{i,j}$ son ponderaciones tal que $w_{1,j} + \cdots + w_{K-j+1,j} = 1$.

Luego, debido al supuesto 2,

$$w_{i,j} = \frac{\frac{L_{i,j-1}}{\sigma_j^2}}{\sum_{i=1}^{K-j+1} \frac{L_{i,j-1}}{\sigma_j^2}} = \frac{L_{i,j-1}}{\sum_{i=1}^{K-j+1} L_{i,j-1}}$$

Finalmente,

$$\hat{f}_j = \sum_{i=1}^{K-j+1} \frac{L_{i,j-1}}{\sum_{i=1}^{K-j+1} L_{i,j-1}} \cdot \frac{L_{i,j}}{L_{i,j-1}} = \frac{\sum_{i=1}^{K-j+1} L_{i,j}}{\sum_{i=1}^{K-j+1} L_{i,j-1}}$$

Anexo II

Dado que

$\mathbb{P}(S_N \leq z) = e^{-\lambda} + \sum_{n=1}^{\infty} \mathbb{P}(N = n). \mathbb{P}(S_N \leq z)$, $z \geq 0$

Entonces,

$\mathbb{P}(S_N = 0) = \mathbb{P}(N = 0) = e^{-\lambda}$

Donde $N \sim Poisson(\lambda)$ y $Z_i \sim Gamma(\alpha, \gamma)$.

Anexo III

```
## Aplicaci□e GLM en Loss Reserving. ##
library(tidyverse)
library(ChainLadder)

## Preparaci□e datos.
data_2011 <-
readxl::read_excel("ssn_20162017_desarrollo_siniestros_automotores.xlsx",
sheet = 8)[15,c(7,11,15,19,23)]
data_2010 <-
readxl::read_excel("ssn_20162017_desarrollo_siniestros_automotores.xlsx",
sheet = 9)[15,c(7,11,15,19,23)]
data_2009 <-
readxl::read_excel("ssn_20162017_desarrollo_siniestros_automotores.xlsx",
sheet = 10)[15,c(7,11,15,19,23)]
data_2008 <-
readxl::read_excel("ssn_20162017_desarrollo_siniestros_automotores.xlsx",
sheet = 11)[15,c(7,11,15,19,23)]
data_2007 <-
readxl::read_excel("ssn_20162017_desarrollo_siniestros_automotores.xlsx",
sheet = 12)[15,c(7,11,15,19,23)]

matrix_data <- matrix(c(data_2007,data_2008,data_2009,data_2010,data_2011),
5, 5, byrow = TRUE)

rownames(matrix_data) <- c(2007:2011)
colnames(matrix_data) <- c(1:5)

true_data <- as.triangle(matrix_data)

model_data <- true_data

for(i in 5:2){
  for(j in (7-i):5){
    model_data[i,j] <- NA
  }
}

## M鴯do Chain-Ladder.
mack_reserves <- MackChainLadder(model_data)

## Chain-Ladder como caso particular de GLM. Over-dispersed Poisson, i.e,
tweedie p = 1.
od_poisson_reserves <- glmReserve(model_data)

## GLM - Distribuci□weedie con "p" estimado, donde 1 < p < 2, es decir, es
una distribuci□ompuesta Poisson/Gamma.
poisson_gamma_reserves <- glmReserve(model_data, var.power = NULL)

## GLM - Distribuci□ormal, i.e, tweedie p = 0.
gaussian_reserves <- glmReserve(model_data, var.power = 0)

## Validaci□or Out Of Sample validation en el contexto de Loss Reserving.
triangle_rmse <- function(true_data, predicted_data){
  triangle_rmse <- list("pointwise_rmse"=c(), "total_rmse" = "")
```

```
  triangle_rmse[[1]] <- seq(1,5)
  if(class(predicted_data)[1] == "MackChainLadder"){
    for(l in 1:5){
      triangle_rmse[[1]][l] <- sqrt((true_data[l,5]-
predicted_data[[3]][l,5])^2/5)
    }
  } else if (class(predicted_data) == "glmReserve"){
    for(l in 1:5){
      triangle_rmse[[1]][l] <- sqrt((true_data[l,5]-
predicted_data[[4]][l,5])^2/5)
    }
  } else{print("error. predicted data is not 'MackChainLadder' nor
'glmReserve'")}

  triangle_rmse[[2]] <- sum(triangle_rmse[[1]])

  triangle_rmse
}

mack_rmse <- triangle_rmse(true_data, mack_reserves)

odp_rmse <- triangle_rmse(true_data, od_poisson_reserves)

pg_rmse <- triangle_rmse(true_data, poisson_gamma_reserves)

gaussian_rmse <- triangle_rmse(true_data, gaussian_reserves)
```