

Dependencia en el modelo colectivo de riesgo de una compañía de seguros en México

Dependence in the Collective Risk Model of an Insurance Company in Mexico

David Conaly Martínez Vázquez*

Christian Bucio Pacheco**

Héctor Alonso Olivares Aguayo***

(Fecha de recepción: 1º de noviembre de 2019. Fecha de aceptación: 30 de diciembre de 2019)

RESUMEN

En la literatura actuarial, el modelo colectivo de riesgo es una importante herramienta en el análisis de la distribución del riesgo para una compañía de seguros. No obstante, por comodidad en los textos actuariales se supone un comportamiento independiente entre el monto y número de reclamaciones. El objetivo de este artículo es demostrar que bajo ciertas condiciones hay evidencia de dependencia entre dichas variables; para demostrar lo anterior, se hace uso de funciones cópula, elípticas y arquimedeanas. La evidencia empírica proviene de una aplicación a uno de los portafolios de reclamaciones del ramo daños de la compañía de seguros *Afirme Seguros* en la Ciudad de México, dicha evidencia demuestra que existe dependencia

* Unidad Académica Profesional Huehuetoca,
Universidad Autónoma del Estado de México,
Huehuetoca, Edo. de México, México
davidconaly@icloud.com, ORCID: 0000-0003-3337-7165

** Unidad Académica Profesional Huehuetoca,
Universidad Autónoma del Estado de México,
Huehuetoca, Edo. de México, México
christianbucio@yahoo.com.mx, ORCID: 0000-0002-0860-199X

*** Facultad de Negocios,
Universidad La Salle, Ciudad de México, México
hectoralonso.olivares@lasalle.mx, ORCID: 0000-0003-2609-8627

entre el número de reclamaciones y el monto de las mismas, lo que contrapone lo estipulado por la teoría del modelo colectivo de riesgo.

Clasificación JEL: G22, D81, C15.

Palabras clave: modelo colectivo de riesgo, seguros, cópula, reclamaciones dependientes.

ABSTRACT

The collective risk model is defined in the actuarial literature as an important risk distribution analysis tool for insurance companies. Actuarial textbooks assume an independent behavior between the number of claims and their amount. The main objective of this paper is to show that under certain circumstances evidence of dependency between the variables studied may be found. To ascertain this objective copula functions, such as Elliptical and Archimedean, were used. A Complaints Portfolio on the damage section of Afirme Seguros Company from Mexico City, Mexico, was analyzed. The empirical evidence found showed the existence of dependency between the number of claims and their amount, this finding contradicts what is stated in the Collective Risk Model Theory.

JEL Classification: G22, D81, C15.

Keywords: *Collective Risk Model, Insurances, Copula, Dependent Claims.*

Introducción

El impacto de las reclamaciones en una compañía de seguros puede desequilibrar la estabilidad de la misma. Por esta razón, es fundamental una adecuada administración, evaluación y previsión de la siniestralidad dentro de un horizonte de tiempo finito; particularmente considerando las condiciones económicas y sociales de los asegurados, para garantizar un correcto nivel de reservas y cálculo de primas (cumpliéndose el principio de ganancia neta dentro del seguro).¹

El modelo colectivo de riesgo, describe el agregado de reclamaciones como un fenómeno adverso para el patrimonio de una aseguradora, que puede presentarse durante un período de tiempo $[0, T]$. Uno de los supuestos, que generalmente se considera por comodidad, es que existe independencia

¹ La esperanza de pérdida para la compañía debe ser menor a lo que se cobra en primas (Klugman, 2012).

entre el número de reclamaciones N , y el monto de las mismas Y_i , ambas siendo variables aleatorias. Este fenómeno puede ser modelado mediante un proceso de Poisson Compuesto, que es un proceso estocástico a tiempo continuo, donde el número de reclamaciones que ocurren en el intervalo de tiempo $[0, T]$ no está relacionado con el monto de las mismas. A partir de este hecho, se construyen diversas estructuras para el cálculo de la distribución del riesgo (comportamiento de reclamaciones) para un portafolio. El objetivo de este artículo es demostrar que bajo ciertas condiciones en México esta independencia no sucede, debido a las características de los asegurados, número de pólizas y el periodo de observación. Por lo tanto, se incluyeron en el análisis de riesgo de una compañía de seguros, elementos característicos del portafolio que afectan de manera considerable el cálculo de primas, reservas y reaseguros, mismos que son esenciales tanto para la compañía de seguros como para los asegurados. Al existir dependencia entre el monto de las reclamaciones y el número de las mismas, resulta necesario construir funciones de densidad condicionales que puedan capturar, de una forma más precisa, la severidad o frecuencia de la dependencia dentro del portafolio. Por esta razón, se han elegido las funciones cópula para el análisis de la dependencia entre ambas variables, puesto que las formas tradicionales de cálculo de correlación sólo miden en términos reales este fenómeno y las funciones cópula muestran una estructura de dependencia.

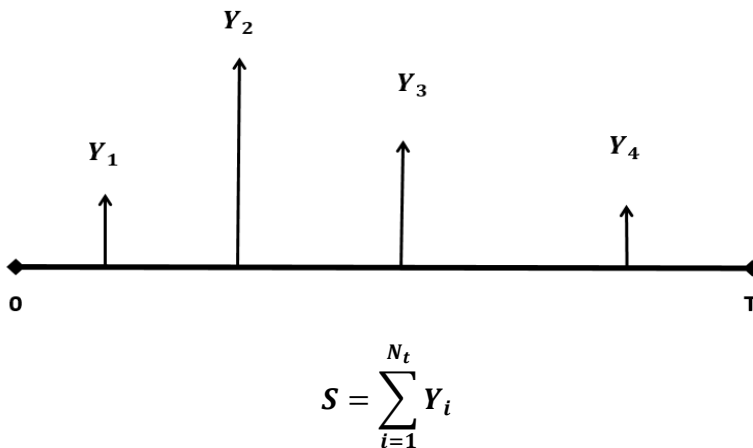
La hipótesis fundamental dentro de la investigación, es que existen factores que propician un mayor número de reclamaciones dentro de una compañía de seguros en un determinado periodo de tiempo y que la relación de éstas respecto a sus montos es diversa, es decir, no nula; en otras palabras, dependiendo el periodo de tiempo bajo análisis y los diversos factores que envuelven a las reclamaciones y a los montos establecidos por las aseguradoras, existe relación entre las reclamaciones y sus respectivos montos. Para ello, en este trabajo se analiza un portafolio de reclamaciones de la compañía de seguros *AFIRME SEGUROS* en México; cabe señalar que el portafolio bajo análisis sólo corresponde al ramo de Daños, Automóviles de Servicio Público. La información considerada pertenece al periodo del 2 de enero de 2017 al 31 de diciembre de 2018. La evidencia empírica demuestra que en el periodo de estudio se tiene la existencia de dependencia entre el número de reclamaciones y el monto de las mismas; en particular en determinados periodos de tiempo (finales de año) por lo tanto, no se cumple el supuesto del modelo colectivo de riesgo referente a la independencia entre estas variables.

El trabajo está estructurado de la manera siguiente, en el subsecuente apartado se presenta el marco teórico, así como su relevancia para la estructura del mismo. Posteriormente, se enuncia la metodología utilizada para el análisis de dependencia estipulada en esta investigación, dependencia entre el número de reclamaciones y el monto de las mismas. A continuación se exponen los resultados en el apartado de evidencia empírica. Y finalmente se ofrecen las conclusiones obtenidas en esta investigación.

1. Marco teórico

El modelo colectivo de riesgo parte de la posibilidad de un arribo de distintos montos de reclamaciones (Y_i) en un horizonte de tiempo finito y a corto plazo.² El agregado de Y_i , representa el riesgo para la compañía de seguros, evidentemente son variables aleatorias y el número total de reclamaciones también es aleatorio.

Gráfica 1. Diagrama modelo colectivo



Fuente: elaboración propia.

La esperanza de pérdida para la compañía, es el producto de la distribución del número de reclamaciones por las expectativas del monto de las mismas. En este modelo es indiferente el número total de pólizas dentro del portafolio, ya que lo único que interesa al administrador del riesgo, es el nú-

² Para mayor referencia sobre el modelo colectivo de riesgo véase Klugman (2012), Bowers (1997), Kaas, R., Goovaerts, M., Dhaene, J. y Denuit, M., (2008).

mero de reclamaciones y la distribución de las Y_i en un intervalo de tiempo $[0, T]$. Por esta razón, la pérdida esperada suponiendo independencia, es el producto de la esperanza de ambas variables aleatorias.

$$E(S) = E(N)E(Y) \quad (1)$$

Donde $E(S)$ es el riesgo para la aseguradora en dicho horizonte de tiempo, $E(N)$ el número de reclamaciones esperadas y $E(Y)$ la expectativa de las variables aleatorias Y_i (o monto de reclamaciones). En este sentido, es entendible el por qué en la mayoría de los análisis actuariales se considera la independencia de ambas variables, si los montos de reclamaciones se consideran variables aleatorias independientes e idénticamente distribuidas, el número de ellas que puedan efectuarse en un intervalo de tiempo no tendrían relación.

Para contrarrestar el supuesto de independencia entre las variables: número de reclamaciones y monto de las mismas; es decir, entre N y Y_i , se necesita de una metodología que cuantifique el grado de dependencia existente entre estas variables. En este sentido, se han desarrollado trabajos que han abordado dependencia dentro del modelo, ya sea entre las reclamaciones en el monto de las variables y el número de las mismas; entre ellos Dhaene y Goovaerts (1996), consideran la existencia de dependencia en pares de reclamaciones de un portafolio de pólizas de vida, lo cual afecta directamente a la prima stop-loss para el asegurado. Klugman, Panjer y Willmot (2012), que desarrollan el modelo colectivo de forma tradicional, considerando independencia entre la distribución del número de reclamaciones y el monto de las mismas, destacando algunas ventajas de modelar de forma independiente ambas variables; como por ejemplo la variación del número esperado de reclamaciones considerando que también cambia el número de pólizas. Asimismo, resaltan que el impacto de fenómenos económicos se ve reflejado de mejor forma en las pérdidas sufridas por los asegurados y la compañía de seguros, y que el efecto de cambiar los deducibles individuales y los límites de las políticas resulta más eficiente; asimismo, que se cuenta con una captación más precisa de las variables, tanto en las reclamaciones como en los deducibles en los modelos para las pérdidas no cubiertas a los asegurados y en los costos de las reclamaciones a los reaseguradores, considerando que todos éstos pueden ser consistentes. Estipulan también que la forma de la distribución de S depende de las distribuciones de N y Y_i .

Escalante y Arango (2004) analizan la importancia del modelo colectivo de riesgo dentro del campo actuarial así como la administración de riesgos para una compañía de seguros partiendo de las características de la distribución del riesgo y de la independencia de las variables. Por su parte, Lin y Willmot (2014), abordan diversos modelos agregados de reclamaciones cuando existe reemplazo en las pólizas considerando independencia y dependencia, primordialmente analizando las consecuencias de la dependencia en las variables aleatorias, número de reclamaciones y monto de las mismas. Hernández-Bastida, *et al.* (2011), suponen independencia entre el monto de las reclamaciones y el número de las mismas para analizar la distribución de las primas colectivas con un enfoque Bayesiano, bajo el principio de la prima de la varianza, en el clásico modelo de riesgo colectivo exponencial Poisson. Gómez y Calderín (2013), estudian el modelo colectivo de riesgo aplicado a daños en automóviles asumiendo la pérdida de Erlang, cuando la frecuencia del reclamo sigue la distribución discreta generalizada de Lindley suponiendo independencia. Por su parte, Ricotta y Clemente (2016), analizan el riesgo de reserva en la suscripción de seguros diferentes a seguros de vida mediante el modelo colectivo, incorporando técnicas Bayesianas a la estimación de los parámetros de dicho modelo. Asimismo, investigan la correlación entre las variables del modelo suponiendo dependencia.

Recientemente, Bazyari y Roozgar (2019) analizan el modelo Poisson compuesto, suponiendo dependencia entre el monto de los reclamos y el tiempo entre llegadas de los mismos, para posteriormente hacer una aproximación a la probabilidad de ruina en tiempo finito, graficando para el valor del umbral fijo asociado al tiempo entre llegadas de reclamo y también para el parámetro de dependencia fija en cópulas. Oh, R., Ahn, J. Y. y Lee, W. (2019) analizan la dependencia flexible entre las variables frecuencia y severidad de las reclamaciones para una compañía de seguros en el ramo daños mediante funciones de cópula bivariadas.

Se puede constatar, dada la revisión de la literatura, que la independencia entre las variables que comprenden al modelo colectivo de riesgo es cuestionable y por lo tanto debe ser analizada antes de asumir lo que regularmente se admite en el modelo colectivo de riesgo, i.e., que las variables monto de las reclamaciones y el número de reclamaciones, son independientes. Por ello, es de suma importancia, poder elegir aquellas herramientas que hagan un análisis más específico de la correlación entre variables aleatorias. La búsqueda de nuevas y mejores metodologías que ofrezcan resultados más precisos en la medición de riesgos en el mercado actuarial, y toma

de decisiones, brindan una mayor estabilidad y eficiencia tanto para el mercado asegurador, como para los reguladores.

La metodología tradicional utilizada para cuantificar la relación que existe entre variables aleatorias es el coeficiente de correlación de Pearson, el cual permite cuantificar la relación lineal que se tiene entre variables aleatorias. Sin embargo, se tiene registrado que dependiendo de las características que tengan las variables aleatorias éstas pueden no corresponder a una función gaussiana, la cual precede a la linealidad. Por ende, en la mayoría de los casos se requiere de una metodología más robusta con respecto a la cuantificación del grado de relación existente entre variables aleatorias. Para ello, se cuenta con la metodología de cópulas.³ Cabe hacer mención que las cópulas estiman patrones de dependencia que tienen como características la no linealidad y la dependencia asintótica; a diferencia del análisis de correlación lineal de Pearson. Asimismo, es primordial enfatizar que vía la estimación a través de cópulas se ha demostrado la obtención de mejores resultados respecto a la descripción de la relación existente entre variables aleatorias con características de no linealidad.

La metodología de cópulas está sustentada en el Teorema de Sklar, el cual es el siguiente:

Teorema de Sklar

Sea una función de distribución n-dimensional F con distribuciones marginales continuas F_1, F_2, \dots, F_n , pero existe una única n-cópula $C: [0,1]^n \rightarrow [0,1]$, tal que:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (2)$$

Por tanto, la función cópula une las marginales para generar una función de distribución multivariada. Cabe hacer mención que en esta investigación se estima sólo el caso de dependencia bivariada debido a que interesa conocer el patrón de dependencia entre el monto de los siniestros y el número de reclamaciones, lo cual procede de solamente dos variables aleatorias, por lo cual pertenece al caso bivariado.

³ El término cópula fue utilizado por primera vez en el Teorema de Sklar (Sklar, 1959).

Existen un número considerable de cópulas, por simplicidad en este trabajo sólo se estiman cópulas de la familia elíptica (cópula Gaussiana y cópula t-Student) y arquimediana (cópula Clayton, cópula Gumbel y cópula Frank).⁴

Las cópulas elípticas juegan un papel esencial dentro de las finanzas y economía, debido a que constituyen una fuente de distribuciones multivariadas que poseen propiedades de la distribución normal y facilitan la obtención de modelos multivariantes para valores extremos y otras formas de dependencia. Dentro de esta familia de cópulas las curvas de nivel de las variables aleatorias tienen un comportamiento semejante al de las elipses. Las dos cópulas más importantes de esta familia de cópulas son la cópula Gaussiana (o Normal) y la cópula t-Student, las cuales se derivan de las funciones de distribución multivariada que poseen estos mismos nombres. Su rango más característico es que representan relaciones de dependencia simétricas sin importar que se analice la cola izquierda o derecha de las distribuciones implicadas (Maldonado y Pazmiño, 2008).

La cópula bivariada Gaussiana es,

$$C(u_1, u_2) = \Phi_{\rho,2}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \quad (3)$$

Donde, Φ denota la distribución Normal (acumulada) y $\Phi_{\rho,2}$ denota la distribución Gaussiana -dimensional con matriz de correlación ρ .

La cópula bivariada t-Student es,

$$C_{2,\rho,v}(u_1, u_2) = T_{2,\rho,v}(T_v^{-1}(u_1), T_v^{-1}(u_2)) \quad (4)$$

Donde T_v es la distribución t-Student (acumulada) con v grados de libertad y $T_{2,\rho,v}$ denota la distribución t-Student 2-dimensional con matriz de correlación ρ y v grados de libertad.

En lo que respecta a la familia de cópulas arquimedianas, éstas son de gran utilidad para representar relaciones que dan mayor énfasis entre los sucesos extremos de las distribuciones marginales. La familia Arquimediana es una clase especial e importante de cópulas que se caracterizan por surgir de funciones seudoinvertibles, han sido utilizadas ampliamente en el

⁴ Para mayor referencia respecto a Teoría de Cópulas, familias de cópulas, estimación, etcétera, véase, Nelsen (1999).

cálculo actuarial, hidrología y pruebas de resistencia bancaria (bank stress test). En general, son cópulas capaces de capturar amplios rangos de dependencia, lo cual les ha permitido diferenciarse de las cópulas elípticas (simétricas).

Existen varias cópulas que pertenecen a la familia Arquimediana, en lo referente a este trabajo sólo se enuncian tres cópulas de esta familia, las cuales como en el caso elíptico son las más usuales dada su aparentemente fácil estimación y características; las cuales son: la cópula Clayton, la cópula Gumbel, y la cópula Frank.

La distribución bivariada perteneciente a la familia de las cópulas Arquimedianas está representada como,

$$C_{\alpha}(u_1, u_2) = \phi_{\alpha}^{-1}[\phi_{\alpha}(u_1) + \phi_{\alpha}(u_2)], \quad 0 \leq u_1, u_2 \leq 1 \quad (5)$$

donde ϕ_{α} es convexa y decreciente tal que $\phi_{\alpha} \geq 0$. A la función ϕ_{α} se le denomina generador de la cópula C_{α} y la inversa del generador ϕ_{α}^{-1} es la transformada de Laplace de una variable latente denotada γ , la cual induce la dependencia α . De esta manera, la selección de un generador da como resultado diferentes cópulas de la familia arquimediana.

La cópula bivariada Clayton es,

$$C_{\alpha}(u_1, u_2) = \{u_1^{1-\alpha} + u_2^{1-\alpha} - 1\}^{1/(1-\alpha)}, \quad \alpha > 1 \quad (6)$$

La cópula bivariada Gumbel es,

$$C_{\alpha}(u_1, u_2) = \exp\{-[(-\ln u_1)^{1/\alpha} + (-\ln u_2)^{1/\alpha}]^{\alpha}\}, \quad 0 < \alpha < 1 \quad (7)$$

Y la cópula bivariada Frank es,

$$C_{\alpha}(u_1, u_2) = \ln[1 + (\alpha^{u_1} - 1)(\alpha^{u_2} - 1)/(\alpha - 1)]/\ln \alpha, \quad \alpha > 0 \quad (8)$$

Complementando la estimación de dependencia, la metodología de cópulas cuenta con su respectiva implementación a través de la correlación de

rango, dicha correlación de rango está estipulada vía las medidas de concordancia Tau de Kendall y Rho de Spearman. En este trabajo por simplicidad y por viabilidad de interpretación sólo se hace uso de la medida de concordancia Tau de Kendall.

La τ de Kendall es una medida de dependencia no paramétrica que identifica los pares concordantes y discordantes de dos variables. Para su estimación se asigna una clasificación a las observaciones de cada variable, es decir, se generan bloques de observaciones y posteriormente se estudia la relación de dependencia existente entre las variables dadas. La τ de Kendall es una medida de concordancia que puede obtenerse a partir de la siguiente función cópula:

$$\tau = 4 \int_0^1 \int_0^1 C_\varphi(u_1, u_2) dC_\varphi(u_1, u_2) - 1 \quad (9)$$

Y respecto a la estimación de los parámetros asociados a una cópula. En este trabajo se hace uso de la estimación a través de máxima verosimilitud.⁵ Este mecanismo obtiene la estimación de los parámetros de la cópula a través de la maximización de su función de log-verosimilitud.

$$l(\varphi) = \sum_{j=1}^n \ln c\left(F_1(x_{j,1}), \dots, F_n(x_{j,n})\right) + \sum_{j=1}^n \sum_{i=1}^d \ln f_i(x_{j,i}) \quad (10)$$

en donde φ es el conjunto de parámetros tanto de las marginales como de la cópula. De esta manera, dados el conjunto de marginales y una cópula, la función de log-verosimilitud puede ser maximizada obteniendo de esta forma el estimador de máxima verosimilitud,

$$\hat{\varphi}_{MLE} = \max_{\varphi \in \Theta} l(\varphi) \quad (11)$$

La Tabla 1 muestra los intervalos en donde están determinados los parámetros estipulados por las cópulas de la familia elíptica y arquimediana

⁵ Para mayor referencia sobre el mecanismo de estimación a través de máxima verosimilitud así como para mayor referencia sobre otros mecanismos de estimación en teoría de cópulas, véase, Sklar (1959) y Nelsen, R. (1999).

Tabla 1. Parámetro y Tau de Kendall
de las Cópulas Elípticas y Arquimedianas

| | Cópula | Parámetro | τ de Kendall |
|----------------------|-----------|--------------------------|--|
| Familia Elíptica | Normal | $\varphi \in [-1,1]$ | $\frac{2}{\pi} \arcseno \varphi$ |
| | t-Student | | |
| Familia Arquimediana | Clayton | $\varphi > -1$ | $\frac{\varphi}{\varphi + 2}$ |
| | Gumbel | $\varphi > 1$ | $\frac{\varphi - 1}{\varphi}$ |
| | Frank | $\varphi \in \mathbb{R}$ | $1 - \frac{4}{\varphi} [1 - D_1(\varphi)]$ |

Fuente: elaboración propia.

y su consecuente interpretación a través de la medida de concordancia de τ Kendall; cabe decir, la τ de Kendall se encuentra dentro del intervalo $[-1,1]$, i.e., $\tau \in [-1,1]$, para todas las familias de cópulas y por ende su viabilidad e interpretación es fácil y factible.

Por ende, la relevancia del análisis propuesto en este trabajo, reside en analizar un portafolio de reclamaciones de dos años, del ramo de Daños, Automóviles Servicio Público de la compañía de seguros *AFIRME SEGUROS* en México, comprobar que en este portafolio en específico no se cumple la independencia entre las variables monto de las reclamaciones *versus* número de reclamaciones, para posteriormente proponer esta perspectiva en la evaluación de primas y cálculo de riesgos en esta compañía de seguros, y por consiguiente sugerir un análisis similar en las demás compañías de seguros.

Del mismo modo debe considerarse, que este tipo de seguro, no es un elemento aislado dentro del mercado, ya que se tienen diferentes ramos dentro del mercado asegurador, lo cual remite a los administradores de riesgo dentro de las compañías de seguros a cuestionarse la no independencia entre las variables que conforman el modelo de riesgo colectivo para los

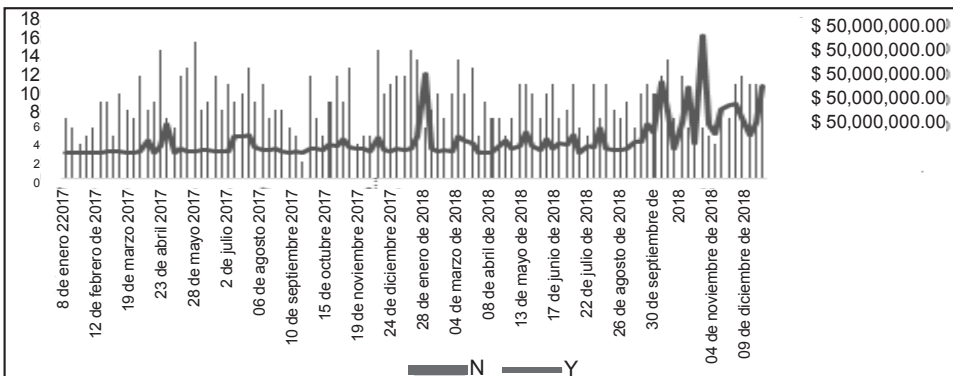
diferentes ramos del sector asegurador, y en este sentido considerar variables y funciones de distribución condicionales para el cálculo de reservas y primas. Se pueden encontrar elementos de cada región así como grupos de asegurados que presentan más reclamaciones, pero se han elegido las más recurrentes y fácilmente identificables, un análisis más detallado sobre este fenómeno excede los propósitos de esta investigación.

2. Metodología y evidencia empírica

En la primera etapa de análisis, se hace un estudio estadístico de la base de datos, la cual corresponde a un portafolio de 928 reclamaciones dentro del ramo Daños de la compañía de seguros *Afirme Seguros*, el periodo bajo análisis de esta investigación es del 8 de enero de 2017 al 30 de diciembre de 2018, lo cual representa 2 años de estudio. El análisis se llevó a cabo mediante la agrupación tanto del número de reclamaciones como de los montos (lo anterior debido a la periodicidad de los siniestros, había días donde no existían reportes de los mismos), la agrupación fue realizada de forma semanal, dando un total de 104 observaciones. La Gráfica 2 muestra el número de reclamaciones y sus respectivos montos.

A través de la Gráfica 2 se observa que en este caso no se cumple con lo estipulado por la teoría del modelo colectivo de riesgo; existen ciertos periodos de tiempo donde hay evidencia de dependencia temporal entre las variables monto de los siniestros y número de reclamaciones, especial-

Gráfica 2. Número de reclamaciones versus el monto de las reclamaciones



Fuente: elaboración propia.

mente se puede observar esto mediante la severidad. Particularmente, es posible observar este comportamiento dependiente a finales de año, donde existe un incremento en ambas series a lo largo del periodo de tiempo analizado.

Cabe decir que la razón principal por la cual se propone el uso de la metodología de cópulas para el estudio de la dependencia entre el número de reclamaciones *versus* la severidad de las mismas, es debido a que se presume que el comportamiento de estas variables no cuenta con características de linealidad y por ende de métodos tradicionales como el coeficiente de correlación de Pearson únicamente subestimarían o sobreestimarían la relación existente entre estas variables (generando así resultados espurios), así como al hecho de no verificar con certeza la independencia de las mismas.

En una segunda etapa del análisis, se generan los estadísticos descriptivos de la información, lo anterior para corroborar las características de no linealidad de las series, número de reclamaciones y monto de las mismas, se complementa el análisis mediante la estimación de diversos estadísticos descriptivos que apoyan en la determinación de dicha característica de no normalidad. Es de suma importancia identificar las variables que no cumplen con estricta linealidad dado que la propuesta de esta investigación principalmente está sustentada en que el comportamiento de las variables es dependiente, pero también en que dichas variables tienen características de no linealidad y por ende la estimación de su grado de dependencia debe ser vía mecanismos que capturen de mejor manera estas características.

En la Tabla 2 se muestran algunos estadísticos descriptivos a partir de los cuales se puede aseverar que la distribución estadística de las variables monto de las reclamaciones (Y_i), así como el número de las mismas (N), distan de la normalidad. El monto de las reclamaciones tiene una mayor concentración de los datos alrededor de la media (leptocurtosis), así como un sesgo hacia la cola derecha de la distribución. Por el contrario el número de reclamaciones tiene menor concentración en la media de su distribución lo que implica un comportamiento platicúrtico, asimismo, presenta también asimetría hacia la cola derecha aunque en menor medida que el monto de las reclamaciones.

En la tercera etapa del análisis, se procede a generar un análisis más específico de la normalidad de los datos. Por lo que respecta a una prueba de normalidad más específica, en esta investigación se hace uso de la prueba

de normalidad Jarque-Bera, se observa que este estadístico no es apropiado al 95% de confianza para el caso del monto de las reclamaciones; sin embargo, para el caso del número de reclamaciones sí pasa dicha prueba, aunque

Tabla 2. Estadísticos descriptivos del número de reclamaciones y del monto de las reclamaciones

| Estadísticos descriptivos | N | Y_i | Estadísticos descriptivos | N | Y_i |
|---------------------------|--------|------------|---------------------------|---------|------------|
| Media | 8.9231 | 4762439.62 | Curtosis | -0.6056 | 10.2172 |
| Error típico | 0.2891 | 738425.627 | Asimetría | 0.1787 | 2.9121 |
| Mediana | 9 | 1623701.02 | Rango | 14 | 45944855.7 |
| Moda | 7 | 0 | Mínimo | 2 | -16060 |
| Desviación estándar | 2.9484 | 7530493.37 | Máximo | 16 | 45928795.7 |
| Varianza | 8.6931 | 5.6708E+13 | Jarque-Bera | 2.2797 | 548.43 |

Fuente: elaboración propia, datos estimados con el software R.

es importante resaltar que la asimetría y curtosis presentada por esta última variable no es la apropiada para una distribución normal.⁶

En la cuarta etapa del análisis, se generan las funciones marginales para la aplicación de las funciones cópulas, es importante señalar que para esto se hace un suavizamiento de los datos. Cabe recordar que en este trabajo se estipuló que el análisis será realizado mediante el uso de cópulas bivariadas, es decir, cópulas compuestas sólo por dos funciones de distribución marginal. En este sentido, es importante señalar que a partir de la característica de no linealidad de los datos, i.e., de la variable número de reclamaciones y de la variable monto de las reclamaciones, se propone una suavización de las funciones de distribución marginales que generan cada una de estas variables, para un mejor ajuste de las características no lineales de estas mismas.

La suavización de las marginales propuesta en esta investigación, es generada a través de un ajuste vía una distribución Kernel. Cabe mencionar que la distribución Kernel es una representación no paramétrica de una función de densidad de una variable aleatoria y se utiliza cuando una distri-

⁶ El supuesto de comportamiento distribucional normal se cumple a partir de la verificación de los siguientes estadísticos: Asimetría = 0, Curtosis=3 y Jarque-Bera < 5.99.

bución paramétrica no puede describir correctamente los datos. La distribución Kernel está definida por una función de suavizamiento y un valor de ancho de banda, que controlan la suavidad de la curva de densidad resultante.⁷ En sí, sea X una variable aleatoria, el estimador de densidad Kernel está dado por,

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (11)$$

donde x_1, \dots, x_n son los valores aleatorios de una función de distribución desconocida, n es el tamaño de la muestra, $K(\cdot)$ es la función de suavizamiento Kernel y h es el ancho de banda.

En este trabajo para el suavizamiento Kernel se emplea una función de distribución Gaussiana, por ende el ajuste (suavización) a las funciones de distribución marginal, es estipulado como un ajuste Kernel Gaussiano.

Determinadas las funciones marginales a ser utilizadas en la metodología de cópula, es necesario recordar qué cópulas serán estimadas en este trabajo. Estas cópulas son las cópulas bivariadas establecidas en el apartado de metodología de este trabajo, y son las cópulas pertenecientes a la familia de cópulas elípticas: la cópula normal y la cópula t-Student, y a la familia de cópulas arquimedianas: la cópula Clayton, la cópula Gumbel y la cópula Frank. Estas dos familias se han elegido por las características antes mencionadas en el marco teórico.

En la quinta etapa del análisis, para cada cópula propuesta se estiman tanto su parámetro, el cual funge como indicador del nivel de dependencia que se tiene entre las variables número de reclamaciones y monto de dichas reclamaciones, así como también su parámetro de concordancia Tau de Kendall, el cual funge también como un indicador del grado de dependencia entre las variables antes mencionadas, pero en este caso éste es un indicador que contempla una estimación más refinada al ser una estimación estipulada en rangos, lo cual quiere decir que para su estimación se requiere que los datos de este indicador estén basados en rangos (bloques de datos) y no en los datos originales (datos en crudo, datos no ordenados en bloques). De

⁷ Para mayor referencia véase, Hill (1985).

este modo, el parámetro de concordancia Tau de Kendall se ve poco afectado ante la presencia de pequeños valores atípicos (extremos) y pone mayor énfasis en la conjunción de los bloques; cabe decir, las variables que tengan la característica de presentar clusters debiesen ser medidas a través de un indicador que capture de mejor manera dichas características. En este trabajo, el parámetro de concordancia Tau de Kendall estimado vía las cópulas arquimedianas a comparación de su respectivo parámetro cópula, sirve para generar una interpretación símil de la dependencia entre las variables analizadas.

La Tabla 3 muestra las estimaciones de los parámetros y de los Tau's de Kendall de las cópulas elípticas y arquimedianas propuestas. Inicialmente, se observa en el caso elíptico que los parámetros estimados para la cópula normal y la cópula t-Student distan de ser cero o cercanos a cero, siendo valores alrededor del 0.2. Asimismo, realizando un análisis de la dependencia pero visto a través del parámetro de concordancia Tau de Kendall, es posible

Tabla 3. Parámetros y Tau's de Kendall de las Cópulas Elípticas y Arquimedianas generadas entre el número de reclamaciones y el monto de las reclamaciones

| Cópulas Elípticas | | | |
|--------------------------|----------------|------------------|----------------|
| Cópula Normal | | Cópula t-Student | |
| Parámetro | Tau de Kendall | Parámetro | Tau de Kendall |
| 0.211 | 0.135 | 0.2547 | 0.164 |

| Cópulas Arquimedianas | | | | | |
|------------------------------|----------------|--------------|----------------|---------------|----------------|
| Cópula Clayton | | Cópula Frank | | Cópula Gumbel | |
| Parámetro | Tau de Kendall | Parámetro | Tau de Kendall | Parámetro | Tau de Kendall |
| 0.654 | 0.246 | 1.733 | 0.187 | 1.1564 | 0.135 |

Fuente: elaboración propia, datos estimados con el software R.

observar que todos estos resultados distan de ser cero o en su caso cercanos a cero, cabe decir que el parámetro de concordancia Tau de Kendall nos permite hacer una interpretación símil entre todas las cópulas estimadas teniendo un estimador $\tau \in [-1,1]$ el cual tiene una interpretación similar al parámetro de correlación de Pearson, $\rho \in [-1,1]$; es decir, valores cercanos

al 1, i.e., ($1 > \tau > 0.7$), representan una buena relación positiva; valores cercanos al -1, i.e., ($-1 > \tau > -0.7$), representan una buena relación negativa; valores cercanos al cero, i.e., ($0 > \tau > -0.3$) ó ($0 < \tau < 0.3$), representan poca relación; y el valor cero, i.e., ($\tau = 0$), representa nula relación.

Los resultados mostrados en la Tabla 3 confirman que existe dependencia entre las variables monto de las reclamaciones Y_i y número de estas reclamaciones N ; aunque esta dependencia es poca dado que los valores Tau de Kendall mostrados en la Tabla 3 son valores menores al 0.25, la dependencia existe, no es nula o cerca a ser nula. De los resultados mostrados en la Tabla 3 resalta el caso del parámetro Tau de Kendall de la cópula Clayton siendo un valor alrededor del 0.25, lo cual indica que se tiene dependencia asimétrica, porque aunque ésta sea pequeña, existe, por lo que se tiene mayor concentración de valores en la cola izquierda generada a través de esta función cópula.

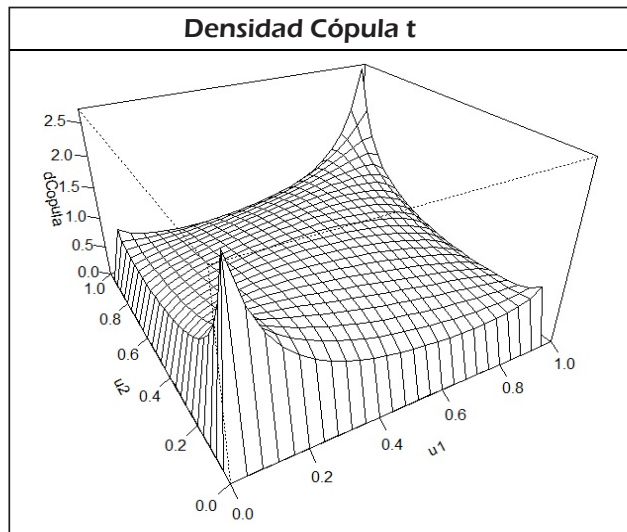
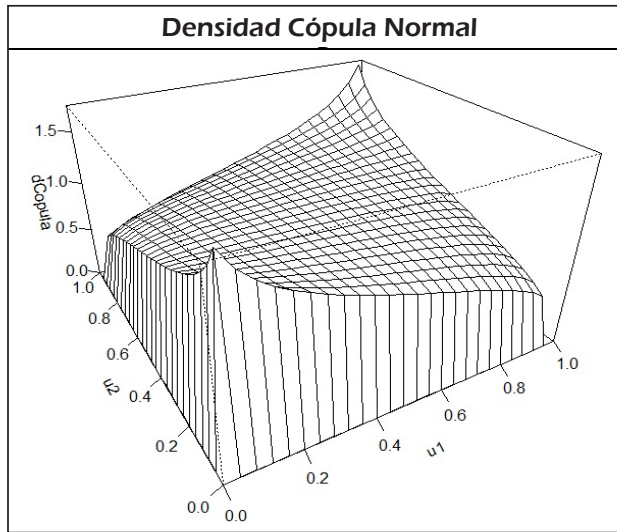
Complementando el estudio de la dependencia vía cópulas entre las variables monto de las reclamaciones Y_i y número de estas reclamaciones N , en la sexta etapa del análisis se hacen mil simulaciones en tres dimensiones, con las respectivas funciones de densidad cópula que fueron propuestas y estimadas, para de esta manera tener un panorama visual del comportamiento de dichas funciones cópula. La Gráfica 3 muestra esta realización de mil simulaciones en 3 dimensiones a través del software estadístico R de cada una de las cópulas estimadas: la cópula Normal, la cópula t-Student, la cópula Clayton, la cópula Gumbel, y la cópula Frank.

La manera de determinar visualmente vía la función de densidad cópula que tiene dependencia entre las variables monto de las reclamaciones Y_i y número de reclamaciones N , es a través de la visualización no completa de la función de densidad cópula, en otras palabras, la independencia entre variables aleatorias vista a través de una función de densidad conjunta se comprueba cuando el espacio está completamente saturado, en este momento, el caso de tres dimensiones sería que el cubo de densidad generado esté completamente lleno.

Mediante la Gráfica 3 se puede reafirmar que los resultados obtenidos en esta investigación indican que existe dependencia entre las variables número de reclamaciones N y monto de las reclamaciones Y_i , ya que se tiene que ninguno de los cubos en donde se plasman las funciones de densidad cópula propuestas tiene la característica de estar saturado, todos los gráficos de densidad de las cópulas estimadas muestran diversas estructuras de dependencia, i.e., diversas curvaturas; las cópulas elípticas y la

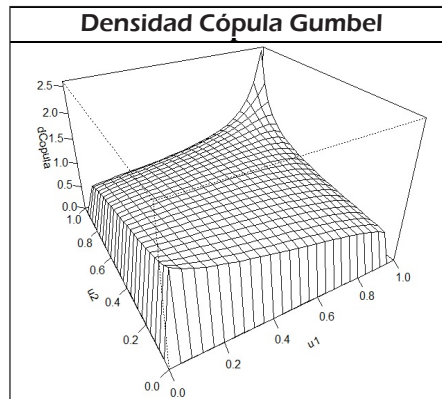
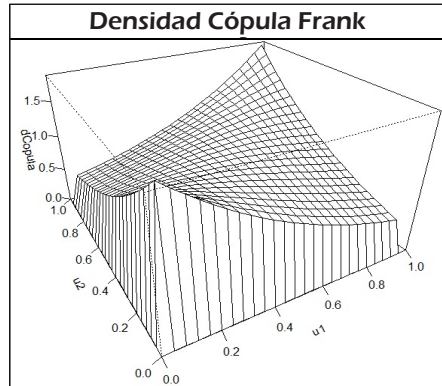
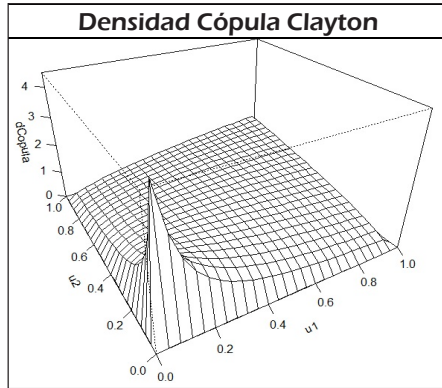
Gráfica 3. Densidad de las Cópulas Elípticas y Arquimedianas generadas entre el número de reclamaciones y el monto de las reclamaciones

Densidad Cópulas Elípticas



Fuente: elaboración propia, simulaciones de las funciones de densidad cópula realizadas con el software R.

Densidad Cópulas Arquimedianas



Fuente: elaboración propia, simulaciones de las funciones de densidad cópula realizadas con el software R.

cópula Frank muestran mayor concentración de densidad en las esquinas del cubo, y las cópulas Clayton y Gumbel muestran mayor concentración en la esquina izquierda y derecha respectivamente, lo cual es una característica específica de estas cópulas dado que éstas se centran en analizar dichos extremos.

En síntesis, los resultados obtenidos en esta investigación confirman que existe dependencia entre las variables monto de las reclamaciones Y_i y el número de estas reclamaciones N , del ramo daños de la compañía de seguros *Afirme Seguros* en el periodo de 2017 a 2018. Aunque cabe decir, que si bien esta dependencia es pequeña, dado los resultados presentados, sí existe, no es nula o cerca a ser nula, lo cual es el objetivo de este trabajo.

Conclusiones

El objetivo de este trabajo fue demostrar que bajo ciertas condiciones hay evidencia de dependencia entre las variables número de reclamaciones N y el monto de las reclamaciones Y_i , reflejando que existen ciertos factores que intervienen en el grado de siniestralidad para una compañía de seguros, que hacen que la distribución de ambas variables esté relacionada en determinados periodos de tiempo. Para demostrar esto se llevó a cabo un análisis de ramos de Daños, Automóviles de Servicio Público, de la compañía de seguros *Afirme Seguros* tomando como periodo de análisis los años 2017 y 2018.

La estimación y comprobación de la dependencia se realizó a través de la metodología de cópulas, dicha metodología es una herramienta de gran magnitud en el análisis de dependencia ya que a diferencia de la medición tradicional (la correlación de Pearson), ésta estima la dependencia asintótica, es decir, pone énfasis dependiendo de qué cópulas se trate, particularmente estimando la dependencia en los extremos de las funciones multivariadas generadas a través de dichas cópulas. La metodología de cópulas se propuso y se utilizó considerando las características de los datos analizados; otro aspecto considerado, fue que actualmente dicha metodología es ampliamente utilizada en el análisis de dependencia.

Los resultados obtenidos en esta investigación confirmaron que existe dependencia entre las variables monto de las reclamaciones Y_i y número de estas reclamaciones N , en el caso específico del ramo daños de la compañía de seguros *Afirme Seguros* en los años 2017 al 2018. Aunque, dicha depen-

dencia es poca, dado que los valores estimados tau de Kendall están por debajo del 0.25 para todas las cópulas analizadas, dicha dependencia existe.

Asimismo, un resultado esencial de esta investigación es proponer que en la estimación del grado de exposición al riesgo de cualquier compañía de seguros (riesgo en el modelo colectivo), debe considerarse una densidad conjunta entre la severidad de los siniestros (reclamaciones) y el número de estos mismos. Por comodidad, la mayoría de las investigaciones suponen que existe independencia, para la generación de los primeros momentos del modelo colectivo de riesgo, pero hay situaciones y portafolios donde esto no se cumple.

La propuesta es que los administradores de riesgo en cualquier compañía de seguros consideren estas singularidades, puesto que de eso depende el cobro de primas, reaseguro y reservas. Igualmente, la importancia de este análisis, reside en incluir esta perspectiva en la evaluación de primas y cálculo de riesgos para cualquier tipo de seguro. Debe considerarse que el tipo de seguro analizado en esta investigación, no es el único tipo de seguro, existe una amplia gama de seguros (seguro de vida, seguro de gastos médicos, etcétera). Por ende, cabe mencionar que la metodología utilizada en este artículo podría ser aplicada a otros tipos de seguros.

Referencias bibliográficas

- Bazyari, A. y Roozegar, R. (2019). "Finite Time Ruin Probability and Structural Density Properties in the Presence of Dependence in Insurance Risk Model". *Communications in Statistics-Theory and Methods*, vol. 48, núm. 5, pp. 1284-1304.
- Bowers, N.L., Gerber, H. U., Hickman J. C., Jones, D. A. y Nesbitt, C. J. (1997). *Actuarial Mathematics* (2da. Ed.). Illinois, EEUU: The Society of Actuaries.
- Dhaene, J. y Goovaerts, M. J. (1996). "Dependency of Risks and Stop-Loss Order 1". *ASTIN Bulletin: The Journal of the IAA*, vol. 26, núm. 2, pp. 201-212.
- Escalante, C. y Arango, G. (2004). "Aspectos básicos del modelo de riesgo colectivo". *Matemáticas: Enseñanza Universitaria*, vol. 12, núm. 2, pp. 3-15.
- Gomez Déniz, E. y Calderín Ojeda, E. (2013). "The Compound DGL/Erlang Distribution in the Collective Risk Model". *Revista de Métodos Cuantitativos para la Economía y la Empresa*, núm. 16, pp. 121-142.

- Hill, P. D. (1985). "Kernel Estimation of a Distribution Function". *Communications in Statistics-Theory and Methods*, vol. 14, núm. 3, pp. 605-620.
- Hernández-Bastida, A., Fernández-Sánchez, M. y Gómez-Déniz, E. (2011). "A Desirable Aspect in the Variance Premium in a Collective Risk Model". *Estudios de Economía Aplicada*, vol. 29, núm. 1, pp. 1-17.
- Kaas, R., Goovaerts, M., Dhaene, J. y Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R* (Vol. 128). Springer Science & Business Media.
- Klugman, Stuart A., Panjer, Harry H. y Willmot, Gordon E. (2012). *Loss Models: From Data to Decisions*, (5a Ed). Nueva York, EEUU: Wiley.
- Lin, X. S., y Willmot, G. E. (2014). *Collective Risk Models*. In Wiley StatsRef: Statistics Reference Online (eds N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri and J. L. Teugels).
- Maldonado, D. y Pazmiño M. (2008). "Nuevas herramientas para la administración del riesgo crediticio: el caso de una cartera crediticia ecuatoriana". *Cuestiones Económicas*, Ecuador, vol. 24, núm. 2.
- Nelsen, R. (1999). *An Introduction to Copulas*, Springer-Verlag, New York.
- Oh, R., Ahn, J. Y. y Lee, W. (2019). "On Copula-based Collective Risk Models". arXiv preprint arXiv: 1906.03604.
- Ricotta, A., y Clemente, G. P. (2016). "An Extension of Collective Risk Model for Stochastic Claim Reserving". *Journal of Applied Finance and Banking*, núm. 6. pp. 45-62.
- Sklar, M. (1959). "Functions de repartition a dimensions et leurs marges". *Publ. inst. statist. univ. Paris*, núm. 8, pp. 229-231.