

**DEPENDENCE BETWEEN AUTO-INSURANCE  
GUARANTEES. ANALYSIS BY COPULES THEORY.**

**DEPENDENCIA ENTRE GARANTÍAS EN EL RAMO  
AUTOS DE UNA EMPRESA ASEGURADORA. UN  
ANÁLISIS A TRAVÉS DE LA TEORÍA DE  
CÓPULAS.**

Carmen Ruiz Arellano<sup>1</sup>

*Actuaria, Plus Ultra Seguros Generales y Vida, S.A.*

Fecha de recepción: 12/04/2021

Fecha de aceptación: 16/11/2021

**Abstract**

The present work tries to analyse the relationships between different auto-insurance guarantees through the application of the copula theory. The analysis of the data allows to shed light on the dependency relationships existing between the risk of insurance policies through various statistical functions. The results show how important are the effects derived from the analysis on the segmentation of customers and on the ways of pricing.

**Keywords:** Copula theory, empirical copula, statistical distributions, insurance, vehicles.

**Resumen**

En el presente trabajo se trata de analizar las relaciones existentes entre distintas garantías del seguro de autos a través de la aplicación de la teoría de cópulas. El análisis de los datos permite arrojar luz sobre las relaciones de dependencia existentes entre los riesgos cubiertos por las pólizas a través de diversas funciones y, cómo de ellas se derivan importantes efectos sobre la segmentación de los clientes y las formas de tarificación.

---

<sup>1</sup> Email: [ruizarellanocarmen@gmail.com](mailto:ruizarellanocarmen@gmail.com)

**Palabras clave:** Teoría de cópulas, cópula empírica, distribuciones estadísticas, seguros autos.

## **1. Introducción**

Realizando un recorrido por la historia (Ver Cuadro en Anexo) se puede visualizar un breve resumen sobre la estadística que conocemos hoy día y cómo se propicia el origen de la Teoría de las Cópulas. Ello nos ayudará a una mejor comprensión del estudio central de este trabajo, basado en el análisis de la dependencia multivariante entre variables aleatorias definidas a partir de datos sobre diferentes garantías del seguro de autos.

Hoeffding (1940-1941) publica la primera aportación a lo que más tarde se conocerá como teoría de cópulas. Este autor trata las cotas para las funciones cópulas, y sobre todo y muy importante, estudia la dependencia. Determina que, ante transformaciones crecientes, la dependencia no varía. Más tarde en 1951 Fréchet observa resultados parecidos, lo que le hace dar nombre a dichas cotas y de ese modo es como hoy las conocemos “cotas de Fréchet”. Éste último autor junto con Féron, aportaron soluciones a la cuestión que se plateaban. Dicha cuestión era el modo de determinar la conexión entre una función de distribución multivariante y las distribuciones marginales (de menor dimensión). Abe Sklar dio solución a ello, por lo que surge el teorema que lleva su nombre en el año 1959. Lo que Sklar (1959) declara en este teorema es que existe una función  $C$ , la cual establece la relación funcional entre la distribución conjunta y sus marginales unidimensionales.

Joe en el año 1996 publica un análisis más en profundidad acerca de los modelos multivariantes y conceptos de dependencia. En los últimos años (y hasta la fecha) se han publicado gran cantidad de monografías sobre este tema relacionadas, muy especialmente, con el campo de las finanzas y en particular con el actuarial. En ellas se analiza (y se pone en práctica) el estudio de riesgos asociados con las actividades vinculadas tanto con el sector de seguros, como con el sector financiero.

El estudio de la dependencia entre variables aleatorias ha sido y es una cuestión de profundos estudios desde mediados del siglo XIX. Galton (1885) fue el primer autor que trató la dependencia entre dos variables a través de la regresión en el modelo que presentó en la International Health Exhibition (Londres). A mediados del siglo XX, gracias al matemático Alfred Rényi

*Dependencia entre garantías en el ramo autos de una empresa...*

(1959) se definieron las propiedades fundamentales que debe cumplir una medida de dependencia. Unos años más tarde, en los años ochenta, uno de los principales temas de mayor auge entre expertos investigadores de las matemáticas era la dependencia entre variables aleatorias. En este contexto se han publicado distintas medidas de dependencia. Poseen diferentes características que, a su vez, dan lugar a mejoras en el estudio de los datos. Véase la siguiente tabla:

*Tabla 1.  
Medidas de Dependencia*

<b>Medida de Dependencia</b>	<b>Ventajas</b>	<b>Desventajas</b>
Coeficiente de Correlación de Pearson	Método comúnmente utilizado de fácil comprensión.	La distribución de las variables debe ser elíptica. Invariancia.
Coeficiente de Spearman	La relación entre las variables no tiene que ser necesariamente lineal aunque sí que debe ser monótonica.	Puede ser sustituido por el Coeficiente de Kendall que tiene mejores propiedades.
Coeficiente ( $\tau$ ) de Kendall	Mejores propiedades estadísticas y una mejor interpretación a modo probabilístico	Es menos sensible a los errores y discrepancias en los datos.
Correlación de Distancias	Tiene la posibilidad de ampliación del estudio a vectores y no solamente a escalares.	Si las variables están correlacionadas, se puede obtener información redundante.
Medida D de Hoeffding	No es necesario aplicar supuestos a la función de distribución	Se aplica principalmente para variables Bernoulli.
Información mutua (MI)	Calcula la reducción de la incertidumbre.	Calificar su valor significativamente distinto a cero
Coeficiente de información maximal (MIC)	Muestras con más de 250 datos.	No acepta variables aleatorias multidimensionales

Fuente: Elaboración Propia.

La teoría de cópulas aporta soluciones a los inconvenientes presentados en la

Tabla 1. Realizando una pequeña síntesis, estas desventajas son:

- Invarianza. Recordemos que el coeficiente de correlación lineal no es invariante ante transformaciones estrictamente crecientes de las variables aleatorias.
- Valor del intervalo del coeficiente de correlación lineal (Entre -1 y 1). En esta teoría se puede dar el caso en el que la variable no esté entre esos valores.
- Coeficiente correlación lineal igual a cero, necesariamente implica independencia (lineal, pero no de otras formas).
- La función cópula permite que toda esa información no se pierda y es capaz de aportarla siempre y cuando se den unas ciertas condiciones. Es conveniente indicar algunos de los diferentes tipos de cópulas en función de una de las posibles clasificaciones, como se puede ver en la ilustración 1.

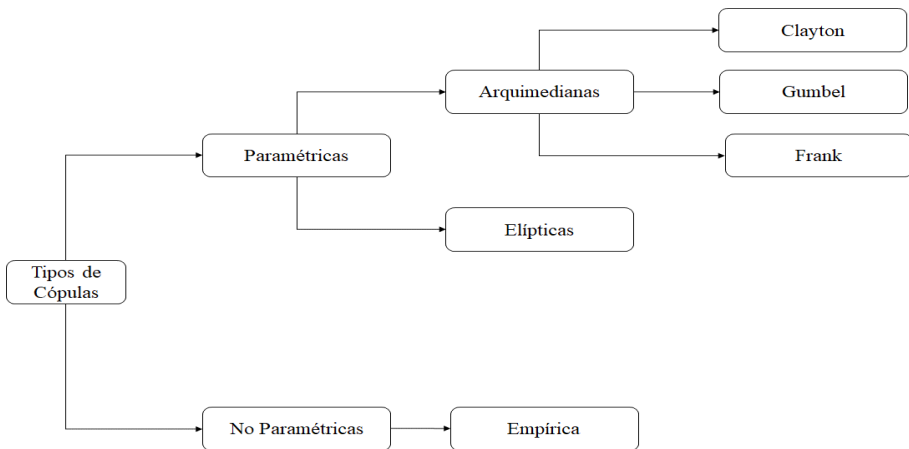


Ilustración 1. Tipos de Cópulas. Fuente: Elaboración Propia.

Aprovechando las ventajas de la aplicación de teorías de cópulas, con una muestra de datos real, se estudiarán las relaciones entre las diferentes variables (riesgos cubiertos por un seguro a terceros de automóvil) para llegar a concluir que existen interrelaciones entre éstas mismas. Por lo tanto, la tarificación con GLM's (Generalized Linear Models) podría estar incurriendo en una pérdida de información y como consecuencia impactando en la rentabilidad de la propia empresa.

El planteamiento que se realiza es el que se describe a continuación.

Actualmente, a la hora de realizar la tarificación ésta se realiza por garantía/cobertura cada una de ellas (con su prima/tasa base, estructura de factores, su distribución de frecuencia, prima de riesgo/pura, etc...) pero siempre independientes entre sí. Por lo tanto, la prima final es la suma de las primas por garantías cubiertas. En la práctica no se mide la dependencia que puede existir entre dichas garantías.

Luego, ¿Qué relación de dependencia existe (o no) entre las garantías implicadas en la tarificación de seguros?

Tal y como se acaba de indicar, las aseguradoras tarifican en función de las garantías basadas en los riesgos cubiertos en cada póliza que comercializan en los diferentes ramos en los que llevan a cabo su actividad (ya sea del negocio de vida o de no vida). La mayoría de estos riesgos son estudiados con independencia unos de otros y lo que se pretende analizar es si realmente se da la independencia entre riesgos o por el contrario existe una relación. Si esta relación existiera, ¿cuál sería su grado de dependencia?

Con teoría de cópulas, lo que se pretende es estudiar las relaciones existentes en su conjunto, es decir, no estudiar dependencias bivariadas sino llevarlas al campo n-dimensional para acercarnos a vislumbrar lo que ocurre en la realidad. En otras palabras, utilizar datos históricos de siniestralidad de una compañía aseguradora para estudiar las relaciones que se dan entre ellos.

En este caso se dispone de cuatro variables las cuales responden a cuatro riesgos distintos: responsabilidad civil, defensa jurídica, cobertura de lunas y asistencia en viajes. Lo que en este trabajo empírico trata de estudiar es: en base a un seguro a terceros de automóvil que, producido un siniestro (implica como mínimo la cobertura a la responsabilidad civil del tomador de la póliza o en su caso el asegurado de la misma) hasta qué punto puede hacerse extensible (afectar) a las coberturas tanto de asistencia en viaje y/o daños en lunas y/o defensa jurídica. Para un mayor conocimiento de las implicaciones que puede ocasionar que se dé uno u otro riesgo, a continuación se indican algunas coberturas<sup>2</sup>:

- Responsabilidad Civil de Suscripción Obligatoria.

---

<sup>2</sup> Depende de la condición que se firme en el contrato de la póliza según la aseguradora. Este caso son las condiciones de Plus Ultra Seguros.

- Responsabilidad Civil Complementaria de la Suscripción Obligatoria.
- Rotura de lunas del Vehículo Asegurado.
- Asistencia en Viaje.

## 2. Metodología. Datos Estudiados.

En primer lugar, se plantea la definición de Illanes (2013) acerca de las funciones cópulas que es el pilar central del estudio para llegar a la solución del mismo. Según el autor mencionado:

Definición. (Illanes, Cópulas paramétricas y no paramétricas., 2013):  
*“Las cópulas pueden pensarse de dos maneras. Una manera es pensar las cópulas como la función que junta o enlaza (el significado de la palabra “cópula” en latín)  $n$  funciones de distribución unidimensionales  $F_i$ , para generar una función de distribución conjunta  $n$ -dimensional con marginales  $F_i$ . La otra manera es considerar a las cópulas como una función de distribución conjunta  $n$ -dimensional con marginales uniformes  $U(0; 1)$ .”*

Estos dos modos de entender las cópulas nos servirán de gran ayuda ya que, dependiendo de los datos de los que se disponga, se podrá aplicar uno u otro. Es decir, si disponemos de datos (en definitiva, variables aleatorias compuestas de observaciones reales) de los cuales queremos extraer un análisis, como a continuación se define, habrá de seguir un método de estimación no paramétrico de la cópula y de forma empírica extraer la mayor información posible. Si a priori se conoce la distribución de las diferentes variables, de las cuales quiere estudiarse el modo en el que se relacionan, con la aplicación de ciertas condiciones (en apartados posteriores se procederá a detallarlas) nos puede llevar a la identificación de la cópula óptima.

Se dispone de una base de datos compuesta por 6030 pólizas de un seguro de coche a terceros. Estas pólizas cubren los siguientes riesgos:

- Responsabilidad Civil
- Cobertura de Lunas
- Defensa Jurídica
- Asistencia en Viajes

Por lo tanto, nuestra base de datos se compone de 6030 observaciones de cada

## Dependencia entre garantías en el ramo autos de una empresa...

una de las cuatro variables mencionadas. Al ser datos empíricos, nos encontramos con que dicha muestra en muchas de las observaciones su valor es cero.

En un primer paso se realiza un análisis de esto mismo, lo que implica una segmentación de la muestra en función de la cantidad de valores positivos que contiene cada una de las variables. La variable de responsabilidad civil (CRC, en adelante) contiene todas las observaciones con valores positivos, es decir, ocurrido un siniestro el asegurado debe ser cubierto por la responsabilidad civil que implica el suceso ocurrido. Al contrario, la variable de cobertura de lunas (CLUNAS, en adelante), solamente es incluida en la obtención de la cópula multivariante. Esto es porque solamente cuenta con 17 valores positivos lo que implicaría que en los análisis bivariados nos encontraríamos con modelos poco fiables. En el caso de defensa jurídica (en adelante CDEF), se obtienen 235 datos positivos. Y, por último, la variable asistencia en viaje (CASIST en adelante) cuenta con 338 casos positivos.

Una vez analizado lo anterior, se procede a estudiar la dependencia de las tres variables elegidas (CRC, CDEF y CASIST) en la totalidad de la muestra:

Tabla 2.

*Matrices de Correlación de la muestra total.*

Matriz Correlaciones Pearson <sup>3</sup>			
	CRC	CDEF	CASIS
CRC	1	-0,1426284	0,09859058
CDEF	-0,1426284	1	0,0495341
CASIS	0,09859058	0,0495341	1

Matriz Correlaciones Spearman <sup>4</sup>			
	CRC	CDEF	CASIS
CRC	1	-0,34796147	0,02592672
CDEF	-0,34796147	1	0,06634437
CASIS	0,02592672	0,06634437	1

Fuente: Elaboración Propia

<sup>3</sup>Si  $\rho = -1$  ó  $\rho = 1$  estamos ante una situación de correlación lineal total (negativa o positiva respectivamente). Por otro lado si  $\rho = 0$  implica la inexistencia de relación.

<sup>4</sup> Si  $0 \leq |\rho Spearman(X, Y)| \leq 1$  a medida que el valor sea cercano a 1 mayor dependencia se producirá y viceversa.

En las tablas anteriores se presentan los resultados de las medidas de dependencia tanto lineal (Pearson) como no lineal (Spearman). Los resultados nos demuestran que la estructura de dependencia de estas variables puede ser no lineal por lo que, con la ayuda de la teoría de cópulas, nos acercaremos a la estimación de la estructura de dependencia que relaciona estas variables.

Una vez se ha comprobado que existe relación entre las variables, y por lo tanto se descarta la hipótesis de que se da independencia entre las mismas, se procede a estudiar la muestra tanto en su totalidad (n=6030) como en diferentes particiones. Las particiones que se realizan se basan en el criterio de prescindir de las observaciones que sean cero en las diferentes variables para así poder obtener mejores análisis. En otros casos se utilizan 1000 observaciones, para un análisis más enriquecido, pudiendo observar qué ocurre con la inclusión de valores nulos. Por ello los análisis realizados serían:

*Tabla 3.*  
*Segmentación de la Muestra.*

<b>Nº Observaciones</b>	<b>Criterio de Orden de la Muestra(De Mayor a Menor)</b>	<b>Variables Estudiadas</b>
6030	NA	CRC , CDEF
6030	NA	CRC , CASIST
6030	NA	TODAS
1000	CRC	CRC , CDEF
1000	CDEF	CRC , CDEF
1000	CASIST	CRC , CASIST
338	CASIST	CRC , CASIST
235	CDEF	CRC , CASIST
235	CDEF	CRC , CDEF

Fuente: Elaboración Propia

Por ejemplo, la interpretación de la cuarta fila de la tabla 3 sería: análisis de una muestra que contiene 1000 observaciones (o pólizas), siendo estas ordenadas de mayor a menor por la variable CDEF (lo que quiere decir que contiene 235 datos positivos y 765 con valor nulo) y la cópula estimada es en relación a las variables CRC y CDEF.

Llegados a este punto, se han de realizar las diferentes estimaciones en base a las muestras anteriores, para así poder seleccionar la cópula que mejor



represente la estructura de dependencia (objetivo de este trabajo). Para ello se ha contado con la ayuda del software R-Studio y más concretamente con algunas de sus librerías<sup>5</sup> (readxl, qcc, copula, VineCopula, VC2copula, CDVine, scatterplot3d).

Una breve descripción de las librerías que más importancias toman en este trabajo son:

- Library ("copula"): Clases (S4) de elípticas de uso común, Arquimedianas, familias de cópula de valor extremo y otras, así como sus rotaciones, mezclas y asimetrías. Cópulas arquimedianas anidadas, relacionadas, herramientas y funciones especiales. Métodos de densidad, distribución y aleatorio. Generación de números, medidas de dependencia bivariadas, transformación de Rosenblat, Función de distribución de Kendall, perspectiva y gráficos de contorno. Montaje de modelos de cópula con parámetros potencialmente parcialmente fijos, incluidos errores estándar, pruebas de independencia en serie, pruebas de especificación de cópula (independencia, intercambiabilidad, simetría radial, valor extremo dependencia, bondad de ajuste) y selección del modelo basado en validación cruzada. Cópula empírica, versiones suavizadas y estimadores no paramétricos de la función de dependencia de Pickands. (Marius Hofert, 2020).
- Library ("VineCopula"): Proporciona herramientas para el análisis estadístico de modelos de cópula vine. El paquete incluye herramientas para la estimación de parámetros, selección de modelos, simulación, pruebas de bondad de ajuste y visualización. También se proporcionan herramientas para la estimación, selección y análisis exploratorio de datos de modelos de cópula bivariada
- Library ("VC2copula"): Proporciona nuevas clases para (rotados) BB1, BB6, BB7, BB8 y Tawn cópulas, extiende las familias existentes de Gumbel y Clayton con rotaciones, y permite configurar un modelo de cópula de vid usando la API 'cópula'. Los objetos correspondientes de la API 'VineCopula' se pueden convertir fácilmente.
- Library ("CDVine"): Provee de funciones para la inferencia estadística de la cópula vine canónica (C-vine) y cópulas D-vine. Herramientas para el análisis de datos exploratorios bivariados y para multivariados, así como la selección de cópula de vid. Los modelos pueden ser estimados

---

<sup>5</sup> Ver (Brechmann y Schepsmeier., 2013) y (Scrucca, L. 2004).

ya sea secuencialmente o por estimación conjunta de máxima verosimilitud. También se incluyen algoritmos de muestreo y métodos de trazado. Se supone que los datos se encuentran en la unidad hipercubo (llamada cópulas).

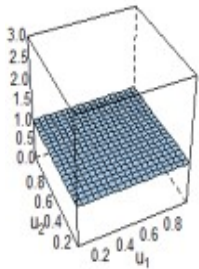
### 3. Resultados

Tal y como se ha comentado en el punto anterior, se procede a mostrar los resultados obtenidos. Esto se hará en orden creciente, es decir, se comenzará mostrando los análisis compuestos de la muestra que contiene el menor número de observaciones (235 pólizas) y se concluirá con la totalidad de la muestra en su conjunto. Esto permite ir enriqueciendo el análisis según se aumenta el número de pólizas estudiadas.

#### 3.1. Muestra compuesta por 235 Pólizas

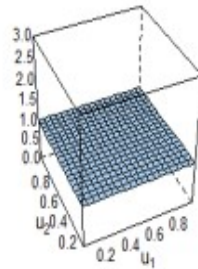
Para este tamaño de muestra se han realizado dos análisis. Por un lado, las variables CRC, CASIST y el segundo estudio con el par de variables compuesto por CRC, CDEF. Las cópulas obtenidas son:

Cópula Independiente



CRC y CASIST ; N=235DEF

Cópula Independiente



CRC y CDEF ; N=235DEF

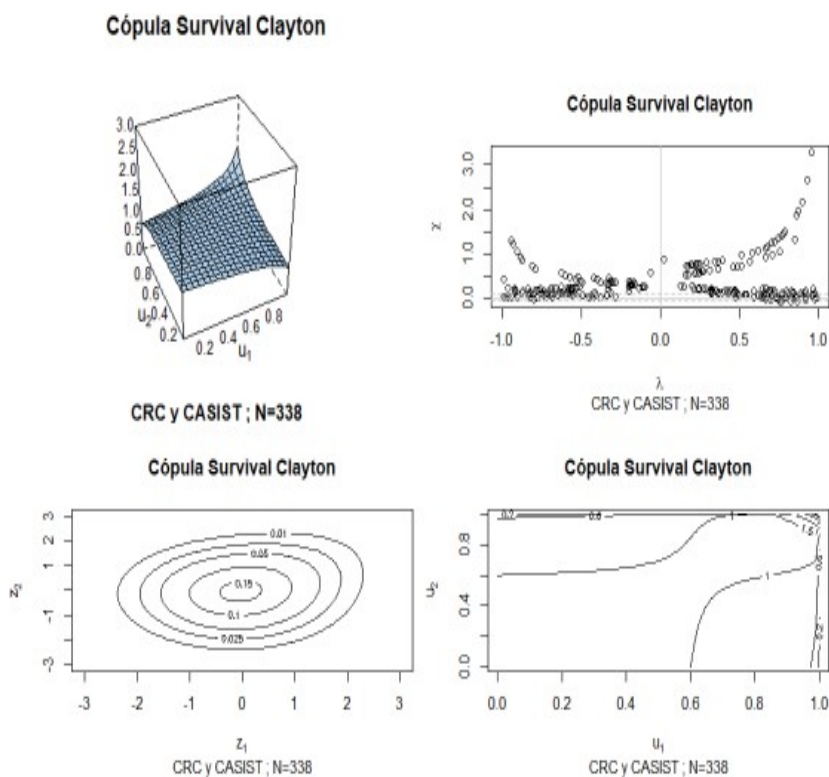
*Ilustración 2.* Cópula Independiente. Fuente: Elaboración Propia.

Como puede observarse en la imagen anterior, la cópula que se obtiene de los dos anteriores análisis es la independiente. Es decir, con la muestra de las 235

pólizas ordenadas según los valores positivos de la variable CDEF, no existe relación entre los pares de variables estudiados.

### 3.2. Muestra compuesta por 338 Pólizas

En esta ocasión, la muestra ha sido seleccionada de manera que la variable CASIST tenga valores mayores que cero en todas las observaciones. Es decir, se debe tener en cuenta que para la variable CASIST siempre se contará con un valor positivo, aunque éste se encuentre próximo a cero. La investigación en este caso se realiza para la variable de ordenación y CRC. Se debe considerar que para todas las observaciones se dispondrán de datos en ambas variables<sup>6</sup>. La cópula seleccionada es:



*Ilustración 3.* Cópula Clayton; N=338. Fuente: Elaboración Propia.

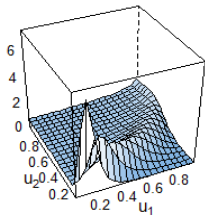
<sup>6</sup> Como ya se mencionó anteriormente, la variable CRC cuenta con valores positivos en la totalidad de la muestra. Esto no quita que puedan ser cercanos a cero.

En el primer gráfico (arriba a la izquierda), se observa que esta cópula con respecto a las dos anteriores tiene cierta similitud ya que se muestra como si las tres fuesen una superficie plana. La diferencia de esta última cópula es que (por ejemplo) en el vértice (1,1) tiene una clara elevación. Lo que puede apreciarse en el gráfico de punto que muestra cómo se distribuyen los puntos de dicha cópula.

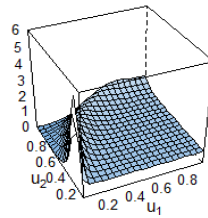
Esta cópula, es una cópula de supervivencia de Clayton modelo que suele aplicarse a estudios de estructuras de tiempo de vida bivariados con respecto a la supervivencia. La estructura de la cópula Clayton es muy parecida, en lo único que difiere es en la elevación de la densidad que se produce en el punto (0,0) en lugar de en el punto (1,1) que es el caso estudiado (Survival Clayton).

### 3.3. Muestra compuesta por 1000 Pólizas

Cópula Rotated Tawn type 1 180 degrees      Cópula Rotated Tawn type 2 180 degrees

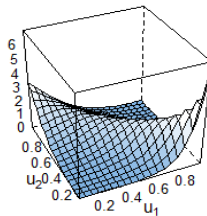


CRC y CASIST ; N=ASIST1000

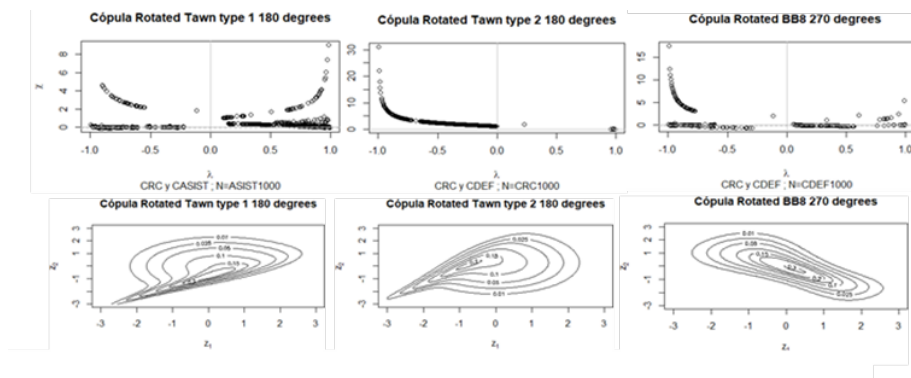


CRC y CDEF ; N=CRC1000

Cópula Rotated BB8 270 degrees



CRC y CDEF ; N=CDEF1000



El resultado de las cópulas que mejor representa cada uno de los tres análisis anteriores son cópulas rotadas. ¿Cuál es el significado de cópulas rotadas? Existen ciertas cópulas cuyos resultados sólo pueden determinar correlaciones positivas, por ello, si la relación entre las variables de estudio es negativa, serviría con aplicar una rotación a los ejes y así la estimación no se vería alterada<sup>7</sup>. En los modelos de cópulas presentados, a los dos primeros se les aplica una rotación de 180 grados y al tercero 270 (rotación máxima).

Tawn cópula es una extensión asimétrica de la cópula Gumbel que es muy utilizada en la modelización de la relación entre observaciones extremas. El hecho de que esta cópula sea representativa en los dos primeros casos es porque para la primera fila de la muestra existen 662 valores nulos en la variable ASIST. Y en la fila central hay un total de 765 con valor cero.

Con respecto a la tercera y última de este apartado, la cópula que mejor se adapta a la muestra ordenada según la variable CDEF estudiando las variables CRC, CDEF, es la BB8 aplicada una rotación de 270 grados. Esta cópula es una mixtura compuesta por una Joe-Frank cópula. La Joe pertenece a la familia de las cópulas de valor extremo que su dependencia está en la cola superior, y para reforzar el argumento de la fuerte (positiva o negativa) dependencia de la que se compone la BB8 es justificada por la cópula Frank.

A continuación, a modo de resumen y ayuda para la mejor comprensión del siguiente apartado, se muestra una tabla resumen de e las cópulas comentadas:

<sup>7</sup> Véase (Li y Kang, 2018)

Tabla 4.

Cuadro resumen tipos de cópulas.

Fuente: (Brechmann y Schepsmeter, 2013)

#	Name	Generator function	Parameter range	Kendall's $\tau$	Tail dependence (lower, upper)
3	Clayton	$\frac{1}{\theta}(t^{-\theta} - 1)$	$\theta > 0$	$\frac{\theta}{\theta+2}$	$(2^{-\frac{1}{\theta}}, 0)$
4	Gumbel	$(-\log t)^\theta$	$\theta \geq 1$	$1 - \frac{1}{\theta}$	$(0, 2 - 2^{\frac{1}{\theta}})$
5	Frank	$-\log\left[\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right]$	$\theta \in \mathbb{R} \setminus \{0\}$	$1 - \frac{4}{\theta} + 4\frac{D_1(\theta)}{\theta}$	$(0, 0)$
6	Joe	$-\log[1 - (1-t)^\theta]$	$\theta > 1$	$1 + \frac{4}{\theta^2} \int_0^1 t \log(t)(1-t)^{2(1-\theta)/\theta} dt$	$(0, 2 - 2^{\frac{1}{\theta}})$
7	BB1	$(t^{-\theta} - 1)^\delta$	$\theta > 0, \delta \geq 1$	$1 - \frac{2}{\delta(\theta+2)}$	$(2^{-\frac{1}{\theta\delta}}, 2 - 2^{\frac{1}{\theta\delta}})$
8	BB6	$(-\log[1 - (1-t)^\theta])^\delta$	$\theta \geq 1, \delta \geq 1$	$1 + \frac{4}{\theta\delta} \int_0^1 (-\log(1 - (1-t)^\theta)) \times (1-t)(1 - (1-t)^{-\theta}) dt$	$(0, 2 - 2^{\frac{1}{\theta\delta}})$
9	BB7	$(1 - (1-t)^\theta)^{-\delta} - 1$	$\theta \geq 1, \delta > 0$	$1 + \frac{4}{\theta\delta} \int_0^1 (-1 - (1-t)^\theta)^{\delta+1} \times \frac{(1 - (1-t)^\theta)^{-\delta} - 1}{(1-t)^{\theta-1}} dt$	$(2^{-\frac{1}{\theta}}, 2 - 2^{\frac{1}{\theta}})$
10	BB8	$-\log\left[\frac{1 - (1-t\delta)^\theta}{1 - (1-\delta)^\theta}\right]$	$\theta \geq 1, \delta \in (0, 1]$	$1 + \frac{4}{\theta\delta} \int_0^1 (-\log\left(\frac{1-t\delta)^\theta - 1}{(1-\delta)^\theta - 1}\right)) \times (1-t\delta)(1 - (1-t\delta)^{-\theta}) dt$	$(0, 0)$

### 3.4. Muestra compuesta por 6030 Pólizas

Por último, se presentan los resultados obtenidos al analizar la muestra total:

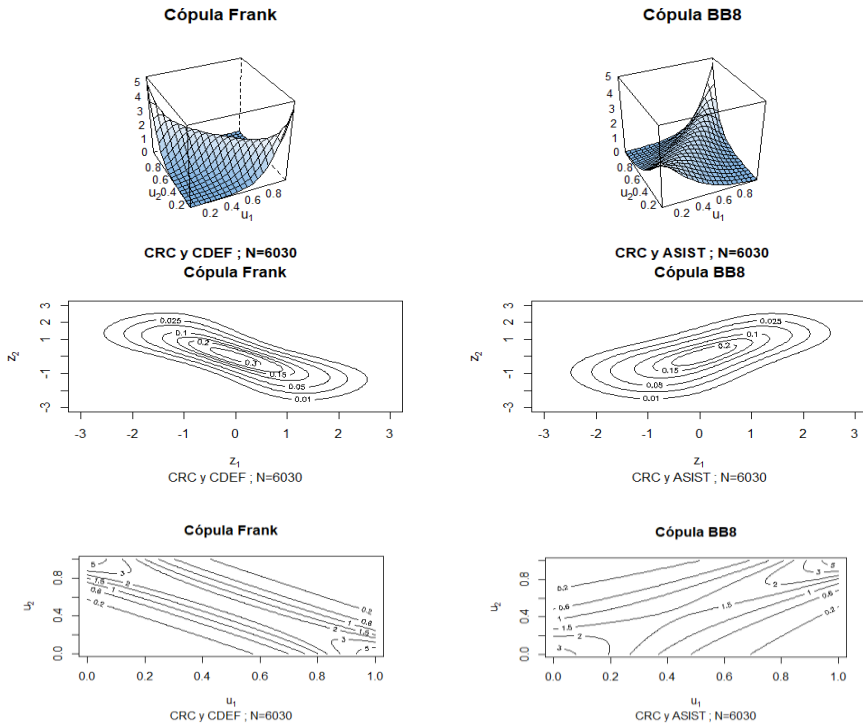


Ilustración 5. Cópula Fran y BB8; N=6030. Fuente: Elaboración Propia

La cópula Frank se ha obtenido de las variables CRC y CDEF lo que nos indica que la variable CDEF, en este caso cuenta con solamente 235 datos positivos siendo el resto cero. Por ello una cópula perteneciente a la familia de valores extremos entraría en la correcta selección de la misma, lo cual se da en éste caso.

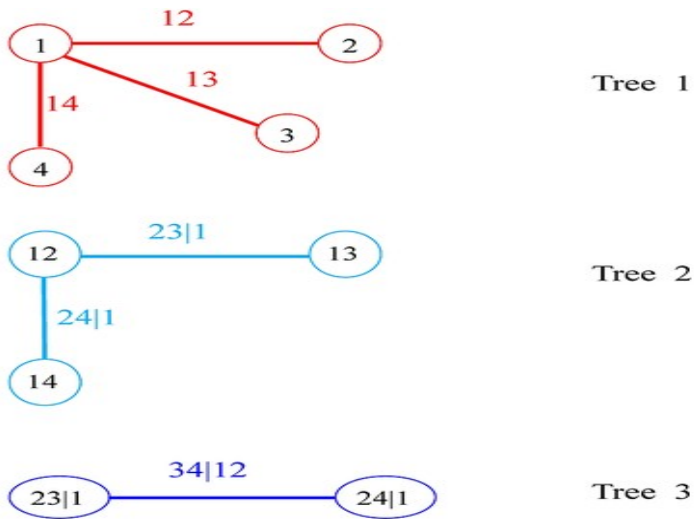
En la columna de la derecha se obtiene de nuevo la cópula BB8 lo cual es coherente con el apartado anterior, como se ha expuesto. En este caso se analiza CRC y CASIST, y como se ha comentado a lo largo de todo este apartado, ésta última variable cuenta con 338 observaciones con valores mayores que cero. Lo que, al aumentar el número de observaciones hace que se pueda disponer de una mayor precisión de estimación de la cópula adecuada.

Finalmente, y para concluir este punto, se obtuvo la cópula n-dimensional. Ésta se construye en base a cópulas bivariadas estudiadas por árboles de decisión. La traducción según (Liu, Zhou, Chen y Guan, 2015) describe:

Tomando como ejemplo la cópula de vine C tetra dimensional, la ilustración 6 se muestra un modelo gráfico de una vine C con tres árboles. Esta representación gráfica ilustra el orden de las variables. En el primer árbol, la variable 1 juega un papel fundamental. La variable 2 tiene este papel fundamental en el segundo árbol, ya que todos los pares posibles de la variable 2 con las variables restantes se modelan condicionalmente a la variable 1, y de manera similar para todos los demás árboles. La densidad multivariada de  $x_1, x_2, x_3$  y  $x_4$  se puede formular de la siguiente manera:

$$f_{1234} = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot c_{12} \cdot c_{13} \cdot c_{14} \cdot c_{23|1} \cdot c_{24|1} \cdot c_{34|21}$$

donde  $c_{1,2}$  ( $F_1(x_1), F_2(x_2)$ ) simplemente se escribe como  $c_{12}$ .



$$f_{1234} = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot c_{12} \cdot c_{13} \cdot c_{14} \cdot c_{23|1} \cdot c_{24|1} \cdot c_{34|12}$$

Ilustración 6. Arboles Cópula vine C. Fuente: (Liu et al., 2015)

Con ésta última explicación podremos interpretar los resultados obtenidos del análisis realizado.

Resultan 6 cópulas, que representa los resultados obtenidos para la cópula conjunta. Para una mejor comprensión se ha adaptado la ilustración 6 dando lugar a la ilustración 7:



## Dependencia entre garantías en el ramo autos de una empresa...

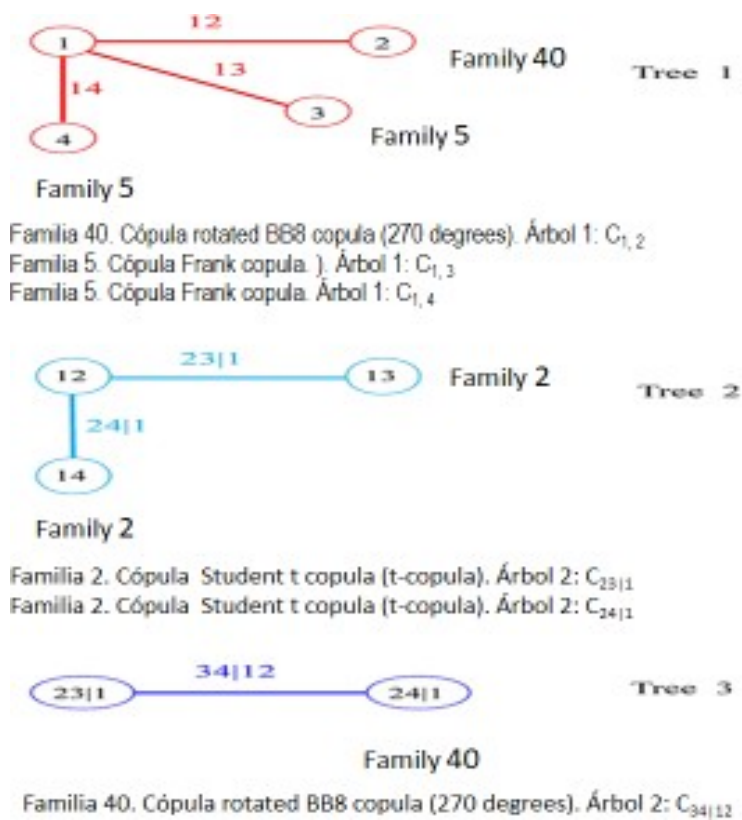


Ilustración 7. Cópula Conjunta (Árbol). Fuente: Elaboración Propia.

Las interrelaciones que se producen no hacen más que evidenciar los resultados que se han ido mostrando a lo largo de todo este punto ya que pertenecen a la familia Arquimediana y ello, como ya se comentó, facilita la comparación entre las diferentes cópulas empíricas como así se ha realizado.

### 4. Discusión y Conclusiones

A la vista de los resultados anteriores ha de tenerse en cuenta que la muestra estudiada de la cuantía de siniestros de una cartera no se puede generalizar y no es posible hablar de una única estructura de dependencia. Es decir, se observan distintas relaciones en función de diversas aproximaciones: tamaño de la muestra, ordenación de los datos, etc.

Tal es así que, en este caso y con la muestra de menor tamaño, las cópulas obtenidas han sido las que se caracterizan como independientes. Es decir, un

plano en el que su dimensión puede calcularse efectuando el producto de lado por lado (lo que equivaldría a multiplicar la función de densidad de cada una de las variables) así se obtendría la función de distribución conjunta sin adicción (suma) de la función cópula que representaría la dependencia de las variables. Por lo tanto, y aplicándolo al caso central que nos ocupa (actuarial), si se desea tarificar un seguro el cuál pertenezca a una cartera de un número no muy elevado de pólizas, se puede presuponer la independencia de las variables y aplicar GLMs para la obtención del precio de la prima. Este método es el comúnmente utilizado ya que su premisa principal y fundamental es la independencia de las variables observadas.

El siguiente caso analizado fue con una sub-muestra de 338 pólizas las cuales estaban compuestas por valores positivos tanto en la variable CRC como en la variable CASIST. El resultado obtenido fue una cópula de supervivencia, es decir su función de distribución es  $1 - F(x)$ , lo que implica que si una Clayton normal concentra su densidad en torno al punto  $(0,0)$ , ésta será mayor cercana al punto  $(1,1)$ . Esta situación es muy útil ya que presenta una dependencia en la cola superior, que en términos del ámbito actuarial nos indica que si se produce un siniestro en uno de los riesgos (en este caso, CRC) tienda a estar acompañado de una situación desfavorable en otro de los riesgos (en este caso, CASIST).

Siguiendo con el análisis de estas dos mismas variables (CRC, CASIST), se puede comprobar que para una muestra de 1000 pólizas la cópula obtenida ha sido Rotated Tawn type 2 180 degrees. Como se puede comprobar esta cópula también cuenta con su mayor densidad en el punto  $(1,1)$  lo cual apoya lo anteriormente explicado que hace que estas dos variables impliquen que, si se da un siniestro, se produzcan costes en las dos garantías cubiertas. Tanto es así que analizando la muestra completa compuesta por 6030 póliza se obtiene la llamada cópula BB8, la cual es una mixtura de la Joe-Frank. Ésta pertenece a las cópulas de valores extremos y refuerza el argumento según el cual tiene su mayor dependencia, y con una fuerte intensidad, en el punto  $(1,1)$ . Por lo tanto, si se desea tarificar no se puede presuponer la independencia de estas dos variables ya que así se ha demostrado. Se deberían tener en cuenta las medidas de dependencia para una mejor eficiencia a la hora de tarificar. Valorando entre un 14% y un 20% como medida de dependencia.

Con respecto a los análisis de las variables CRC y CDEF se han obtenido cópulas que se caracterizan por su asimetría. Por ejemplo, la cópula Frank es

la que más radicalidad presenta con respecto a esta característica ya que sus densidades se reparten igualmente tanto en torno al punto (0, 0) como el punto (1, 1) y justamente se obtiene del análisis de la muestra total del estudio. Reduciendo el número de pólizas objeto de estudio a 1000, se obtuvo la cópula Rotated BB8 270 degrees (muestra ordenada según la variable CDEF) que como se ha comentado anteriormente es una mixtura Joe-Frank. Y también la cópula Rotated Tawn type 2 180 degrees (orden de la muestra según CRC). ¿Cómo se interpretaría esto aplicado al ámbito actuarial? La simetría implica que se considera la misma dependencia entre las cuantías de los grandes siniestros como de los pequeños. Siendo CRC y CDEF las variables analizadas tendría su lógica ya que automáticamente cuando se da un siniestro en un seguro de autos se incurre en una responsabilidad civil por parte del conductor que normalmente implica la defensa jurídica de la misma. Por lo tanto, aunque este tipo de cópulas no es muy aconsejable en el sector asegurador por aplicar la diferencia entre los grandes y los pequeños siniestros, para el caso de éstas dos variables se descarta esta conjetura. Se considera que no se puede aplicar la independencia para obtener el precio de la prima pura y considerar para este caso de estudio entre un 50% y un 60% de dependencia negativa. Este aspecto es muy importante ya que significa que, si por ejemplo CRC aumenta, CDEF disminuirá. Lo cual tiene sentido ya que si se da un siniestro en el cual el conductor culpable deja siniestro el otro coche implicado en el suceso, se incurrirá en un alto coste de responsabilidad civil por parte del conductor culpable sin ser necesaria la defensa jurídica del mismo por las circunstancias ocurridas.

Las 6 familias de cópulas obtenidas no hacen nada más que evidenciar los resultados obtenidos en las segmentaciones aplicadas. Es decir, la muestra estudiada en su conjunto presenta por un lado relaciones simétricas y por el otro, dependencia en colas. No hay más que observar que la cópula obtenida del árbol 3, es una cópula rotada de una cópula mixta Joe-Frank lo que hace que no se deba considerar una única estructura de dependencia para todas las observaciones en su conjunto, sino que habría que realizar estudios segmentados para conseguir una mayor eficiencia y aplicabilidad.

Por último y a modo de conclusiones, en este estudio se ha tratado de analizar las relaciones entre riesgos del sector de autos para un conjunto de pólizas a través de la teoría de cópulas. Los resultados obtenidos muestran, por un lado, la importante flexibilidad de las cópulas a la hora de ajustar diversas funciones a los datos, lo cual es una ventaja para estudios ad hoc aplicable a diversos

ámbitos del seguro y, por otro, la relevancia que tiene para conseguir una adecuada tarificación.

A raíz de las últimas líneas del apartado anterior, se podría concluir que habría que aplicar una adecuada segmentación de la muestra de la que se dispone y con ello poder conseguir diferentes grupos de clientes que presenten características similares. Esto implicaría la consecución de un precio con mayor personalización dirigida al cliente y mayores niveles de eficiencia para la aseguradora.

Con este trabajo de investigación se ha conseguido demostrar que las relaciones entre los riesgos de un seguro de automóvil no son independientes, por lo tanto, a la hora de tarificar se ha de descartar la hipótesis de que existe independencia entre los mismo. Esto es un gran inconveniente para los modelos GLM que tradicionalmente se utilizan en las aseguradoras para obtener el precio de la prima pura.

## 5. Futuras Líneas de investigación

A raíz de lo estudiado a lo largo de este trabajo se pueden plantear unas posibles futuras líneas de investigación. Una de ellas podría ser el cambio en el modo actual de tarificación. Es decir, actualmente se construyen modelos de tarificación basados en GLM's para los cuales su principal premisa es la independencia entre sus variables. Se debería tomar en cuenta el grado de dependencia entre los riesgos que asumen las aseguradoras para poder cubrir las garantías que los asegurados contratan. La aplicación de las funciones cópulas en este campo podría ser de gran ayuda ya que evitaría la pérdida de información.

Por otra parte, y para una mejor eficiencia e imagen de cara al cliente, se debería estudiar el modo de contar con un "ranking de clientes" el cuál refleje el clúster de los mismos, ya que con ello se lograría conseguir un precio más personalizado a cada uno de ellos lo que ayudaría a controlar el fraude y podrían obtenerse distintos patrones de peligrosidad entre los mismos.

## 6. Referencias

Brechmann, E. y Schepsmeier, U. (2013). Cdvine: Modeling dependence with c-and d-vine copulas in r. *Journal of Statistical Software*, 52(3), 1-

27.

Galton, F. (1885). International Health Exhibition. mimeo.

Illanes, G. (2013). Cópulas paramétricas y no paramétricas. *Obtenido de* [https://www.fing.edu.uy/sites/default/files/biblio/9476/c%20c3%20b3pula%20s%20param%20c3%20a9tricas%20y%20no%20param%20c3%20a9tricas\\_2013.pdf](https://www.fing.edu.uy/sites/default/files/biblio/9476/c%20c3%20b3pula%20s%20param%20c3%20a9tricas%20y%20no%20param%20c3%20a9tricas_2013.pdf)

Li, F., y Kang, Y. (2018). Improving forecasting performance using covariate-dependent copula models. *International Journal of Forecasting*, 34(3), 456-476.

Liu, Z., Zhou, P., Chen, X., y Guan, Y. (2015). A multivariate conditional model for streamflow prediction and spatial precipitation refinement. *Journal of Geophysical Research: Atmospheres*, 120(19), 10-116.

Marius Hofert, I. K. (19 de 05 de 2020). <https://cran.r-project.org/web/packages/copula/copula.pdf>. *Obtenido de* <https://cran.r-project.org/web/packages/copula/copula.pdf>

Scrucca, L. (2004). qcc: an R package for quality control charting and statistical process control. *dim (pistonrings)*, 1(200), 3.

Anexo

El siguiente cuadro muestra la evolución histórica del análisis de trabajos relevantes (desde mi punto de vista) relacionados con la teoría de cópulas.

Año	Publicación	Autor/Autores
1904	"The proof and measurement of association between two things"	Spearman, C.
1938	"A new measure of rank correlation"	Kendall, M.G.
1940	"Masstabinvariante Korrelationstheorie"	Hoeffding, W.
1950	"On a measure of dependence between tow random variables"	Blomqvist, N.
1951	"Sur les tableaux de corrélation dont les marges sont données"	Fréchet, M.
1956	"Sur les tableaux de corrélation dont les marges sont données"	Féron, R.
1957	"Les tableaux de correlation et les programme linéaires"	Fréchet, M.
1959	"On measures of dependence"	Rényi, A.
1959	"Fonctions de répartition á n dimensions et leurs marges"	Sklar, A.
1960	"The performance of some correlation coe_cients for a general bivariate distribution."	Farlie, D.J.G.
1960	"Bivariate exponential distributions"	Gumbel, E.J.
1966	"Some concepts of dependence"	Lehmann, E.L.
1967	"A multivariate exponential distribution"	Marshall, A.W., Olkin, I.
1972	"Fréchet classes and compatibility of distribution functions"	Dall'Aglio, C.
1981	"On nonparametric measures of dependence for random variables"	Schweizer, B., Woff, E.F.
1991	"A new approach to dependence in multivariate distributions"	Kotz, S., Seeger, J.P.
1991	"Probabilistic interpretations of copulas and their convex sums"	Mikusinski, P., Sherwood, H., Taylor, M.D.
1994	"Some advances in the study of the compatibility of three bivariate copulas"	Quesada-Molina, J.J., Rodríguez-Lallena, J.A.
1995	"Copulas, Characterization, Correlation, and Counterexamples"	Nelsen, R.B.
1996	"Multivariate models and dependence concepts"	Joe, H.
1996	"Copulas, Marginals, and Joint Distributions"	Marshall, A.W.
1996	"Random Variables, Distribution Functions, and Copulas - A Personal Look Backward and Forward"	Sklar, A.
2004	"Multidimensional dependency measures"	Fernández, B., González-Barrios, J.M.
2005	"Actuarial Theory for Dependent Risks"	Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R.
2010	"Application of copulas as a new geostatistical tool"	Jing, L.
2015	Modeling Dependence in High Dimensions With Factor Copulas	Dong Hwan Oh , Andrew J. Patton
2018	Single-index copulas	Jean-David Fermanian, J.D., Lopez, D.