

PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ

Escuela de Posgrado



Clasificación de riesgo para frecuencias y severidades en un seguro de
automóviles usando modelos GAMLSS

Tesis para obtener el grado académico de Magíster en Estadística
que presenta:

Diana Patricia Hernández Bello

Asesor:

Dr. Luis Hilmar Valdivieso Serrano

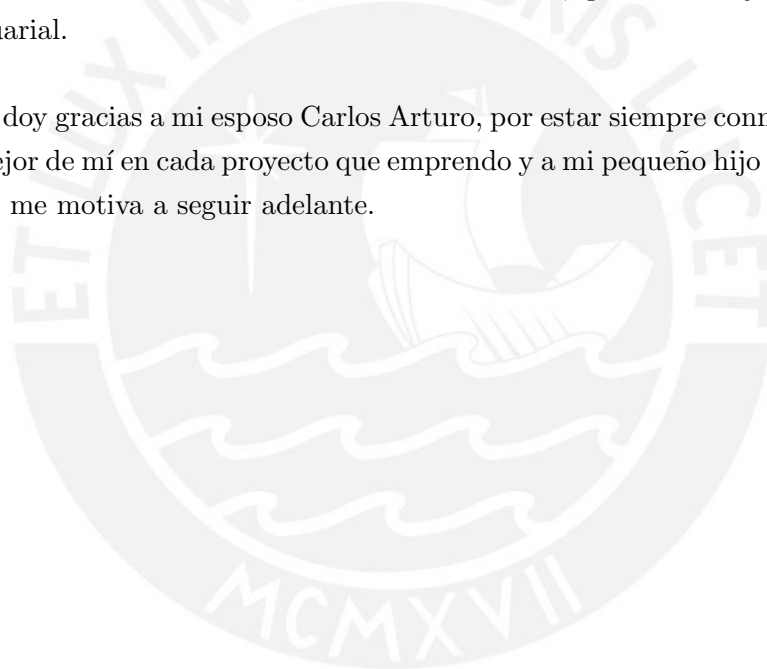
Lima, 2021

Agradecimientos

Quiero manifestar mi más sincero agradecimiento al profesor Dr. Luis Valdivieso por todo su apoyo, por sus comentarios y valiosos aportes que me llevaron a culminar uno de mis proyectos más retadores y que finalmente se sintetiza en esta tesis.

Agradezco a la Dra. Zaida Quiroz y el Dr. Christian Bayes por sus valiosas observaciones y comentarios. Por otra parte, agradezco a todos los profesores de la Maestría de Estadística de quienes aprendí temas totalmente nuevos e interesantes, que motivan y complementan mi profesión actuarial.

Finalmente le doy gracias a mi esposo Carlos Arturo, por estar siempre conmigo, apoyándome para dar lo mejor de mí en cada proyecto que emprendo y a mi pequeño hijo Alejandro Gabriel quien siempre me motiva a seguir adelante.



Resumen

En la tarificación de seguros generales, en particular en seguros de vehículos, es valioso incorporar toda la información disponible del asegurado, del bien asegurado y de los siniestros que se han presentado, con el fin de obtener modelos que consideren las variables relevantes en la estimación y así generar una prima de riesgo adecuada para el riesgo que se está analizando.

Los modelos a considerar están contruidos con base en las reclamaciones que ha presentado el asegurado y su estimación se obtiene mediante distribuciones del número y monto de siniestros dando como resultado tarifas que incluyen recargos y descuentos en base a la experiencia siniestral, lo que se conoce como Sistema Bonus-Malus. Adicionalmente se han analizado modelos de regresión que incluyen información tanto del asegurado como del vehículo y cuya estimación de la prima de riesgo se realiza a través de la media tanto de la frecuencia como de la severidad. Sin embargo, dado que los riesgos en la cartera expuesta son heterogéneos, se plantean también modelos de regresión en los que la estimación de la frecuencia y la severidad se realiza a través de parámetros como: la media, la varianza, el sesgo y la curtosis, estos últimos son denominados modelos aditivos generalizados de localización, escala y forma (GAMLSS).

Palabras-clave: Seguros de vehículos, Sistema Bonus–Malus, GAMLSS.

Abstract

In the pricing of general insurance, particularly in vehicle insurance, it is valuable to incorporate all available information about the insured, the insured goods and the claims that have been filed, in order to obtain models that consider the relevant variables in the estimation and thus generate a risk premium that is appropriate to the risk being analyzed.

The models to be considered are built based on the claims that the insured has presented and their estimation is obtained through distributions of the number and amount of the claims, resulting in rates that include surcharges and discounts based on the experience of the claims, known as the Bonus-Malus System. In addition, regression models have been analyzed that include both insured and vehicle information and whose risk premium is estimated through the mean of both frequency and severity. However, since the risks of the exposed portfolio are heterogeneous, regression models are also proposed in which the estimation of frequency and severity is done through parameters such as: mean, variance, skewness and kurtosis, they are known as generalized additive models of location, scale and shape (GAMLSS).

Keywords: Automobile insurance, Bonus–Malus System, GAMLSS.

Índice general

Índice de figuras	VII
Índice de cuadros	VIII
1. Introducción	1
1.1. Consideraciones preliminares	1
1.2. Objetivos	2
1.3. Organización del trabajo	2
2. Conceptos y modelos con variables a posteriori	3
2.1. Introducción	3
2.2. Modelo BMS para la frecuencia y la severidad basados en información a posteriori	4
2.2.1. Frecuencias bajo el modelo binomial negativo	5
2.2.2. Frecuencias bajo el modelo de Sichel	7
2.2.3. Frecuencias bajo el modelo Poisson inversa gaussiana	10
2.2.4. Severidad bajo el modelo de Pareto	11
2.2.5. Estimación de la prima de acuerdo al principio de prima neta	13
3. Modelos BMSG y GAMLSS con información a priori	15
3.1. Modelos BMSG para la frecuencia y severidad basados en información a priori y a posteriori	15
3.1.1. Frecuencia bajo el modelo binomial negativo	16
3.1.2. Frecuencia bajo el modelo de Sichel	17
3.1.3. Severidad bajo el modelo Pareto	18
3.1.4. Estimación de la prima neta usándose un BMSG	20
3.2. Modelos BMSG GAMLSS	20
3.2.1. Frecuencia bajo el modelo binomial negativo tipo II	21
3.2.2. Frecuencia bajo el modelo de Delaporte	22
3.2.3. Frecuencia bajo el modelo de Sichel	23
3.2.4. Frecuencia bajo el modelo Zero-inflated Poisson (ZIP)	24
3.2.5. Severidad bajo el modelo Gamma	25
3.2.6. Severidad bajo el modelo Weibull	25
3.2.7. Severidad bajo el modelo de Weibull tipo III	26
3.2.8. Severidad bajo el modelo Gamma generalizado	27
3.2.9. Severidad bajo el modelo Pareto generalizado	27

3.3. Estimación de primas bajo los principios del valor esperado y de la desviación estándar	28
4. Aplicación	29
4.1. Descripción de los datos	29
4.1.1. Información a posteriori	30
4.1.2. Información a priori	32
4.2. Modelos BMS con información a posteriori	36
4.2.1. Modelo binomial negativo	37
4.2.2. Modelo Poisson inverse Gaussian (PIG)	37
4.2.3. Modelo de Frecuencia de Sichel	38
4.2.4. Modelo de severidad de Pareto	39
4.3. Modelo BMSG con variables a priori y posteriori	40
4.3.1. Frecuencia bajo el modelo Binomial Negativo	41
4.3.2. Frecuencia bajo el modelo Sichel	44
4.3.3. Resultados de los modelos BMSG con variables a priori	47
4.4. Modelos BMSG-GAMLSS basados en información a priori	48
4.4.1. Frecuencia bajo el modelo NBII	48
4.4.2. Frecuencia bajo el modelo Delaporte	50
4.4.3. Frecuencia bajo el modelo de Sichel	52
4.4.4. Severidad bajo el modelo Gamma	54
4.4.5. Severidad bajo el modelo Weibull	56
4.4.6. Severidad bajo el modelo Gamma generalizada	57
4.5. Resultados de los modelos BMSG GAMMLS	58
5. Conclusiones	61
5.1. Conclusiones	61
5.2. Sugerencia para investigaciones futuras	61
A. Código R	62
Bibliografía	82

Índice de figuras

4.1. Histograma y gráfico de los valores ordenados del número de siniestros	30
4.2. Diagrama de caja del número de siniestros por asegurado	31
4.3. Histograma y gráfico de los valores ordenados del monto total de siniestros	32
4.4. Diagrama de caja del monto de siniestro por asegurado	32
4.5. Histogramas de edad asegurado, edad del vehículo y suma asegurada	34
4.6. Diagramas de caja de siniestros de acuerdo a las variables categóricas región, marca, uso y grupo.	34
4.7. Interacción de variables cuantitativas y categóricas del asegurado en el periodo de t años	35
4.8. Interacción de variables cuantitativas y categóricas del vehículo en el periodo de t años	35
4.9. Diagrama de dispersión multivariado considerando las variables número de siniestros, edad del asegurado, edad del vehículo y suma asegurada	36
4.10. Histograma del número de siniestros y curva de la distribución binomial negativa ajustada.	37
4.11. Distribución del número de siniestros y ajuste de la curva $PIG(\hat{\alpha}, \hat{\xi}, -1/2)$	38
4.12. Histograma del monto total siniestros y curva de la distribución pareto ajustada.	39
4.13. Worm plot de los residuales modelo Binomial Negativa por periodo $j = 1, 2, 3$	43
4.14. Gráficos de residuales del modelo Sichel para $j = 1, 2, 3$	46
4.15. Worm plot de los residuales modelos NBII por periodo $j = 1, 2, 3$	49
4.16. Gráfico de residuales del modelo Delaporte para $j = 1, 2, 3$	51
4.17. Worm plot del modelo de severidad distribución Gamma para el tercer año	55
4.18. Worm plot del modelo de severidad distribución Weibull para el tercer año	57
4.19. Worm plot del modelo de severidad distribución Gamma Generalizada para el tercer año	58

Índice de cuadros

4.1. Variables a priori	33
4.2. Prima neta en unidades de 100 dólares	40
4.3. Coeficientes de regresión Modelo Binomial Negativa para los periodos $j = 1, 2, 3$	42
4.4. Criterios del Modelo Binomial Negativa para los periodos $j = 1, 2, 3$	42
4.5. Resumen por año de los residuales por periodo $j = 1, 2, 3$	43
4.6. Coeficientes de regresión Modelo Sichel para los periodos $j = 1, 2, 3$	44
4.7. Criterios del Modelo Sichel para los periodos $j = 1, 2, 3$	45
4.8. Resumen de los residuales por periodo j bajo modelo Sichel	47
4.9. Tabla de segmentos	47
4.10. Frecuencias obtenidas modelando μ con los modelos Binomial Negativa y Sichel	48
4.11. Coeficientes de regresión Modelo GAMLSS Binomial negativo II para los periodos $j = 1, 2, 3$	48
4.12. Resumen por año de los residuales modelo NBII por periodo $j = 1, 2, 3$	49
4.13. Coeficientes de regresión Modelo GAMLSS Delaporte para los periodos $j = 1, 2, 3$	50
4.14. Resumen por año de los residuales modelo Delaporte por periodo $j = 1, 2, 3$	50
4.15. Coeficientes de regresión Modelo GAMLSS Sichel II para los periodos $j = 1, 2, 3$	52
4.16. Resumen por año de los residuales modelo Sichel por periodo $j = 1, 2, 3$	52
4.17. Comparación de los modelos de frecuencia por año	53
4.18. Coeficientes de regresión Modelo GAMMA para los periodos $j = 1, 2, 3$	54
4.19. Resumen por año de los residuales modelo Gamma por periodo $j = 1, 2, 3$	55
4.20. Coeficientes de regresión Modelo Weibull para los periodos $j = 1, 3$	56
4.21. Resumen por año de los residuales modelo Weibull por periodo $j = 1, 3$	56
4.22. Coeficientes de regresión Modelo Gamma Generalizada para los periodos $j = 1, 2, 3$	57
4.23. Resumen por año de los residuales modelo Gamma Generalizada por periodo $j = 1, 2, 3$	58
4.24. Comparación de los modelos de severidad por año	58
4.25. Comparación de los modelos de frecuencia por año	59
4.26. Primas para el tercer año con el modelo de frecuencia <i>NBII</i> y los diferentes modelos de severidad en unidades de 100 dólares	59
4.27. Primas para el tercer año con el modelo de frecuencia <i>DEL</i> y los diferentes modelos de severidad en unidades de 100 dólares	60

4.28. Primas para el tercer año con el modelo de frecuencia *SICH* y los diferentes
modelos de severidad en unidades de 100 dólares 60



Capítulo 1

Introducción

1.1. Consideraciones preliminares

En el ámbito actuarial, el determinar una tarifa de seguro adecuada para cada uno de los clientes de un portafolio, considerando el nivel de riesgo que cada uno de ellos representa, es una de las tareas que demanda mayor análisis y que en muchos casos se simplifica utilizando modelos agregados. Por esta razón, en diferentes tipos de seguros, los modelos aplicados han ido evolucionando a fin de utilizar la mayor cantidad de información disponible que agregue valor y permita construir una tarifa adecuada que refleje el riesgo de cada asegurado.

En el caso de seguros de automóviles se utilizan modelos de tarificación basados en la experiencia de siniestros del individuo, uno de estos es el denominado sistema de Bonus–Malus, donde se otorga un premio (descuento) o castigo (recargo) al asegurado sobre la tarifa del seguro. Este descuento o recargo se obtiene de acuerdo con la experiencia siniestral del asegurado, en el periodo en el que ha estado expuesto dentro del portafolio de seguros.

Dentro de la literatura, de los sistemas Bonus–Malus (BMS) para seguros de automóviles, se acostumbra plantear modelos que basan esta segmentación considerando únicamente el número de siniestros (o la frecuencia de siniestros), dejando de lado el componente de la magnitud de la pérdida (severidad). Este último componente es importante, pues permite realizar una mejor segmentación de los asegurados, dado que la tarifa no tendrá incrementos abruptos para asegurados que presentan siniestros de costo bajo y los riesgos deseables permanecerán en el portafolio.

Así mismo, dentro de los BMS es importante y enriquecedor, complementar la información a posteriori del asegurado (frecuencia y severidad) con información a priori, que permite realizar la clasificación del riesgo, incorporando características del asegurado y del vehículo, por ejemplo: edad, sexo y lugar de residencia del asegurado, año y tipo del vehículo, etc.

En el artículo propuesto por Frangos y Vrontos (2001), se contempla un sistema Bonus–Malus Generalizado, que integra tanto la información a posteriori (frecuencia y severidad) como la información a priori (características del individuo y del vehículo), en el cálculo de la prima o tarifa del seguro de automóviles.

Por otro lado, en la mayoría de los modelos se plantea la estimación de la tarifa a través de la media tanto para la frecuencia como para la severidad. Sin embargo, estos modelos implícitamente proveen un modelo para la escala y la forma, que pueden no ser totalmente adecuados dado que no incorporan la heterogeneidad del riesgo que tiene el portafolio. Es por ello que Tzougas et al. (2015), abordan la construcción del modelo de tarificación incorporando los principios del valor esperado y de la varianza para la frecuencia y la severidad.

Considerando el contexto del mercado asegurador peruano, que cuenta con información de seguros de automóviles de varios años y dado que estos seguros están en constante crecimiento, sería de utilidad aplicar estos modelos a un portafolio de seguros real para ver como se realizaría su aplicación y presentar resultados con datos reales. Esto permitirá contribuir al desarrollo de la elaboración de modelos de tarificación más robustos.

1.2. Objetivos

El objetivo general de la tesis es estudiar la clasificación del riesgo en un seguro de automóviles a través de los sistemas Bonus-Malus y Bonus-Malus Generalizado incluyendo el análisis de la media y varianza de la frecuencia y severidad de los siniestros a través de un modelo aditivo generalizado de localización, escala y forma (GAMLSS) y aplicarlo a un portafolio de seguro de automóviles de una compañía de seguros peruana.

Los objetivos específicos son:

- Revisar la literatura sobre los diferentes modelos Bonus - Malus propuestos.
- Estudiar las propiedades de éstos modelos.
- Implementar en R la estimación de los modelos, con ayuda de los modelos GAMLSS.
- Aplicar el modelo a un conjunto de datos locales.

1.3. Organización del trabajo

En el capítulo 2 se presentan los conceptos preliminares y los modelos del sistema Bonus - Malus, sobre la frecuencia y la severidad basado en información a posteriori. En el capítulo 3 se desarrollarán los sistemas Bonus - Malus de la frecuencia y la severidad basados en información a priori y a posteriori y los modelos de regresión para localización, escala y forma de la frecuencia y de la severidad. Finalmente en el capítulo 4 mostraremos una aplicación de los modelos estudiados a la cartera de un seguro de automóviles del medio local.

Capítulo 2

Conceptos y modelos con variables a posteriori

2.1. Introducción

Las compañías aseguradoras determinan la tarifa del seguro (prima) mediante modelos que combinan la clasificación del riesgo del individuo y la experiencia siniestral. Es decir, antes de que un individuo ingrese al portafolio asegurado, no se conoce cuál será su comportamiento al conducir, ni cual será la verdadera distribución de los accidentes que este tendrá, por lo que en una primera etapa solo se pueden considerar variables a priori dentro del modelo de tarificación. Una vez que el individuo hace parte de la cartera de asegurados, se evidenciará en sus siniestros cual es su comportamiento al conducir.

Generalmente, cuando se tiene construida la tarifa generada con variables a priori, se aplica un factor de recargo, en base a la siniestralidad observada total del portafolio. Esta metodología podría ser inadecuada, dado que combina una tarificación específica para la categoría de riesgo de cada asegurado con un factor de siniestralidad global de la cartera, y no se estaría reflejando el aporte del riesgo individual de cada uno de los asegurados al portafolio.

En este capítulo se presentan algunos conceptos y consideraciones que serán usadas en el desarrollo del modelo de clasificación de riesgo de frecuencia y severidad, aplicado al seguro de automóviles. Se abordarán los modelos clásicos de frecuencia y severidad a través del sistema Bonus - Malus aplicando inferencia bayesiana.

Los modelos a detallar serán los siguientes:

- Modelo Sistema Bonus – Malus para la frecuencia basado en información a posteriori (Modelo BMS de frecuencia).
- Modelo Sistema Bonus – Malus para la severidad basado en información a posteriori (Modelo BMS de severidad).

Teniendo en consideración que abordaremos la tarificación de un producto de seguros, a continuación, se presentan algunos términos clave, los cuales permitirán delimitar el entorno en el que se desarrollarán y aplicarán los modelos a estudiar. Así mismo, algunos de estos términos se ampliarán posteriormente obteniendo las fórmulas relacionadas.

Definición 2.1.1. La prima neta o prima de riesgo, es el monto necesario que establece la compañía de seguros para dar cobertura únicamente por el riesgo que se quiere asegurar, no incluye recargos por gastos de comercialización o gastos de gestión, ni margen de utilidad. Generalmente se obtiene como el producto de la frecuencia y la severidad, bajo el supuesto que estos dos componentes son independientes.

Definición 2.1.2. La frecuencia de los siniestros se define como el cociente (ratio) entre el número de siniestros que ha presentado un asegurado o un portafolio asegurado y su exposición. La exposición se mide en unidades de tiempo.

Cada uno de los siniestros que se han presentado, tienen un monto de pérdida asociado, lo que se conoce como severidad del siniestro.

Definición 2.1.3. Las variables a priori o de tarificación son aquellas que se conocen previamente, es decir, en el momento en que el asegurado manifiesta su intención de asegurarse y brinda información a la compañía de seguros, Estas permiten clasificar a los asegurados en grupos homogéneos.

Por otro lado, las variables que permiten evidenciar el comportamiento del individuo al conducir se conocen como variables a posteriori.

Definición 2.1.4 (Sistema Bonus–Malus). El sistema de Bonus–Malus (BMS) es un sistema de tarificación basado en la experiencia de siniestros de un individuo o grupo asegurado, que busca penalizar a los asegurados que presenten uno o más accidentes, recargando su tarifa (maluses) o premiando a los asegurados que no tienen reclamos, otorgándoles descuentos (bonuses).

Una observación importante es que se debe procurar que el sistema Bonus–Malus sea óptimo, es decir, que sea justo para el asegurado y financieramente balanceado para la compañía. Será justo, si en la renovación el asegurado debe pagar una prima acorde a la propia frecuencia y severidad de sus reclamos y se incorpora su historia pasada, y será financieramente balanceado, si los incrementos o reducciones en la prima efectuada a los asegurados en el portafolio garantizan el equilibrio total de la prima.

Los modelos que se abordaran inicialmente contemplan la estimación de la frecuencia y la severidad, a partir del comportamiento de siniestros que ha presentado el asegurado. Para esto se utilizarán las distribuciones tanto para las variables aleatorias definidas como para alguno de sus parámetros. Estos modelos permitirán, con base en la información histórica del asegurado, inferir cual será la tarifa adecuada, de acuerdo al comportamiento que este presente durante la duración del contrato de seguro.

2.2. Modelo BMS para la frecuencia y la severidad basados en información a posteriori

Los modelos construidos con información a posteriori se construyen únicamente con base en la información de siniestros. Por lo que este primer análisis, consiste en estudiar las variables de interés para la derivación de la prima neta, es decir, la frecuencia y la severidad de siniestros.

2.2.1. Frecuencias bajo el modelo binomial negativo

El modelo de Poisson, es uno de los más utilizados en la literatura para datos de conteo. En este, la v.a. K , que representa el número de reclamos, tiene condicionalmente como función de probabilidad a:

$$P(K = k | \Lambda = \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots,$$

donde la v.a. $\Lambda > 0$, denota el riesgo subyacente de cada asegurado a tener un accidente. Si se asume que este riesgo es una v.a. $\Lambda \sim \text{Gamma}(a, \tau)$ con función de densidad

$$u(\lambda) = \frac{\lambda^{a-1} \tau^a \exp(-\tau \lambda)}{\Gamma(a)},$$

siendo $a > 0$ y $\tau > 0$; se tiene entonces el resultado siguiente.

Proposición 2.2.1. *Si $K | \Lambda = \lambda \sim \text{Poisson}(\lambda)$ y $\Lambda \sim \text{Gamma}(a, \tau)$, entonces la distribución marginal del número de reclamos K tendrá como función de probabilidad a*

$$P(K = k) = \frac{\Gamma(a+k)}{\Gamma(a)k!} \left(\frac{\tau}{1+\tau} \right)^a \left(\frac{1}{1+\tau} \right)^k, \quad k = 0, 1, 2, \dots,$$

con media $E(K) = a/\tau$ y varianza $\text{Var}(K) = (a/\tau)(1 + 1/\tau)$. Esta distribución es conocida como binomial negativa de parámetros $a > 0$ y $\tau > 0$ y se le denota por $K \sim \text{BN}(a, \tau)$.

Demostración. Sea $K | \Lambda = \lambda \sim \text{Poisson}(\lambda)$ y $\Lambda \sim \text{Gamma}(a, \tau)$, entonces tenemos que condicionando a Λ , la distribución marginal de K es

$$\begin{aligned} P(K = k) &= \int_0^\infty \left(\frac{e^{-\lambda} \lambda^k}{k!} \right) \cdot \left(\frac{\lambda^{a-1} \tau^a e^{-\tau \lambda}}{\Gamma(a)} \right) d\lambda \\ &= \frac{\tau^a}{\Gamma(a)k!} \int_0^\infty \lambda^{a+k-1} e^{-\lambda(1+\tau)} d\lambda \\ &= \frac{\tau^a}{\Gamma(a)k!} \cdot \frac{\Gamma(a+k)}{(1+\tau)^{a+k}} \cdot \int_0^\infty \frac{(1+\tau)^{a+k}}{\Gamma(a+k)} \lambda^{a+k-1} e^{-\lambda(1+\tau)} d\lambda \\ &= \frac{\tau^a}{\Gamma(a)k!} \cdot \frac{\Gamma(a+k)}{(1+\tau)^{a+k}}. \end{aligned}$$

Así,

$$P(K = k) = \frac{\tau^a}{\Gamma(a)k!} \cdot \frac{\Gamma(a+k)}{(1+\tau)^{a+k}} = \frac{\Gamma(a+k)}{\Gamma(a)k!} \cdot \left(\frac{\tau}{1+\tau} \right)^a \left(1 - \frac{\tau}{1+\tau} \right)^k.$$

Esta es la función de densidad de una distribución binomial negativa con parámetros (a, τ) . \square

Este resultado nos indica que el número de siniestros sigue una distribución binomial negativa bajo el supuesto que el riesgo implícito latente de presentar siniestros se distribuya como una gamma.

Bajo esta consideración y dado que poseemos información del asegurado, entonces podemos determinar la frecuencia de siniestros del asegurado durante el periodo de tiempo que ha estado expuesto. Definimos, $K' = \sum_{j=1}^t K_j$ como el número total de reclamos observado de un asegurado en un periodo de t años, donde K_j es el número de reclamos que el individuo ha efectuado en el periodo j .

Por el teorema de Bayes, podemos obtener la siguiente función de densidad a posteriori de Λ , para un asegurado o grupo de asegurados dado su historial de reclamos observado k_1, \dots, k_t , donde se asume que condicionalmente al nivel de riesgo, los reclamos de los asegurados son independientes entre si.

Proposición 2.2.2. *La función de densidad a posteriori de Λ es*

$$\pi(\lambda | k_1, \dots, k_t) = \frac{(\tau + t)^{k'+a} \lambda^{k'+a-1} e^{-(t+\tau)\lambda}}{\Gamma(a + k')},$$

donde $k' = \sum_{j=1}^t k_j$. Esta corresponde a la función de densidad de una v.a. $\text{Gamma}(a+k', t+\tau)$.

Demostración. Sea $K | \Lambda = \lambda \sim \text{Poisson}(\lambda)$ y $\Lambda \sim \text{Gamma}(a, \tau)$, entonces por el teorema de Bayes

$$\begin{aligned} \pi(\lambda | k_1, \dots, k_t) &= \pi(\lambda) \frac{P(K_1 = k_1, \dots, K_t = k_t | \Lambda = \lambda)}{P(K_1 = k_1, \dots, K_t = k_t)} \\ &= \left(\frac{\lambda^{a-1} \tau^a \exp(-\tau\lambda)}{\Gamma(a)} \right) \left(\lambda^{\sum_{j=1}^t k_j} e^{-t\lambda} \right) c(k_1, \dots, k_t, a, \tau) \\ &= \tau^a \cdot \lambda^{a + \sum_{j=1}^t k_j - 1} \exp(-(t + \tau)\lambda) \cdot c(k_1, \dots, k_t, a, \tau), \end{aligned}$$

donde $c(k_1, \dots, k_t, a, \tau)$ es un término que no depende de λ . Por lo tanto,

$$\Lambda | K_1 = k_1, \dots, K_t = k_t \sim \text{Gamma}(a + k', t + \tau).$$

□

Con el objeto de incorporar el conocimiento que tenemos con los reclamos observados, es necesario calcular el mejor estimador del riesgo del asegurado en el año $t + 1$, usando la experiencia de los t años anteriores. Este estimador lo denotaremos como $\hat{\Lambda}_{t+1}$.

En el contexto bayesiano y usándose una función de pérdida cuadrática, es conocido que la elección óptima de $\hat{\Lambda}_{t+1}$ para un asegurado con historia de reclamos K_1, \dots, K_t es la media de la distribución a posteriori de Λ ; es decir,

$$\begin{aligned}\hat{\Lambda}_{t+1}(K_1, \dots, K_t) &= \frac{a + K'}{t + \tau} \\ &= \bar{\lambda} \left(\frac{a + K'}{a + t\bar{\lambda}} \right),\end{aligned}\quad (2.1)$$

donde $\bar{\lambda} = a/\tau$.

Este resultado nos permitirá obtener la estimación de la prima, basada en las frecuencias con base en información previa del asegurado.

2.2.2. Frecuencias bajo el modelo de Sichel

Siguiendo lo expuesto por Tzougas y Frangos (2014) realizamos la siguiente extensión para el modelo de frecuencias.

Dado que el portafolio es heterogéneo, pues todos los asegurados tienen un riesgo subyacente diferente de tener un accidente, y bajo el supuesto que $K \mid \Lambda = \lambda$, se distribuye de acuerdo a una $Poisson(\lambda)$, asumimos ahora que el parámetro Λ sigue una distribución gaussiana inversa generalizada, denotada por $GIG(\mu, \sigma, \nu)$, con función de densidad:

$$\pi(\lambda) = \frac{\left(\frac{c}{\mu}\right)^\nu \lambda^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda + \frac{\mu}{c}\frac{1}{\lambda}\right)\right]}{2B_\nu\left(\frac{1}{\sigma}\right)}, \quad \lambda > 0 \quad (2.2)$$

donde $\mu > 0$, $\sigma > 0$, $-\infty < \nu < \infty$ y $c = \frac{B_{\nu+1}\left(\frac{1}{\sigma}\right)}{B_\nu\left(\frac{1}{\sigma}\right)}$, siendo

$$B_\nu(z) = \frac{1}{2} \int_0^\infty x^{\nu-1} \exp\left[-\frac{1}{2}z\left(x + \frac{1}{x}\right)\right] dx,$$

la función de Bessel modificada de tercera especie de orden ν con argumento z .

La media y varianza de Λ está dada por $E(\Lambda) = \mu$ y

$$Var(\Lambda) = \mu^2 \left(\frac{2\sigma(\nu+1)}{c} + \frac{1}{c^2} - 1 \right).$$

Proposición 2.2.3. Si $K \mid \Lambda = \lambda \sim Poisson(\lambda)$ y $\Lambda \sim GIG(\mu, \sigma, \nu)$, entonces la distribución marginal del número de reclamos K , tiene como función de probabilidad a

$$P(K = k) = \frac{\left(\frac{\mu}{c}\right)^k B_{k+\nu}(\alpha)}{k!(\alpha\sigma)^{k+\nu} B_\nu\left(\frac{1}{\sigma}\right)}, \quad k = 0, 1, 2, \dots \quad (2.3)$$

Esta es conocida como la distribución de Sichel de parámetros μ, σ, ν , y se le denota por $K \sim Sichel(\mu, \sigma, \nu)$.

Demostración. Si consideramos la parametrización

$$\mu = \frac{\alpha\theta}{2\sqrt{1-\theta}} \cdot \frac{B_{\nu+1}(\alpha\sqrt{1-\theta})}{B_\nu(\alpha\sqrt{1-\theta})} \quad \text{y} \quad \sigma = \frac{1}{\alpha\sqrt{1-\theta}}$$

obtenemos que la función de densidad de la distribución inversa gaussiana generalizada se puede expresar como

$$\pi(\lambda) = \frac{(1 - \theta)^{\nu/2} \left(\frac{2}{\alpha\theta}\right)^\nu \lambda^{\nu-1}}{2B_\nu(\alpha\sqrt{1-\theta})} \exp \left[\left(1 - \frac{1}{\theta}\right) \lambda - \frac{\alpha^2\theta}{4\lambda} \right]$$

donde $\alpha > 0$, $0 < \theta < 1$, $-\infty < \nu < \infty$ y $B_\nu(\cdot)$ es la función de Bessel de tercera especie. Hallamos ahora la distribución marginal de K

$$P(K = k) = \frac{(1 - \theta)^{\nu/2}}{B_\nu(\alpha\sqrt{1-\theta})} \cdot \frac{(\alpha\theta/2)^k}{k!} B_{\nu+k}(\alpha). \tag{2.4}$$

Veamos primero que las ecuaciones (2.3) y (2.4) son equivalentes.

$$P(K = k) = \frac{\left(\frac{\mu}{c}\right)^k B_{k+\nu}(\alpha)}{k!(\alpha\sigma)^{k+\nu} B_\nu\left(\frac{1}{\sigma}\right)}$$

usando la parametrización correspondiente y teniendo en cuenta que

$$c = \frac{B_{\nu+1}(\alpha\sqrt{1-\theta})}{B_\nu(\alpha\sqrt{1-\theta})},$$

reemplazamos en la ecuación (2.3)

$$\begin{aligned} P(K = k) &= \frac{\left(\frac{\alpha\theta}{2\sqrt{1-\theta}} \cdot \frac{B_{\nu+1}(\alpha\sqrt{1-\theta})}{B_\nu(\alpha\sqrt{1-\theta})} \right)^k B_{k+\nu}(\alpha)}{k! \left(\alpha \frac{1}{\alpha\sqrt{1-\theta}} \right)^{k+\nu} B_\nu(\alpha\sqrt{1-\theta})} \\ &= \frac{\left(\frac{\alpha\theta}{2\sqrt{1-\theta}} \right)^k}{k! \left(\frac{1}{\sqrt{1-\theta}} \right)^{k+\nu}} \cdot \frac{1}{B_\nu(\alpha\sqrt{1-\theta})} \cdot B_{k+\nu}(\alpha) \\ &= \frac{(1 - \theta)^{\nu/2}}{B_\nu(\alpha\sqrt{1-\theta})} \cdot \frac{(\alpha\theta/2)^k}{k!} B_{\nu+k}(\alpha). \end{aligned}$$

De esta forma, obtenemos la equivalencia. Ahora procedemos a probar el resultado principal,

$$\begin{aligned} P(K = k) &= \int_0^\infty \frac{e^{-\lambda} \lambda^k}{k!} \cdot \frac{(1 - \theta)^{\nu/2} \left(\frac{2}{\alpha\theta}\right)^\nu}{2B_\nu(\alpha\sqrt{1-\theta})} \cdot \lambda^{\nu-1} \exp \left[\left(1 - \frac{1}{\theta}\right) \lambda - \frac{\alpha^2\theta}{4\lambda} \right] d\lambda \\ &= \frac{(1 - \theta)^{\nu/2} \left(\frac{2}{\alpha\theta}\right)^\nu}{2B_\nu(\alpha\sqrt{1-\theta})k!} \cdot \int_0^\infty \lambda^{k+\nu-1} \exp \left[-\lambda + \left(1 - \frac{1}{\theta}\right) \lambda - \frac{\alpha^2\theta}{4\lambda} \right] d\lambda \\ &= \frac{(1 - \theta)^{\nu/2} \left(\frac{2}{\alpha\theta}\right)^\nu}{2B_\nu(\alpha\sqrt{1-\theta})k!} \cdot \int_0^\infty \lambda^{k+\nu-1} \exp \left[-\frac{\lambda}{\theta} - \frac{\alpha^2\theta}{4\lambda} \right] d\lambda. \end{aligned}$$

Aplicando la sustitución $u = \frac{2\lambda}{\alpha\theta}$, $\frac{\alpha\theta}{2} du = d\lambda$, además, si $\lambda = 0$, entonces $u = 0$ y conforme

λ tiene a infinito, u también tiende a infinito, de esta forma,

$$\begin{aligned} P(K = k) &= \frac{(1 - \theta)^{\nu/2} \left(\frac{\alpha\theta}{2}\right)^{-\nu}}{2B_\nu(\alpha\sqrt{1 - \theta})k!} \cdot \int_0^\infty \left(\frac{\alpha\theta u}{2}\right)^{k+\nu-1} \exp\left[-\frac{\alpha u}{2} - \frac{\alpha^2}{4} \frac{2}{\alpha u}\right] \cdot \frac{\alpha\theta}{2} du \\ &= \frac{(1 - \theta)^{\nu/2} \left(\frac{\alpha\theta}{2}\right)^{-\nu}}{2B_\nu(\alpha\sqrt{1 - \theta})k!} \cdot \left(\frac{\alpha\theta}{2}\right)^{k+\nu} \int_0^\infty u^{k+\nu-1} \exp\left[-\frac{1}{2}\alpha\left(u + \frac{1}{u}\right)\right] du \\ &= \frac{(1 - \theta)^{\nu/2} \left(\frac{\alpha\theta}{2}\right)^k}{B_\nu(\alpha\sqrt{1 - \theta})k!} B_{k+\nu}(\alpha) \end{aligned}$$

□

Si $K \sim \text{Sichel}(\mu, \sigma, \nu)$, entonces la media de K es igual a $E(K) = \mu$ y su varianza igual a

$$\text{Var}(K) = \mu + \mu^2 \left(\frac{2\sigma(\nu + 1)}{c} + \frac{1}{c^2} - 1 \right).$$

Al igual que en el modelo con frecuencia binomial negativo, se busca determinar la frecuencia dados los reclamos observados del asegurado. Para esto consideramos un individuo con una experiencia de t períodos, y sean:

- k_j el número de reclamos observados del asegurado en el periodo j con $j = 1, 2, \dots, t$, que son realizaciones de las variables aleatorias respectivas K_j . Dado un nivel de riesgo, estas últimas se asumen independientes.
- $K' = \sum_{j=1}^t K_j$ el número total de reclamos del asegurado en los t períodos.
- $K_j \mid \Lambda = \lambda$ se distribuye Poisson(λ), para $j = 1, 2, \dots, t$.

Se cumple entonces el siguiente resultado.

Proposición 2.2.4. *La distribución a posteriori de Λ para un asegurado dado su historial de reclamos k_1, \dots, k_t es $GIG(w_1, w_2, k' + \nu)$, donde $k' = \sum_{j=1}^t k_j$, $w_1 = \frac{c}{\sigma\mu} + 2t$, $w_2 = \frac{\mu}{\sigma c}$,*

$$\sigma > 0, \text{ y } c = \frac{B_{\nu+1}\left(\frac{1}{\sigma}\right)}{B_\nu\left(\frac{1}{\sigma}\right)}.$$

Demostración. Dado que $K \mid \Lambda \sim \mathcal{P}(\lambda)$ y $\Lambda \sim GIG(\mu, \sigma, \nu)$, entonces

$$\begin{aligned} \pi(\Lambda \mid K_1, K_2, \dots, K_t) &\propto f(k \mid \Lambda)\pi(\lambda) \\ &\propto \left[\prod_{i=1}^t \frac{e^{-\lambda} \lambda^{k_i}}{k_i!} \right] \cdot \frac{\left(\frac{c}{\mu}\right)^\nu \lambda^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda + \frac{\mu}{c}\frac{1}{\lambda}\right)\right]}{2B_\nu\left(\frac{1}{\sigma}\right)} \\ &\propto e^{-t\lambda} \lambda^{\sum_{i=1}^t k_i} \left(\frac{c}{\mu}\right)^\nu \lambda^{\nu-1} \frac{\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda + \frac{\mu}{c}\frac{1}{\lambda}\right)\right]}{2B_\nu\left(\frac{1}{\sigma}\right)} \\ &\propto \lambda^{\nu-1+\sum_{i=1}^t k_i} \left(\frac{c}{\mu}\right)^\nu \frac{\exp\left[-\lambda t - \frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda + \frac{\mu}{c}\frac{1}{\lambda}\right)\right]}{2B_\nu\left(\frac{1}{\sigma}\right)}, \end{aligned}$$

considerando a $w_1 = \frac{c}{\sigma\mu} + 2t$, $w_2 = \frac{\mu}{\sigma c}$, $\sigma > 0$, obtenemos

$$\propto \lambda^{\nu-1+\sum_{i=1}^t k_i} \left(\frac{c}{\mu}\right)^\nu \frac{\exp\left[-\frac{1}{2}\left(w_1\lambda + w_2\frac{1}{\lambda}\right)\right]}{2B_\nu\left(\frac{1}{\sigma}\right)}.$$

Aplicando las propiedades de la función de Bessel modificada de tercera especie (ver Abramowitz y Stegun (1964) sección 9.6 y 9.7)

$$\propto \frac{\lambda^{k'+\nu-1} \left(\frac{w_1}{w_2}\right)^{\frac{k'+\nu}{2}}}{2B_{k'+\nu}(\sqrt{w_1 w_2})} \exp\left[-\frac{1}{2}\left(w_1\lambda + w_2\frac{1}{\lambda}\right)\right].$$

Luego, $\Lambda | K \sim GIG(w_1, w_2, k' + \nu)$. □

Usando una función de pérdida cuadrática, el mejor estimador de la tasa de riesgo en el año $t + 1$, $\hat{\Lambda}_{t+1}$, es la media de la distribución anterior, es decir,

$$\begin{aligned} \hat{\Lambda}_{t+1}(K_1, \dots, K_t) &= \int_0^\infty \lambda \pi(\lambda | k_1, \dots, k_t) d\lambda, \\ &= \sqrt{\frac{w_2}{w_1}} \frac{B_{K'+\nu+1}(w_1 w_2)}{B_{K'+\nu}(w_1 w_2)}. \end{aligned}$$

2.2.3. Frecuencias bajo el modelo Poisson inversa gaussiana

El modelo de frecuencia Poisson inversa gaussiana es un caso particular del modelo de Sichel, cuando $\nu = -1/2$. En efecto, usando las propiedad de funciones de Bessel $B_{-\frac{1}{2}}(x) = B_{\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}} e^x$, a la distribución $GIG(\mu, \sigma, -1/2)$ del riesgo Λ se le conoce como una distribución inversa gaussiana, a denotarse por $IG(\mu, \sigma)$, y su función de densidad viene dada por:

$$u(\lambda) = \frac{\sqrt{\mu}}{\sqrt{2\pi\sigma\lambda^3}} \exp\left[-\frac{1}{2\sigma\mu\lambda}(\lambda - \mu)^2\right],$$

donde $\lambda > 0$, $\sigma > 0$.

Proposición 2.2.5. Si $K | \Lambda = \lambda \sim Poisson(\lambda)$ y $\Lambda \sim IG(\mu, \sigma)$, entonces K tiene como función de probabilidad a:

$$P(K = k) = \left(\frac{2a}{\pi}\right)^{\frac{1}{2}} \frac{\mu^k e^{\frac{1}{\sigma}} B_{k-\frac{1}{2}}(a)}{(a\sigma)^k k!}, \quad k = 0, 1, 2, 3, \dots, \quad (2.5)$$

donde $a^2 = \sigma^{-2} + 2\frac{\mu}{\sigma}$. Esta es conocida como una distribución Poisson inversa gaussiana y la denotaremos por $K \sim PIG(\mu, \sigma)$.

La media y la varianza de la distribución $PIG(\mu, \sigma)$ está dada por $E(K) = \mu$ y $Var(K) = \mu + \mu^2\sigma$, respectivamente.

Incorporando la historia de los siniestros del asegurados en los t periodos anteriores, obtenemos el siguiente mejor estimador de la prima para el periodo $t + 1$:

$$\hat{\Lambda}_{t+1}(K_1, \dots, K_t) = \sqrt{\frac{h_2 B_{K'+\frac{1}{2}}(h_1 h_2)}{h_1 B_{K'-\frac{1}{2}}(h_1 h_2)}}, \quad (2.6)$$

donde $h_1 = \frac{1}{\sigma\mu} + 2t$, $h_2 = \frac{\mu}{\sigma}$, $\sigma > 0$ y $K' = \sum_{j=1}^t K_j$.

Hasta este momento se han definido los modelos que permiten estimar el riesgo subyacente de un asegurado a través de las distribuciones Binomial negativa, Sichel y Poisson inversa gaussiana. Ahora abordaremos el análisis de las severidades de los siniestros mediante la distribución de Pareto.

2.2.4. Severidad bajo el modelo de Pareto

Sea X el monto de reclamo de cada asegurado e Y la media del monto de reclamos de cada asegurado. Supongamos que $X | Y = y \sim \text{Exp}(y)$, es decir, la densidad condicional de cada reclamo dada la media del reclamo es:

$$f(x | y) = \frac{1}{y} e^{-x/y} \quad x > 0,$$

donde recordemos que la media condicional de esta v.a. es $E(X | Y = y) = y$ y su varianza $\text{Var}(X | Y = y) = y^2$.

El modelo propuesto por Frangos y Vrontos (2001) asume que la distribución a priori de la media de las severidades es gamma inversa (GI), denotada por:

$$Y \sim GI(s, m),$$

la cual tiene como función de densidad a:

$$g(y) = \frac{\frac{1}{m} e^{-\frac{m}{y}}}{\left(\frac{y}{m}\right)^{s+1} \Gamma(s)},$$

y valor esperado

$$E(Y) = \frac{m}{s-1},$$

donde $s > 0$ y $m > 0$ son hiperparámetros del modelo.

Proposición 2.2.6. Si $X | Y = y \sim \text{Exp}(y)$ e $Y \sim GI(s, m)$, entonces la densidad marginal de las severidades de los reclamos X es igual a

$$f(x) = s \cdot m^s \cdot (x + m)^{-s-1} \quad x > 0,$$

que es la función de densidad de una distribución de Pareto con parámetros s y m , a la que denotaremos por $\text{Par}(s, m)$.

Demostración. Sea $X | Y = y \sim \text{Exp}(y)$ e $Y \sim GI(s, m)$, la distribución marginal de la

severidad es

$$\begin{aligned}
 f(x) &= \int_0^\infty f(x | y) \cdot g(y) dy \\
 &= \int_0^\infty \left(\frac{1}{y} e^{-\frac{x}{y}} \right) \left(\frac{e^{-\frac{x}{y}} \cdot m^s}{y^{s+1} \Gamma(s)} \right) dy \\
 &= m^s \int_0^\infty \frac{e^{-\frac{1}{y}(m+x)}}{y^{s+1+1} \cdot \Gamma(s)} dy \\
 &= \frac{m^s \cdot s}{(m+x)^{s+1}} \int_0^\infty \frac{(m+x)^{s+1} e^{-\frac{1}{y}(m+x)}}{y^{s+2} \cdot s \cdot \Gamma(s)} dy \\
 &= \frac{m^s \cdot s}{(m+x)^{s+1}}.
 \end{aligned}$$

Esta es la función de densidad de una distribución de Pareto con parámetros (s, m) . □

Es importante tener en cuenta que, el sistema Bonus–Malus considera el monto de pérdida de cada reclamo y se debe encontrar la distribución a posteriori de la media del monto de reclamo Y , dada la información de la historia de reclamos de cada asegurado.

Supongamos que el asegurado ha estado en el portafolio t periodos y que el número de reclamos que el ha tenido en el periodo j es K_j . Sea $K' = \sum_{j=1}^t K_j$ el número total de reclamos que ha presentado el asegurado y sea X_k el monto para el k -ésimo de estos reclamos. Entonces la información de la historia de reclamos se presentará de forma vectorial como $(x_1, x_2, \dots, x_{K'})$ y el monto total de reclamo para un asegurado específico en un periodo de t años será igual a $\sum_{k=1}^{K'} x_k$.

Proposición 2.2.7. *La distribución a posteriori de Y dada la historia de reclamos es*

$$Y | X_1 = x_1, \dots, X_{K'} = x_{K'} \sim GI \left(s + k', m + \sum_{k=1}^{k'} x_k \right),$$

donde $k' = \sum_{j=1}^t k_j$.

Demostración. Dado que $X | Y = y \sim \exp(y)$, $y \sim GI(s, m)$ el teorema de Bayes implica

que

$$\begin{aligned} \pi(y \mid x_1, x_2, \dots, x_{k'}) &\propto f(x_1, x_2, \dots, x_{k'})q(y) \\ &\propto \left[\prod_{i=1}^{k'} \frac{1}{y} e^{-\frac{x_i}{y}} \right] \cdot \frac{m^s e^{-\frac{m}{q}}}{q^{s+1} \Gamma(s)} \\ &\propto \frac{1}{y^n} e^{-\sum_{i=1}^{k'} \frac{x_i}{y}} \cdot \frac{1}{q^{s+1}} \cdot \frac{m^s e^{-\frac{m}{q}}}{\Gamma(s)} \\ &\propto \frac{1}{y^{s+n+1}} e^{-(m + \sum_{i=1}^{k'} \frac{x_i}{q})} \frac{1}{\Gamma(s)}. \end{aligned}$$

$$Y \mid X_1 = x_1, \dots, X_{k'} = x_{k'} \sim GI \left(s + k', m + \sum_{k=1}^{k'} x_k \right). \quad \square$$

A fin de determinar la severidad dados los montos de siniestros observados en los t periodos anteriores, es de interés calcular el mejor estimador del monto esperado por reclamo en el periodo $t+1$, denotado por \hat{Y}_{t+1} . Usando la función de pérdida cuadrática la severidad óptima para un asegurado con historial de monto de reclamos $X_1, X_2, \dots, X_{K'}$ será la media de la distribución a posteriori, es decir,

$$\hat{Y}_{t+1}(X_1, X_2, \dots, X_{K'}) = \frac{m + \sum_{k=1}^{K'} X_k}{s + K' - 1} \quad (2.7)$$

Hemos así obtenido el estimador de la severidad, incluyendo la información disponible del asegurado y esto ahora nos permitirá obtener la prima neta.

2.2.5. Estimación de la prima de acuerdo al principio de prima neta

Asumiendo independencia entre la frecuencia y severidad, la prima neta se obtiene como el producto de la frecuencia y la severidad de los siniestros.

La frecuencia de reclamos $\Lambda_{t+1}(K_1, \dots, K_t)$ para un asegurado o un grupo de asegurados con t periodos en el portafolio, los cuales producen K' reclamos con montos totales de reclamo $\sum_{k=1}^{K'} X_k$ está dada en (2.1) y la severidad $\hat{Y}_{t+1}(X_1, \dots, X_k)$ viene dada por (2.7). Así, de acuerdo con el modelo de frecuencia utilizado tenemos los siguientes estimadores para la prima neta:

- Con frecuencia binomial negativa y severidad Gamma inversa

$$\text{Prima neta} = \left(\frac{a + K'}{t + \tau} \right) \left(\frac{m + \sum_{k=1}^{K'} X_k}{s + K' - 1} \right)$$

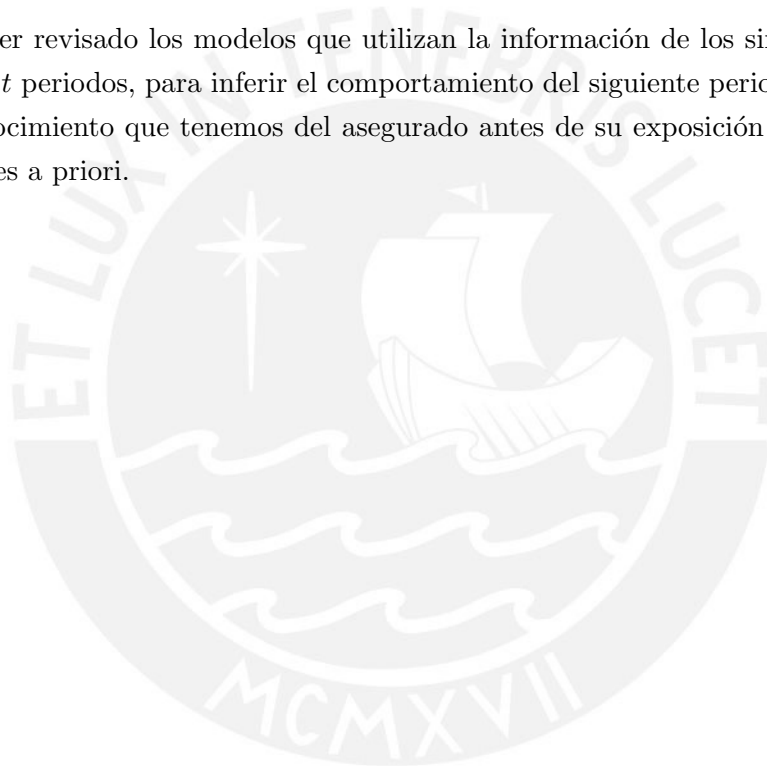
- Con frecuencia PIG y severidad Gamma inversa

$$\text{Prima neta} = \sqrt{\frac{h_2}{h_1}} \left(\frac{B_{K'+1/2}(h_1 h_2)}{B_{K'-1/2}(h_1 h_2)} \right) \left(\frac{m + \sum_{k=1}^{K'} X_k}{s + K' - 1} \right)$$

- Con frecuencia Sichel y severidad Gamma inversa

$$\text{Prima neta} = \left(\sqrt{\frac{w_2}{w_1}} \right) \left(\frac{B_{K'+\nu+1}(w_1 w_2)}{B_{K'+\nu}(w_1 w_2)} \right) \left(\frac{m + \sum_{k=1}^{K'} X_k}{s + K' - 1} \right)$$

Luego de haber revisado los modelos que utilizan la información de los siniestros que se ha recopilado en t periodos, para inferir el comportamiento del siguiente periodo, procedemos a incluir el conocimiento que tenemos del asegurado antes de su exposición al riesgo a través de las variables a priori.



Capítulo 3

Modelos BMSG y modelos GAMLSS basados en información a priori

El contar con información histórica completa de una cartera de asegurados que incluye no solo la información de siniestros, sino también las variables a priori o variables de tarificación nos permitirá obtener modelos más robustos con mayor cantidad de información.

En este capítulo se estudiarán sistemas Bonus - Malus generalizados, los cuales incorporan la información a priori y a posteriori del asegurado, con el fin de explicar la media y varianza del número de reclamos y de las severidades, bajo el supuesto que las distribuciones a considerar posean una solución analítica para la distribución del riesgo del asegurado.

También se incluirá algunos modelos que permiten estimar no solo la media sino también otros parámetros de escala, forma, sesgo de la frecuencia y de la severidad. La prima que se obtendrá en estos casos será solo la prima a priori, dada la complejidad y la forma no analítica que supone calcular la primas a posteriori cuando no se tienen distribuciones conjugadas, como en el caso anterior. Para esto último haremos uso de los modelos aditivos generalizados de localización escala y forma (GAMLSS).

Los modelos por detallar serán los siguientes:

- Modelos BMSG para la frecuencia y severidad basado en información a priori y a posteriori.
- Modelo BMS GAMLSS para la frecuencia y severidad basado en información a priori bajo el enfoque GAMLSS.

3.1. Modelos BMSG para la frecuencia y severidad basados en información a priori y a posteriori

Denominaremos modelos de sistemas Bonus-Malus Generalizados (BMSG) a los que incorporan variables a priori y variables a posteriori en el sistema Bonus-Malus. En los modelos que se presentan a continuación se considera que la variable de interés, número de siniestros o severidad, se pueden explicar mediante un modelo de regresión con las variables a priori.

3.1.1. Frecuencia bajo el modelo binomial negativo

El BMSG para la frecuencia bajo el modelo binomial negativo será desarrollado de acuerdo a Dionne y Vanasse (1989).

Considere un individuo i con una experiencia en el portafolio de t periodos. Suponga que el número de reclamos del individuo i para el período j , denotado como K_i^j , sigue una distribución de Poisson con parámetro λ_i^j y las variables aleatorias K_i^j son independientes.

La frecuencia esperada de reclamos del individuo i , λ_i^j , se considerará ahora como una función de un vector de h características individuales, denotado por $\mathbf{c}_i^j = (c_{i,1}^j, \dots, c_{i,h}^j)$, las cuales representan las variables a priori. Asumiremos que $\lambda_i^j = \exp(\mathbf{c}_i^j \boldsymbol{\beta}^j)$, donde $\boldsymbol{\beta}^j$ es el vector de los coeficientes de regresión y por tanto,

$$P(K_i^j = k \mid \lambda_i^j) = \frac{e^{-\exp(\mathbf{c}_i^j \boldsymbol{\beta}^j)} (\exp(\mathbf{c}_i^j \boldsymbol{\beta}^j))^k}{k!}.$$

Este modelo asume que las h características individuales proveen suficiente información para determinar el número de reclamos esperados. Si esto es limitado será conveniente introducir una variable aleatoria ε_i bajo el siguiente modelo de regresión:

$$\begin{aligned} \Lambda_i^j &= \exp(\mathbf{c}_i^j \boldsymbol{\beta}^j + \varepsilon_i) \\ &= \exp(\mathbf{c}_i^j \boldsymbol{\beta}^j) U_i, \end{aligned}$$

donde $U_i = \exp(\varepsilon_i)$ incorpora una componente de aleatoriedad a la media de la frecuencia de reclamos, haciendo que está sea ahora aleatoria.

Si asumimos que U_i sigue una distribución Gamma con $E(U_i) = 1$ y $Var(U_i) = 1/a$, entonces Λ_i^j tiene una distribución gamma y por la proposición (2.2.1) se sigue que marginalmente $K_i^j \sim BN(a, \exp(-\mathbf{c}_i^j \boldsymbol{\beta}^j))$,

Note que hemos escogido $E(U_i) = 1$ con el fin de asegurar que $E(\varepsilon_i) = 0$. Así,

$$E(K_i^j) = \exp(\mathbf{c}_i^j \boldsymbol{\beta}^j) \text{ y } Var(K_i^j) = \exp(\mathbf{c}_i^j \boldsymbol{\beta}^j) \left[1 + \frac{\exp(\mathbf{c}_i^j \boldsymbol{\beta}^j)}{a} \right].$$

Usando ahora un argumento similar a lo visto en la proposición (2.2.2) tenemos que la distribución a posteriori para un asegurado con una historia de k_i^1, \dots, k_i^t reclamos y características c_i^1, \dots, c_i^{t+1} es

$$Gamma \left(a + \sum_{j=1}^t k_i^j, \frac{a + \sum_{j=1}^t \exp(\mathbf{c}_i^j \boldsymbol{\beta}^j)}{\exp(\mathbf{c}_i^{t+1} \boldsymbol{\beta}^{t+1})} \right).$$

Para determinar la frecuencia en el año $t + 1$ dado el historial de reclamos de t periodos y conociendo las características individuales hasta el periodo $t + 1$, buscamos el estimador $\hat{\Lambda}_i^{t+1}$.

Mediante la función de pérdida cuadrática, el estimador óptimo $\hat{\Lambda}_i^{t+1}$, dadas las observaciones

K_i^1, \dots, K_i^t y covariables c_i^1, \dots, c_i^{t+1} es igual a

$$\hat{\Lambda}_i^{t+1}(K_i^1, \dots, K_i^t; c_i^1, \dots, c_i^{t+1}) = \exp(c_i^{t+1} \beta^{t+1}) \left[\frac{a + \sum_{j=1}^t K_i^j}{a + \sum_{j=1}^t \exp(c_i^j \beta^j)} \right], \quad (3.1)$$

donde $\sum_{j=1}^t K_i^j$ denota el número total de reclamos de un asegurado i en los t periodos.

3.1.2. Frecuencia bajo el modelo de Sichel

Otro de los modelos de la frecuencia que incorporan información a priori y a posteriori es el expuesto en Tzougas et al. (2015).

Consideramos un asegurado i con t años de experiencia en el portafolio. Sea K_i^j el número de reclamos de el asegurado i ha tenido en el periodo j y asumamos que los K_i^j son variables aleatorias independientes, con distribución de Poisson de parámetro $\lambda_i^j = \exp(c_i^j \beta^j)$ donde $\mathbf{c}_i^j = (c_{i,1}^j, \dots, c_{i,h}^j)$ es un vector de h características o variables a priori y β^j es el vector de coeficientes de regresión. Esto es,

$$\begin{aligned} P(K_i^j = k \mid \lambda_i^j) &= \frac{\exp(-\lambda_i^j) (\lambda_i^j)^k}{k!} \\ &= \frac{\exp(-\exp(c_i^j \beta^j)) (\exp(c_i^j \beta^j))^k}{k!}. \end{aligned}$$

Si bien las h características o variables a priori brindan información, éstas podrían no contener toda la información necesaria para explicar el número total de reclamos. Por esto sería conveniente introducir una variable aleatoria ε_i bajo el modelo de regresión.

$$\begin{aligned} \Lambda_i^j &= \exp(c_i^j \beta^j + \varepsilon_i) \\ &= \exp(c_i^j \beta^j) U_i, \end{aligned}$$

donde $U_i = \exp(\varepsilon_i)$ incorpora una componente de aleatoriedad a la media del número de reclamos, haciendo que la media de los reclamos sea aleatoria.

Si asumimos que $U_i \sim \text{GIG}(1, \sigma, \nu)$, entonces $\Lambda_i^j \sim \text{GIG}(\exp(c_i^j \beta^j), \sigma, \nu)$ y por la proposición (2.2.3) se tiene que marginalmente $K_i^j \sim \text{Sichel}(\exp(c_i^j \beta^j), \sigma, \nu)$.

La parametrización indicada de la distribución de Sichel asegura que el parámetro de localización es la media

$$E(K_i^j) = \mu_i^j = \exp(c_i^j \beta^j).$$

Además, la varianza de K_i^j es

$$\text{Var}(K_i^j) = \mu_i^j = \exp(c_i^j \beta^j) + (\exp(c_i^j \beta^j))^2 \left[\frac{2\sigma(\nu + 1)}{c} + \frac{1}{c^2} - 1 \right],$$

donde $c = \frac{B_{\nu+1} \left[\frac{1}{\sigma} \right]}{B_{\nu} \left[\frac{1}{\sigma} \right]}$ y $B_{\nu}(z)$ es la función modificada de Bessel de tercera especie de orden ν con argumento z .

Si tenemos un asegurado i con historia de reclamos K_i^1, \dots, K_i^t , covariables a priori c_i^1, \dots, c_i^{t+1} , y $K' = \sum_{j=1}^t K_i^j$ el número total de reclamos del asegurado i , entonces la media de la frecuencia del individuo i para el periodo $t+1$ será una función de las características en $t+1$ del individuo y del factor U_i , donde $U_i \sim \text{GIG}(1, \sigma, \nu)$.

Por otro lado, aplicando un argumento similar al empleado en la proposición (2.2.4), se tiene que marginalmente la distribución a posteriori de la frecuencia de reclamos esperada es:

$$\Lambda_i^t \mid K_i^1 = k_i^1, \dots, K_i^t = k_i^t; c_i^1, \dots, c_i^t \sim \text{GIG}(w_1, w_2, k' + \nu)$$

donde $k' = \sum_{j=1}^t k_j$, $w_1 = \frac{c + 2\sigma \sum_{j=1}^t \exp(c_i^j \beta^j)}{\sigma \exp(c_i^{t+1} \beta^{t+1})}$, $w_2 = \frac{\exp(c_i^{t+1} \beta^{t+1})}{\sigma c}$ y $c = \frac{B_{\nu+1} \left[\frac{1}{\sigma} \right]}{B_{\nu} \left[\frac{1}{\sigma} \right]}$.

Es necesario calcular el mejor estimador del riesgo en el año $t+1$, usando la experiencia de los t años anteriores. Este estimador lo denotaremos por $\hat{\Lambda}_i^{t+1}$.

Para determinar la frecuencia estimada en el año $t+1$ se tiene que considerando una función de pérdida cuadrática, el estimador óptimo de $\hat{\Lambda}_i^{t+1}$ es la media de la distribución $\text{GIG}(w_1, w_2, K' + \nu)$

$$\hat{\Lambda}_i^{t+1}(K_i^1, \dots, K_i^t; c_i^1, \dots, c_i^{t+1}) = \left(\sqrt{\frac{w_2}{w_1}} \right) \frac{B_{K'+\nu+1}(w_1 w_2)}{B_{K'+\nu}(w_1 w_2)}.$$

3.1.3. Severidad bajo el modelo Pareto

Considere un individuo i con una experiencia de t períodos y definamos

- K_i^j igual al número de reclamos del asegurado i en el año j .
- $K' = \sum_{j=1}^t K_i^j$ el número total de reclamos del asegurado i .
- $X_{i,k}^j$ igual al monto incurrido o severidad del k -ésimo reclamo del asegurado i en el periodo j .

Entonces la información de la historia de reclamos se presenta de forma vectorial

$$X_{i,1}, X_{i,2}, \dots, X_{i,K'},$$

y el monto total de reclamo para un asegurado específico en t periodos será igual a $\sum_{k=1}^{K'} X_{i,k}$.

Supongamos ahora que los $X_{i,k}^j$ siguen una distribución exponencial con parámetro y_i^j , donde y_i^j denota la severidad media del reclamo de un asegurado i en el periodo j .

Consideramos ahora que el monto esperado de reclamo es una función de un vector de las h características del individuo, denotadas como $\mathbf{d}_i^j = (d_{i,1}^j, \dots, d_{i,h}^j)$ que representan las diferentes variables a priori. Específicamente asumimos que $y_i^j = \exp(\mathbf{d}_i^j \boldsymbol{\gamma}^j)$, donde $\boldsymbol{\gamma}^j$ es el vector de los coeficientes de regresión y así,

$$f(x_{i,k}^j | y_i^j) = \frac{1}{\exp(d_i^j \boldsymbol{\gamma}^j)} e^{-\frac{x}{\exp(d_i^j \boldsymbol{\gamma}^j)}}.$$

En este modelo se asume que las h características individuales proveen la información necesaria para determinar el monto de reclamos esperado. Sin embargo, si ellas no cubren toda la información, será conveniente introducir una variable aleatoria Z_i en el componente de regresión, tal que,

$$\begin{aligned} Y_i^j &= \exp(d_i^j \boldsymbol{\gamma}^j + Z_i) \\ &= \exp(d_i^j \boldsymbol{\gamma}^j) W_i, \end{aligned}$$

donde $W_i = \exp(Z_i)$ incorpora una componente de aleatoriedad a la media de número de reclamos que ahora es una variable aleatoria.

Si $W_i \sim \text{GI}(s, s-1)$ entonces Y_i^j tiene distribución $\text{GI}(s, (s-1) \exp(d_i^j \boldsymbol{\gamma}^j))$ y por la proposición (2.2.7) la función de densidad marginal para $X_{i,k}^j$ es

$$f(x) = s \cdot [(s-1) \exp(d_i^j \boldsymbol{\gamma}^j)]^s \cdot (x + (s-1) \exp(d_i^j \boldsymbol{\gamma}^j))^{-s-1},$$

la cual es una distribución de Pareto con parámetros s y $(s-1) \exp(d_i^j \boldsymbol{\gamma}^j)$.

Esto muestra que la parametrización anterior no afecta los resultados si hay un término constante en la regresión. Note que hemos escogemos $E(W_i) = 1$ con el fin de obtener $E(Z_i) = 0$ y también tenemos que

$$\begin{aligned} E(X_{i,k}^j) &= \exp(d_i^j \boldsymbol{\gamma}^j) \\ \text{Var}(X_{i,k}^j) &= \frac{[(s-1) \exp(d_i^j \boldsymbol{\gamma}^j)]^2}{s-1} \left(\frac{2}{s-2} - \frac{1}{s-1} \right). \end{aligned}$$

Para encontrar la severidad estimada, es necesario calcular el mejor estimador del monto esperado de reclamo en el año $t+1$. Usando la información de la experiencia pasada del monto de reclamo sobre los t periodos y las características del individuo hasta el periodo $t+1$, se puede probar, similarmente a lo visto en la proposición (2.2.7) que la distribución a posteriori para la media del monto del reclamo de un asegurado dados sus montos de reclamaciones $x_{i,1}, \dots, x_{i,K'}$ en t periodos y características d_i^1, \dots, d_i^{t+1} es Gamma inversa (GI). Más explícitamente, ella viene dada por:

$$\hat{Y}_i^{t+1} | X_{i,1} = x_{i,1}, \dots, X_{i,K'} = x_{i,K'}; d_i^1, \dots, d_i^{t+1} \sim GI \left(s + K', (s-1) \exp(d_i^j \boldsymbol{\gamma}^j) + \sum_{k=1}^{K'} X_{i,k} \right).$$

Usándose una función de pérdida cuadrática, el mejor estimador de la media del monto del reclamo para el año $t + 1$ dada las severidades de $X_{i,1}, \dots, X_{i,K'}$ y covariables a priori d_i^1, \dots, d_i^{t+1} , viene dada por la media a posteriori

$$\hat{Y}_i^{t+1}(X_{i,1}, \dots, X_{i,K'}; d_i^1, \dots, d_i^{t+1}) = \frac{(s-1)\frac{1}{t} \sum_{j=1}^t \exp(d_i^j \gamma^j) + \sum_{k=1}^{K'} X_{i,k}}{s + K' - 1}. \quad (3.2)$$

Note que si $t = 0$, únicamente una tarificación a priori será usada en el primer periodo $\hat{Y}_i^1 = \exp(d_i^1 \gamma^1)$.

Hasta el momento hemos cubierto modelos que incorporan el historial de siniestros del asegurado y que permiten utilizarlos a fin de obtener la prima, actualizando la información del asegurado; para dichos modelos hemos usado el hecho que existe una distribución a priori conjugada. Sin embargo, esto no se cumple en la mayoría de los casos, por lo que en los modelos que se proponen a continuación solamente obtendremos la prima a priori bajo el principio de valor esperado y el principio de desviación estándar. Se deja abierto como posible desarrollo futuro la estimación de primas a posteriori, mediante la aplicación de métodos no analíticos. Por ejemplo métodos de simulación por cadenas de Markov de MonteCarlo.

3.1.4. Estimación de la prima neta usándose un BMSG

Resumiendo, los estimadores de la prima neta derivados usando un modelo BMSG para la frecuencia y severidad Pareto se definen como:

- Bajo el modelo de frecuencia binomial negativa,

$$\text{Prima neta} = \exp(c_i^{t+1} \beta^{t+1}) \left[\frac{a + \sum_{j=1}^t K_i^j}{a + \sum_{j=1}^t \exp(c_i^j \beta^j)} \right] \left(\frac{(s-1)\frac{1}{t} \sum_{j=1}^t \exp(d_i^j \gamma^j) + \sum_{j=1}^{K'} x_{i,k}}{s + K' - 1} \right)$$

- Bajo el modelo de frecuencia de Sichel,

$$\text{Prima neta} = \left(\sqrt{\frac{w_2}{w_1}} \right) \left(\frac{B_{K'+\nu+1}(w_1 w_2)}{B_{K'+\nu}(w_1 w_2)} \right) \left(\frac{(s-1)\frac{1}{t} \sum_{j=1}^t \exp(d_i^j \gamma^j) + \sum_{j=1}^{K'} x_{i,k}}{s + K' - 1} \right)$$

3.2. Modelos BMSG GAMLSS

Los modelos GAMLSS son usados para modelar una variable respuesta que no sigue una distribución de la familia exponencial o que es heterogénea.

En algunos casos cuando se tiene una distribución con dos parámetros, se suele modelar solo la media (μ) en función de covariables, pero hay momentos en que asumir un parámetro de

escala (σ) constante no es adecuado. Adicionalmente, en distribuciones de dos parámetros el sesgo y la curtosis son fijos para modelos sobre (μ) y (σ). Para resolver el problema se podrían modelar los parámetros de escala, sesgo y curtosis como función de las variables explicativas.

El modelo GAMLSS asume variables aleatorias independientes Y_i con función de densidad $f(y_i | \boldsymbol{\theta})$ condicional a $\boldsymbol{\theta}$, donde $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)^\top$ es un vector, cada uno de los cuales puede ser una función de variables explicativas.

Definición 3.2.1. Si la forma de la distribución asumida para la variable respuesta Y es $\mathcal{D}(y | \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$ el modelo GAMLSS asume que

$$Y \sim \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}),$$

donde

$$\begin{aligned} \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(\mathbf{x}_{11}) + \cdots + s_{1J_1}(\mathbf{x}_{1J_1}) \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(\mathbf{x}_{21}) + \cdots + s_{2J_2}(\mathbf{x}_{2J_2}) \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(\mathbf{x}_{31}) + \cdots + s_{3J_3}(\mathbf{x}_{3J_3}) \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(\mathbf{x}_{41}) + \cdots + s_{4J_4}(\mathbf{x}_{4J_4}), \end{aligned}$$

\mathbf{X}_k son las matrices de diseño que incorporan los términos aditivos lineales al modelo, $\boldsymbol{\beta}_k$ son los coeficientes de regresión lineales y $s_{kj}(\mathbf{x}_{kj})$ representan funciones suaves para las variables explicativas, \mathbf{x}_{kj} con $j = 1, \dots, J_k$. Note que las matrices \mathbf{X}_k pueden contener columnas que son iguales o diferentes a los \mathbf{x}_{kj} . Los vectores $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4$ son llamados los predictores de $\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}$ respectivamente.

A continuación, se exponen cada uno de los modelos a usar tanto en la estimación de la frecuencia, como en la estimación de la severidad. En algunos de estos se considera la estimación de los parámetros de localización y escala, en otros modelos el parámetro forma. La prima neta se obtendrá como el producto de la frecuencia y la severidad, asumiendo independencia en estos componentes.

Los modelos de la frecuencia y severidad a considerar, utilizando GAMLSS, se muestran en Tzougas et al. (2015).

3.2.1. Frecuencia bajo el modelo binomial negativo tipo II

Sea K_i la variable aleatoria del número de reclamos que presenta el asegurado i con función de probabilidad

$$P(K_i = k_i) = \frac{\Gamma\left(k_i + \frac{\mu_i}{\sigma_i}\right) \sigma_i^{k_i}}{(k_i)! \Gamma\left(\frac{\mu_i}{\sigma_i}\right) (1 + \sigma_i)^{k_i + \frac{\mu_i}{\sigma_i}}},$$

para $\mu_i > 0$ y $\sigma_i > 0$.

Se asume que los parámetros μ_i y σ_i se modelan como función de las variables a priori.

$$\begin{aligned}\mu_i &= e_i \exp(c_{1i}\beta_1) \\ \sigma_i &= \exp(c_{2i}\beta_2)\end{aligned}$$

donde $c_{li} = (c_{li,1}, \dots, c_{li,h})$ es un vector de características o covariables y $\beta_l^\top = (\beta_{l,1}, \dots, \beta_{l,h})$ es un vector columna de coeficientes de regresión, para $l = 1, 2$ y e_i es la exposición de la póliza i . En nuestro caso, se asume que $e_i = 1$.

La media y la varianza de K_i están dadas por

$$E(K_i) = e_i \exp(c_{1i}\beta_1) \text{ y } Var(K_i) = e_i \exp(c_{1i}\beta_1)[1 + \exp(c_{2i}\beta_2)].$$

3.2.2. Frecuencia bajo el modelo de Delaporte

Ahora estudiaremos un modelo obtenido a partir de una convolución entre una distribución de Poisson y una distribución binomial negativa, conocida como distribución de Delaporte, que denotaremos por DEL cuyos parámetros son μ, σ y ν .

Sea K_i la variable aleatoria del número de reclamos que presenta el asegurado i con función de probabilidad

$$P(K_i = k_i) = \frac{e^{\mu_i \nu_i}}{\Gamma\left(\frac{1}{\sigma_i}\right)} [1 + \mu_i \sigma_i (1 - \nu_i)]^{-\frac{1}{\sigma_i}} S,$$

donde $\sigma_i > 0$ y $0 \leq \nu_i < 1$ son parámetros del modelo y donde

$$S = \sum_{m=0}^{k_i} \binom{k_i}{m} \frac{\mu_i^{k_i} \nu_i^{k_i-m}}{(k_i)!} \left[\mu_i + \frac{1}{\sigma_i(1-\nu_i)} \right]^{-m} \Gamma\left(\frac{1}{\sigma_i} + m\right).$$

Se asume que

$$\begin{aligned}\mu_i &= e_i \exp(c_{1i}\beta_1) \\ \sigma_i &= \exp(c_{2i}\beta_2) \\ \nu_i &= \frac{\exp(c_{3i}\beta_3)}{1 + \exp(c_{3i}\beta_3)}\end{aligned}$$

donde $c_{li} = (c_{li,1}, \dots, c_{li,h})$ es un vector de características o covariables y $\beta_l^\top = (\beta_{l,1}, \dots, \beta_{l,h})$ es un vector columna de coeficientes de regresión, para $l = 1, 2, 3$ y e_i es la exposición de la póliza i . En nuestro caso, se asume que $e_i = 1$.

La media y la varianza de K_i están dadas por

$$E(K_i) = e_i \exp(c_{1i}\beta_1)$$

y

$$\text{Var}(K_i) = e_i \exp(c_{1i}\beta_1) + [e_i \exp(c_{1i}\beta_1)]^2 \exp(c_{2i}\beta_2) \left[1 - \frac{\exp(c_{3i}\beta_3)}{1 + \exp(c_{3i}\beta_3)} \right]$$

3.2.3. Frecuencia bajo el modelo de Sichel

Si bien este modelo ya se ha tratado y posee una fórmula explícita, el desarrollo aquí considerado, extiende la utilidad de incorporar la información a priori en la explicación de su tercer parámetro ν mediante el uso de los modelos GAMLSS.

Sea K_i la variable aleatoria del número de reclamos que presenta el asegurado i con función de probabilidad

$$P(K_i = k_i) = \frac{\left(\frac{\mu_i}{c_i}\right)^{k_i} B_{k_i + \nu_i}(a_i)}{(k_i)! (a_i \sigma_i)^{k_i + \nu_i} B_{\nu_i}\left(\frac{1}{\sigma_i}\right)}$$

donde $\sigma_i > 0$ y $-\infty < \nu_i < \infty$, $a_i^2 = \sigma_i^{-2} + 2\mu_i(c_i \sigma_i)^{-1}$ y $c_i = \frac{B_{\nu_i+1}\left(\frac{1}{\sigma_i}\right)}{B_{\nu_i}\left(\frac{1}{\sigma_i}\right)}$,

$$2B_\nu(z) = \int_0^\infty x^{\nu-1} \exp\left[-\frac{1}{2}z\left(x + \frac{1}{x}\right)\right] dx,$$

se conoce como la función de Bessel modificada de tercera especie de orden ν con argumento z . Se asume que

$$\mu_i = e_i \exp(c_{1i}\beta_1)$$

$$\sigma_i = \exp(c_{2i}\beta_2)$$

$$\nu_i = c_{3i}\beta_3$$

donde $c_{li} = (c_{li,1}, \dots, c_{li,h})$ es un vector de características o covariables y $\beta_l^\top = (\beta_{l,1}, \dots, \beta_{l,h})$ es un vector columna de coeficientes de regresión, para $l = 1, 2, 3$ y e_i es la exposición de la póliza i . En nuestro caso, se asume que $e_i = 1$.

La media y la varianza de K_i están dadas por

$$E(K_i) = e_i \exp(c_{1i}\beta_1)$$

y

$$\text{Var}(K_i) = e_i \exp(c_{1i}\beta_1) + [e_i \exp(c_{1i}\beta_1)]^2 \left[\frac{2 \exp(c_{2i}\beta_2) [c_{3i}\beta_3 + 1]}{c_i} \frac{1}{c_i^2} - 1 \right],$$

donde

$$c_i = \frac{B_{c_{3i}\beta_3+1}\left(\frac{1}{\exp(c_{2i}\beta_2)}\right)}{B_{c_{3i}\beta_3}\left(\frac{1}{\exp(c_{2i}\beta_2)}\right)}.$$

3.2.4. Frecuencia bajo el modelo Zero-inflated Poisson (ZIP)

Sea K_i la variable aleatoria del número de reclamos que presenta el asegurado i con función de probabilidad

$$P(K_i = k_i) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\mu_i}, & \text{si } k_i = 0 \\ (1 - \pi_i)\frac{e^{-\mu_i}\mu_i^{k_i}}{(k_i)!}, & \text{si } k_i = 1, 2, 3, \dots \end{cases}$$

donde $0 < \pi_i < 1$ y $\mu_i > 0$.

Se asume que

$$\begin{aligned} \mu_i &= e_i \exp(c_{1i}\beta_1) \\ \pi_i &= \frac{\exp(c_{2i}\beta_2)}{1 + \exp(c_{2i}\beta_2)} \end{aligned}$$

donde $c_{li} = (c_{li,1}, \dots, c_{li,h})$ es un vector de características o covariables y $\beta_l^\top = (\beta_{l,1}, \dots, \beta_{l,h})$ es un vector columna de coeficientes de regresión, para $l = 1, 2$ y e_i es la exposición de la póliza i . En nuestro caso, se asume que $e_i = 1$.

La media y la varianza de K_i están dadas por

$$E(K_i) = e_i \exp(c_{1i}\beta_1)[1 - \exp(c_{2i}\beta_2)]$$

y

$$Var(K_i) = e_i \exp(c_{1i}\beta_1)[1 - \exp(c_{2i}\beta_2)][1 + e_i \exp(c_{1i}\beta_1) \exp(c_{2i}\beta_2)].$$

Dado que en seguros es común encontrar que las observaciones de la variable costo de siniestro es igual a cero, en una proporción elevada, este modelo podría ser de mucha utilidad, sin embargo, en el contexto de los modelos GAMLSS, debemos poder incorporar este caso y hacer algunos ajustes.

De acuerdo a lo indicado en (M Stasinopoulos (2017)), para las distribuciones definidas en $(0, \infty)$ que incluyen un punto de probabilidad no nula, se sugiere que a las regresiones sobre los parámetros, μ , σ , ν y τ para la distribución que excluye el punto de probabilidad 0, se considera una regresión sobre $\xi_0 = p_0$ para la probabilidad cero.

$$Y \sim \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}, \xi_0),$$

donde

$$\begin{aligned}\eta_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(\mathbf{x}_{11}) + \cdots + s_{1J_1}(\mathbf{x}_{1J_1}) \\ \eta_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(\mathbf{x}_{21}) + \cdots + s_{2J_2}(\mathbf{x}_{2J_2}) \\ \eta_3 &= g_3(\mathbf{v}) = \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(\mathbf{x}_{31}) + \cdots + s_{3J_3}(\mathbf{x}_{3J_3}) \\ \eta_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(\mathbf{x}_{41}) + \cdots + s_{4J_4}(\mathbf{x}_{4J_4}) \\ \eta_5 &= g_5(\boldsymbol{\xi}_0) = \mathbf{X}_5\boldsymbol{\beta}_5 + s_{51}(\mathbf{x}_{51}) + \cdots + s_{5J_5}(\mathbf{x}_{5J_5}),\end{aligned}$$

donde $\mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\tau}, \boldsymbol{\xi}_0)$, es la distribución de respuesta cero-inflacionada definida en $[0, \infty)$.

3.2.5. Severidad bajo el modelo Gamma

Sea $X_{i,k}$ el costo del k -ésimo reclamo reportado por el asegurado i , con $i = 1, 2, \dots, n$ y asumimos que $X_{i,1}, X_{i,2}, \dots$ son variables aleatorias independientes e idénticamente distribuidas gamma con parámetros s_i y m_i con función de densidad

$$f(x) = \frac{1}{(s_i^2 m_i)^{\frac{1}{s_i^2}}} \frac{x^{\frac{1}{s_i^2}-1} \exp\left(-\frac{x}{s_i^2 m_i}\right)}{\Gamma\left(\frac{1}{s_i^2}\right)}$$

para $X_{i,k} > 0$, donde $m_i > 0$ y $s_i > 0$.

Asumimos que

$$\begin{aligned}m_i &= \exp(d_{1i}\gamma_1) \\ s_i &= \exp(d_{2i}\gamma_2)\end{aligned}$$

donde $d_{li} = (d_{li,1}, \dots, d_{li,h})$ es un vector de características o covariables y $\boldsymbol{\gamma}_l^\top = (\gamma_{l,1}, \dots, \gamma_{l,h})$ es un vector columna de coeficientes de regresión, para $l = 1, 2$.

La media y la varianza de $X_{i,k}$ están dadas por

$$E(X_{i,k}) = \exp(d_{1i}\gamma_1)$$

y

$$Var(X_{i,k}) = [\exp(d_{2i}\gamma_2)]^2 [\exp(d_{1i}\gamma_1)]^2$$

3.2.6. Severidad bajo el modelo Weibull

Sea $X_{i,k}$ el costo del k -ésimo reclamo reportado por el asegurado i , con $i = 1, 2, \dots, n$ y asumimos que $X_{i,1}, X_{i,2}, \dots$ son independientes e idénticamente distribuidos Weibull con parámetros s_i y m_i con función de densidad

$$f(x) = \frac{s_i x^{s_i-1}}{m_i^{s_i}} \exp\left[-\left(\frac{x}{m_i}\right)^{s_i}\right],$$

para $X_{i,k} > 0$, donde $m_i > 0$ y $s_i > 0$.

Asumimos que

$$\begin{aligned} m_i &= \exp(d_{1i}\gamma_1) \\ s_i &= \exp(d_{2i}\gamma_2) \end{aligned}$$

donde $d_{li} = (d_{li,1}, \dots, d_{li,h})$ es un vector de características o covariables y $\gamma_l^\top = (\gamma_{l,1}, \dots, \gamma_{l,h})$ es un vector columna de coeficientes de regresión, para $l = 1, 2$.

La media y la varianza de $X_{i,k}$ están dadas por

$$E(X_{i,k}) = \exp(d_{1i}\gamma_1) \Gamma\left(\frac{1}{\exp(d_{2i}\gamma_2)} + 1\right)$$

y

$$Var(X_{i,k}) = [\exp(d_{1i}\gamma_1)]^2 \left\{ \Gamma\left(\frac{2}{\exp(d_{2i}\gamma_2)} + 1\right) - \left[\Gamma\left(\frac{1}{\exp(d_{2i}\gamma_2)} + 1\right) \right]^2 \right\}$$

3.2.7. Severidad bajo el modelo de Weibull tipo III

Sea $X_{i,k}$ el costo del k -ésimo reclamo reportado por el asegurado i , con $i = 1, 2, \dots, n$ y asumimos que $X_{i,1}, X_{i,2}, \dots$ son independientes e idénticamente distribuidos Weibull tipo III con parámetros s_i y m_i con función de densidad

$$f(x) = \frac{s_i}{m_i} \Gamma\left(\frac{1}{s_i} + 1\right) \left[\frac{x}{m_i} \Gamma\left(\frac{1}{s_i} + 1\right) \right]^{s_i - 1} \exp\left\{ - \left[\frac{x}{m_i} \Gamma\left(\frac{1}{s_i} + 1\right) \right]^{s_i} \right\},$$

para $X_{i,k} > 0$, donde $m_i > 0$ y $s_i > 0$.

Asumimos que

$$\begin{aligned} m_i &= \exp(d_{1i}\gamma_1) \\ s_i &= \exp(d_{2i}\gamma_2) \end{aligned}$$

donde $d_{li} = (d_{li,1}, \dots, d_{li,h})$ es un vector de características o covariables y $\gamma_l^\top = (\gamma_{l,1}, \dots, \gamma_{l,h})$ es un vector columna de coeficientes de regresión, para $l = 1, 2$.

La media y la varianza de $X_{i,k}$ están dadas por

$$E(X_{i,k}) = \exp(d_{1i}\gamma_1)$$

y

$$Var(X_{i,k}) = [\exp(d_{1i}\gamma_1)]^2 \left\{ \Gamma\left(\frac{2}{\exp(d_{2i}\gamma_2)} + 1\right) \left[\Gamma\left(\frac{1}{\exp(d_{2i}\gamma_2)} + 1\right) \right]^{-2} - 1 \right\}$$

3.2.8. Severidad bajo el modelo Gamma generalizado

Sea $X_{i,k}$ el costo del k -ésimo reclamo reportado por el asegurado i , con $i = 1, 2, \dots, n$ y asumimos que $X_{i,1}, X_{i,2}, \dots$ son independientes e idénticamente distribuidos Gamma generalizada con parámetros s_i, m_i y n_i con función de densidad

$$f(x) = \frac{|n_i| \theta_i^{\theta_i} \left(\frac{x}{m_i}\right)^{n_i \theta_i} \exp\left[-\theta_i \left(\frac{x}{m_i}\right)^{n_i}\right]}{\Gamma(\theta_i) x}$$

para $X_{i,k} > 0$, donde $m_i > 0$ y $s_i > 0$, $-\infty < n_i < \infty$, y $\theta_i = \frac{1}{s_i^2 n_i^2}$

Asumimos que

$$\begin{aligned} m_i &= \exp(d_{1i} \gamma_1) \\ s_i &= \exp(d_{2i} \gamma_2) \\ n_i &= d_{3i} \gamma_3 \end{aligned}$$

donde $d_{li} = (d_{li,1}, \dots, d_{li,h})$ es un vector de características o covariables y $\gamma_l^\top = (\gamma_{l,1}, \dots, \gamma_{l,h})$ es un vector columna de coeficientes de regresión, para $l = 1, 2, 3$.

La media y la varianza de $X_{i,k}$ están dadas por

$$E(X_{i,k}) = \frac{\exp(d_{1i} \gamma_1) \Gamma\left(\theta_i + \frac{1}{d_{3i} \gamma_3}\right)}{\theta^{\frac{1}{d_{3i} \gamma_3}} \Gamma(\theta_i)}$$

y

$$Var(X_{i,k}) = \frac{[\exp(d_{1i} \gamma_1)]^2 \left\{ \Gamma(\theta_i) \Gamma\left(\theta_i + \frac{2}{d_{3i} \gamma_3}\right) - \left[\Gamma\left(\theta_i + \frac{1}{d_{3i} \gamma_3}\right) \right]^2 \right\}}{\theta^{\frac{2}{d_{3i} \gamma_3}} [\Gamma(\theta_i)]^2}$$

donde $\theta_i = \frac{1}{s_i^2 n_i^2} = \frac{1}{(\exp(d_{2i} \gamma_2))^2 (d_{3i} \gamma_3)^2}$

3.2.9. Severidad bajo el modelo Pareto generalizado

Sea $X_{i,k}$ el costo del k -ésimo reclamo reportado por el asegurado i , con $i = 1, 2, \dots, n$ y asumimos que $X_{i,1}, X_{i,2}, \dots$ son independientes e idénticamente distribuidos Gamma generalizada con parámetros m_i, n_i y t_i con función de densidad

$$f(x) = \frac{\Gamma(n_i + t_i)}{\Gamma(n_i) \Gamma(t_i)} \frac{m_i^{t_i} x^{n_i - 1}}{(x + m_i)^{n_i + t_i}}$$

para $X_{i,k} > 0$, donde $m_i > 0$, $n_i > 0$ y $t_i > 0$. Asumimos que

$$\begin{aligned} m_i &= \exp(d_{1i}\gamma_1) \\ n_i &= \exp(d_{2i}\gamma_2) \\ t_i &= \exp(d_{3i}\gamma_3) \end{aligned}$$

donde $d_{li} = (d_{li,1}, \dots, d_{li,h})$ es un vector de características o covariables y $\gamma_l^\top = (\gamma_{l,1}, \dots, \gamma_{l,h})$ es un vector columna de coeficientes de regresión, para $l = 1, 2, 3$.

La media y la varianza de $X_{i,k}$ están dadas por

$$E(X_{i,k}) = \frac{\exp(d_{1i}\gamma_1) \exp(d_{2i}\gamma_2)}{\exp(d_{3i}\gamma_3) - 1}$$

y

$$Var(X_{i,k}) = \frac{[\exp(d_{1i}\gamma_1)]^2 \exp(d_{2i}\gamma_2)}{\exp(d_{3i}\gamma_3) - 1} \left\{ \frac{\exp(d_{2i}\gamma_2) + \exp(d_{3i}\gamma_3) - 1}{[\exp(d_{3i}\gamma_3) - 1][\exp(d_{3i}\gamma_3) - 2]} \right\}$$

3.3. Estimación de primas bajo los principios del valor esperado y de la desviación estándar

Utilizando los modelos indicados anteriormente, que permiten modelar tanto la frecuencia esperada como la severidad, hemos conseguido estimar su prima neta. En la práctica, sin embargo, se deben tomar en cuenta otros aspectos como los márgenes de utilidad y la exclusión de ruina. Para su incorporación se han propuesto en la literatura, varios principios de valuación, los cuales, por la naturaleza de nuestros modelos, calzan perfectamente a nuestros propósitos. Los principios que usaremos para un asegurado i , el cual presenta un número K_i de reclamos y severidades independientes e idénticamente distribuidas $X_{i,1}, X_{i,2}, \dots, X_{i,m}$ para algún m en \mathbb{N} , serán:

- El principio de valor esperado nos brinda la prima

$$\text{Prima} = (1 + \rho)E(K_i)E(X_{i,k})$$

donde $\rho > 0$ es un factor de seguridad.

- El principio de la desviación estándar nos brinda la prima

$$\text{Prima} = E(K_i) + p\sqrt{Var(K_i)} \left(E(X_{i,k}) + p\sqrt{Var(X_{i,k})} \right),$$

donde $\rho > 0$ es un factor de seguridad.

Capítulo 4

Aplicación

En los capítulos anteriores se han detallado modelos de tarificación considerando tanto información a priori como a posteriori del asegurado sobre las distribuciones de frecuencia y severidad de los siniestros. A continuación, se presentará un conjunto de datos reales, así como la aplicación de los modelos descritos a este conjunto de datos que comprende información de un portafolio de asegurados de una compañía de seguros local, expuestos en un periodo de tiempo específico, plazo en el cual estos asegurados estuvieron expuestos al riesgo de presentar un accidente vehicular que pudo afectar el vehículo o a la persona asegurada.

El interés en la aplicación de estos modelos a un conjunto real de datos consiste en validar los supuestos establecidos, comparar los resultados entre estos y obtener conclusiones que nos sirvan para como punto de partida para el desarrollo de nuevos modelos que incorporen el conocimiento del portafolio de asegurados y los siniestros.

4.1. Descripción de los datos

Los datos corresponden a un Seguro de Automóviles de una compañía de seguros peruana, la cual ha permitido su uso. Los asegurados analizados han estado dentro del portafolio durante un periodo de tiempo específico, en el cual todos ellos han tenido la misma exposición. En este periodo, algunos de los asegurados han presentado uno o más siniestros con distintos montos y otros no han presentado reclamaciones.

Consideraciones a tener en cuenta en la información:

1. El periodo de análisis será de tres años.
2. Los asegurados que tienen tres años de exposición continua en el seguro ingresaran al análisis.
3. Se incluyen únicamente los siniestros que se deriven de un accidente del vehículo asegurado.
4. Se trabajarán con aquellos casos en donde se han visto afectadas las coberturas del seguro correspondientes a: Daños al vehículo, Lesiones a ocupantes, Responsabilidad Civil Daños y Responsabilidad Civil Lesiones.
5. Los montos de cada siniestro se trabajarán en moneda dólar americano.

4.1.1. Información a posteriori

La información a posteriori es aquella que comprende la información del número de asegurados, número de siniestros del portafolio y montos relacionados a las reclamaciones. Durante el periodo de análisis se cuenta con un total de 49,921 asegurados quienes han presentado 20,141 siniestros lo que indica que en el portafolio expuesto existe un alto riesgo de sufrir accidente vehicular con lesiones leves o graves del asegurado y/o daño del vehículo. El monto asociado a las pérdidas que cubre el seguro pueden variar de acuerdo al grado de la lesión del asegurado o daño del vehículo.

En el portafolio analizado, se observa que el 72 % de los asegurados no han presentado siniestros, el 19 % ha presentado 1 siniestro, el 6 % ha presentado 2 siniestros y el 3 % ha presentado 3 siniestros o más. Por cada año de exposición, se tiene que el 22 % de los siniestros se han presentado en el primer año, 46 % durante el segundo año y 36 % en el tercer año, considerando que durante el periodo de análisis y por cada año de exposición, el asegurado, puede tener uno o más siniestros.

Dado que el evento de interés consiste en observar que un asegurado presente un accidente vehicular, de forma que se afecte la póliza de seguro, se analizarán las variables a posteriori: número de siniestros que presenta el asegurado y el monto de reclamación de cada uno de los siniestros.

De acuerdo al histograma de siniestros podemos observar que hay una gran cantidad de asegurados que no presenta siniestros o presenta un siniestro durante el periodo de análisis, por lo que se puede observar que la variable número de siniestros presenta una asimetría positiva y una curtosis alta.

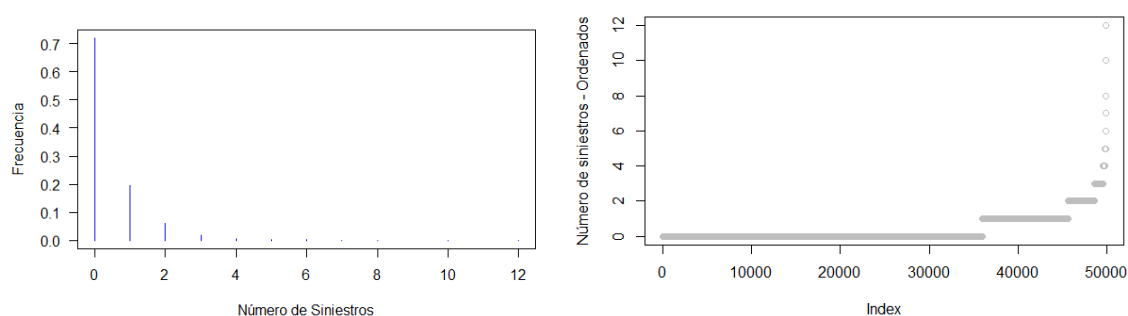


Figura 4.1: Histograma y gráfico de los valores ordenados del número de siniestros

En la figura 4.1 se observa que la distribución de la variable número de siniestros es asimétrica a la derecha y leptocúrtica.

Por otro lado, se presentan varios casos extremos en la variable número de siniestros, de entre más de 5 siniestros y hasta un máximo de 12 siniestros, lo cual indica que estos asegurados tienen un comportamiento diferente del resto de la cartera. Sin embargo, dado que son datos reales observados, se mantendrán en el análisis sin realizar algún tratamiento adicional.

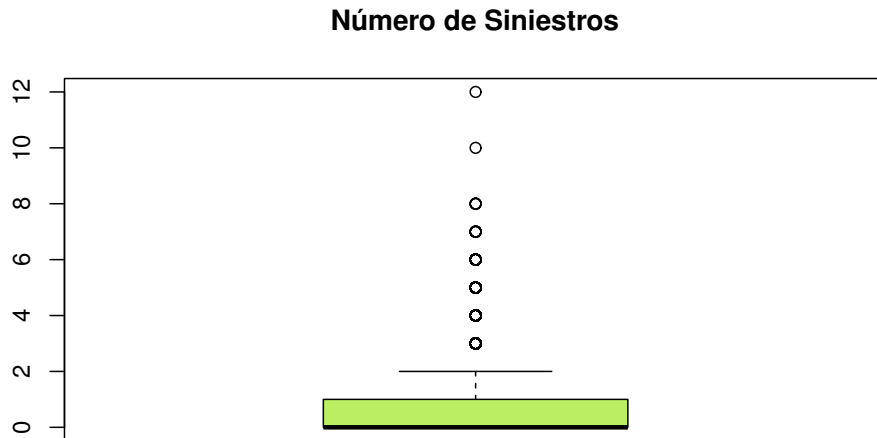


Figura 4.2: Diagrama de caja del número de siniestros por asegurado

Teniendo en cuenta, que la exposición de las pólizas consideradas es de tres años continuos, por cada año la frecuencia se define como el cociente entre el número de siniestros del portafolio y su exposición, sobre el conjunto de datos analizado se ha cuantificado y se obtiene:

$$\text{Frecuencia anual} = 0.134$$

Esto significa que por cada 1,000 asegurados expuestos en un año se estima que se tendrá 134 siniestros en el año con una desviación estándar de la frecuencia anual de 259 casos.

Adicionalmente, cuando un asegurado presenta un siniestro, la compañía de seguros resarce la pérdida económica mediante la indemnización acordada en la póliza de seguro, dicha indemnización corresponde al monto de reclamo.

En el portafolio analizado, este monto de reclamo es igual a cero para el 72 % de los asegurados. El 10 % de los siniestros tienen un monto menor o igual a \$700 dólares, el 15 % presentó siniestros que varían entre \$700 y \$5,000 dólares y el 2 % presentó reclamaciones mayores a \$5,000 dólares.

La variable monto de siniestro, es también asimétrica y leptocúrtica.

El monto promedio de siniestro o severidad se estima como el cociente entre el monto total de siniestros que ha presentado el asegurado o portafolio y el número de siniestros relacionados a estas reclamaciones. El monto promedio de siniestro del portafolio observado es

$$\text{Monto promedio} = \$535$$

Con una desviación estándar del monto promedio de \$1,980.

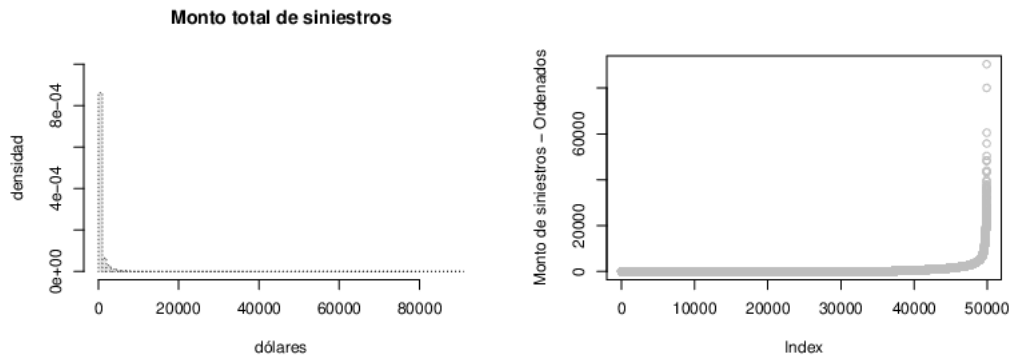


Figura 4.3: Histograma y gráfico de los valores ordenados del monto total de siniestros

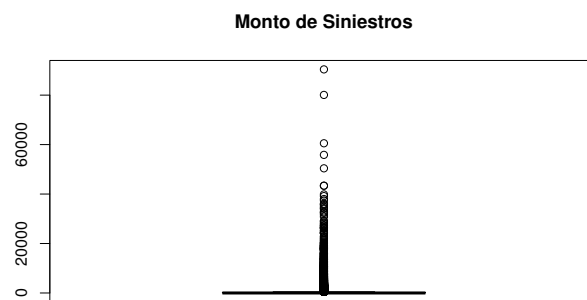


Figura 4.4: Diagrama de caja del monto de siniestro por asegurado

Como podemos ver en la figura 4.4, se presentan valores extremos en la variable monto promedio de siniestros por asegurado, con montos superiores a \$60,000. Dado que estamos considerando las diferentes coberturas de seguros de automóviles este es un comportamiento típico del portafolio, por lo que, los estos valores se mantendrán en los análisis.

4.1.2. Información a priori

En la sección anterior se han presentado las variables a posteriori, número de siniestros, costo de siniestros, frecuencia y severidad que nos muestran el comportamiento del portafolio únicamente desde el punto de vista de las reclamaciones que han ocurrido.

Con el fin de introducir variables relacionadas al riesgo que se desea medir, antes de que este se materialice, se han examinado las variables a priori, que se tienen disponibles. Estas variables corresponden al asegurado y al vehículo, dentro de las variables del asegurado se tienen: edad, género y región, mientras que para el vehículo se tienen: marca, uso, clase, grupo, edad del vehículo y suma asegurada.

Las variables que identifican al asegurado son edad y genero, así como la región donde reside. Por otra parte, las variables del vehículo son: marca, la cual se ha considerado agrupada dado la gran cantidad de marcas disponibles. La clase que identifica el tipo de vehículo, el uso: comercial o particular y el grupo que indica si se trata de un vehículo liviano o pesado.

En el cuadro 4.1 se pueden observar las variables consideradas indicando su media y des-

viación estándar (para las variables cuantitativas) y frecuencia relativa (para las variables categóricas)

Variable	Descripción	Nombre de variable	Media / Frec.	Desv. Estandar
NUM SIN	Número de siniestros	Num_siniestros	0.40	0.78
GENERO	Genero del Asegurado 1 = masculino 2 = femenino 3 = empresarial	genero_m genero_f genero_e	0.52 0.27 0.20	
EDAD_ASEG	Edad del asegurado en años	edad_a	44.5	14.46
REGION	Región 0 = otros 1 = lima	region_o region_l	0.10 0.90	
EDAD_VEH	Edad del vehiculo en años	edad_v	4.62	5.03
SUMA_ASE	Suma asegurada en dolares	suma_a	18,548.0	15,217.4
CLASE	Clasificación del Vehiculo 1 = automoviles 2 = camiones 3 = pick up 4 = rurales 5 = otros	clase_a clase_c clase_p clase_r clase_o	0.53 0.06 0.03 0.36 0.03	
USO	Uso del vehiculo 0 = comercial 1 = particular	uso_c uso_p	0.16 0.84	
GRUPO	Grupo del vehiculo 0 = livianos 1 = pesados	grupo_l grupo_p	0.94 0.06	
MARCA	Marca del Vehiculo 1 = toyota 2 = hyundai 3 = volkswagen 4 = honda 5 = mazda 6 = otros	marca_toy marca_hyu marca_vol marca_hon marca_maz marca_otr	0.22 0.13 0.05 0.05 0.04 0.51	

Cuadro 4.1: Variables a priori

Las variables dependientes son el número de reclamos o frecuencia de siniestros y monto promedio de siniestro o severidad. En cuanto a las covariables cuantitativas tenemos: edad del asegurado, edad del vehículo, suma asegurada siendo, esta última continua. Asimismo, las covariables categóricas a considerar son: genero, región, marca, clase, uso y grupo.

En la figura 4.5 se muestra el histograma de las variables cuantitativas. En la variable edad del asegurado se observa que hay un alto porcentaje de asegurados en el rango de 30 a 35 años, para los demás rangos de edad se observa que el portafolio es homogéneo. Por otro lado, en la variable la edad del vehículo se observa que el portafolio está concentrado en vehículos de hasta 5 años de antigüedad y las sumas aseguradas se concentran en el rango de 10,000 a 30,000 dólares.

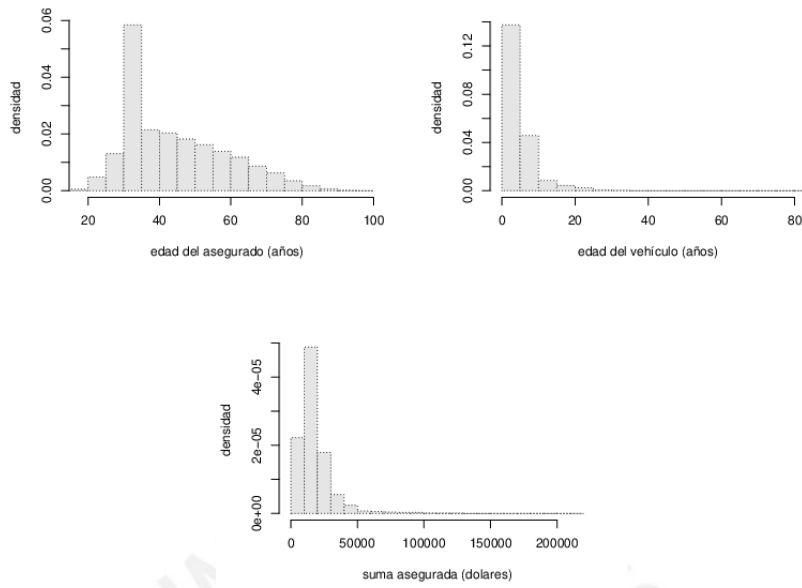


Figura 4.5: Histogramas de edad asegurado, edad del vehículo y suma asegurada

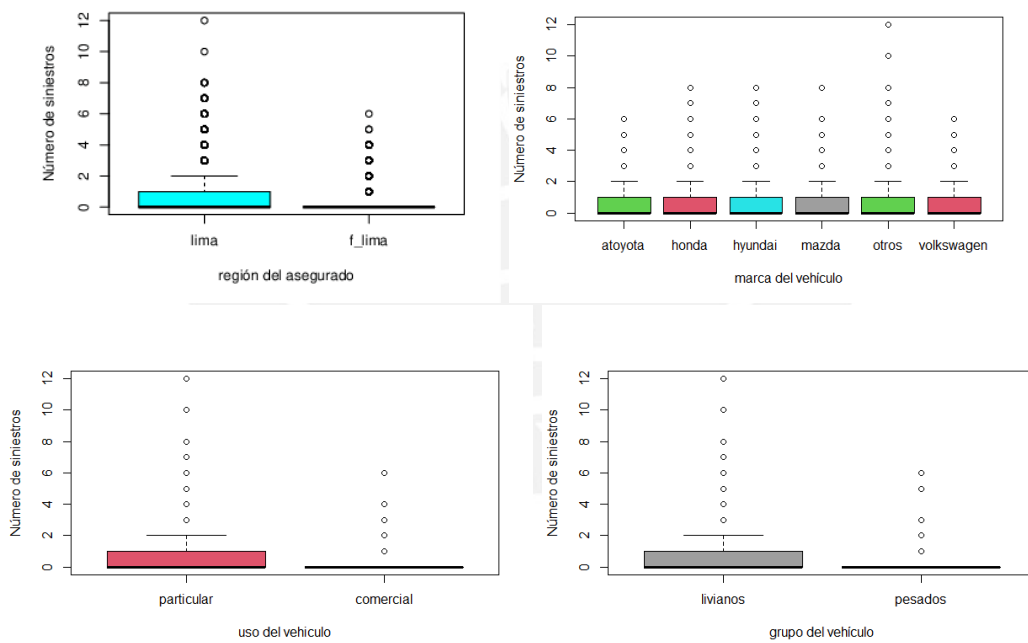


Figura 4.6: Diagramas de caja de siniestros de acuerdo a las variables categóricas región, marca, uso y grupo.

Por otro lado, al observar la interacción de la variable dependiente número de siniestros con las variables regresoras categóricas, vemos que estas variables son asimétricas y presentan outliers.

En cuanto las variables del asegurado, en la siguiente figura se observa que en las edades jóvenes se tiene mayor número de siniestros, tanto para Lima como para fuera de Lima (f_lima)

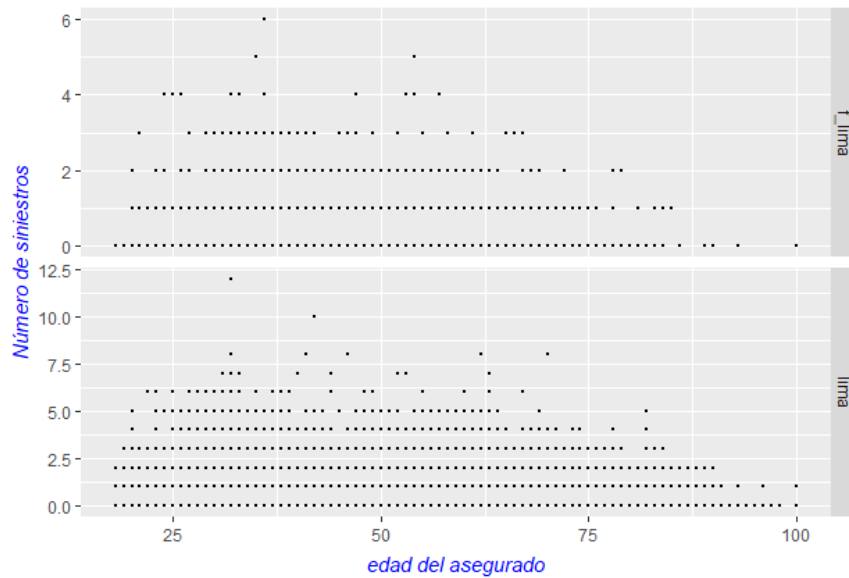


Figura 4.7: Interacción de variables cuantitativas y categóricas del asegurado en el periodo de t años

Adicionalmente podemos observar la interacción entre las variables del vehículo asegurado.

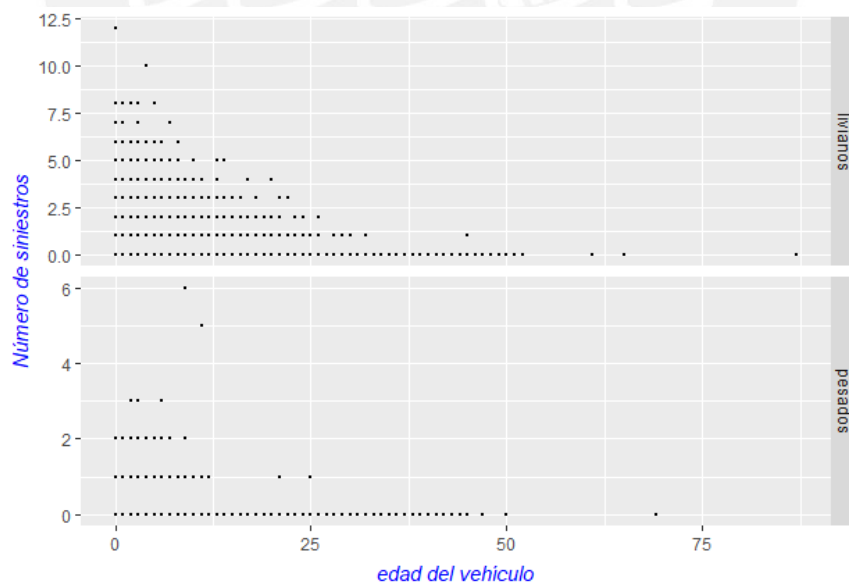


Figura 4.8: Interacción de variables cuantitativas y categóricas del vehículo en el periodo de t años

A continuación, se presenta un diagrama de dispersión multivariado considerando las variables cuantitativas, número de casos, edad del vehículo, edad asegurado y suma asegurada.

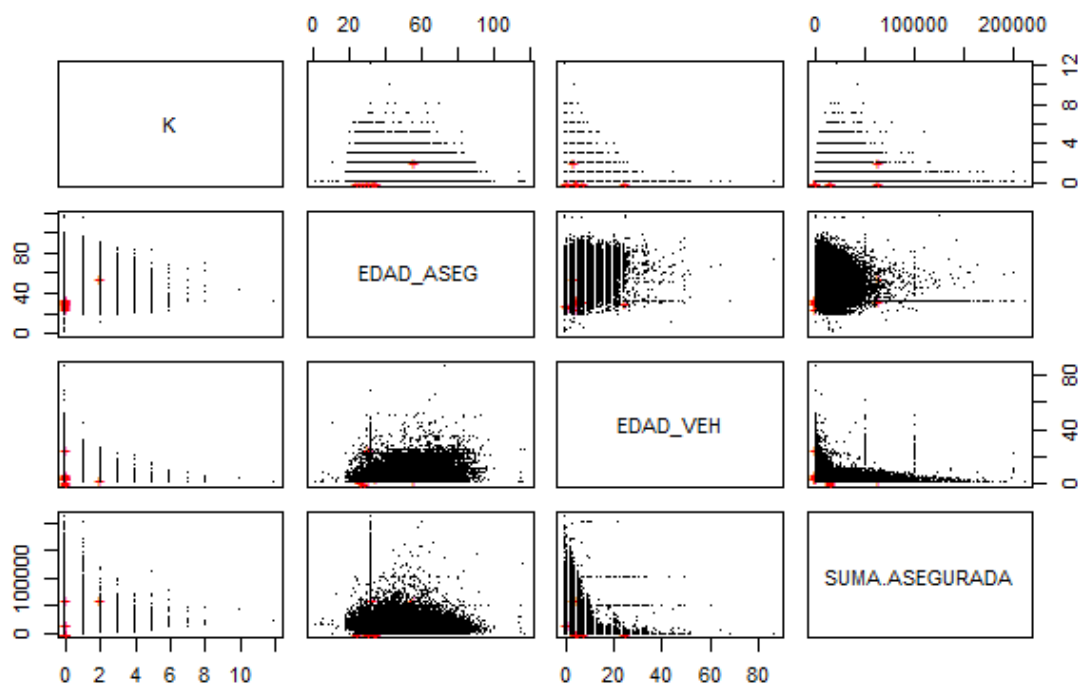


Figura 4.9: Diagrama de dispersión multivariado considerando las variables número de siniestros, edad del asegurado, edad del vehículo y suma asegurada

4.2. Modelos BMS con información a posteriori

Los modelos propuestos en esta sección se concentran en determinar la frecuencia de siniestros y la severidad de dichos siniestros, para esto se analizarán tres modelos diferentes en el caso de frecuencia y un único modelo en el caso de la severidad.

Los modelos para estimar la frecuencia de siniestros serán:

- Modelo de frecuencia binomial negativo.
- Modelo de frecuencia Poisson inversa Gaussiana.
- Modelo de frecuencia Sichel.

Para estimar la severidad se considerará el Modelo de severidad Pareto

Dentro de cada modelo se han estimado los parámetros, estos se han obtenido mediante máxima verosimilitud y en algunos casos se aplicaron métodos numéricos.

4.2.1. Modelo binomial negativo

En este modelo se asume que el número de siniestros se distribuye $BN(a, \tau)$ Para estimar los parámetros de la distribución, consideramos el paquete en R, `fitdistrplus`, en donde obtenemos los EMV para a y τ .

Obtenemos $\hat{a} = 0.8274424$ y $\hat{\mu} = 0.4032333$ y $\hat{\tau} = \frac{\hat{a}}{\hat{\mu}} = 2.052$. Observamos el ajuste que tiene el aplicar al número de siniestros la distribución $BN(a, \tau)$.

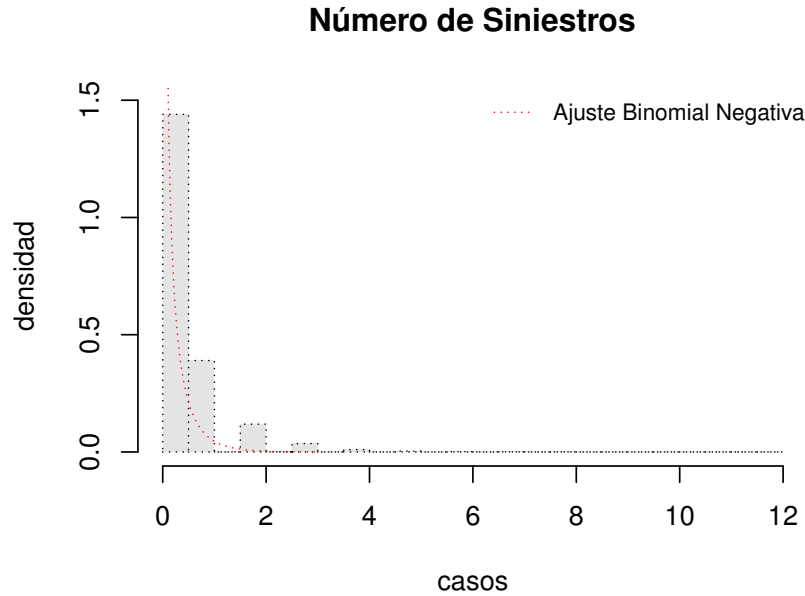


Figura 4.10: Histograma del número de siniestros y curva de la distribución binomial negativa ajustada.

Dado que, hemos encontrado estimadores de \hat{a} y $\hat{\tau}$ procedemos a calcular la frecuencia para el periodo, $t + 1$.

4.2.2. Modelo Poisson inverse Gaussian (PIG)

Por otro lado, para el modelo $PIG(\mu, \sigma)$ se han revisado los parámetros utilizando las derivadas de la función de log-verosimilitud, según lo indicado en Stein et al. (1987).

Para ν igual a $-1/2$ las ecuaciones de verosimilitud son:

$\hat{\xi} = \bar{x}$ y $\hat{\alpha}$, la cual es la solución de

$$\left[\hat{\alpha}(\hat{\xi}^2 + \hat{\alpha}^2)^{-1/2} \right] \left[1 + \frac{\bar{x}}{\hat{\omega}} \right] - \frac{1}{n} \sum_{i=1}^n R_{x_i - 1/2}(\hat{\alpha}) = 0,$$

donde $\hat{\xi} = \mu$, $\gamma = \nu$, $\hat{\omega} = (\hat{\xi}^2 + \hat{\alpha}^2)^{1/2} - \hat{\xi}$, $\hat{\sigma} = \frac{1}{\hat{\omega}}$ y $R_\nu(z) =$ es el cociente de las funciones de Bessel.

Para cada una de estas ecuaciones se han estimado los estimadores con la librería `optim`, en el anexo 1 se incluye el código en R.

Los estimadores de los parámetros en el modelo PIG son $\hat{\alpha} = 1.846714$, $\hat{\xi} = 0.4035$ y dado que estamos trabajando con la PIG tenemos que $\nu = -1/2$

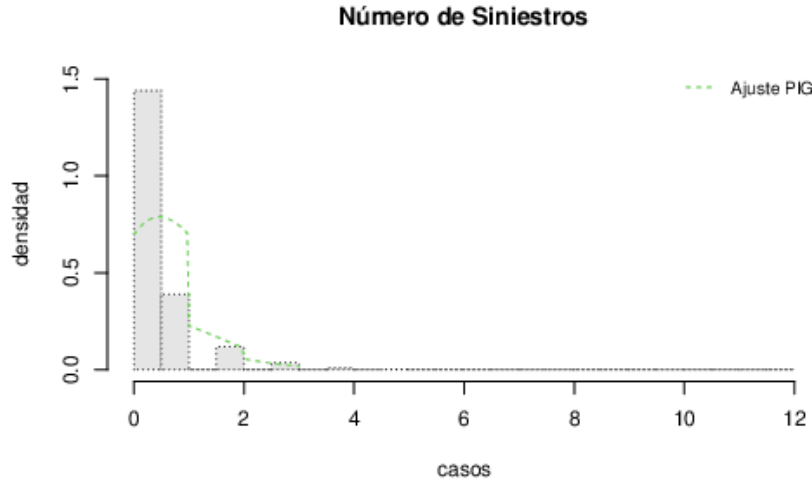


Figura 4.11: Distribución del número de siniestros y ajuste de la curva $PIG(\hat{\alpha}, \hat{\xi}, -1/2)$

y de acuerdo con 2.6, obtenemos la frecuencia en el periodo $t + 1$, considerando los t periodos anteriores.

4.2.3. Modelo de Frecuencia de Sichel

Para el modelo de frecuencia de Sichel con parámetros $(\mu, \sigma$ y $\nu)$ se han obtenido los estimadores de máxima verosimilitud de μ y σ mediante iteración de Newton-Raphson para un ν fijo y posteriormente se ha buscado maximizar la función de log-verosimilitud con los estimadores de μ y σ encontrados, para luego aplicar el algoritmo indicado en el apéndice B de Stein et al. (1987).

Tomando $\gamma = \nu$, $\mu = \xi R_\gamma(\omega)$, $\xi = \frac{\alpha\theta}{2\sqrt{1-\theta}}$, $w = \alpha\sqrt{1-\theta}$ y $\sigma = \frac{1}{\omega}$. Tenemos que para un γ fijo las ecuaciones de verosimilitud son:

$$\hat{\alpha}(\hat{\xi}^2 + \hat{\alpha}^2)[R_\gamma(\hat{\omega}) + \bar{x}/\hat{\omega}] - \frac{1}{n} \sum_{i=1}^n R_{x_i+\gamma}(\hat{\alpha}) = 0 \text{ y } \hat{\xi} R_\gamma(\hat{\omega}) = \bar{x}$$

y las segundas derivadas de función de Log verosimilitud de la distribución $S(\alpha, \xi, \gamma)$ dadas las observaciones x_1, x_2, \dots, x_n son:

$$\begin{aligned} \frac{\partial^2}{\partial \alpha^2}(\log L) &= x \cdot (\xi^2 \omega - \alpha^2 \eta) / (\omega^2 \eta^3) + n R_\gamma(\omega) [\xi^2 \omega - \alpha^2 \eta (2\gamma + 1) / (\omega \eta^3)] \\ &\quad + n \alpha^2 (R_\gamma^2(\omega) - 1) / \eta^2 + (2/\alpha) \sum_{i=1}^n x_i R_{x_i+\gamma}(\alpha) \\ &\quad + [(2\gamma + 1)/\alpha] \sum_{i=1}^n R_{x_i+\gamma}(\alpha) - \sum_{i=1}^n R_{x_i+\gamma}^2(\alpha) + n \\ \frac{\partial^2}{\partial \xi \partial \alpha}(\log L) &= \alpha x / \eta^3 + n \alpha R_\gamma(\omega) [(2\gamma + 1) \eta - \alpha \xi] / \eta^3 - n \alpha \omega (R_\gamma^2(\omega) - 1) / \eta^2 \\ \frac{\partial^2}{\partial \xi^2}(\log L) &= -x \cdot (\xi^{-2} - \xi / \eta^3) + n \{ \omega^2 \eta (R_\gamma^2(\omega) - 1) - R_\gamma(\omega) [(2\gamma + 1) \omega \eta - \alpha^2] \} / \eta^3, \end{aligned}$$

donde $\eta = \sqrt{\xi^2 + \alpha^2}$ y $x. = \sum_{i=1}^n x_i$, estas ecuaciones son utilizadas para el método Newton-Raphson.

Dado que, hemos obtenido los valores de $\hat{\alpha}$ y $\hat{\xi}$ que son $\hat{\alpha} = 1.82$ y $\hat{\xi} = 3,250.24$ procedemos a calcular valor de γ .

El estimador de γ se calcula máximando la función de log-verosimilitud que contempla en su formulación las funciones de Bessel $B_\gamma(\omega)$ y $B_{x_i+\gamma}(\alpha)$, pero dado que estas funciones solo se pueden aproximar para valores grandes de ω y α Abramowitz y Stegun (1964), se deja como un posible desarrollo futuro el contemplar un método de aproximación de estas funciones para valores pequeños, que contribuyan a la resolución del problema.

4.2.4. Modelo de severidad de Pareto

En el modelo de severidad de Pareto, se han estimado los parámetros de la distribución de la severidad mediante el método momentos. Bajo el supuesto que la severidad se distribuye Pareto(s, m), los estimadores de momentos son $\hat{s} = 2.157$ y $\hat{m} = 618.91$. Siguiendo a Hogg y Klugman (1984) estimamos con el método de Newton los parámetros \hat{s} y \hat{m} y obtenemos 2.74 y 487.81 respectivamente.

En la figura 4.12, se puede observar el histograma de la frecuencia de los siniestros, con la distribución Pareto ajustada

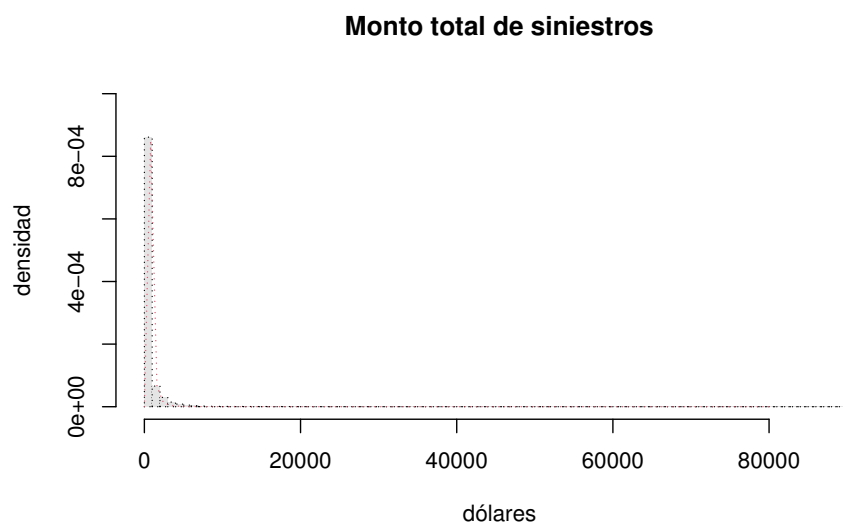


Figura 4.12: Histograma del monto total siniestros y curva de la distribución Pareto ajustada.

Dado que hemos obtenido los estimadores de las distribuciones de la frecuencia de siniestros y severidad, en base a su historia de los t periodos anteriores bajo el modelo de frecuencia binomial negativa y severidad gamma inversa, podemos encontrar para cada asegurado la estimación de la frecuencia λ^{t+1} de acuerdo con (2.1) y del costo promedio y^{t+1} de acuerdo con (2.5), brindándonos ellos la prima neta de acuerdo al número de siniestros que ha presentado el asegurado en el periodo de análisis.

Siniestros Historicos	Frecuencia BN y Severidad Pareto	Frecuencia PIG y Severidad Pareto
0	100	100
1	390	252
2	607	345
3	866	468
4	942	500
5	1,209	637
6	2,115	1,110
7	1,498	786

Cuadro 4.2: Prima neta en unidades de 100 dólares

La estimación de la severidad para el periodo $t + 1$ de acuerdo con (2.7) muestra que al asegurado que tiene 0 siniestros, se le asigna un monto esperado diferente de cero, dado que únicamente estamos teniendo en consideración en el modelo la información de siniestros, en consecuencia se obtienen primas elevadas.

4.3. Modelo BMSG con variables a priori y posteriori

En la sección 4.1.2 se presentaron las variables de tarificación o variables a priori disponibles para el seguro de automóviles. Sin embargo, al revisar estas variables en los modelos se observó que algunas de estas no son significativas, por lo que se optó por retirarlas del análisis.

Las variables regresoras consideradas son: edad del asegurado, región, suma asegurada, edad del vehículo, grupo, uso y marca.

Dentro de los modelos indicados con información a priori se revisarán:

- Frecuencia bajo el modelo binomial negativa
- Frecuencia bajo el modelo Sichel
- Severidad bajo el modelo Pareto

En todos los modelos se ha tomado una consideración sobre el comportamiento de las variables de un año a otro y este supuesto es que únicamente las variables relacionadas con la edad cambiarán, las demás se mantienen constantes de periodo a otro.

Por otro lado, en estos modelos se considera que la media será el único parámetro que se estimará como función de las variables explicativas.

Para los modelos que revisaremos a continuación se tendrán en cuenta los siguientes aspectos para su diagnóstico y calibración.

Cuantiles residuales normalizados (aleatorizados)

Usaremos los cuantiles residuales con el fin de validar la adecuación del ajuste del modelo GAMLSS, dado que tienen la ventaja que para cualquier distribución de la variable de

respuesta, los residuales tendrán una distribución $N(0, 1)$ bajo el supuesto que el modelo es correcto.

Si tenemos que la distribución $f(y | \theta)$ es ajustada a las observaciones y_i , para $i = 1, \dots, n$, los cuantiles residuales normalizados están dados por:

$$\hat{r}_i = \Phi^{-1}(\hat{u}_i),$$

donde Φ^{-1} función de distribución inversa de la normal estándar. Según lo indicado en Stasinopoulos et al. (2017)

En adelante, a los cuantiles residuales normalizados, los llamaremos residuales.

En los gráficos de los residuales, si se comportan bien, se observa que los valores ajustados de μ y los residuales, muestran una dispersión aleatoria alrededor de la línea horizontal que es cero. Mientras que el gráfico de Kernel muestra una estimación de los residuales que es aproximadamente normal y por otro lado, el Q-Q plot presenta un comportamiento lineal.

En cuanto a las estadísticas de los cuantiles residuales, si la media está cerca a cero, la varianza cerca a 1, el coeficiente de skewness (asimetría) es cercano a cero y el coeficiente de curtosis cercano a 3 entonces estas estadísticas sugieren que los residuales están distribuidos $N(0, 1)$ y esto indicaría que el modelo es adecuado.

El comportamiento de los residuales los podemos evidenciar gráficamente, mediante la función `plot` (del paquete `GAMLSS` de R) y mediante las estadísticas de los residuales que se generan.

Por otro lado, los gráficos `worm plot` nos permiten ver en que regiones de una variable explicativa, el modelo no se ajusta adecuadamente a la data. Se usa como herramienta de diagnóstico para validar los residuales para diferentes rangos de una o 2 variables explicativas.

Se han calibrado los modelos considerando el criterio general de Akaike (GAIC), también se ha tenido en cuenta el criterio de Akaike (AIC) y el criterio Schwartz Bayesian (SCB).

$$GAIC = \hat{D} + \kappa \times df,$$

donde $\hat{D} = -2\hat{l}$ es la devianza ajustada, \hat{l} es la verosimilitud ajustada, df son los grados de libertad usados en el modelo y κ es una constante. El criterio de Akaike (AIC) se obtiene cuando $\kappa = 2$ y el criterio Schwartz Bayesian (SCB) se obtiene cuando $\kappa = \log(n)$.

4.3.1. Frecuencia bajo el modelo Binomial Negativo

En el modelo de frecuencia obtenemos los coeficientes de la regresión para cada uno de los j periodos observados ($j = 1, 2, 3$), aquí consideramos 7 variables a priori y se obtienen 11 regresores, dado que 4 de ellas son categóricas.

Las variables que se considerarán son:

- Edad Asegurado
- Edad Vehículo
- Uso
 - Particular
 - Comercial
- Grupo
 - Pesados
 - Livianos
- Región
 - Lima
 - Fuera de Lima
- Suma Asegurada
- Marca
 - Honda
 - Hyundai
 - Mazda
 - Volkswagen
 - Otros

En el modelo binomial negativo, se procede a obtener los coeficientes de la regresión utilizando modelos GAMLSS, para estimar el parámetro μ , como función de los coeficientes de regresión y el parámetro γ se obtiene considerando únicamente el intercepto, para cada uno de los j periodos ($j = 1, 2, 3$), teniendo las $h = 11$ variables a priori.

		Año 1		Año 2		Año 3	
μ	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	intercepto	-3.04200	$< 2e - 16$ ***	-2.84500	$< 2e - 16$ ***	-3.31600	$< 2e - 16$ ***
	edad_a_año j	-0.01060	$< 2e - 16$ ***	-0.00533	1.60E-09 ***	-0.00387	4.91E-05 ***
	edad_v_año j	-0.03865	4.04E-16 ***	-0.05890	$< 2e - 16$ ***	-0.04122	$< 2e - 16$ ***
	uso_p	0.92300	$< 2e - 16$ ***	1.13700	$< 2e - 16$ ***	1.23500	$< 2e - 16$ ***
	grupo_p	-0.83300	2.87E-08 ***	-0.94650	5.61E-15 ***	-0.67590	6.18E-08 ***
	region_l	0.33400	1.73E-07 ***	0.33250	2.19E-13 ***	0.32860	3.08E-11 ***
	suma_a	0.00001	6.79E-13 ***	0.00001	$< 2e - 16$ ***	0.00001	$< 2e - 16$ ***
	marca_hon	-0.08756	0.2947	0.27030	7.65E-07 ***	0.36110	1.10E-09 ***
	marca_hyu	0.01947	0.75188	0.14410	0.000398 ***	0.19250	0.0000254 ***
	marca_maz	0.09844	0.25412	0.22030	0.000152 ***	0.36410	8.19E-09 ***
	marca_otr	-0.13110	0.00464 **	0.05172	0.101224	0.14340	0.000054 ***
	marca_vol	-0.09121	0.27747	0.11160	0.051289	0.15580	0.0146 *
σ	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	(Intercept)	0.57537	7.39E-16 ***	-0.25193	0.00015 ***	-0.0198	0.762

Cuadro 4.3: Coeficientes de regresión Modelo Binomial Negativa para los periodos $j = 1, 2, 3$.

Se ha revisado la significancia de las variables con el test de Wald y el test de verosimilitud generalizado, para asegurarnos que las variables son significativas.

	Año 1	Año 2	Año 3
df	13	13	13
Devianza Global	29,646	46,797	41,954
AIC	29,672	46,823	41,980
SBC	29,787	46,937	42,094

Cuadro 4.4: Criterios del Modelo Binomial Negativa para los periodos $j = 1, 2, 3$.

Los siniestros $K_i^j \sim BN(a, \exp(c_i^j \beta^j))$.

Para cada año j , $j = 1, 2, 3$ y asegurado i , tenemos

$$\begin{aligned} \hat{\mu}_i^1 &= \exp(c_i^1 \beta^1) = \exp(\beta_0^1 + c_{i,1}^1 \beta_1^1 + c_{i,2}^1 \beta_2^1 + \dots + c_{i,11}^1 \beta_{11}^1) \\ \hat{\mu}_i^2 &= \exp(c_i^2 \beta^2) = \exp(\beta_0^2 + c_{i,1}^2 \beta_1^2 + c_{i,2}^2 \beta_2^2 + \dots + c_{i,11}^2 \beta_{11}^2) \\ \hat{\mu}_i^3 &= \exp(c_i^3 \beta^3) = \exp(\beta_0^3 + c_{i,1}^3 \beta_1^3 + c_{i,2}^3 \beta_2^3 + \dots + c_{i,11}^3 \beta_{11}^3) \\ \hat{\sigma}_i^1 &= \exp(\gamma_0^1) \quad \hat{\sigma}_i^2 = \exp(\gamma_0^2) \quad \hat{\sigma}_i^3 = \exp(\gamma_0^3) \end{aligned}$$

Con las estadísticas de los residuales en cada año, se puede verificar que el modelo tiene un buen ajuste, dado que estas son muy similares a la de la normal estándar.

	Año 1	Año 2	Año 3
media	-0.003034	0.001156	-0.003494
varianza	1.003289	0.995185	1.002366
coef. de asimetría	-0.003570	-0.004732	0.009300
coef. de curtosis	3.000733	3.025289	2.991582

Cuadro 4.5: Resumen por año de los residuales por periodo $j = 1, 2, 3$.

Por otro lado, si observamos el gráfico del worm plot, de los residuales, vemos que los puntos están cercanos a la línea horizontal en los tres periodos analizados, lo que indica que la distribución de los residuales, es cercana a una normal estándar.

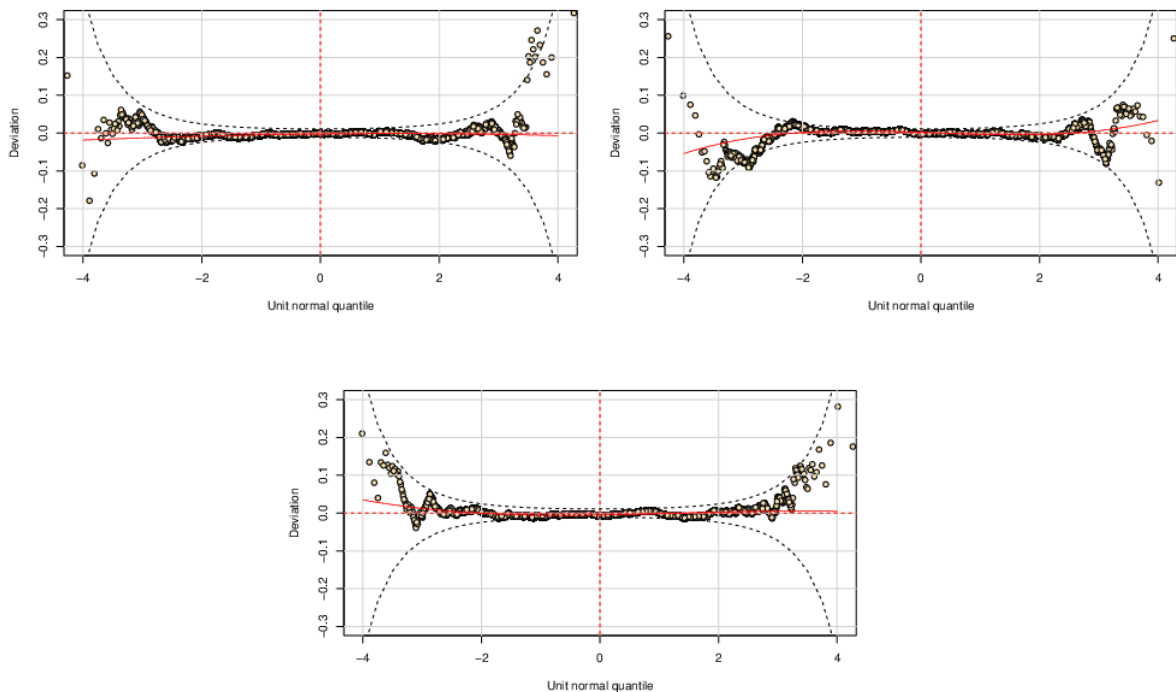


Figura 4.13: Worm plot de los residuales modelo Binomial Negativa por periodo $j = 1, 2, 3$

4.3.2. Frecuencia bajo el modelo Sichel

Otro de los modelos de frecuencia con variables a priori es el de Sichel, aquí debemos estimar los parámetros μ , σ y ν . Con el fin de estimar los parámetros usaremos el paquete GAMLSS de R.

Para este modelo se han considerado las mismas variables indicadas en el modelo de frecuencia Binomial negativa y adicionalmente se ha incorporado una variable categórica más que corresponde al vehículo asegurado. En este caso consideramos 8 variables a priori. La variable adicional considera la clasificación de los vehículos de acuerdo a:

- Automóviles.
- Camiones.
- Pick-up.
- Rurales.
- Otros.

En el modelo Sichel, procedemos a obtener los coeficientes de la regresión para cada uno de los j periodos ($j = 1, 2, 3$), teniendo las $h = 15$ variables a priori, que incluyen la apertura de las variables categóricas.

μ	Variable	Año 1		Año 2		Año 3	
		Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	intercepto	-2.8960	$< 2e - 16$ ***	-2.7000	$< 2e - 16$ ***	-3.2140	$< 2e - 16$ ***
	edad_a_año j	-0.0110	$< 2e - 16$ ***	-0.0057	1.33E-10 ***	-0.0042	3.01E-05 ***
	edad_v_año j	-0.0362	1.68E-12 ***	-0.0571	$< 2e - 16$ ***	-0.0402	$< 2e - 16$ ***
	uso_p	0.8148	$< 2e - 16$ ***	1.0140	$< 2e - 16$ ***	1.1430	$< 2e - 16$ ***
	clase_c	-1.0520	0.000203 ***	-0.9828	3.05E-05 ***	-0.6843	2.00E-03 **
	clase_o	-0.3521	0.028259 *	-0.3508	0.010354 *	-0.1731	0.21467
	clase_p	-0.5166	0.000411 ***	-0.5101	2.44E-06 ***	-0.3556	1.35E-03 **
	clase_r	-0.0333	0.368213	0.0201	0.435304	0.0518	0.0621
	grupo_p	-0.0942	0.696045	-0.2223	0.282334	-0.1461	0.4338
	region_l	0.2979	9.38E-06 ***	0.3060	1.65E-11 ***	0.3113	2.13E-10 ***
	suma_a	0.0000	1.60E-10 ***	0.0000	$< 2e - 16$ ***	0.0000	$< 2e - 16$ ***
	marca_hon	-0.0906	0.270215	0.2755	9.12E-07 ***	0.3670	1.07E-09 ***
	marca_hyu	0.0291	0.618366	0.1449	0.000381 ***	0.1852	4.54E-05 ***
	marca_maz	0.1142	0.165074	0.2348	9.41E-05 ***	0.3732	5.48E-09 ***
	marca_otr	-0.1147	0.009929 **	0.0606	0.056145	0.1426	4.50E-05 ***
	marca_vol	-0.0715	0.384451	0.1375	0.017096 *	0.1778	0.00485 **
σ	Variable	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	(Intercept)	0.67372	$< 2e - 16$ ***	0.4791	0.0121 *	0.3923	6.47E-04 ***
ν	Variable	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	(Intercept)	-0.3376	0.0345 *	0.9701	2.01E-14 ***	0.4456	0.000834 ***

Cuadro 4.6: Coeficientes de regresión Modelo Sichel para los periodos $j = 1, 2, 3$.

Se ha revisado la significancia de las variables con el test de Wald y el test de verosimilitud generalizada, para asegurarnos que las variables son significativas. Estos test se obtienen en el paquete GAMLSS usando `summary()` y `drop1()`.

	Año 1	Año 2	Año 3
df	18	18	18
Devianza Global	29,618	46,754	41,931
AIC	29,654	46,790	41,967
SBC	29,813	46,949	42,125

Cuadro 4.7: Criterios del Modelo Sichel para los periodos $j = 1, 2, 3$.

Los siniestros $K_i^j \sim Sichel(\exp(c_i^j \beta^j), \sigma, \nu)$. Para cada año j , $j = 1, 2, 3$ y asegurado i , tenemos

$$\begin{aligned}
 \hat{\mu}_i^1 &= \exp(c_i^1 \beta^1) = \exp(\beta_0^1 + c_{i,1}^1 \beta_1^1 + c_{i,2}^1 \beta_2^1 + \cdots + c_{i,15}^1 \beta_{15}^1) \\
 \hat{\mu}_i^2 &= \exp(c_i^2 \beta^2) = \exp(\beta_0^2 + c_{i,1}^2 \beta_1^2 + c_{i,2}^2 \beta_2^2 + \cdots + c_{i,15}^2 \beta_{15}^2) \\
 \hat{\mu}_i^3 &= \exp(c_i^3 \beta^3) = \exp(\beta_0^3 + c_{i,1}^3 \beta_1^3 + c_{i,2}^3 \beta_2^3 + \cdots + c_{i,15}^3 \beta_{15}^3) \\
 \hat{\sigma}_i^1 &= \exp(\gamma_0^1) & \hat{\sigma}_i^2 &= \exp(\gamma_0^2) & \hat{\sigma}_i^3 &= \exp(\gamma_0^3) \\
 \hat{\nu}_i^1 &= \exp(\eta_0^1) & \hat{\nu}_i^2 &= \exp(\eta_0^2) & \hat{\nu}_i^3 &= \exp(\eta_0^3)
 \end{aligned}$$

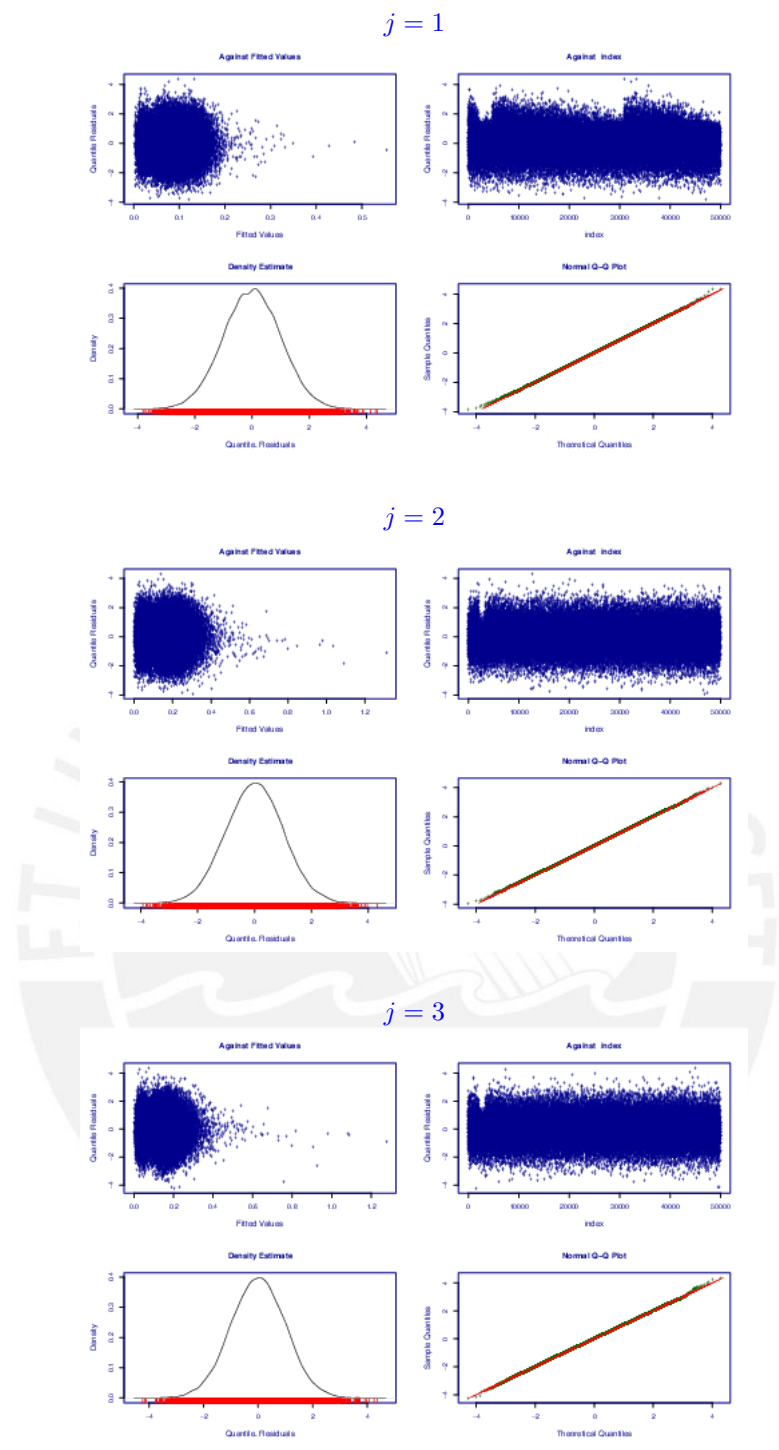


Figura 4.14: Gráficos de residuales del modelo Sichel para $j = 1, 2, 3$

Al analizar el conjunto de datos se ha observa que muchos asegurados no han presentado siniestros y por lo tanto presentan severidades iguales a cero, por lo que, al ajustar los modelos se presentaba error: este problema se puede abordar considerando los modelos cero - inflacionados, los cuales realizan un tratamiento especial de los siniestros con monto igual a cero.

	Año 1	Año 2	Año 3
media	-0.001978	0.004768	0.003048
varianza	1.001605	1.000243	0.994609
coef. de asimetría	0.006046	0.007743	-0.010985
coef. de curtosis	2.979241	2.973763	3.027899

Cuadro 4.8: Resumen de los residuales por periodo j bajo modelo Sichel

En el conjunto real de datos, al tratar de modelar la severidad total de los siniestros con el modelo de Pareto se genera error, de la misma forma si se busca modelar la severidad promedio de siniestro con la distribución Gamma Inversa; en ninguno de los casos se logró obtener resultados del modelo y es por esto que este resultado no se presentará.

4.3.3. Resultados de los modelos BMSG con variables a priori

A pesar de que no se obtuvo resultado de la severidad, a continuación se presentará una tabla resumen con las estimaciones de frecuencia a fin de compararla con los resultados obtenidos en los modelos GAMLSS propuestos. Para esto, a continuación se muestra la tabla de segmentos que se construyen considerando rangos de valores representativos dentro de cada variables a priori que se utilizarán como referencia para presentar los resultados obtenidos.

Segmento	Edad Asegurado 36 - 45	Edad vehiculo 0 - 5	Uso Particular	Grupo Livianos	Región Lima	Rango Suma Asegurada 12,000 - 20,000
1	si	si	si	si	si	si
2	no	si	si	si	si	si
3	no	si	si	si	si	si
4	no	si	si	si	si	si
5	si	no	si	si	si	si
6	si	no	si	si	si	si
7	si	si	si	no	si	si
8	si	si	si	si	no	si
9	si	si	si	si	si	no
10	si	si	si	si	si	no
11	no	si	si	si	si	no

Cuadro 4.9: Tabla de segmentos

Segmento	Frecuencia BN	Frecuencia Sichel
1	0.17	0.20
2	0.18	0.21
3	0.16	0.19
4	0.15	0.18
5	0.14	0.17
6	0.12	0.16
7	0.04	0.06
8	0.13	0.16
9	0.14	0.17
10	0.22	0.25
11	0.23	0.26

Cuadro 4.10: Frecuencias obtenidas modelando μ con los modelos Binomial Negativa y Sichel

4.4. Modelos BMSG-GAMLSS basados en información a priori

En los modelos GAMLSS se considera que no solo la media sino que otros parámetros de la distribución se modelarán como función de las variables explicativas, en particular, para el conjunto de datos analizado, nos estamos refiriendo a la varianza y el sesgo.

4.4.1. Frecuencia bajo del modelo NBII

En este modelo para μ se han considerado 7 variables a priori: edad asegurado, edad vehículo, uso, grupo, región, suma asegurada y marca. En el caso de sigma tenemos 3 variables a priori: edad vehículo, grupo y suma asegurada.

En el cuadro 4.11, se presentan los coeficientes y p-valores respectivos.

μ	Variable	Año 1		Año 2		Año 3	
		Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	intercepto	-3.0310	$< 2e - 16$ ***	-2.8800	$< 2e - 16$ ***	-3.3900	$< 2e - 16$ ***
	edad_a_año j	-0.0103	$< 2e - 16$ ***	-0.0049	1.84E-08 ***	-0.0036	1.67E-04 ***
	edad_v_año j	-0.0359	1.94E-12 ***	-0.0562	$< 2e - 16$ ***	-0.0355	$< 2e - 16$ ***
	uso_p	0.9045	$< 2e - 16$ ***	1.1000	$< 2e - 16$ ***	1.1980	$< 2e - 16$ ***
	grupo_p	-0.8320	8.02E-09 ***	-0.9985	$< 2e - 16$ ***	-0.7127	2.15E-09 ***
	region_l	0.2835	6.07E-06 ***	0.3258	4.19E-13 ***	0.3256	3.21E-11 ***
	suma_a	0.0000	6.21E-14 ***	0.0000	$< 2e - 16$ ***	0.0000	$< 2e - 16$ ***
	marca_hon	-0.1149	0.153216	0.2712	4.56E-07 ***	0.3513	1.14E-09 ***
	marca_hyu	0.0251	0.644881	0.1426	0.000336 ***	0.1913	1.74E-05 ***
	marca_maz	0.0823	0.289872	0.2326	4.01E-05 ***	0.3683	1.30E-09 ***
	marca_otr	-0.1393	9.51E-04 ***	0.0539	0.082158	0.1422	3.68E-05 ***
	marca_vol	-0.0929	0.2414	0.1203	0.031879 *	0.1570	1.14E-02 *
σ	Variable	Año 1		Año 2		Año 3	
		Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	(Intercept)	-2.1520	$< 2e - 16$ ***	-2.5380	$< 2e - 16$ ***	-2.7330	$< 2e - 16$ ***
	edad_v_año j	-0.0050	0.808	-0.0001	0.99603	0.0335	7.05E-02
	grupo_p	-0.5926	2.59E-01	-1.8630	3.28E-03 **	-1.2010	7.44E-04 ***
	suma_a	0.00002	4.88E-06 ***	0.00003	6.23E-15 ***	0.00003	$< 2e - 16$ ***

Cuadro 4.11: Coeficientes de regresión Modelo GAMLSS Binomial negativo II para los periodos $j = 1, 2, 3$.

Los siniestros $K_i^j \sim NBII(\exp(c_{1i}^j, \beta_1^j), \exp(c_{2i}^j, \beta_2^j))$. Por cada parámetro $l = 1, 2$, para cada

año $j = 1, 2, 3$ y asegurado i tenemos,

$$\begin{aligned} \hat{\mu}_i^1 &= \exp(c_{1i}^1 \beta_1^1) = \exp(\beta_{10}^1 + c_{1i,1}^1 \beta_{11}^1 + c_{1i,2}^1 \beta_{12}^1 + \dots + c_{1i,11}^1 \beta_{111}^1) \\ \hat{\sigma}_i^1 &= \exp(c_{2i}^1 \beta_2^1) = \exp(\beta_{20}^1 + c_{2i,1}^1 \beta_{21}^1 + c_{2i,2}^1 \beta_{22}^1 + c_{2i,3}^1 \beta_{23}^1) \\ \hat{\mu}_i^2 &= \exp(c_{1i}^2 \beta_1^2) = \exp(\beta_{10}^2 + c_{1i,1}^2 \beta_{11}^2 + c_{1i,2}^2 \beta_{12}^2 + \dots + c_{1i,11}^2 \beta_{111}^2) \\ \hat{\sigma}_i^2 &= \exp(c_{2i}^2 \beta_2^2) = \exp(\beta_{20}^2 + c_{2i,1}^2 \beta_{21}^2 + c_{2i,2}^2 \beta_{22}^2 + c_{2i,3}^2 \beta_{23}^2) \\ \hat{\mu}_i^3 &= \exp(c_{1i}^3 \beta_1^3) = \exp(\beta_{10}^3 + c_{1i,1}^3 \beta_{11}^3 + c_{1i,2}^3 \beta_{12}^3 + \dots + c_{1i,11}^3 \beta_{111}^3) \\ \hat{\sigma}_i^3 &= \exp(c_{2i}^3 \beta_2^3) = \exp(\beta_{20}^3 + c_{2i,1}^3 \beta_{21}^3 + c_{2i,2}^3 \beta_{22}^3 + c_{2i,3}^3 \beta_{23}^3) \end{aligned}$$

En cuanto a las estadísticas de los residuales, vemos que son cercanos a las de una normal estandar, por lo que podemos concluir que, el modelo es adecuado.

	Año 1	Año 2	Año 3
media	-0.001572	0.003149	-0.003175
varianza	1.004854	0.991578	1.007460
coef. de asimetría	-0.000655	0.005236	0.002005
coef. de curtosis	3.002223	2.984310	2.953256

Cuadro 4.12: Resumen por año de los residuales modelo NBII por periodo $j = 1, 2, 3$.

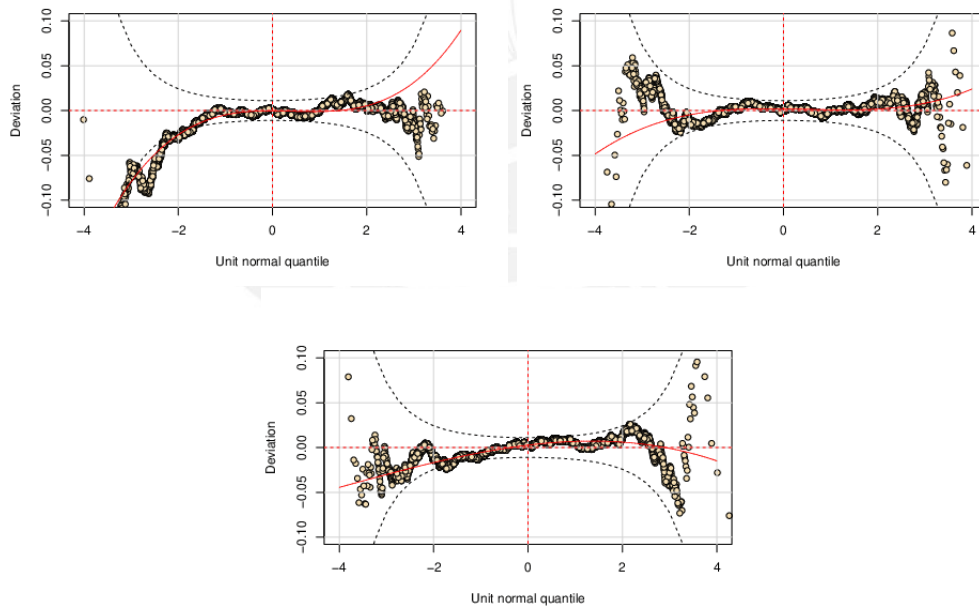


Figura 4.15: Worm plot de los residuales modelos NBII por periodo $j = 1, 2, 3$

Se observa en la figura 4.15 que los residuales tienen un comportamiento cercano a la normal.

4.4.2. Frecuencia bajo del modelo Delaporte

En este modelo para μ se han considerado 7 variables a priori: edad asegurado, edad vehículo, uso, grupo, región, suma asegurada y marca, en el caso de sigma tenemos 3 variables a priori: edad vehículo, grupo y suma asegurada y para ν únicamente se ha modelado con un coeficiente β_0 . A continuación, en el cuadro 4.13, se presentan los coeficientes y p-valores respectivos.

μ	Variable	Año 1		Año 2		Año 3	
		Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	intercepto	-3.0770	< 2e - 16 ***	-2.9000	< 2e - 16 ***	-3.4260	< 2e - 16 ***
	edad_a_año j	-0.0107	< 2e - 16 ***	-0.0054	1.79E-09 ***	-0.0039	1.08E-04 ***
	edad_v_año j	-0.0359	9.01E-12 ***	-0.0555	< 2e - 16 ***	-0.0338	1.2E-15 ***
	uso_p	0.9191	< 2e - 16 ***	1.1330	< 2e - 16 ***	1.2190	< 2e - 16 ***
	grupo_p	-0.8090	7.76E-08 ***	-0.9708	1.67E-15 ***	-0.6339	0.000000316 ***
	region_l	0.3338	6.44E-07 ***	0.3335	1.35E-13 ***	0.3298	1.69E-11 ***
	suma_a	0.0000	1.27E-11 ***	0.0000	< 2e - 16 ***	0.0000	< 2e - 16 ***
	marca_hon	-0.1054	0.20608	0.2549	7.96E-06 ***	0.3380	3.10E-08 ***
	marca_hyu	0.0199	0.72944	0.1484	0.000225 ***	0.2015	5.97E-06 ***
	marca_maz	0.0953	0.24288	0.2206	2.06E-04 ***	0.3710	3.04E-09 ***
	marca_otr	-0.1368	0.00176 **	0.0481	0.125809	0.1384	6.58E-05 ***
	marca_vol	-0.0900	0.27095	0.1187	0.037372 *	0.1642	8.56E-03 **
σ	Variable	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	(Intercept)	0.8146	2.96E-06 ***	-0.741	1.86E-05 ***	-0.6667	0.000905 ***
	edad_v_año j	0.0392	0.0834 .	0.05907	0.00304 **	0.0862	3.93E-05 ***
	grupo_p	1.1130	6.41E-02 .	0.2338	7.33E-01	1.2170	5.92E-04 ***
	suma_a	0.00001	0.0222 *	0.00002	1.87E-05 ***	0.00002	0.000000128 ***
ν	Variable	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	(Intercept)	-1.0343	5.55E-15 ***	-2.5198	4.04E-10 ***	-1.7535	2e-16 ***

Cuadro 4.13: Coeficientes de regresión Modelo GAMLSS Delaporte para los periodos $j = 1, 2, 3$.

Los siniestros $K_i^j \sim DEL \left(\exp(c_{1i}^j \beta_1^j), \exp(c_{2i}^j \beta_2^j), \frac{\exp(c_{3i}^j \beta_3^j)}{1 + \exp(c_{3i}^j \beta_3^j)} \right)$. Por cada parámetro $l = 1, 2, 3$, para cada año $j = 1, 2, 3$ y asegurado i tenemos.

$$\begin{aligned} \hat{\mu}_i^1 &= \exp(c_{1i}^1 \beta_1^1) = \exp(\beta_{10}^1 + c_{1i,1}^1 \beta_{11}^1 + c_{1i,2}^1 \beta_{12}^1 + \dots + c_{1i,11}^1 \beta_{111}^1) \\ \hat{\sigma}_i^1 &= \exp(c_{2i}^1 \beta_2^1) = \exp(\beta_{20}^1 + c_{2i,1}^1 \beta_{21}^1 + c_{2i,2}^1 \beta_{22}^1 + c_{2i,3}^1 \beta_{23}^1) \\ \hat{\mu}_i^2 &= \exp(c_{1i}^2 \beta_1^2) = \exp(\beta_{10}^2 + c_{1i,1}^2 \beta_{11}^2 + c_{1i,2}^2 \beta_{12}^2 + \dots + c_{1i,11}^2 \beta_{111}^2) \\ \hat{\sigma}_i^2 &= \exp(c_{2i}^2 \beta_2^2) = \exp(\beta_{20}^2 + c_{2i,1}^2 \beta_{21}^2 + c_{2i,2}^2 \beta_{22}^2 + c_{2i,3}^2 \beta_{23}^2) \\ \hat{\mu}_i^3 &= \exp(c_{1i}^3 \beta_1^3) = \exp(\beta_{10}^3 + c_{1i,1}^3 \beta_{11}^3 + c_{1i,2}^3 \beta_{12}^3 + \dots + c_{1i,11}^3 \beta_{111}^3) \\ \hat{\sigma}_i^3 &= \exp(c_{2i}^3 \beta_2^3) = \exp(\beta_{20}^3 + c_{2i,1}^3 \beta_{21}^3 + c_{2i,2}^3 \beta_{22}^3 + c_{2i,3}^3 \beta_{23}^3) \\ \hat{\nu}_i^1 &= \frac{\exp(\beta_{30}^1)}{1 + \exp(\beta_{30}^1)} \quad \hat{\nu}_i^2 = \frac{\exp(\beta_{30}^2)}{1 + \exp(\beta_{30}^2)} \quad \hat{\nu}_i^3 = \frac{\exp(\beta_{30}^3)}{1 + \exp(\beta_{30}^3)} \end{aligned}$$

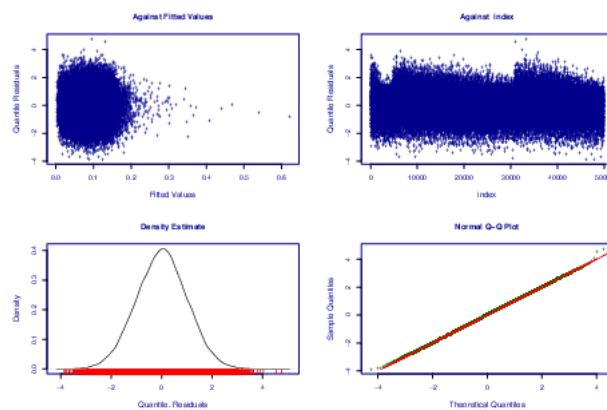
En cuanto a los residuales vemos que el modelo es adecuado para cada uno de los tres años analizados.

	Año 1	Año 2	Año 3
media	0.000665	0.003263	0.000315
varianza	1.005836	0.995983	0.998317
coef. de asimetría	-0.009419	0.003631	0.004891
coef. de curtosis	2.987624	2.997959	3.002019

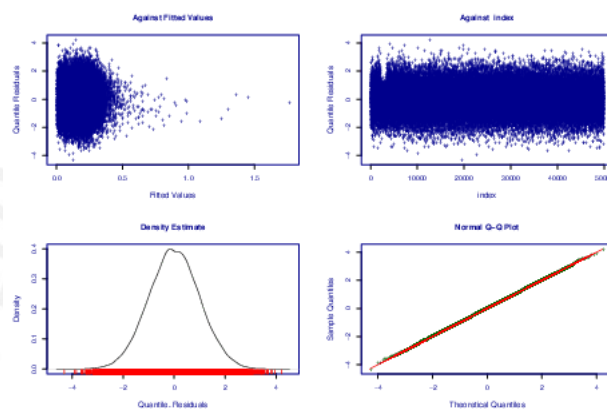
Cuadro 4.14: Resumen por año de los residuales modelo Delaporte por periodo $j = 1, 2, 3$.

En el figura 4.16 se observa que los residuales tienen un comportamiento cercano a la distribución normal

$j = 1$



$j = 2$



$j = 3$

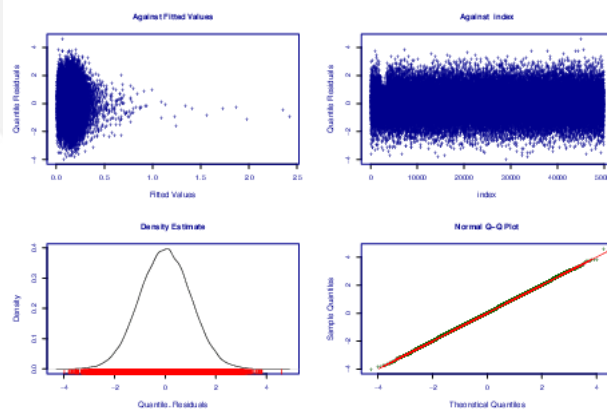


Figura 4.16: Gráfico de residuales del modelo Delaporte para $j = 1, 2, 3$

estándar por lo que se considera que el modelo es adecuado para estimar el número de reclamos.

4.4.3. Frecuencia bajo el modelo de Sichel

Al igual que los modelos de NBII y Delaporte, se han considerado las mismas variables a priori para μ : edad asegurado, edad vehículo, uso, grupo, región, suma asegurada y marca, en el caso de σ se considera: edad vehículo, grupo y suma asegurada y para ν únicamente se ha modelado el intercepto. A continuación, en el cuadro 4.15, se presentan los coeficientes y p-valores respectivos.

μ	Variable	Año 1		Año 2		Año 3	
		Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	intercepto	-3.1010	< 2e - 16 ***	-2.9110	< 2e - 16 ***	-3.4760	< 2e - 16 ***
	edad_a_año j	-0.0107	< 2e - 16 ***	-0.0054	1.85E-09 ***	-0.0039	9.84E-05 ***
	edad_v_año j	-0.0344	8.78E-11 ***	-0.0550	< 2e - 16 ***	-0.0306	6.28E-13 ***
	uso_p	0.9165	< 2e - 16 ***	1.1310	< 2e - 16 ***	1.2120	< 2e - 16 ***
	grupo_p	-0.7831	0.000000301 ***	-0.9697	1.77E-15 ***	-0.5807	5.39E-06 ***
	region_l	0.3335	6.45E-07 ***	0.3336	1.13E-13 ***	0.3285	1.93E-11 ***
	suma_a	0.0000	3.36E-13 ***	0.0000	< 2e - 16 ***	0.0000	< 2e - 16 ***
	marca_hon	-0.1140	0.17157	0.2549	7.02E-06 ***	0.3275	7.91E-08 ***
	marca_hyu	0.0196	0.73263	0.1482	0.000215 ***	0.2020	5.00E-06 ***
	marca_maz	0.0927	0.25368	0.2225	1.59E-04 ***	0.3704	2.42E-09 ***
	marca_otr	-0.1400	0.00131 **	0.0481	0.123673	0.1365	7.78E-05 ***
	marca_vol	-0.0903	0.26759	0.1197	0.034852 *	0.1659	0.00764 **
σ	Variable	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	(Intercept)	0.0347	0.836339	-1.0510	2.26E-08 ***	-1.3500	1.23E-10 ***
	edad_v_año j	0.0432	0.05816 .	0.0652	0.00206 **	0.1018	3.61E-06 ***
	grupo_p	1.1460	3.09E-02 *	0.1721	8.10E-01	1.3420	0.000364 ***
	suma_a	0.00002	5.75E-04 ***	0.00002	1.52E-06 ***	0.00003	1.17E-13 ***
ν	Variable	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$	Coficiente	$Pr(> t)$
	(Intercept)	-0.632	0.0183 *	-0.5205	0.277	-0.7635	9.16E-14 ***

Cuadro 4.15: Coeficientes de regresión Modelo GAMLSS Sichel II para los periodos $j = 1, 2, 3$.

Los siniestros $K_i^j \sim Sichel \left(\exp(c_{1i}^j \beta_1^j), \exp(c_{2i}^j \beta_2^j), \frac{\exp(c_{3i}^j \beta_3^j)}{1 + \exp(c_{3i}^j \beta_3^j)} \right)$. Por cada parámetro $l = 1, 2, 3$, para cada año $j = 1, 2, 3$ y asegurado i tenemos,

$$\begin{aligned}
 \hat{\mu}_i^1 &= \exp(c_{1i}^1 \beta_1^1) = \exp(\beta_{10}^1 + c_{1i,1}^1 \beta_{11}^1 + c_{1i,2}^1 \beta_{12}^1 + \dots + c_{1i,11}^1 \beta_{111}^1) \\
 \hat{\sigma}_i^1 &= \exp(c_{2i}^1 \beta_2^1) = \exp(\beta_{20}^1 + c_{2i,1}^1 \beta_{21}^1 + c_{2i,2}^1 \beta_{22}^1 + c_{2i,3}^1 \beta_{23}^1) \\
 \hat{\mu}_i^2 &= \exp(c_{1i}^2 \beta_1^2) = \exp(\beta_{10}^2 + c_{1i,1}^2 \beta_{11}^2 + c_{1i,2}^2 \beta_{12}^2 + \dots + c_{1i,11}^2 \beta_{111}^2) \\
 \hat{\sigma}_i^2 &= \exp(c_{2i}^2 \beta_2^2) = \exp(\beta_{20}^2 + c_{2i,1}^2 \beta_{21}^2 + c_{2i,2}^2 \beta_{22}^2 + c_{2i,3}^2 \beta_{23}^2) \\
 \hat{\mu}_i^3 &= \exp(c_{1i}^3 \beta_1^3) = \exp(\beta_{10}^3 + c_{1i,1}^3 \beta_{11}^3 + c_{1i,2}^3 \beta_{12}^3 + \dots + c_{1i,11}^3 \beta_{111}^3) \\
 \hat{\sigma}_i^3 &= \exp(c_{2i}^3 \beta_2^3) = \exp(\beta_{20}^3 + c_{2i,1}^3 \beta_{21}^3 + c_{2i,2}^3 \beta_{22}^3 + c_{2i,3}^3 \beta_{23}^3) \\
 \hat{\nu}_i^1 &= \frac{\exp(\beta_{30}^1)}{1 + \exp(\beta_{30}^1)} \quad \hat{\nu}_i^2 = \frac{\exp(\beta_{30}^2)}{1 + \exp(\beta_{30}^2)} \quad \hat{\nu}_i^3 = \frac{\exp(\beta_{30}^3)}{1 + \exp(\beta_{30}^3)}.
 \end{aligned}$$

En cuanto a los residuales vemos que el modelo es adecuado,

	Año 1	Año 2	Año 3
media	0.004375	-0.004061	-0.002209
varianza	1.005892	1.007309	0.997153
coef. de asimetría	-0.009955	-0.002756	-0.006383
coef. de curtosis	3.019514	3.001996	3.010976

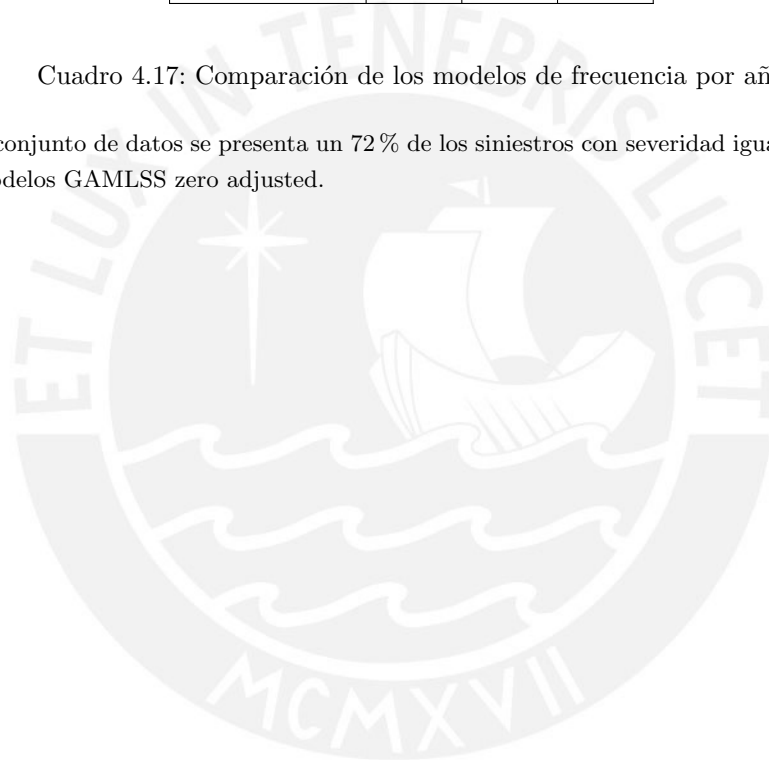
Cuadro 4.16: Resumen por año de los residuales modelo Sichel por periodo $j = 1, 2, 3$.

Se han comparado los modelos de Binomial negativa, Delaporte y Sichel y se tiene que el modelo de Sichel presenta un mejor comportamiento de la devianza global.

NBII	Año 1	Año 2	Año 3
df	16	16	16
Devianza Global	29,646	46,796	41,930
AIC	29,678	46,828	41,962
SBC	29,819	46,969	42,103
DEL	Año 1	Año 2	Año 3
df	17	17	17
Devianza Global	29,639	46,788	41,928
AIC	29,673	46,822	41,962
SBC	29,823	46,972	42,112
SICH	Año 1	Año 2	Año 3
df	17	17	17
Devianza Global	29,637	46,794	41,923
AIC	29,671	46,828	41,957
SBC	29,821	46,978	42,107

Cuadro 4.17: Comparación de los modelos de frecuencia por año

Dado que en el conjunto de datos se presenta un 72% de los siniestros con severidad igual a cero, procedemos a utilizar los modelos GAMLSS zero adjusted.



4.4.4. Severidad bajo el modelo Gamma

En este modelo para μ se han considerado 8 variables a priori: edad asegurado, edad vehículo, uso, clase, grupo, región, suma asegurada, en el caso de σ tenemos 6 variables a priori: edad del asegurado, edad vehículo, uso, grupo, región y suma asegurada y para ξ_0 tenemos 4 variables a priori: edad del asegurado, edad del vehículo, uso y suma asegurada. A continuación, se presentan los coeficientes

μ	Variable	Año 1		Año 2		Año 3	
		Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	intercepto	8.0590	$< 2e - 16$ ***	8.2120	$< 2e - 16$ ***	7.8400	$< 2e - 16$ ***
	edad_a_año j	-0.0076	5.36E-09 ***	-0.0054	4.83E-09 ***	-0.0057	1.26E-08 ***
	edad_v_año j	-0.0056	2.59E-01	-0.0214	1.73E-08 ***	-0.0066	0.108899
	uso_p	-0.6989	3.81E-13 ***	-0.6691	$< 2e - 16$ ***	-0.3180	0.000243 ***
	clase_c	0.1918	0.49369	-0.3554	2.02E-01	0.8866	1.28E-03 **
	clase_o	-0.4512	0.01401 *	-0.3323	0.02036 *	-0.1695	0.247028
	clase_p	0.4251	0.00657 **	0.1861	1.02E-01	0.2565	2.94E-02 *
	clase_r	0.0173	0.65588	-0.0093	0.73499	-0.0543	0.060961
	grupo_p	-0.0061	0.97937	0.6503	0.00831 **	-0.1882	0.443255
	region_l	-0.0914	2.08E-01	-0.2451	2.44E-06 ***	-0.2367	6.12E-06 ***
	suma_a	0.0000	1.23E-11 ***	0.0000	$< 2e - 16$ ***	0.0000	$< 2e - 16$ ***
σ	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	(Intercept)	0.14810	0.02581 *	0.08348	0.105274	0.01490	0.80143
	edad_a_año j	-0.00094	0.22271	-0.00041	0.461513	-0.00138	0.02287 *
	edad_v_año j	-0.00934	3.46E-03 **	-0.00559	1.62E-02 *	-0.00174	4.86E-01
	uso_p	-0.03488	4.54E-01	0.02033	5.91E-01	0.04775	2.84E-01
	grupo_p	-0.24150	0.00853 **	0.23690	0.000729 ***	0.02911	0.70016
	region_l	-0.06310	0.09495	-0.06469	0.021081 *	0.00160	0.95855
	suma_a	0.00000	1.60E-06 ***	0.00000	2.19E-03 **	0.00000	1.31E-03 **
ξ_0	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	(Intercept)	3.07400	$< 2e - 16$ ***	2.79700	$< 2e - 16$ ***	3.18500	$< 2e - 16$ ***
	edad_a_año j	0.01004	1.26E-14 ***	0.00429	9.70E-06 ***	0.00350	6.23E-04 ***
	edad_v_año j	0.04542	$< 2e - 16$ ***	0.06513	$< 2e - 16$ ***	0.04417	$< 2e - 16$ ***
	uso_p	-1.22800	$< 2e - 16$ ***	-1.52300	$< 2e - 16$ ***	-1.58900	$< 2e - 16$ ***
	suma_a	-0.00001	1.91E-05 ***	-0.00001	$< 2e - 16$ ***	-0.00001	$< 2e - 16$ ***

Cuadro 4.18: Coeficientes de regresión Modelo GAMMA para los periodos $j = 1, 2, 3$.

La severidad $X_{i,k}^j \sim \text{Gamma} \left(\exp(d_{1i}^j \gamma_1^j), \exp(d_{2i}^j \gamma_2^j), \frac{\exp(d_{3i}^j \gamma_3^j)}{1 + \exp(d_{3i}^j \gamma_3^j)} \right)$. Por cada parámetro $l = 1, 2, 3$ y asegurado i tenemos,

$$\begin{aligned} \hat{\mu}_i^1 &= \exp(d_{1i}^1 \gamma_1^1) = \exp(\gamma_{10}^1 + d_{1i,1}^1 \gamma_{11}^1 + d_{1i,2}^1 \gamma_{12}^1 + \dots + d_{1i,10}^1 \gamma_{110}^1) \\ \hat{\mu}_i^2 &= \exp(d_{1i}^2 \gamma_1^2) = \exp(\gamma_{10}^2 + d_{1i,1}^2 \gamma_{11}^2 + d_{1i,2}^2 \gamma_{12}^2 + \dots + d_{1i,10}^2 \gamma_{110}^2) \\ \hat{\mu}_i^3 &= \exp(d_{1i}^3 \gamma_1^3) = \exp(\gamma_{10}^3 + d_{1i,1}^3 \gamma_{11}^3 + d_{1i,2}^3 \gamma_{12}^3 + \dots + d_{1i,10}^3 \gamma_{110}^3) \\ \hat{\sigma}_i^1 &= \exp(d_{2i}^1 \gamma_2^1) = \exp(\gamma_{20}^1 + d_{2i,1}^1 \gamma_{21}^1 + d_{2i,2}^1 \gamma_{22}^1 + \dots + d_{2i,6}^1 \gamma_{26}^1) \\ \hat{\sigma}_i^2 &= \exp(d_{2i}^2 \gamma_2^2) = \exp(\gamma_{20}^2 + d_{2i,1}^2 \gamma_{21}^2 + d_{2i,2}^2 \gamma_{22}^2 + \dots + d_{2i,6}^2 \gamma_{26}^2) \\ \hat{\sigma}_i^3 &= \exp(d_{2i}^3 \gamma_2^3) = \exp(\gamma_{20}^3 + d_{2i,1}^3 \gamma_{21}^3 + d_{2i,2}^3 \gamma_{22}^3 + \dots + d_{2i,6}^3 \gamma_{26}^3) \\ \hat{\xi}_{0i}^1 &= \frac{\exp(\gamma_{30}^1 + d_{3i,1}^1 \gamma_{31}^1 + d_{3i,2}^1 \gamma_{32}^1 + \dots + d_{3i,4}^1 \gamma_{34}^1)}{1 + \exp(\gamma_{30}^1 + d_{3i,1}^1 \gamma_{31}^1 + d_{3i,2}^1 \gamma_{32}^1 + \dots + d_{3i,4}^1 \gamma_{34}^1)} \\ \hat{\xi}_{0i}^2 &= \frac{\exp(\gamma_{30}^2 + d_{3i,1}^2 \gamma_{31}^2 + d_{3i,2}^2 \gamma_{32}^2 + \dots + d_{3i,4}^2 \gamma_{34}^2)}{1 + \exp(\gamma_{30}^2 + d_{3i,1}^2 \gamma_{31}^2 + d_{3i,2}^2 \gamma_{32}^2 + \dots + d_{3i,4}^2 \gamma_{34}^2)} \\ \hat{\xi}_{0i}^3 &= \frac{\exp(\gamma_{30}^3 + d_{3i,1}^3 \gamma_{31}^3 + d_{3i,2}^3 \gamma_{32}^3 + \dots + d_{3i,4}^3 \gamma_{34}^3)}{1 + \exp(\gamma_{30}^3 + d_{3i,1}^3 \gamma_{31}^3 + d_{3i,2}^3 \gamma_{32}^3 + \dots + d_{3i,4}^3 \gamma_{34}^3)} \end{aligned}$$

Se puede observar que la estadística del coeficiente de curtosis muestra valores mayores que tres, lo que muestra que no se obtiene un buen ajuste de la distribución Gamma.

	Año 1	Año 2	Año 3
media	0.003334	-0.001800	-0.001636
varianza	0.996003	0.999519	0.995048
coef. de asimetría	0.047059	0.075149	0.089155
coef. de curtosis	3.314850	3.473087	3.483856

Cuadro 4.19: Resumen por año de los residuales modelo Gamma por periodo $j = 1, 2, 3$.

Este comportamiento se puede evidenciar en el gráfico worm plot.

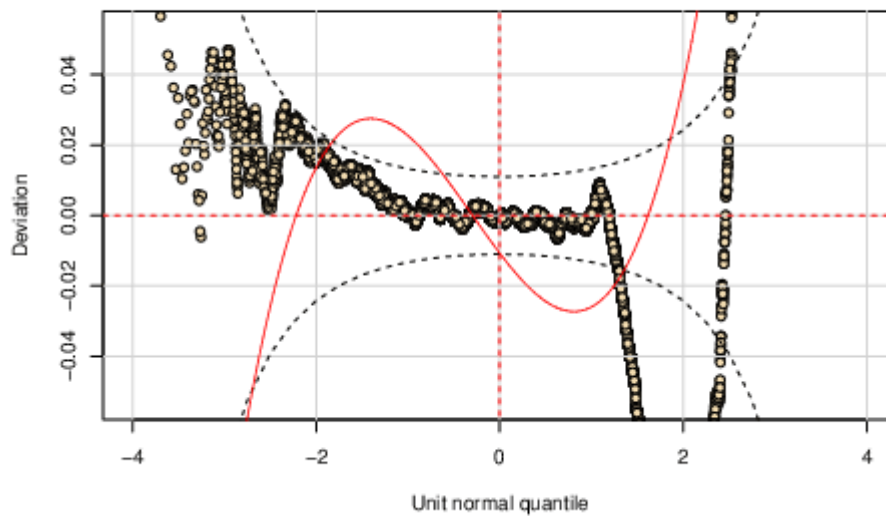


Figura 4.17: Worm plot del modelo de severidad distribución Gamma para el tercer año

4.4.5. Severidad bajo el modelo Weibull

En este modelo para μ se han considerado 8 variables a priori: edad asegurado, edad vehículo, uso, clase, grupo, región, suma asegurada, en el caso de σ tenemos 6 variables a priori: edad del asegurado, edad vehículo, uso, grupo, región y suma asegurada y para ξ_0 tenemos 4 variables a priori: edad del asegurado, edad del vehiculo, uso y suma asegurada. A continuación, se presentan los coeficientes

μ	Variable	Año 1		Año 3	
		Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	intercepto	7.8940	$< 2e - 16$ ***	7.7580	$< 2e - 16$ ***
	edad_a_año j	-0.0062	0.00000576 ***	-0.0045	2.71E-05 ***
	edad_v_año j	0.0018	7.29E-01	-0.0036	0.403053
	uso_p	-0.6802	4.01E-11 ***	-0.3557	0.000127 ***
	clase_c	0.2642	0.36693	0.8884	2.54E-03 **
	clase_o	-0.3926	0.04404 *	-0.1677	0.283778
	clase_p	0.4418	0.00885 **	0.2516	4.66E-02 *
	clase_r	0.0165	0.69034	-0.0466	0.133896
	grupo_p	0.1010	0.68193	-0.2048	0.435107
	region_l	-0.0477	5.41E-01	-0.2266	6.30E-05 ***
	suma_a	0.0000	4.05E-06 ***	0.0000	$< 2e - 16$ ***
σ	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	(Intercept)	-0.2407	0.00295 **	-0.1135	0.1088
	edad_a_año j	0.0019	0.04428 *	0.0018	0.0154 *
	edad_v_año j	0.0104	5.94E-03 **	0.0040	1.76E-01
	uso_p	0.0300	5.96E-01	-0.0598	2.60E-01
	grupo_p	0.2678	0.01424 *	-0.0336	0.7112
	region_l	0.0646	0.16204	0.0053	0.887
	suma_a	0.0000	6.21E-06 ***	0.0000	1.18E-02 *
ξ_0	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	(Intercept)	3.0740	$< 2e - 16$ ***	3.1850	$< 2e - 16$ ***
	edad_a_año j	0.0100	1.26E-14 ***	0.0035	0.000623 ***
	edad_v_año j	0.0454	$< 2e - 16$ ***	0.0442	$< 2e - 16$ ***
	uso_p	-1.2280	$< 2e - 16$ ***	-1.5890	$< 2e - 16$ ***
	suma_a	0.0000	0.0000191 ***	0.0000	$< 2e - 16$ ***

Cuadro 4.20: Coeficientes de regresión Modelo Weibull para los periodos $j = 1, 3$.

La estimación de los parámetros $\hat{\mu}$, $\hat{\sigma}$ y $\hat{\xi}_0$ se realiza de forma análoga a lo presentado en sección anterior.

En el modelo de Weibull el año no converge por lo que no se obtiene resultados para este periodo y no se presentan en el cuadro.

	Año 1	Año 3
media	-0.008926	-0.003603
varianza	1.003059	1.005914
coef. de asimetría	0.032520	0.046012
coef. of kurtosis	3.193850	3.274495

Cuadro 4.21: Resumen por año de los residuales modelo Weibull por periodo $j = 1, 3$.

Se puede observar que la estadística del coeficiente de curtosis muestra valores mayores que tres, lo que muestra que no se obtiene un buen ajuste de la distribución Weibull.

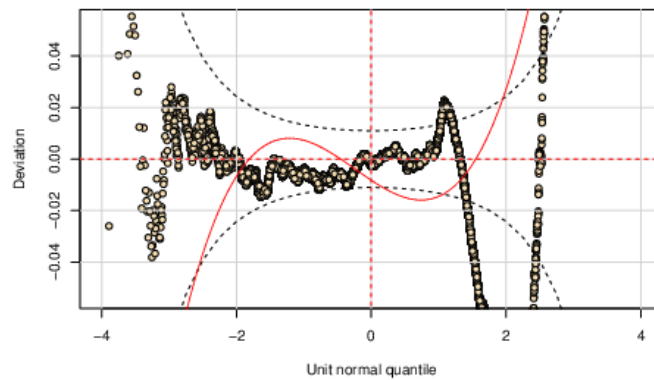


Figura 4.18: Worm plot del modelo de severidad distribución Weibull para el tercer año

4.4.6. Severidad bajo el modelo Gamma generalizada

En este modelo para μ se han considerado 8 variables a priori: edad asegurado, edad vehículo, uso, clase, grupo, región, suma asegurada, en el caso de σ tenemos 6 variables a priori: edad del asegurado, edad vehículo, uso, grupo, región y suma asegurada, para ν el intercepto y para ξ_0 tenemos 4 variables a priori: edad del asegurado, edad del vehículo, uso y suma asegurada. A continuación, se presentan los coeficientes

		Año 1		Año 2		Año 3	
μ	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	intercepto	7.6850	$< 2e - 16$ ***	7.8560	$< 2e - 16$ ***	7.4900	$< 2e - 16$ ***
	edad_a_año j	-0.0064	5.75E-06 ***	-0.0047	3.36E-06 ***	-0.0046	1.69E-05 ***
	edad_v_año j	-0.0008	8.74E-01	-0.0170	5.75E-05 ***	-0.0038	0.3996
	uso_p	-0.6659	2.68E-10 ***	-0.7034	$< 2e - 16$ ***	-0.3904	2.51E-05 ***
	clase_c	0.3026	0.30923	-0.1961	5.49E-01	0.7273	1.48E-02 *
	clase_o	-0.3990	0.0459 *	-0.2806	0.06755 .	-0.2061	0.1885
	clase_p	0.4505	0.00848 **	-0.0583	6.31E-01	0.1721	1.77E-01
	clase_r	0.0102	0.80891	-0.0084	0.77729	-0.0411	0.192
	grupo_p	0.0591	0.81388	0.4545	0.12467	-0.1070	0.6872
	region_l	-0.0675	4.05E-01	-0.1749	1.49E-03 **	-0.2294	3.83E-05 ***
	suma_a	0.0000	2.94E-06 ***	0.0000	3.43E-12 ***	0.0000	$< 2e - 16$ ***
σ	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	(Intercept)	0.23860	0.00103 **	0.09575	0.0914 .	0.03088	0.638
	edad_a_año j	-0.00022	0.79287	0.00021	0.7266	-0.00151	0.0245 *
	edad_v_año j	-0.00981	4.53E-03 **	-0.00283	2.65E-01	0.00179	5.18E-01
	uso_p	-0.05321	2.96E-01	0.03747	3.67E-01	0.05588	2.57E-01
	grupo_p	-0.28430	0.00425 **	0.41920	5.65E-08 ***	0.03177	0.7051
	region_l	-0.09795	0.0178 *	-0.03986	0.1989	0.02111	0.5381
	suma_a	0.00001	9.27E-07 ***	0.00000	5.41E-02 .	0.00000	1.31E-05 ***
ν	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	(Intercept)	0.50351	$> 2e-16$ ***	0.45984	$< 2e - 16$ ***	0.38630	$< 2e - 16$ ***
ξ_0	Variable	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$	Coefficiente	$Pr(> t)$
	(Intercept)	3.07400	$< 2e - 16$ ***	2.79700	$< 2e - 16$ ***	3.18500	$< 2e - 16$ ***
	edad_a_año j	0.01004	1.26E-14 ***	0.00429	9.70E-06 ***	0.00350	6.23E-04 ***
	edad_v_año j	0.04542	$< 2e - 16$ ***	0.06513	$< 2e - 16$ ***	0.04417	$< 2e - 16$ ***
	uso_p	-1.22800	$< 2e - 16$ ***	-1.52300	$< 2e - 16$ ***	-1.58900	$< 2e - 16$ ***
	suma_a	-0.00001	1.91E-05 ***	-0.00001	$< 2e - 16$ ***	-0.00001	$< 2e - 16$ ***

Cuadro 4.22: Coeficientes de regresión Modelo Gamma Generalizada para los periodos $j = 1, 2, 3$.

La estimación de los parámetros $\hat{\mu}$, $\hat{\sigma}$, $\hat{\nu}$ y $\hat{\xi}_0$ se realiza de forma análoga a lo presentado en sección anterior.

	Año 1	Año 2	Año 3
media	-0.001703	-0.003748	-0.000443
varianza	0.993243	1.000846	0.997213
coef. de asimetría	0.021135	0.024236	0.031646
coef. de curtosis	3.101991	3.124428	3.071438

Cuadro 4.23: Resumen por año de los residuales modelo Gamma Generalizada por periodo $j = 1, 2, 3$.

Aquí se puede observar que el coeficiente de curtosis se mejora considerablemente respecto a los dos modelos anteriores.

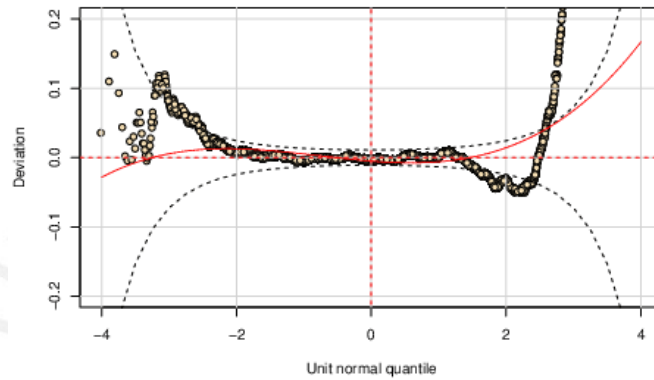


Figura 4.19: Worm plot del modelo de severidad distribución Gamma Generalizada para el tercer año

GA	Año 1	Año 2	Año 3
df			
Devianza Global	88,450	157,117	137,624
AIC	88,496	157,163	137,670
SBC	88,699	157,366	137,872
WEI	Año 1	Año 2	Año 3
df	23		23
Devianza Global	88,335		137,490
AIC	88,381		137,536
SBC	88,584		137,739
GG	Año 1	Año 2	Año 3
df			
Devianza Global	88,126	156,480	136,918
AIC	88,174	156,528	136,966
SBC	88,386	156,739	137,177

Cuadro 4.24: Comparación de los modelos de severidad por año

Dado los criterios de diagnostico, la distribución que más se ajusta a los datos es la Gamma Generalizada.

4.5. Resultados de los modelos BMSG GAMMLS

A continuación se presentan los resultados de las frecuencias, usando los modelos de BMSG GAMMLS y aplicando el principio de la desviación estándar. Se ha aplicado un recargo de $\rho = 0.1$ como factor de seguridad en la frecuencia y los resultados se presentan bajo la segmentaciones definidas en la tabla 4.9.

Se observa que para el segundo y tercer año se han obtenido resultados similares que los modelos de regresión donde se estima la media como función de las variables explicativas.

El año 1 presenta una frecuencia más baja que los años dos y tres en todos los segmentos.

Segmento	Frecuencia								
	NBII Año 1	DEL Año 1	SICH Año 1	NBII Año 2	DEL Año 2	SICH Año 2	NBII Año 3	DEL Año 3	SICH Año 3
1	0.15	0.15	0.15	0.27	0.27	0.27	0.22	0.22	0.22
2	0.16	0.17	0.17	0.29	0.29	0.29	0.23	0.23	0.23
3	0.13	0.13	0.13	0.25	0.25	0.25	0.21	0.21	0.21
4	0.11	0.11	0.11	0.23	0.23	0.23	0.20	0.20	0.19
5	0.12	0.12	0.12	0.21	0.21	0.21	0.19	0.19	0.19
6	0.10	0.10	0.10	0.15	0.15	0.15	0.15	0.15	0.15
7	0.07	0.07	0.07	0.10	0.10	0.10	0.07	0.07	0.07
8	0.11	0.11	0.11	0.20	0.20	0.20	0.17	0.17	0.17
9	0.13	0.13	0.13	0.23	0.23	0.23	0.19	0.19	0.19
10	0.17	0.17	0.17	0.34	0.34	0.34	0.29	0.29	0.30
11	0.19	0.19	0.19	0.35	0.35	0.36	0.30	0.30	0.31

Cuadro 4.25: Comparación de los modelos de frecuencia por año

Por otro lado, hemos obtenido las primas considerando los mismos segmentos y se observan diferencias importantes en los valores obtenidos.

Estas primas son más bajas que las presentadas en el cuadro (4.2) que no consideraba información a priori en aproximadamente un 50 %, dado que incorporan variables a priori y a posteriori del asegurado, lo que permite hacer una mejor segmentación del riesgo.

Segmentos	NBII - GA	NBII - WEI	NBII - GG
1	126.56	121.46	100.28
2	148.35	140.88	114.94
3	109.46	106.05	88.58
4	91.63	89.87	76.36
5	85.13	82.97	69.27
6	52.11	51.51	43.59
7	18.59	18.39	17.59
8	124.30	118.26	98.56
9	89.80	87.00	73.00
10	258.03	241.73	195.40
11	314.75	289.83	234.20

Cuadro 4.26: Primas para el tercer año con el modelo de frecuencia *NBII* y los diferentes modelos de severidad en unidades de 100 dólares

Segmentos	DEL - GA	DEL - WEI	DEL - GG
1	126.55	121.44	100.27
2	148.68	141.19	115.20
3	109.16	105.76	88.34
4	89.31	75.89	75.89
5	85.56	83.39	69.62
6	52.75	52.14	44.12
7	17.94	17.76	16.98
8	123.21	117.23	97.69
9	89.50	86.70	72.75
10	262.43	245.69	198.61
11	325.10	298.77	241.73

Cuadro 4.27: Primas para el tercer año con el modelo de frecuencia *DEL* y los diferentes modelos de severidad en unidades de 100 dólares

Segmentos	SICH - GA	SICH - WEI	SICH - GG
1	126.21	121.12	100.00
2	148.14	140.68	114.78
3	108.93	105.54	88.15
4	90.89	89.14	75.74
5	86.21	84.03	70.15
6	53.85	53.23	45.05
7	17.87	17.68	16.91
8	122.73	116.78	97.32
9	88.37	85.61	71.83
10	279.30	260.86	210.94
11	364.97	332.79	270.72

Cuadro 4.28: Primas para el tercer año con el modelo de frecuencia *SICH* y los diferentes modelos de severidad en unidades de 100 dólares

Capítulo 5

Conclusiones

5.1. Conclusiones

Se ha observado que los modelos analizados son aplicables al conjunto de datos indicado, obteniéndose los distintos estimadores de la frecuencia y severidad y la prima neta.

Las estimaciones de la frecuencia y la severidad obtenidas con la información a posteriori únicamente consideran la información del número de siniestros que ha presentado el asegurado, sin tener en consideración otras variables que pueden agravar el riesgo.

Por otro lado, las estimaciones de la frecuencia y severidad aplicando modelos con información a priori y posteriori, si contemplan incluir más información disponible, lo cual permite encontrar una tarifa que puede mejor adecuarse al riesgo particular que presenta cada asegurado.

De la aplicación de los distintos modelos se observa que las primas se reducen de forma considerable cuando se tienen en cuenta las variables a priori del asegurado y del vehículo, dado que no solamente estamos considerando medir el riesgo desde el punto de vista del siniestro y la experiencia en el seguro sino que incorporamos covariables que nos permiten mejor modelar el número esperado de siniestros y el monto esperado de estos mediante modelos de regresión sobre su media, varianza y sesgo.

Los modelos que tuvieron mejor ajuste fueron los de frecuencia, como se evidencia a través del diagnóstico presentado. En el caso de los modelos de severidad los ajustes son menores, incluso en algunos casos no se llegó a obtener resultado del modelo.

Ha sido de utilidad estudiar los modelos de regresión bajo el enfoque GAMLSS con un conjunto de datos reales, dado que se puede corroborar que estos modelos se ajustan en manera más flexible a los datos y nos brindan nuevas oportunidades de poder generar primas más bajas que pueden ser más competitivas en el mercado asegurador.

5.2. Sugerencia para investigaciones futuras

Es importante mencionar que en la aplicación de los modelos GAMLSS, solo se obtiene una prima a priori y queda abierto como posible desarrollo futuro la estimación de frecuencias y severidades a posteriori para las distintas distribuciones tratadas, mediante la aplicación de métodos no analíticos. Por ejemplo, la simulación de cadenas de Markov de Montecarlo.

Apéndice A

Código R

ESTIMACIÓN DE LOS PARÁMETROS DE MÁXIMA VEROSIMILITUD DE LA BINOMIAL NEGATIVA

```
setwd("C:/DianaH/M15. Tesis II/BASES R")
Siniestros<-read.csv("SINIESTROS.csv")
summary(Siniestros)
FSIN<-Siniestros$Frecuencia
y<-Siniestros$K

library(fitdistrplus)

plotdist(y,histo=TRUE,demp = TRUE, ylim=c(0,4))
descdist(y,boot=1000)

ajuste1.lme<-fitdist(y,"nbinom")
summary(ajuste1.lme)

hist(y,prob=TRUE,
      xlab="Casos",
      ylab="Densidad",
      main="Número de siniestros",
      col = 'grey90', border="grey", lty=1,breaks = 30,
      ylim=c(0,2))
curve(dnbinom(x,a3,p3,log=FALSE),0,12,n=100,col=2,lty=3, ylim=c(0,2),
      add=TRUE)
legend("topright",c("Ajuste Binomial Negativa"),col=2,lty=2,lwd=1,box
      .lty=0,box.lwd = 0,cex=0.8)

gofstat(ajuste1.lme)
denscomp(ajuste1.lme, fitlty = 1)
ajuste1.mme<-fitdist(y,"nbinom",method = "mme")
summary(ajuste1.mme)
gofstat(ajuste1.lme)
```

CÁLCULO DE LA FRECUENCIA BAJO EL MODELO BINOMIAL NEGATIVO

```

SiniestrosK<-read.csv("SiniestrosK.csv")
summary(SiniestrosK)

y<-SiniestrosK$Kprim

a<-0.82744
p<-0.6723492
tao<-2.05203
lam1<-a/tao
t<-3

sigma.e<-a
mu.e<-a/tao
library(nlirms)
lam.s0<-enc.PGA(time=3,claim=y,mu=mu.e,sigma = sigma.e)
SiniestrosK$lam.s<-lam.s0
write.csv(SiniestrosK, "SiniprimaBN2.csv")

```

ESTIMACIÓN DE LOS PARÁMETROS DE LA DISTRIBUCIÓN PIG

```

setwd("C:/DianaH/M15. Tesis II/BASES R")
Siniestros<-read.csv("SINIESTROS.csv")
summary(Siniestros)
FSIN<-Siniestros$Frecuencia

library(HyperbolicDist)

y<-Siniestros$K
e1<-mean(y)
nu<- -0.5
n<-length(y)
xi<-e1

floglik<-function(alpha){

  n<-length(y)
  xi<-e1
  eta<-(((xi^2)+(alpha^2))^0.5)
  w<-eta-xi
  RBay<-besselRatio(alpha,y+nu, 1, useExpScaled = 700)

  fl1<-(alpha*((xi^2+alpha^2)^(-0.5)))*(1+(e1/w))-((1/n)*(sum(RBay)))
  fl1
}

optim(5,floglik,lower = 0.001, upper=100,method = "L-BFGS-B")
nlminb(5,floglik,lower=0.001, upper=100)

```


CÁLCULO DE LA FRECUENCIA BAJO EL MODELO PIG

```

setwd("C:/DianaH/M15. Tesis II/BASES R oct21")

SiniestrosK<-read.csv("SiniestrosK.csv")
summary(SiniestrosK)

y<-SiniestrosK$Kprim

mu.e<-0.4034575
sigma.e<-0.6725786
alpha.e<-1.846714
t<-3

library(nlirms)
lam.01<-enc.PGIG(time=3,claim = y, mu=mu.e, sigma = sigma.e, nu=0.5)
SiniestrosK$lam.s<-lam.01

write.csv(SiniestrosK, "SiniPrimaPIG2.csv")

```

ESTIMACIÓN DE MÁXIMA VEROSIMILITUD DE LOS PARÁMETROS DE LA SICHEL

```

setwd("D:/Maestria/M15. Tesis II/BASES R")
Siniestros<-read.csv("SINIESTROS.csv")
summary(Siniestros)
FSIN<-Siniestros$Frecuencia

##Ratio Bessel###
library(HyperbolicDist)
alpha=0.911
besselRatio(alpha,1-0.5, 1, useExpScaled = 700)
#####

y<-Siniestros$K
e1<-mean(y)
nu<- -0.5 #####AQUI CAMBIO POR NU=-0.5###
n<-length(y)
alpha<-1
xsi<-0.4
nu

#Función Score
g<-function(theta,y){
  alpha<-theta[1]
  xsi<-theta[2]
  n<-length(y)
  w<-(((xsi^2+alpha^2)^0.5)-xsi)
  x<-besselRatio(alpha,y+nu, 1, useExpScaled = 700)
}

```

```

x1<-besselRatio(w,nu,1,useExpScaled = 700)
a<-alpha*((xsi^2+alpha^2)^(-0.5))*(x1+(e1/w))-(1/n*sum(x))
b<-xsi*x1-e1
c(a,b)
}

#Hessiana
H<-function(theta,y){
  alpha<-theta[1]
  xsi<-theta[2]
  n<-length(y)
  eta<-((xsi^2+alpha^2)^0.5)#OK pag 943
  w<-eta-xsi
  x<-besselRatio(alpha,y+nu, 1, useExpScaled = 700)
  x1<-besselRatio(w,nu,1,useExpScaled = 700)
  a2<-sum(y)*((xsi^2*w)-(alpha^2*eta)/(w^2*eta^3))
  +n*x1*((xsi^2*w)-((alpha^2*eta*(2*nu+1))/(w*eta^3)))
  +(n*alpha^2*(x1^2-1)/eta^2)+((2/alpha)*(sum(y*x)))
  +((2*nu+1)/alpha)*sum(x)-(sum(x^2))+n
  b2<-sum(y)*(xsi^-2-(xsi/eta^3))
  +n*((w^2*eta*(x1^2)-1)-x1*((2*nu+1)*w*eta-alpha^2))/eta^3
  ab<-alpha*sum(y)/eta^3
  +n*alpha*x1*((2*nu+1)*eta-alpha*xsi)/nu^3
  -n*alpha*w*(x1^2-1)/eta^2
  matrix(c(a2,ab,ab,b2),2,2)
}

#Algoritmo de NR

theta0<-c(0.9,0.5)
res<-matrix(theta0,1,2)
res

cont<-0
h<-0
while(cont==0){
  h<-h+1
  theta.old<-res[h,]
  theta.new<-c(theta.old - solve(H(theta.old,y)) %*% g(theta.old,y))
  res<-rbind(res,theta.new)
  if(sum((theta.new-theta.old)^2)<1e-20){cont<-1}
}

res

```

MODELO DE FRECUENCIA VARIABLES A PRIORI Y POSTERIORI, REGRESIÓN BN

```
setwd("C:/DianaH/M15. Tesis II/Bases R nov21")
```

```

SinVAP3<-read.csv("BASEVAR.csv")
attach(SinVAP3)
summary(SinVAP3)

clase<-factor(SinVAP3$CLASE)
uso<-factor(SinVAP3$USO)
marca<-factor(SinVAP3$MARCA2)
grupo<-factor(SinVAP3$GRUPO)
genero<-factor(SinVAP3$GENERO)
region<-factor(SinVAP3$REGION)
edadA1<-SinVAP3$EDAD_ASEG_1
edadA2<-SinVAP3$EDAD_ASEG_2
edadA3<-SinVAP3$EDAD_ASEG_3
edadV1<-SinVAP3$EDAD_VEH_1
edadV2<-SinVAP3$EDAD_VEH_2
edadV3<-SinVAP3$EDAD_VEH_3
sumaA<-SinVAP3$SUMA.ASEGURADA
ContA<-SinVAP3$ContAse

library(gamlss)

#####
#   SINIESTTROS AÑO 1

modelo61 <- gamlss(Num_Siniestros_año1~edadA1+edadV1+uso+grupo+region
  +sumaA+marca,
                family=NBI, data=SinVAP3, trace=FALSE)
summary(modelo61)
drop1(modelo61)
Rsq(modelo61)

newpar61<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(modelo61, par=newpar61)

wp(modelo61, ylim.all = 0.2)
wp(modelo61,xvar = SinVAP3$EDAD_ASEG_1,n.inter = 9)
wp(modelo61,xvar = SinVAP3$EDAD_VEH_1,n.inter = 9)
wp(modelo61,xvar =~ SinVAP3$MARCA2,n.inter = 9)

dtop(modelo61,xvar = SinVAP3$EDAD_ASEG_1,n.inter = 9)

#####
#   SINIESTTROS AÑO 2

modelo62 <- gamlss(Num_Siniestros_año2~edadA2+edadV2+uso+grupo+region

```

```

+sumaA+marca ,
                                family=NBI , data=SinVAP3 , trace=FALSE)
summary(modelo62)
drop1(modelo62)
Rsq(modelo62)

newpar62<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(modelo62, par=newpar61)

wp(modelo62, ylim.all = 0.2)

#####
#   SINIESTTROS AÑO 3
#####

modelo63 <- gamlss(Num_Siniestros_año3~edadA3+edadV3+uso+grupo+region
+sumaA+marca ,
                  family=NBI , data=SinVAP3 , trace=FALSE)
summary(modelo63)
drop(modelo63)
Rsq(modelo63)

newpar63<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(modelo63, par=newpar61)

wp(modelo63, ylim.all = 0.3)

```

MODELO DE FRECUENCIA VARIABLES A PRIORI Y POSTERIORI, REGRESIÓN SICHEL

```

setwd("C:/DianaH/M15. Tesis II/Bases R nov21")
SinVAP3<-read.csv("BASEVAR.csv")
attach(SinVAP3)
summary(SinVAP3)

clase<-factor(SinVAP3$CLASE)
uso<-factor(SinVAP3$USO)
marca<-factor(SinVAP3$MARCA2)
grupo<-factor(SinVAP3$GRUPO)

```

```

genero<-factor(SinVAP3$GENERO)
region<-factor(SinVAP3$REGION)
edadA1<-SinVAP3$EDAD_ASEG_1
edadA2<-SinVAP3$EDAD_ASEG_2
edadA3<-SinVAP3$EDAD_ASEG_3
edadV1<-SinVAP3$EDAD_VEH_1
edadV2<-SinVAP3$EDAD_VEH_2
edadV3<-SinVAP3$EDAD_VEH_3
sumaA<-SinVAP3$SUMA.ASEGURADA
ContA<-SinVAP3$ContAse

library(gamlss)

#####
#   SINIESTTROS AÑO 1

modelo31 <- gamlss(Num_Siniestros_año1~edadA1+edadV1+uso+clase+grupo+
  region+sumaA+marca,
                family=SICHEL, data=SinVAP3, trace=FALSE)
summary(modelo31)

drop1(modelo31)
Rsq(modelo31)

newpar31<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(modelo31, par=newpar31)

wp(modelo31, ylim.all = 0.3)

#####
#   SINIESTTROS AÑO 2

modelo32 <- gamlss(Num_Siniestros_año2~edadA2+edadV2+uso+clase+grupo+
  region+sumaA+marca,
                family=SICHEL, data=SinVAP3, trace=FALSE)
summary(modelo32)

drop1(modelo32)
Rsq(modelo32)

newpar32<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(modelo32, par=newpar32)

```

```

wp(modelo32, ylim.all = 0.3)

#####
#   SINIESTTROS AÑO 3

modelo33 <- gamlss(Num_Siniestros_año3~edadA3+edadV3+uso+clase+grupo+
  region+sumaA+marca,
                family=SICHEL, data=SinVAP3, trace=FALSE)
summary(modelo33)

drop1(modelo33)
Rsq(modelo33)

newpar33<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(modelo33, par=newpar33)

wp(modelo33, ylim.all = 0.3)

```

MODELO DE SEVERIDAD VARIABLES A PRIORI Y POSTERIORI - REGRESIÓN PARETO

```

setwd("C:/DianaH/M15. Tesis II/Bases R nov21")
SinVAP3<-read.csv("BASEVAR_V2.csv")
attach(SinVAP3)
summary(SinVAP3)

clase<-factor(SinVAP3$CLASE)
uso<-factor(SinVAP3$USO)
marca<-factor(SinVAP3$MARCA2)
grupo<-factor(SinVAP3$GRUPO)
genero<-factor(SinVAP3$GENERO)
region<-factor(SinVAP3$REGION)
edadA1<-SinVAP3$EDAD_ASEG_1
edadA2<-SinVAP3$EDAD_ASEG_2
edadA3<-SinVAP3$EDAD_ASEG_3
edadV1<-SinVAP3$EDAD_VEH_1
edadV2<-SinVAP3$EDAD_VEH_2
edadV3<-SinVAP3$EDAD_VEH_3
sumaA<-SinVAP3$SUMA.ASEGURADA
ContA<-SinVAP3$ContAse

#### MODELOS ZERO INFLATED SON LOS QUE DEBO USAR

```

```

### COMO NO SE PUEDE GENERAR LA DISTRIBUCIÓN DE PARETO, SE GENERO LA
    GAMMA INVERSA CON LOS PARAMETROS REQUERIDOS

library(gamlss)
library(gamlss.inf)

#####
# MONTO PROMEDIO SINIESTROS AÑO 1, AÑO 2 Y AÑO 3

modelo.T11<-gamlssZadj(y1,
                      mu.formula =~edadA1+edadV1+uso+clase+grupo+
                        region+sumaA+marca,
                      sigma.formula =~edadA1+edadV1+sumaA,
                      #nu.formula =~1,
                      xi0.formula =~edadA1+edadV1+sumaA,
                      family=IGAMMA(),data=SinVAP3,trace=TRUE)

dropterm(modelo.T11,test="Chisq")

summary(modelo.T11) # no corre IG, corre GG / Revisar

wp(modelo.T11)
plot(modelo.T11)

modelo.T12<-gamlssZadj(y2,
                      mu.formula =~edadA2+edadV2+uso+clase+grupo+
                        region+sumaA+marca,
                      sigma.formula =~edadA2+edadV2+sumaA,
                      #nu.formula =~1,
                      xi0.formula =~edadA2+edadV2+sumaA,
                      family=IGAMMA(),data=SinVAP3,trace=TRUE)

summary(modelo.T12)

wp(modelo.T12)
plot(modelo.T12)

modelo.T13<-gamlssZadj(y3,
                      mu.formula = ~clase+uso+grupo+region+edadA3+
                        edadV3+sumaA,
                      sigma.formula = ~1,
                      xi0.formula =~1,
                      family=GG(),data=SinVAP3,trace=TRUE)

summary(modelo.T13)

wp(modelo.T13)
plot(modelo.T13)

```

```
#####
#   MONTOS PROMEDIO TOTAL CARTERA

modelo.T2<-gamlssZadj(monto_total,
                    mu.formula = ~clase+uso+grupo+region+edadA1+
                        edadV1+sumaA,
                    sigma.formula = ~edadA1+edadV1+sumaA,
                    xi0.formula = ~1,
                    family=PARETO2(),data=SinVAP3,trace=TRUE)

summary(modelo.T2)

wp(modelo.T2,ylim.all = 0.2) # wp no sale muy bien, explicar

# term.plot(modelo.T2)

# cambia los colores del grafico

newpar.T2<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4",
              ",
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)

plot(modelo.T2, par=newpar.T2)

### Otros graficos ###

Q.stats(modelo.T2)
centiles.Zadj(modelo.T2,xvar = prommont)

#####
#   MONTO SINIESTTROS AÑO 1, AÑO 2 Y AÑO 3           clase+uso+grupo+
            region+edadA1+edadV1+sumaA

modelo.T21<-gamlssZadj(monto_año1,mu.formula = ~clase+uso+grupo+
                    region+edadA1+edadV1+sumaA,
                    sigma.formula = ~1,
                    xi0.formula = ~1,
                    family=PARETO2(),data=SinVAP3,trace=TRUE)

summary(modelo.T21)

modelo.T22<-gamlssZadj(monto_año2,mu.formula = ~clase+uso+grupo+
                    region+edadA2+edadV2+sumaA,
```



```

        sigma.formula = ~1,
        xi0.formula = ~1,
        family=PARETO2(), data=SinVAP3, trace=TRUE)
summary(modelo.T22)

modelo.T23<-gamlssZadj(monto_año3, mu.formula = ~clase+uso+grupo+
  region+edadA3+edadV3+sumaA,
        sigma.formula = ~1,
        xi0.formula = ~1,
        family=PARETO2(), data=SinVAP3, trace=TRUE)
summary(modelo.T23)

```

MODELO DE FRECUENCIA VARIABLES A PRIORI Y POSTERIORI - GAMLSS

```

setwd("C:/DianaH/M15. Tesis II/Bases R nov21")
SinVAP3<-read.csv("BASEVAR_V2.csv")
attach(SinVAP3)
summary(SinVAP3)

##VARIABLES
clase<-factor(SinVAP3$CLASE)
uso<-factor(SinVAP3$USO)
marca<-factor(SinVAP3$MARCA2)
grupo<-factor(SinVAP3$GRUPO)
genero<-factor(SinVAP3$GENERO)
region<-factor(SinVAP3$REGION)
edadA1<-SinVAP3$EDAD_ASEG_1
edadA2<-SinVAP3$EDAD_ASEG_2
edadA3<-SinVAP3$EDAD_ASEG_3
edadV1<-SinVAP3$EDAD_VEH_1
edadV2<-SinVAP3$EDAD_VEH_2
edadV3<-SinVAP3$EDAD_VEH_3
sumaA<-SinVAP3$SUMA.ASEGURADA
ContA<-SinVAP3$ContAse

#deptaseg<-factor(SinVAP3$DEPTO_ASE)

#####
# REGRESIÓN GAMLSS FRECUENCIA #
#####

##### todas las variables:      edadA+genero+region+sumaA+edadV+
      clase+grupo+uso+marca
##### variables significativas:  edadA+region+sumaA+edadV+grupo+
      uso+marca

library(gamlss)

```

```
#####
# MODELO I: con 3 parametros: Sichel año 1

ModeloS1 <- gamlss(Num_Siniestros_año1~edadA1+edadV1+uso+grupo+region
  +sumaA+marca,
  sigma.fo=~edadV1+grupo+sumaA,
  nu.fo=~1,family=SICHEL,
  data=SinVAP3, trace=FALSE)
summary(ModeloS1)

drop1(ModeloS1)

newparS1<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
  col="blue4", col.main="blue4",col.lab="blue4",pch="+",
  cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(ModeloS1, par=newparS1) ##graficar una variable

wp(ModeloS1, ylim.all = 0.1)

##worm plot, chapman Pag. 430
coef12<-wp(ModeloS1,xvar=SinVAP3$edadV,n.inter=9)
coef12<-wp(ModeloS1,xvar=~grupo,n.inter=9) #funciona bien en
  categoricas

##dtop
coef13<-dtop(ModeloS1,xvar=SinVAP3$edadV,n.inter=9)

##estadisticas Z y Q; Pag 436, no grafica los intervalos de edad...
  REVISAR, al parecer
##grafica los montos pero segun el ejemplo no es correcto (NO)
Qestad1<-Q.stats(ModeloS1,xvar=edadV1,n.inter=9)
print(Qestad1, digits=3)

#rqres
rqres.plot(ModeloS1)
#####

#####
# MODELO I: con 3 parametros: Sichel año 2

ModeloS2 <- gamlss(Num_Siniestros_año2~edadA2+edadV2+uso+grupo+region
  +sumaA+marca,
  sigma.fo=~edadV2+grupo+sumaA,
```

```

        nu.fo=~1,family=SICHEL,
        data=SinVAP3, trace=FALSE)
summary(ModeloS2)

drop1(ModeloS2)

newparS2<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
            col="blue4", col.main="blue4",col.lab="blue4",pch="+",
            cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(ModeloS2, par=newparS2)

wp(ModeloS2, ylim.all = 0.1)

#####
# MODELO I: con 3 parametros: Sichel año 3

ModeloS3 <- gamlss(Num_Siniestros_año3~edadA3+edadV3+uso+grupo+region
+sumaA+marca,
                sigma.fo=~edadV3+grupo+sumaA,
                nu.fo=~1,family=SICHEL,
                data=SinVAP3, trace=FALSE)
summary(ModeloS3)

drop1(ModeloS3)

newparS3<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
            col="blue4", col.main="blue4",col.lab="blue4",pch="+",
            cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(ModeloS3, par=newparS3)

wp(ModeloS3, ylim.all = 0.1)

#####
# MODELOS II: con 2 parametros: NBII, ZIP año 1

ModeloNB1 <- gamlss(Num_Siniestros_año1~edadA1+edadV1+uso+grupo+
region+sumaA+marca,
                sigma.fo=~edadV1+grupo+sumaA,
                family=NBII,
                data=SinVAP3, trace=FALSE)
summary(ModeloNB1)

drop1(ModeloNB1)

```

```

newparNB1<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4
",
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(ModeloNB1, par=newparNB1) ##graficar una variable

wp(ModeloNB1,ylim.all = 0.1)

#####
# MODELOS II: con 2 parametros: NBII, ZIP año 2

ModeloNB2 <- gamlss(Num_Siniestros_año2~edadA2+edadV2+uso+grupo+
  region+sumaA+marca,
                sigma.fo=~edadV2+grupo+sumaA,
                family=NBII,
                data=SinVAP3, trace=FALSE)
summary(ModeloNB2)

drop1(ModeloNB2)

newparNB2<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4
",
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(ModeloNB2, par=newparNB2)

wp(ModeloNB2,ylim.all = 0.1)

#####
# MODELOS II: con 2 parametros: NBII, ZIP año 3

ModeloNB3 <- gamlss(Num_Siniestros_año3~edadA3+edadV3+uso+grupo+
  region+sumaA+marca,
                sigma.fo=~edadV3+grupo+sumaA,
                family=NBII,
                data=SinVAP3, trace=FALSE)
summary(ModeloNB3)

drop1(ModeloNB3)

newparNB3<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4
",
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(ModeloNB3, par=newparNB3)

```

```

wp(ModeloNB3,ylim.all = 0.1)

#####
# MODELOS III: con 3 parametros: DELAPORTE año 1

ModeloD1 <- gamlss(Num_Siniestros_año1~edadA1+edadV1+uso+grupo+region
+sumaA+marca,
                 sigma.fo=~edadV1+grupo+sumaA,
                 nu.fo=~1,family=DEL,
                 data=SinVAP3, trace=FALSE)
summary(ModeloD1)

drop1(ModeloD1)

newparD1<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
             col="blue4", col.main="blue4",col.lab="blue4",pch="+",
             cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(ModeloD1, par=newparD1)

wp(ModeloD1,ylim.all = 0.1)

#####
# MODELOS III: con 3 parametros: DELAPORTE año 2

ModeloD2 <- gamlss(Num_Siniestros_año2~edadA2+edadV2+uso+grupo+region
+sumaA+marca,
                 sigma.fo=~edadV2+grupo+sumaA,
                 nu.fo=~1,family=DEL,
                 data=SinVAP3, trace=FALSE)
summary(ModeloD2)

drop1(ModeloD2)

newparD2<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
             col="blue4", col.main="blue4",col.lab="blue4",pch="+",
             cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(ModeloD2, par=newparD2)

wp(ModeloD2,ylim.all = 0.1)

#####
# MODELOS III: con 3 parametros: DELAPORTE año 3

ModeloD3 <- gamlss(Num_Siniestros_año3~edadA3+edadV3+uso+grupo+region

```

```

+sumaA+marca ,
                                sigma.fo=~edadV3+grupo+sumaA ,
                                nu.fo=~1,family=DEL ,
                                data=SinVAP3 , trace=FALSE)
summary(ModeloD3)

drop1(ModeloD3)

newparD3<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)
plot(ModeloD3, par=newparD3)

wp(ModeloD3,ylim.all = 0.1)

```

MODELO DE SEVERIDAD VARIABLES A PRIORI Y POSTERIORI - GAMLSS

```

setwd("C:/DianaH/M15. Tesis II/Bases R nov21")
MontoS<-read.csv("BASEVAR_V2.csv")
attach(MontoS)
summary(MontoS)

##VARIABLES
clase<-factor(MontoS$CLASE)
uso<-factor(MontoS$USO)
marca<-factor(MontoS$MARCA2)
grupo<-factor(MontoS$GRUPO)
genero<-factor(MontoS$GENERO)
region<-factor(MontoS$REGION)
edadA1<-MontoS$EDAD_ASEG_1
edadA2<-MontoS$EDAD_ASEG_2
edadA3<-MontoS$EDAD_ASEG_3
edadV1<-MontoS$EDAD_VEH_1
edadV2<-MontoS$EDAD_VEH_2
edadV3<-MontoS$EDAD_VEH_3
sumaA<-MontoS$SUMA.ASEGURADA
ContA<-MontoS$ContAse

#####
# REGRESIÓN GAMLSS SEVERIDAD      #
#####

##### todas las variables:      edadA+genero+region+sumaA+edadV+
      clase+grupo+uso+marca
##### variables significativas:  edadA+region+sumaA+edadV+clase+
      grupo+uso

```

```

library(gamlss)
library(gamlss.inf)

#####
# MODELOS POR AÑO #
#####

# MODELO I: GAMMA AÑO 1

ModeloG1<-gamlssZadj(monto_año1,
                    mu.formula =~edadA1+edadV1+uso+clase+grupo++
                      region+sumaA,
                    sigma.formula =~edadA1+edadV1+uso+grupo++region+
                      sumaA,
                    xi0.formula =~edadA1+edadV1+uso+sumaA,
                    family=GA(),data=MontoS,trace=TRUE)

summary(ModeloG1)

# MODELO I: GAMMA AÑO 2

ModeloG2<-gamlssZadj(monto_año2,
                    mu.formula =~edadA2+edadV2+uso+clase+grupo++
                      region+sumaA,
                    sigma.formula =~edadA2+edadV2+uso+grupo++region+
                      sumaA,
                    xi0.formula =~edadA2+edadV2+uso+sumaA,
                    family=GA(),data=MontoS,trace=TRUE)

summary(ModeloG2)

# MODELO I: GAMMA AÑO 3

ModeloG3<-gamlssZadj(monto_año3,
                    mu.formula =~edadA3+edadV3+uso+clase+grupo++
                      region+sumaA,
                    sigma.formula =~edadA3+edadV3+uso+grupo++region+
                      sumaA,
                    xi0.formula =~edadA3+edadV3+uso+sumaA,
                    family=GA(),data=MontoS,trace=TRUE)

summary(ModeloG3)

# GRAFICOS DE LOS 3 MODELOS GAMMA

newparG<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4",

```

```

col="blue4", col.main="blue4",col.lab="blue4",pch="+",
cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)

plot(ModeloG3, xvar = edadA3, par=newparG) ##CAMBIAR G1, G2, G3

wp(ModeloG3) ##revusar limites grafico
wp(ModeloG3, xvar = ~MontoS$MARCA)
Q.stats(ModeloG3)

#####
# MODELO 2: WEIBULL AÑO 1 #
#####

ModeloW1<-gamlssZadj(monto_año1,
mu.formula =~edadA1+edadV1+uso+clase+grupo++
region+sumaA,
sigma.formula =~edadA1+edadV1+uso+grupo++region+
sumaA,
xi0.formula =~edadA1+edadV1+uso+sumaA,
family=WEI(),data=MontoS,trace=TRUE)

summary(ModeloW1)

#####
# MODELO 2: WEIBULL AÑO 2 #
#####

ModeloW2<-gamlssZadj(monto_año2,
mu.formula =~edadA2+edadV2+uso+clase+grupo++
region+sumaA,
sigma.formula =~edadA2+edadV2+uso+grupo++region+
sumaA,
xi0.formula =~edadA2+edadV2+uso+sumaA,
family=WEI(),data=MontoS,trace=TRUE)

summary(ModeloW2)

#####
# MODELO 2: WEIBULL AÑO 3 #
#####

ModeloW3<-gamlssZadj(monto_año3,
mu.formula =~edadA3+edadV3+uso+clase+grupo++
region+sumaA,
sigma.formula =~edadA3+edadV3+uso+grupo++region+
sumaA,
xi0.formula =~edadA3+edadV3+uso+sumaA,
family=WEI(),data=MontoS,trace=TRUE)

```



```

summary(ModeloW3)

#####
#   GRAFICOS DE LOS 3 MODELOS WEIBULL#
#####

newparW<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4",
             col="blue4", col.main="blue4",col.lab="blue4",pch="+",
             cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)

plot(ModeloW3, xvar = edadA3, par=newparW) ##cambiar W1, W2 O W3

wp(ModeloW3) ##revusar limites grafico

wp(ModeloG3, xvar = ~MontoS$MARCA)
Q.stats(ModeloG3)

#####
# MODELO 3: GG AÑO 1           #
#####

ModeloGG1<-gamlssZadj(monto_año1,
                      mu.formula =~edadA1+edadV1+uso+clase+grupo+
                        region+sumaA,
                      sigma.formula =~edadA1+edadV1+uso+grupo+region+
                        sumaA,
                      xi0.formula =~edadA1+edadV1+uso+sumaA,
                      family=GG(),data=MontoS,trace=TRUE)

summary(ModeloGG1)

#####
# MODELO 3: GG AÑO 2           #
#####

ModeloGG2<-gamlssZadj(monto_año2,
                      mu.formula =~edadA2+edadV2+uso+clase+grupo+
                        region+sumaA,
                      sigma.formula =~edadA2+edadV2+uso+grupo+region+
                        sumaA,
                      xi0.formula =~edadA2+edadV2+uso+sumaA,
                      family=GG(),data=MontoS,trace=TRUE)

summary(ModeloGG2)

#####

```

```

# MODELO 3: GG AÑO 3                                     #
#####

ModeloGG3<-gamlssZadj(monto_año3,
                      mu.formula =~edadA3+edadV3+uso+clase+grupo+
                        region+sumaA,
                      sigma.formula =~edadA3+edadV3+uso+grupo+region+
                        sumaA,
                      xi0.formula =~edadA3+edadV3+uso+sumaA,
                      family=GG(),data=MontoS,trace=TRUE)

summary(ModeloGG3)

#####
#   GRAFICOS DE LOS 3 MODELOS GAMMA GENERALIZADA
#####

newparGG<-par(mfrow=c(2,2),mar=par("mar")+c(0,1,0,0),col.axis="blue4"
,
              col="blue4", col.main="blue4",col.lab="blue4",pch="+",
              cex=.45, cex.lab=1.2, cex.axis=1, cex.main=1.2)

plot(ModeloGG3, xvar = edadA3, par=newparGG) ##cambiar GG1, GG2 O GG3

wp(ModeloGG3, ylim.all = 0.2) ##revusar limites grafico

wp(ModeloGG1, xvar = ~MontoS$MARCA)
Q.stats(ModeloGG1)

```

Bibliografia

- Abramowitz, M. y Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, ninth dover printing, tenth gpo printing edn, Dover, New York City.
- Dionne, G. y Vanasse, C. (1989). A generalization of actuarial automobile insurance rating models: The negative binomial distribution with a regression component, *ASTIN Bulletin* **19**(2): 199–212.
- Frangos, N. E. y Vrontos, S. D. (2001). Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance, *ASTIN Bulletin* **31**(1): 1 – 22.
- Hogg, R. V. y Klugman, S. A. (1984). *Loss Distributions*, John Wiley & Sons, Nashville, TN.
- M Stasinopoulos, M Enea, R. R. (2017). *Zero adjusted distributions on the positive real line*, mran.microsoft.com.
- Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V. y Bastiani, F. D. (2017). *Flexible regression and smoothing : using GAMLSS in R*, R, Chapman and Hall/CRC.
URL: <https://www.crcpress.com/Flexible-Regression-and-Smoothing-Using-GAMLSS-in-R/Stasinopoulos-Rigby-Heller-Voudouris-Bastiani/p/book/9781138197909>
- Stein, G. Z., Zucchini, W. y Juritz, J. M. (1987). Parameter estimation for the sichel distribution and its multivariate extension, *Journal of the American Statistical Association* **82**(399): 938–944.
URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1987.10478520>
- Tzougas, G. y Frangos, N. (2014). The design of an optimal bonus-malus system based on the sichel distribution, *EAA Series*, Springer International Publishing, pp. 239–260.
- Tzougas, G., Vrontos, S. y Frangos, N. (2015). Risk classification for claim counts and losses using regression models for location, scale and shape, *Variance* **9**(1): 140–157.