

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
ESCUELA DE POSGRADO



PONTIFICIA
**UNIVERSIDAD
CATÓLICA**
DEL PERÚ

**“OPTIMIZACIÓN DE PORTAFOLIOS DE INVERSIÓN A TRAVÉS DEL
VALOR EN RIESGO CONDICIONAL (CVAR) UTILIZANDO CÓPULAS EN
PARES”**

Tesis para optar el grado de Magíster en Matemáticas Aplicadas

AUTOR

Pablo Isaac Navarrete Alvarez

ASESORA

Loretta Betzabe Rosa Gasco Campos

JURADO

Abelardo Jordán Liza

Luis Carlos Chávez-Bedoya Mercado

LIMA – PERÚ

2012

Resumen

OPTIMIZACIÓN DE PORTAFOLIOS DE INVERSIÓN A TRAVÉS DEL VALOR EN RIESGO CONDICIONAL (CVAR) UTILIZANDO CÓPULAS EN PARES

Pablo Isaac Navarrete Alvarez

En la presente tesis se demuestran de manera exhaustiva las principales propiedades del CVaR presentadas en los trabajos de Rockafellar y Uryasev (2000, 2002). En particular, se completan las demostraciones del teorema a través del cual se puede minimizar al CVaR utilizando la función auxiliar F_α . Estos resultados se mantienen cuando la función de distribución de pérdidas presenta discontinuidades e incluso saltos. Además, se demuestra que el CVaR es continuo con respecto al nivel de confianza elegido y se demuestra que es una medida de riesgo coherente.

Por otro lado, se realiza la optimización de un portafolio de inversión utilizando al CVaR como medida de riesgo. Dado que la evidencia estadística muestra que los activos no siguen un comportamiento gaussiano, se utiliza la teoría de cópulas para modelar la dependencia contemporánea de los datos. Finalmente, se comparan los resultados obtenidos de la optimización del modelo media-varianza de Markowitz (M-V) frente a los obtenidos en el modelo media-CVaR (M-CVaR).

Agradecimientos

Deseo agradecer profundamente a mi asesora, la Dra. Loretta Gasco, por su enorme paciencia, comprensión y pertinentes sugerencias que llevaron a enriquecer y completar la tesis. También deseo agradecer al resto del jurado por sus minuciosas revisiones y sugerencias prácticas, las cuales también enriquecieron esta tesis.

Por otro lado, deseo dedicar este trabajo al amor de mi vida, mi esposa Joko, con cuya paciencia y comprensión han hecho posible la culminación de este trabajo.

Índice general

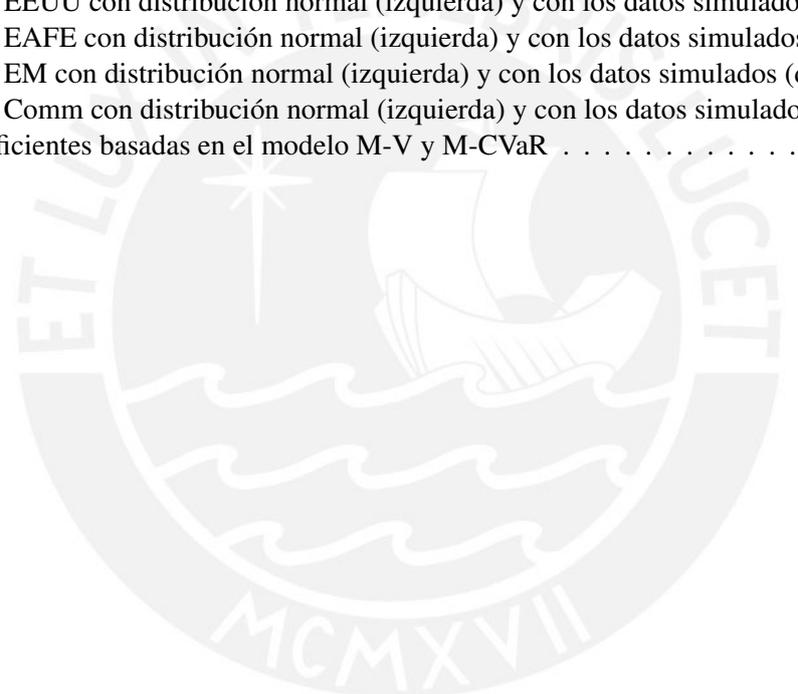
1	Introducción	6
2	Medidas de riesgo y optimización del CVaR	8
2.1	Introducción	8
2.2	Medidas de riesgo	8
2.2.1	Medidas de riesgo basadas en la teoría de portafolio de Markowitz	8
2.2.2	Valor en riesgo (VaR)	9
2.2.3	Medidas de riesgo coherentes	9
2.3	Optimización del CVaR	10
2.3.1	Optimización del CVaR	10
2.3.2	Optimización del CVaR para distribuciones de pérdidas más generales	16
3	Cóputas	27
3.1	Conceptos básicos	27
3.2	Cóputas	28
3.3	El teorema de Sklar	28
3.3.1	Cóputas de sobrevivencia	29
3.4	Medidas de dependencia	30
3.4.1	Tau de Kendall	31
3.4.2	Rho de Spearman	31
3.4.3	Coeficiente de dependencia de colas	31
3.5	Cóputas multivariadas	32
3.5.1	Cóputas elípticas	32
3.5.2	Cóputas arquimedianas	33
3.6	Cóputas bivariadas	33
3.6.1	Familia de cóputas bivariadas elípticas	33
3.6.2	Familias de cóputas bivariadas arquimedianas	34
3.7	Cóputas en pares, vine regular y simulación	45
3.7.1	Cóputas en pares	46
3.7.2	Vines regular	47
3.7.3	Simulación	49
3.8	Herramientas para el análisis exploratorio de los datos	50
3.8.1	Test de independencia	50
3.8.2	Chi plot	50
3.8.3	Kendall plot (K-plot)	52
3.8.4	λ - función	52
3.9	Elección de la vine regular	52
3.10	Estimación de parámetros de las cóputas	54
3.11	Selección del modelo	55
3.11.1	AIC	56
3.11.2	BIC	56

3.11.3	El test de Vuong	56
3.11.4	El test de Clarke	57
3.12	Test de bondad de ajuste	57
3.12.1	Análisis gráfico	57
3.12.2	Análisis estadístico	57
4	Optimización del modelo M-CVaR con retornos no gaussianos	59
4.1	Introducción	59
4.2	Generación y simulación de escenarios	59
4.2.1	Generación de escenarios	59
4.2.2	Simulación de escenarios	60
4.3	Modelos de series de tiempo	61
4.3.1	Modelo ARMA-GARCH	61
4.3.2	Extensiones al modelo GARCH	62
4.3.3	El test de Ljung Box	63
4.4	Optimización del modelo M-V de Markowitz	63
4.5	Tratamiento de los datos	65
4.5.1	Datos	65
4.5.2	Hechos estilizados de los retornos de los activos financieros	65
4.5.3	Elección del modelo de serie de tiempo	73
4.5.4	Elección de la vine cópula	80
4.5.5	Estimación de parámetros de las cópulas en pares de los retornos filtrados	80
4.5.6	Test de bondad de ajuste de las cópulas	86
4.5.7	Test de bondad de ajuste de los retornos simulados de los activos	88
4.6	Optimización del modelo M-CVaR	93
4.6.1	Frontera eficiente	93
4.6.2	Portafolio óptimo	94
5	Conclusiones	95

Índice de figuras

2.1	La ecuación $\Psi(\mathbf{x}, \zeta) = \alpha$ no tiene solución en ζ	18
2.2	La ecuación $\Psi(\mathbf{x}, \zeta) = \alpha$ tiene infinitas soluciones en ζ	18
3.1	Gráficos de dispersión del rango de una cópula bivariada gaussiana. (a): $\rho = -0.7$; (b): $\rho = 0.7$	34
3.2	Gráficos de dispersión del rango de una t-cópula bivariada. (a) $\rho = -0.7, v = 3$; (b) $\rho = 0.7, v = 3$; (c) $\rho = -0.7, v = 10$; (d) $\rho = 0.7, v = 10$	35
3.3	Gráficos de dispersión del rango de una cópula Clayton bivariada. (a): $\theta = 2$; (b): $\theta = 6$	37
3.4	Gráficos de dispersión del rango de una cópula Gumbel bivariada. (a): $\theta = 2$; (b): $\theta = 6$	38
3.5	Gráficos de dispersión del rango de una cópula Frank bivariada. (a): $\theta = 4$; (b): $\theta = 8$	39
3.6	Gráficos de dispersión del rango de una cópula Joe bivariada. (a): $\theta = 4$; (b): $\theta = 8$	40
3.7	Gráficos de dispersión del rango de una cópula Clayton-Gumbel (BB1). (a) $\theta = 0.5, \delta = 3$; (b) $\theta = 2, \delta = 3$; (c) $\theta = 0.5, \delta = 6$; (d) $\theta = 2, \delta = 6$	41
3.8	Gráficos de dispersión del rango de una cópula Joe-Gumbel (BB6). (a) $\theta = 0.5, \delta = 3$; (b) $\theta = 2, \delta = 3$; (c) $\theta = 0.5, \delta = 6$; (d) $\theta = 2, \delta = 6$	42
3.9	Gráficos de dispersión del rango de una cópula Joe-Clayton (BB7). (a) $\theta = 0.5, \delta = 3$; (b) $\theta = 2, \delta = 3$; (c) $\theta = 0.5, \delta = 6$; (d) $\theta = 2, \delta = 6$	44
3.10	Gráficos de dispersión del rango de una cópula Joe-Frank (BB8). (a) $\theta = 3, \delta = 0.7$; (b) $\theta = 6, \delta = 0.7$; (c) $\theta = 3, \delta = 1$; (d) $\theta = 6, \delta = 1$	45
3.11	Gráficos de dispersión del rango de una cópula Joe rotada 0, 90, 180 y 270 grados. Parámetros correspondientes a los valores del tau de Kendall de 0.8 para la dependencia positiva y -0.8 para la dependencia negativa	46
3.12	Ejemplo de árboles C-vine (izquierda) y D-vine (derecha) de dimensión 4	49
3.13	χ -plot basados en muestras aleatorias de tamaño $n = 100$ de un vector normal bivariado. 1 (a) $\rho = 0.05$; 1 (b) χ -plot. 2 (a) $\rho = 0.50$; 2 (b) χ -plot. 3 (a) $\rho = 0.95$; 3 (b) χ -plot	51
3.14	K-plot basados en muestras aleatorias de tamaño $n = 100$ de un vector normal bivariado. 1 (a) $\rho = 0.05$; 1 (b) k-plot. 2 (a) $\rho = 0.50$; 2 (b) k-plot. 3 (a) $\rho = 0.95$; 3 (b) k-plot	53
3.15	(a) λ - función empírica de una cópula Joe simulada. (b) λ - función teórica de una cópula Joe simulada y la λ - función empírica	53
4.1	Proceso para la simulación de escenarios	61
4.2	Histograma de retornos mensuales y distribución normal de USIG, EMIG, EMHY, EEUU, EA-FE y EM	66
4.3	Autocorrelograma de los retornos	69
4.4	Autocorrelograma de los retornos al cuadrado	70
4.5	Los cuatro primeros momentos de los retornos diarios del S&P500 (EE.UU.) varían en el tiempo usando una muestra móvil de 6 meses	71
4.6	Correlaciones de retornos diarios del S&P500 (EE.UU.) y otros activos usando una ventana de 6 meses	72
4.7	Sobre la diagonal principal se muestra el scatter plot de cada par de series mientras que debajo de la diagonal principal se muestran los p values de los test de independencia de cada par de variables	80

4.8	Debajo de la diagonal principal se encuentran los gráficos de contornos; mientras que encima de la diagonal principal se encuentran los gráficos de dispersión de las nueve variables	81
4.9	Sobre la diagonal principal se muestra el scatter plot de cada par de series mientras que debajo de la diagonal principal se muestran los Chi plots para cada par de variables	82
4.10	Sobre la diagonal principal se muestra el scatter plot de cada par de series mientras que debajo de la diagonal principal se muestran los K plots para cada par de variables	83
4.11	λ -función empírica de la cópula en par USIG-EMIG y de las teóricas Gaussiana, t, clayton, Gumbel, Frank, Joe, BB1 y BB6	84
4.12	λ -función de las cópula teóricas BB7, BB8 y las rotadas 180° Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7 y BB8	85
4.13	En la primera fila se observan los gráficos de dispersión de los retornos, de los datos en escala de cópulas y el gráfico de contornos. En la fila inferior, el K-plot, Chi-plot y la función lambda empírica y la teórica Gumbel rotada 180°	87
4.14	QQ-plot de USIG con distribución normal (izquierda) y con los datos simulados (derecha) . . .	89
4.15	QQ-plot de EMIG con distribución normal (izquierda) y con los datos simulados (derecha) . . .	89
4.16	QQ-plot de EMHY con distribución normal (izquierda) y con los datos simulados (derecha) . .	90
4.17	QQ-plot de EEUU con distribución normal (izquierda) y con los datos simulados (derecha) . .	90
4.18	QQ-plot de EAFE con distribución normal (izquierda) y con los datos simulados (derecha) . . .	91
4.19	QQ-plot de EM con distribución normal (izquierda) y con los datos simulados (derecha)	91
4.20	QQ-plot de Comm con distribución normal (izquierda) y con los datos simulados (derecha) . .	92
4.21	Fronteras eficientes basadas en el modelo M-V y M-CVaR	93



Índice de cuadros

4.1	Índices utilizados para el proceso de optimización	65
4.2	Estadísticas descriptivas de los índices	67
4.3	P-values de las pruebas de normalidad de Jarque Bera	67
4.4	P-values de las pruebas de autocorrelación de los retornos de Ljung Box	68
4.5	P-values de las pruebas de autocorrelación de los retornos al cuadrado de Ljung Box	69
4.6	Matriz de correlación y p-value de los retornos	70
4.7	Matriz de correlación y p-value de los retornos en el 2008	72
4.8	Coefficientes estimados, p-values y grados de libertad del índice USIG	73
4.9	Valor de la función de verosimilitud, AIC y BIC	74
4.10	P-values de las pruebas de autocorrelación de Ljung Box de los residuos estandarizados	74
4.11	Coefficientes estimados, p-values y grados de libertad del índice EMIG	75
4.12	Coefficientes estimados, p-values y grados de libertad del índice EMHY	75
4.13	Coefficientes estimados, p-values y grados de libertad del índice EE.UU.	76
4.14	Coefficientes estimados, p-values y grados de libertad del índice EAFE	76
4.15	Coefficientes estimados, p-values y grados de libertad del índice EM	77
4.16	Coefficientes estimados, p-values y grados de libertad del índice Comm	77
4.17	Valor de la función de verosimilitud, AIC y BIC de los índices	78
4.18	Modelos seleccionados	79
4.19	Resumen de los p-values de las pruebas de autocorrelación de los residuos estandarizados	79
4.20	Resumen de los p-values de las pruebas de autocorrelación de los residuos estandarizados al cuadrado	79
4.21	Criterios de selección para la cópula bivariada USIG-EMIG	86
4.22	Valor de la función de verosimilitud, AIC y BIC de la estructura de la vine	88
4.23	P-value de las pruebas de Clarke y Vuong de la vine	88
4.24	P-value de las pruebas de K-S y A-D	92
4.25	Valores de las fronteras eficientes de los modelos M-V y M-CVaR	93
4.26	Asignación de activos de los modelos M-V y M-CVaR para un mismo nivel de retorno	94

Capítulo 1

Introducción

La optimización de los portafolios de inversión es una de las herramientas más importantes para la toma de decisiones de inversión. Dentro de este proceso, la medida de riesgo a utilizar es un punto crítico pues nos permitirá realizar una adecuada asignación de los activos y de los riesgos dentro del portafolio elegido.

Debido a los trabajos de H. Markowitz (1952), la medida de riesgo tradicionalmente usada es la varianza. Además, su popularidad se extendió debido a que es fácil de entenderla y calcularla. Sin embargo, una de las críticas de esta medida de riesgo es su simetría, pues penaliza de igual forma a las grandes ganancias así como las grandes pérdidas en el proceso de optimización. En general, esto es inconsistente con las preferencias de los inversionistas (*behavioral finance*), los cuales tienden a evitar pérdidas en lugar de hacer ganancias (preferencias asimétricas).

Por otro lado, existe amplia evidencia empírica que muestra que los retornos de los activos no tienen un comportamiento gaussiano. Esto hace más relevante la utilización de una medida de riesgo alternativa que incorpore este hecho en el proceso de optimización. El uso de modelos gaussianos, cuando los retornos de los activos no son normales, puede ser peligroso pues tal práctica podría generar una equivocada elección del portafolio, debido a la subestimación de pérdidas extremas.

El valor en riesgo condicional (CVaR, por sus siglas en inglés) es una medida de riesgo alternativa que supera varios de los inconvenientes de la varianza pues captura la asimetría de las preferencias al tomar en cuenta la incidencia de la asimetría y la curtosis inducida por las distribuciones no gaussianas de los activos. Además, el CVaR es una medida de riesgo coherente en el sentido de Artzner, Delbaen, Eber y Heath (1999), tiene buenas propiedades matemáticas para los problemas de optimización, como su convexidad, y tiene como una de sus características que en el proceso de optimización se puede obtener el valor en riesgo (VaR, por sus siglas en inglés), medida de riesgo ampliamente utilizada por inversores y establecida explícitamente por los reguladores como la principal herramienta cuantitativa para calcular sus requerimientos de capital. No obstante, a pesar de su uso extendido, el VaR no está exento de críticas. La principal de ellas es su falta de coherencia, en particular, no cumple el axioma de subaditividad, lo cual significa que la diversificación de los instrumentos que conforman un portafolio no necesariamente resulta en una reducción del riesgo.

De esta manera, teniendo en cuenta lo relevante del CVaR, así como el comportamiento no gaussiano de los precios de los activos, los objetivos de la presente tesis son los siguientes:

1. Enunciar y demostrar de manera exhaustiva las principales propiedades de la medida de riesgo coherente CVaR.
2. Realizar la optimización de un portafolio de inversión utilizando al CVaR como medida de riesgo y tomando en cuenta los hechos estilizados del comportamiento de los distintos tipos de activos. Para capturar los hechos estilizados se utilizará la teoría de cópulas.

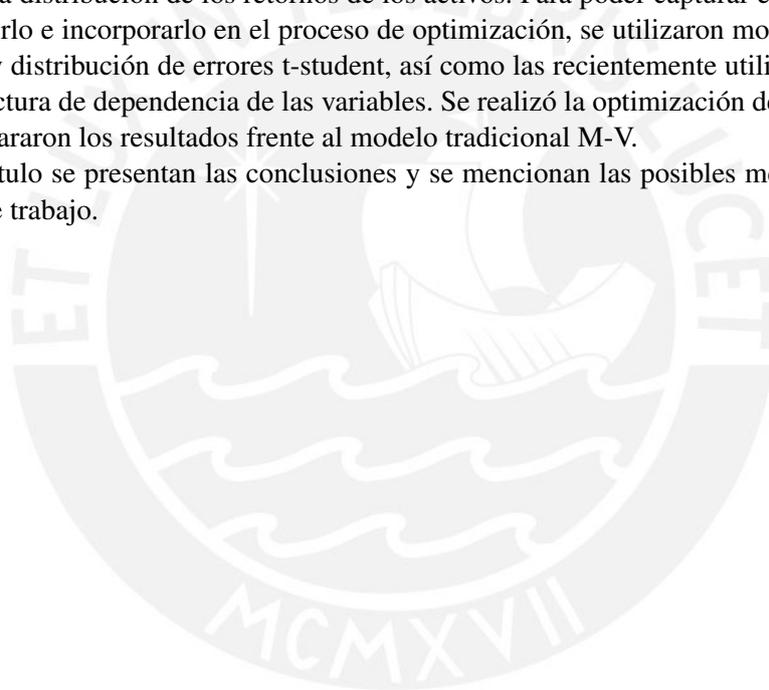
En ese sentido, la presente tesis se encuentra dividida en cinco capítulos. En el segundo capítulo se presenta una breve introducción de las medidas de riesgo, dentro de la cual se muestra los cuatro axiomas de las medidas de riesgo coherente. Luego, se demuestran en forma exhaustiva las propiedades del CVaR presentadas en los trabajos seminales de Rockafellar y Uryasev (2000, 2002). Aquí es donde se encuentra uno de los aportes de la

tesis, pues si bien es cierto en los trabajos de Rockafellar y Uryasev se presentan las principales propiedades del CVaR y su relación con el VaR, las demostraciones no se presentan de manera exhaustiva y completa. La tesis trata de completar las demostraciones de las propiedades del CVaR, y de esa forma ser autocontenida, cuando las distribuciones de pérdidas se comportan de forma continua e incluso cuando éstas presentan discontinuidades o posibles saltos.

En el tercer capítulo se presenta la teoría de cópulas. Esta teoría ha cobrado reciente importancia debido a la flexibilidad para modelar datos multivariados. El teorema central en esta teoría es el de Sklar, el cual permite relacionar a las marginales univariadas con su función de distribución multivariada. Luego se presentarán las cópulas más utilizadas en las aplicaciones financieras, así como las herramientas exploratorias de los datos, la estimación de los parámetros de las cópulas y test de bondad de ajuste.

En el cuarto capítulo, se realiza un ejercicio de optimización de un portafolio de inversión tomando en cuenta los hechos estilizados del comportamiento de los precios de los activos. Previamente, se presenta la teoría de series de tiempo y el modelo de media varianza de Markowitz (M-V). En línea con otros trabajos empíricos, se encuentra que los retornos de los activos presentan exceso de curtosis, sesgo negativo, correlación serial en los retornos así como en las volatilidades y no están distribuidos como una normal. Estos resultados subestimarían el riesgo al utilizar el modelo M-V pues éste asume normalidad al considerar solo los dos primeros momentos de la distribución de los retornos de los activos. Para poder capturar este comportamiento no gaussiano, reproducirlo e incorporarlo en el proceso de optimización, se utilizaron modelos ARMA-GARCH con apalancamiento y distribución de errores t-student, así como las recientemente utilizadas cópulas en pares para modelar la estructura de dependencia de las variables. Se realizó la optimización del modelo media CVaR (M-CVaR) y se compararon los resultados frente al modelo tradicional M-V.

En el último capítulo se presentan las conclusiones y se mencionan las posibles mejoras metodológicas y los futuros campos de trabajo.



Capítulo 2

Medidas de riesgo y optimización del CVaR

2.1 Introducción

En este capítulo se presentarán los axiomas de las medidas de riesgo coherentes y se centrará la discusión en la optimización del CVaR. En ese sentido, en la tercera sección se mostrarán de forma exhaustiva las demostraciones de los teoremas desarrollados en Rockafellar y Uryasev (2000, 2002). En Rockafellar y Uryasev (2000), los autores muestran una nueva técnica para optimizar al CVaR y en forma simultánea calcular el VaR mediante la caracterización de ambas a través de una función F_α . Posteriormente, en Rockafellar y Uryasev (2002), los autores amplían el trabajo previo cuando la función de distribución acumulada de pérdidas Ψ tiene un comportamiento discreto y presenta saltos. Se verá que relajando este supuesto los resultados de su trabajo previo se mantienen. En este último trabajo, los autores demuestran, de acuerdo a su definición de coherencia, que el CVaR es una medida de riesgo coherente y que es continua con respecto al nivel de confianza α , lo cual no se observa en el caso del VaR.

2.2 Medidas de riesgo

Los eventos de crisis han demostrado la necesidad de adecuadas medidas de riesgo. Inadecuadas medidas de riesgo pueden subestimar particulares tipos de riesgos como los de mercado, créditos, etc. y conducir a equivocadas asignaciones de activos. Aunque no se desea ofrecer una completa descripción de la evolución de las medidas de riesgo, desde un punto de vista histórico, se puede identificar ciertos períodos en los cuales éstas han evolucionado:

1. Medidas de riesgo previos a los aportes de H. Markowitz.
2. Medidas de riesgo basadas en la Teoría de Portafolio de Markowitz.
3. Valor en Riesgo.
4. Medidas de Riesgo Coherentes.

2.2.1 Medidas de riesgo basadas en la teoría de portafolio de Markowitz

Markowitz fue el primero en formalizar el riesgo de un portafolio, la diversificación y la selección de activos en un marco teórico matemáticamente consistente. Eligió a la desviación estándar como una medida de riesgo. Sin embargo, la desviación estándar no parece ser una buena elección como medida de riesgo, debido a que ella penaliza de manera simétrica las desviaciones positivas y negativas de la media. Las deficiencias de la desviación estándar como una medida de riesgo fueron reconocidas por Markowitz, quien fue el primero que sugirió al desvío semiestándar como un sustituto (Markowitz, 1959). Debido a esto, el banco de inversión JPMorgan propuso la utilización de una medida de riesgo alternativa llamada valor en riesgo (VaR).

2.2.2 Valor en riesgo (VaR)

En 1994 se publicó el documento técnico Riskmetrics por parte del banco de inversión JP Morgan donde se presentó por primera vez al VaR. Esto implicó un cambio en la forma cómo se medían los riesgos debido a las siguientes razones:

1. El banco de inversión JP Morgan creó al VaR como una medida de riesgo de toda la institución bajo una sola medida de riesgo.
2. Las medidas de riesgo previas se centraban en explicar el retorno de un activo basado en algún modelo teórico de riesgo retorno como el CAPM. Por otro lado, el VaR cambió el centro de atención al centrarse en el riesgo en sí mismo y en términos de pérdidas en lugar de las ganancias esperadas.
3. El Comité de Supervisión Bancaria de Basilea, el cual estandariza la regulación y la práctica bancaria a nivel internacional, estipuló el requerimiento de capital por riesgo de mercado sobre la base del VaR en 1995. Este factor influyó en la popularidad del VaR como medida de riesgo.

2.2.3 Medidas de riesgo coherentes

Un hito significativo en la evolución de las medidas de riesgo fue alcanzado cuando Artzner, P. Delbaen, F. Eber, J. Heath, D (1999) propusieron los primeros axiomas de las medidas de riesgo, de tal forma que si éstas obedecieran estos axiomas serían llamadas medidas de riesgo coherentes. Esto implica que no es posible establecer una función arbitraria como medida de riesgo.

Axiomas de las medidas de riesgo coherente

Los autores establecieron una definición de riesgo, presentaron y justificaron un marco teórico para la construcción de una medida de riesgo. Según los autores, dado que el riesgo está relacionado con la variabilidad del valor futuro de una posición, es mejor considerarla solo como valores futuros. Por ello, el objeto básico de sus estudios serán variables aleatorias sobre el conjunto de los estados de la naturaleza en una fecha futura, los cuales son interpretados como los posibles valores futuros de un portafolio.

Siguiendo a Artzner, P. Delbaen, F. Eber, J. Heath, D (1999), se define el concepto de medida de riesgo coherente, de la siguiente manera.

Definición 1. Sea Ω el conjunto de los estados de la naturaleza, asumimos que es finito. Considerando que Ω es el conjunto de resultados de un experimento, podemos calcular el valor neto final de una posición para cada elemento de Ω . Ésta es una variable aleatoria denotada por X .

Definición 2. Sea \mathcal{F} el conjunto de todos los riesgos, esto es, el conjunto de todas las funciones de valor real definidas sobre Ω . Dado que Ω es asumido finito y no vacío, \mathcal{F} puede ser identificada con \mathbb{R}^n , donde $n = \text{card}(\Omega)$.

Definición 3 (Medida de riesgo). Un medida de riesgo es un mapeo de \mathcal{F} a \mathbb{R} .

Definición 4 (Medida de riesgo coherente). Por lo general, un mapeo $\rho : \mathcal{F} \rightarrow \mathbb{R}$ será llamado medida de riesgo coherente si satisface los siguientes axiomas:

Axioma 5 (M). Monotonicidad: para todo X e $Y \in \mathcal{F}$ con $X \leq Y$, tenemos $\rho(X) \leq \rho(Y)$.

El axioma de monotonicidad implica que si tengo una posición más riesgosa, esto implicará un mayor valor en la medida de riesgo.

Axioma 6 (HP). Homogeneidad positiva: para todo $\lambda \geq 0$ y para todo $X \in \mathcal{F}$, $\rho(\lambda X) = \lambda \rho(X)$.

El axioma de homogeneidad asegura que aumentar (disminuir) el tamaño de la posición de un portafolio aumenta (disminuye) el riesgo de esa posición.

Axioma 7 (S). Subaditividad: para todo X_1 y $X_2 \in \mathcal{F}$, $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$.

Si la subaditividad no se mantiene, entonces en algunos casos se tiene $\rho(X_1 + X_2) > \rho(X_1) + \rho(X_2)$. Esto colisiona con el concepto de diversificación.

Axioma 8 (T). Invarianza por traslación: para todo $X \in \mathcal{F}$ y todo número real α , tenemos $\rho(X + \alpha) = \rho(X) - \alpha$.

Esto implica que al añadir (sustraer) una posición segura α (sin riesgo), simplemente reduce (aumenta) el riesgo en ese monto.

2.3 Optimización del CVaR

La introducción del concepto de medida de riesgo coherente representó un avance en el desarrollo de las medidas de riesgo. En esa línea, debido a sus buenas propiedades para los problemas de optimización, fácil implementación y su relación con el VaR, el CVaR ha ganado popularidad como una de las medidas de riesgo coherentes más analizadas en el mundo académico y recientemente utilizada en problemas aplicados¹.

De esta manera, teniendo en cuenta lo relevante del CVaR, uno de los objetivos de esta tesis es demostrar de manera exhaustiva las principales propiedades de esta medida de riesgo coherente presentadas en los trabajos seminales de Rockafellar y Uryasev (2000, 2002). Aquí es donde se encuentra uno de los aportes de la tesis pues, si bien es cierto en estos trabajos se presentan las principales propiedades del CVaR y su relación con el VaR, las demostraciones no se presentan de manera exhaustiva ni completa. La tesis trata de completar las demostraciones de las propiedades del CVaR.

2.3.1 Optimización del CVaR

La contribución de Rockafellar y Uryasev (2000) fue la de establecer una técnica práctica para minimizar el CVaR y al mismo tiempo calcular el VaR de un portafolio. Para ello, los autores caracterizaron ambas funciones en términos de la función F_α , cuyos detalles se verán más adelante, y bajo ciertos supuestos, como el de la continuidad en la función de distribución acumulada de pérdidas, presentaron los dos teoremas que establecieron que minimizar la función F_α con respecto a ζ daba como resultado el CVaR, lo cual a su vez implica que minimizar F_α con respecto a todos sus argumentos equivalía a minimizar al CVaR. En ese proceso de optimización se puede obtener el VaR del portafolio optimizado. No obstante, antes de presentar los teoremas con sus respectivas demostraciones, se definirán conceptos previos.

Sea \mathbf{x} el vector decisión escogido de un cierto subconjunto $X \subset \mathbb{R}^n$ y sea \mathbf{y} con valores en \mathbb{R}^m un vector aleatorio. El vector \mathbf{x} puede ser interpretado como un portafolio y X como el conjunto de los portafolios disponibles. Por otro lado, \mathbf{y} representa la incertidumbre que puede afectar al portafolio. De esta manera, se dan las siguientes definiciones.

Definición 9. Sea $f : X \times \mathbb{R}^m \rightarrow \mathbb{R}$ la pérdida asociada al vector decisión \mathbf{x} . Para cada \mathbf{x} , la pérdida $f(\mathbf{x}, \cdot) : \mathbb{R}^m \rightarrow \mathbb{R}$ es una variable aleatoria con distribución en \mathbb{R} inducida por \mathbf{y} .

La distribución de probabilidad subyacente de \mathbf{y} en \mathbb{R}^m se asume que tiene densidad, la cual será denotada por $p(\mathbf{y})$ ².

Definición 10 (Distribución acumulada de f). La probabilidad de que $f(\mathbf{x}, \mathbf{y})$ no exceda un cierto umbral ζ está dada por:

$$\Psi(\mathbf{x}, \zeta) = \int_{\{f(\mathbf{x}, \mathbf{y}) \leq \zeta\}} p(\mathbf{y}) d\mathbf{y}. \quad (2.1)$$

¹ Si desea ver un resumen de distintos tipos de medidas de riesgo, ver Chen, Liu y Wang (2004) y Lleo (2009).

² Sin embargo, se mostrará que una expresión analítica $p(\mathbf{y})$ no será necesaria para la implementación de esta técnica.

Como función de ζ para un \mathbf{x} fijo, $\Psi(\mathbf{x}, \zeta)$ es la función de distribución acumulada de las pérdidas asociadas a \mathbf{x} . Por lo tanto, $\Psi(\mathbf{x}, \zeta)$ determina el comportamiento de esta variable aleatoria. Además, $\Psi(\mathbf{x}, \zeta)$, por ser una función de distribución acumulada, es no decreciente con respecto a ζ y continua por la derecha, pero no necesariamente continua por izquierda debido a la posibilidad de saltos. No obstante, se asume que $\Psi(\mathbf{x}, \zeta)$ es continua con respecto a ζ^3 .

Ahora, definimos al VaR y al CVaR.

Definición 11 (VaR y CVaR). El VaR y CVaR para la variable aleatoria asociada con \mathbf{x} y para un determinado nivel de probabilidad $\alpha \in (0, 1)$ serán denotados por $\zeta_\alpha(\mathbf{x})$ y $\phi_\alpha(\mathbf{x})$, respectivamente y definidos por:

$$\zeta_\alpha(\mathbf{x}) = \min \{ \zeta \in \mathbb{R} : \Psi(\mathbf{x}, \zeta) \geq \alpha \} \quad (2.2)$$

$$\phi_\alpha(\mathbf{x}) = (1 - \alpha)^{-1} \int_{\{f(\mathbf{x}, \mathbf{y}) \geq \zeta_\alpha(\mathbf{x})\}} f(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) d\mathbf{y}. \quad (2.3)$$

En la primera fórmula, $\zeta_\alpha(\mathbf{x})$ es el extremo izquierdo de un intervalo no vacío de valores de ζ tales que $\Psi(\mathbf{x}, \zeta) = \alpha$. Esto se sigue del hecho que $\Psi(\mathbf{x}, \zeta)$ es continuo y no decreciente con respecto a ζ . No obstante, el intervalo puede contener más de un solo punto dependiendo del comportamiento de Ψ . En resumen, con respecto a un nivel de probabilidad α , el VaR de un portafolio es el menor valor ζ tal que, con probabilidad α , las pérdidas no excederán ζ ; mientras que el CVaR es el esperado condicional de las pérdidas sobre ese monto $\zeta = \zeta_\alpha(\mathbf{x})$.

Como se mostrará en la siguiente sección, el CVaR es una medida de riesgo coherente, cuya forma funcional está determinada por (2.3). Sin embargo, con el objeto de determinar un \mathbf{x} que minimice el CVaR, Rockafellar y Uryasev (2000) mostraron que no es necesario trabajar directamente con la función $\phi_\alpha(\mathbf{x})$, lo cual puede ser difícil debido a que en su formulación aparece el VaR, el cual no es adecuado para problemas de optimización, como su falta de convexidad. Por ello, los autores proponen una técnica práctica para minimizar el CVaR y al mismo tiempo calcular el VaR de un portafolio, a través de la siguiente caracterización de $\phi_\alpha(\mathbf{x})$ y $\zeta_\alpha(\mathbf{x})$ en términos de la función F_α sobre $X \times \mathbb{R}$ definida como sigue:

$$F_\alpha(\mathbf{x}, \zeta) = \zeta + (1 - \alpha)^{-1} \int_{\mathbf{y} \in \mathbb{R}^m} [f(\mathbf{x}, \mathbf{y}) - \zeta]^+ p(\mathbf{y}) d\mathbf{y}, \quad (2.4)$$

donde $[t]^+ = \max\{t, 0\}$.

La forma de esta última expresión la podemos establecer de la siguiente manera partiendo de la definición del CVaR.

$$\begin{aligned} \phi_\alpha(\mathbf{x}) &= (1 - \alpha)^{-1} \int_{\{f(\mathbf{x}, \mathbf{y}) \geq \zeta_\alpha(\mathbf{x})\}} f(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) d\mathbf{y} \\ &= (1 - \alpha)^{-1} \int_{\{f(\mathbf{x}, \mathbf{y}) \geq \zeta_\alpha(\mathbf{x})\}} [f(\mathbf{x}, \mathbf{y}) + \zeta - \zeta] p(\mathbf{y}) d\mathbf{y} \\ &= (1 - \alpha)^{-1} \left(\int_{\{f(\mathbf{x}, \mathbf{y}) \geq \zeta_\alpha(\mathbf{x})\}} [f(\mathbf{x}, \mathbf{y}) - \zeta] p(\mathbf{y}) d\mathbf{y} + \int_{\{f(\mathbf{x}, \mathbf{y}) \geq \zeta_\alpha(\mathbf{x})\}} \zeta p(\mathbf{y}) d\mathbf{y} \right) \\ &= (1 - \alpha)^{-1} \int_{\{f(\mathbf{x}, \mathbf{y}) \geq \zeta_\alpha(\mathbf{x})\}} [f(\mathbf{x}, \mathbf{y}) - \zeta] p(\mathbf{y}) d\mathbf{y} + (1 - \alpha)^{-1} \zeta (1 - \alpha) \\ &= \zeta + (1 - \alpha)^{-1} \int_{\{f(\mathbf{x}, \mathbf{y}) \geq \zeta_\alpha(\mathbf{x})\}} [f(\mathbf{x}, \mathbf{y}) - \zeta] p(\mathbf{y}) d\mathbf{y}. \end{aligned}$$

Si ahora hacemos $\zeta = \zeta_\alpha(\mathbf{x})$ obtenemos

$$\begin{aligned} \phi_\alpha(\mathbf{x}) &= \zeta_\alpha(\mathbf{x}) + (1 - \alpha)^{-1} \int_{\{f(\mathbf{x}, \mathbf{y}) \geq \zeta_\alpha(\mathbf{x})\}} [f(\mathbf{x}, \mathbf{y}) - \zeta_\alpha(\mathbf{x})] p(\mathbf{y}) d\mathbf{y} \\ &= \zeta_\alpha(\mathbf{x}) + (1 - \alpha)^{-1} \int_{\mathbf{y} \in \mathbb{R}^m} [f(\mathbf{x}, \mathbf{y}) - \zeta_\alpha(\mathbf{x})]^+ p(\mathbf{y}) d\mathbf{y} \\ &= F_\alpha(\mathbf{x}, \zeta_\alpha(\mathbf{x})). \end{aligned}$$

³En la sección 2.3.2 se relajará este supuesto permitiendo que la función de distribución acumulada de pérdidas presente saltos. Incluso en esos casos, las propiedades del CVaR se mantienen.

Con ello tenemos

$$F_\alpha(\mathbf{x}, \zeta_\alpha(\mathbf{x})) = \phi_\alpha(\mathbf{x}). \quad (2.5)$$

Inspirado en este resultado, obtenemos la función vista en (2.4). De esta manera, los autores propusieron la utilización de la función F_α para minimizarla con respecto a todos sus argumentos obteniendo el CVaR mínimo y a la vez el VaR del portafolio óptimo.

Las características de la función F_α están mencionadas en los siguientes teoremas.

Teorema 12. Para cada $\mathbf{x} \in \mathbb{R}^n$, la función $F_\alpha(\mathbf{x}, \cdot) : \mathbb{R} \rightarrow \mathbb{R}$ es convexa y continuamente diferenciable. El CVaR de las pérdidas asociadas con cualquier $\mathbf{x} \in X$ es determinado por la fórmula

$$\phi_\alpha(\mathbf{x}) = \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta).$$

En esta fórmula, el conjunto de los valores de ζ para los cuales el mínimo es alcanzado

$$A_\alpha(\mathbf{x}) := \operatorname{argmin}_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta)$$

es un intervalo cerrado, acotado y no vacío (quizá un único punto) y el VaR de las pérdidas está dado por

$$\zeta_\alpha(\mathbf{x}) = \text{extremo izquierdo de } A_\alpha(\mathbf{x}).$$

En particular, uno siempre tiene que

$$\zeta_\alpha(\mathbf{x}) \in \operatorname{argmin}_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta) \text{ y } \phi_\alpha(\mathbf{x}) = F_\alpha(\mathbf{x}, \zeta_\alpha(\mathbf{x})).$$

De acuerdo a este resultado, el CVaR puede ser calculado sin necesidad de calcular primero el VaR. En su lugar, el VaR es calculado como un subproducto de todo el procedimiento.

Para la demostración del presente teorema utilizaremos el siguiente lema.

Lema 13. Con \mathbf{x} fijo, sea $G(\zeta) = \int_{\mathbf{y} \in \mathbb{R}^m} g(\zeta, \mathbf{y}) p(\mathbf{y}) d\mathbf{y}$, donde $g(\zeta, \mathbf{y}) = [f(\mathbf{x}, \mathbf{y}) - \zeta]^+$. Entonces G es convexa, continuamente diferenciables y con derivada

$$G'(\zeta) = \Psi(\mathbf{x}, \zeta) - 1.$$

Para la prueba completa del lema, procederemos primero demostrando que G es convexa como función de ζ . Luego demostraremos que G es continuamente diferenciable y que su derivada es $G'(\zeta) = \Psi(\mathbf{x}, \zeta) - 1$.

(Demostrando que $G(\zeta)$ es convexa como función de ζ)

Para cada \mathbf{x} e \mathbf{y} , sea $g(\zeta, \mathbf{y}) := [f(\mathbf{x}, \mathbf{y}) - \zeta]^+$. Sabemos que g es convexa en ζ pues la función $[\cdot]^+$ es convexa. Sean $\zeta_1, \zeta_2 \in \mathbb{R}$ y $\forall \lambda \in [0, 1]$, entonces tenemos lo siguiente dada la convexidad de g :

$$g(\zeta_1 \lambda + \zeta_2 (1 - \lambda), \mathbf{y}) \leq \lambda g(\zeta_1, \mathbf{y}) + (1 - \lambda) g(\zeta_2, \mathbf{y}).$$

Integrando en \mathbf{y}

$$\int_{\mathbf{y} \in \mathbb{R}^m} g(\zeta_1 \lambda + \zeta_2 (1 - \lambda), \mathbf{y}) p(\mathbf{y}) d\mathbf{y} \leq \lambda \int_{\mathbf{y} \in \mathbb{R}^m} g(\zeta_1, \mathbf{y}) p(\mathbf{y}) d\mathbf{y} + (1 - \lambda) \int_{\mathbf{y} \in \mathbb{R}^m} g(\zeta_2, \mathbf{y}) p(\mathbf{y}) d\mathbf{y}$$

$$G(\zeta_1 \lambda + \zeta_2 (1 - \lambda)) \leq \lambda G(\zeta_1) + (1 - \lambda) G(\zeta_2).$$

Así, $G(\zeta)$ es convexa. □

(Demostrando que $G(\zeta)$ es continuamente diferenciable con derivada $G'(\zeta) = \Psi(\mathbf{x}, \zeta) - 1$) Vamos a probar primero que $\frac{\partial^+ G(\zeta)}{\partial \zeta} = \Psi(\mathbf{x}, \zeta) - 1$. Para ello basta probar que si $t_n \downarrow 0$, entonces

$$\lim_{n \rightarrow +\infty} \frac{G(\zeta + t_n, \mathbf{y}) - G(\zeta, \mathbf{y})}{t_n} = \Psi(\mathbf{x}, \zeta) - 1.$$

Recordemos que la función de distribución $\Psi(\mathbf{x}, \zeta)$ es continua en ζ y entonces, para cada ζ , el conjunto

$$D_\zeta = \{\mathbf{y} : f(\mathbf{x}, \mathbf{y}) = \zeta\} \subset \mathbb{R}^m$$

tiene medida nula. Notemos además que, por la definición de g , para cada ζ , la derivada $\frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y})$ no está definida solo en los puntos $\mathbf{y} \in D_\zeta$. Fijemos $\zeta \in \mathbb{R}$ y $t_n \downarrow 0$. Para cada $n \geq 1$, sea $\varphi_n : \mathbb{R}^m \rightarrow \mathbb{R}$ la función definida por

$$\varphi_n(\mathbf{y}) = \frac{[g(\zeta + t_n, \mathbf{y}) - g(\zeta, \mathbf{y})]}{t_n}.$$

Por lo visto arriba

$$\lim_{n \rightarrow +\infty} \varphi_n(\mathbf{y}) = \frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) \text{ c.s.}$$

Además, como $g(\cdot, \mathbf{y})$ es convexa, entonces esta convergencia es monótona no creciente. En particular, φ_n es una sucesión de funciones monótonas no crecientes que converge a $\frac{\partial}{\partial \zeta} g(\zeta, \cdot)$. Además, es simple ver que

$$-1 \leq \varphi_n(\mathbf{y}) = \frac{[g(\zeta + t_n, \mathbf{y}) - g(\zeta, \mathbf{y})]}{t_n} \leq 0.$$

Entonces $-\varphi_n(\mathbf{y})$ es una sucesión monótona no decreciente de funciones no negativas que convergen c.s. a la función $-\frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y})$. Así, por el Teorema de la Convergencia Monótona

$$\lim_{n \rightarrow +\infty} E \{-\varphi_n\} = E \left\{ \lim_{n \rightarrow +\infty} -\varphi_n \right\} = E \left\{ -\frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) \right\}. \quad (2.6)$$

De esta manera,

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{G(\zeta + t_n, \mathbf{y}) - G(\zeta, \mathbf{y})}{t_n} &= \lim_{n \rightarrow +\infty} \left\{ \int_{\mathbf{y} \in \mathbb{R}^m} \left[\frac{g(\zeta + t_n, \mathbf{y}) - g(\zeta, \mathbf{y})}{t_n} \right] p(\mathbf{y}) d\mathbf{y} \right\} \\ &= \lim_{n \rightarrow +\infty} E \{\varphi_n(\mathbf{y})\} \\ &= E \left\{ \frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) \right\}. \end{aligned}$$

Por otro lado,

$$E \left\{ \frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) \right\} = \int_{\mathbf{y} \in \mathbb{R}^m} \frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) p(\mathbf{y}) d\mathbf{y}.$$

Analizando el integrando tenemos

$$\frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) = \begin{cases} 0, & \text{si } f(\mathbf{x}, \mathbf{y}) < \zeta \\ -1, & \text{si } f(\mathbf{x}, \mathbf{y}) > \zeta \end{cases},$$

y teniendo en cuenta que D_ζ tiene medida nula, entonces

$$\begin{aligned} \int_{\mathbf{y} \in \mathbb{R}^m} \frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) p(\mathbf{y}) d\mathbf{y} &= \int_{\{f(\mathbf{x}, \mathbf{y}) < \zeta\}} 0 p(\mathbf{y}) d\mathbf{y} + \int_{\{f(\mathbf{x}, \mathbf{y}) > \zeta\}} -1 p(\mathbf{y}) d\mathbf{y} \\ &= - \left(1 - \int_{\{f(\mathbf{x}, \mathbf{y}) \leq \zeta\}} p(\mathbf{y}) d\mathbf{y} \right) \\ &= \Psi(\mathbf{x}, \zeta) - 1. \end{aligned}$$

Esto muestra que

$$\frac{\partial^+ G(\zeta)}{\partial \zeta} = \Psi(\mathbf{x}, \zeta) - 1.$$

Ahora, vamos a probar que $\frac{\partial^- G(\zeta)}{\partial \zeta} = \Psi(\mathbf{x}, \zeta) - 1$. Para ello basta probar que si $t_n \uparrow 0$, entonces

$$\lim_{n \rightarrow +\infty} \frac{G(\zeta + t_n, \mathbf{y}) - G(\zeta, \mathbf{y})}{t_n} = \Psi(\mathbf{x}, \zeta) - 1.$$

Fijemos $\zeta \in \mathbb{R}$ y $t_n \uparrow 0$. Para cada $n \geq 1$, sea $\mu_n : \mathbb{R}^m \rightarrow \mathbb{R}$ la función definida por

$$\mu_n(\mathbf{y}) = \frac{[g(\zeta + t_n, \mathbf{y}) - g(\zeta, \mathbf{y})]}{t_n}.$$

Por lo visto en la parte anterior

$$\lim_{n \rightarrow +\infty} \mu_n(\mathbf{y}) = \frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) \text{ c.s.}$$

Más aun, μ_n es una sucesión de funciones monótonas no decrecientes que converge a $\frac{\partial}{\partial \zeta} g(\zeta, \cdot)$. Además, es simple ver que

$$-1 \leq \mu_n(\mathbf{y}) = \frac{[g(\zeta + t_n, \mathbf{y}) - g(\zeta, \mathbf{y})]}{t_n} \leq 0.$$

Entonces $\mu_n(\mathbf{y})$ es una sucesión monótona no decreciente de funciones que convergen c.s. a la función $\frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y})$. Así, por el Teorema de la Convergencia Monótona

$$\lim_{n \rightarrow +\infty} E \{ \mu_n \} = E \left\{ \lim_{n \rightarrow +\infty} \mu_n \right\} = E \left\{ \frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) \right\}.$$

De esta manera,

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{G(\zeta + t_n, \mathbf{y}) - G(\zeta, \mathbf{y})}{t_n} &= \lim_{n \rightarrow +\infty} \left\{ \int_{\mathbf{y} \in \mathbb{R}^m} \left[\frac{g(\zeta + t_n, \mathbf{y}) - g(\zeta, \mathbf{y})}{t_n} \right] p(\mathbf{y}) d\mathbf{y} \right\} \\ &= \lim_{n \rightarrow +\infty} E \{ \mu_n(\mathbf{y}) \} \\ &= E \left\{ \frac{\partial}{\partial \zeta} g(\zeta, \mathbf{y}) \right\} \\ &= \Psi(\mathbf{x}, \zeta) - 1. \end{aligned}$$

Finalmente, esto muestra que

$$\frac{\partial^+ G(\zeta)}{\partial \zeta} = \frac{\partial^- G(\zeta)}{\partial \zeta} = \Psi(\mathbf{x}, \zeta) - 1.$$

Así, $G(\zeta)$ es continuamente diferenciable con derivada $G'(\zeta) = \Psi(\mathbf{x}, \zeta) - 1$. □

Con estos resultados el lema queda demostrado. Ahora utilizaremos estos resultados para demostrar parte del Teorema 12.

(Demostrando que $F_\alpha(\mathbf{x}, \zeta)$ es convexa y continuamente diferenciable en ζ) Recordemos que

$$\begin{aligned} F_\alpha(\mathbf{x}, \zeta) &= \zeta + (1 - \alpha)^{-1} \int_{\mathbf{y} \in \mathbb{R}^m} [f(\mathbf{x}, \mathbf{y}) - \zeta]^+ p(\mathbf{y}) d\mathbf{y} \\ &= \zeta + (1 - \alpha)^{-1} G(\zeta). \end{aligned}$$

De esta manera, como $\alpha \in (0, 1)$, entonces $F_\alpha(\mathbf{x}, \zeta)$ es convexa con respecto a ζ pues la suma de funciones convexas es una función convexa. □

(Demostrando que $\phi_\alpha(\mathbf{x}) = \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta)$) Utilizando el lema previo tenemos lo siguiente

$$\begin{aligned} \frac{\partial F_\alpha(\mathbf{x}, \zeta)}{\partial \zeta} &= 1 + (1 - \alpha)^{-1} [\Psi(\mathbf{x}, \zeta) - 1] \\ &= \frac{(1 - \alpha) + \Psi(\mathbf{x}, \zeta) - 1}{1 - \alpha} \\ &= \frac{\Psi(\mathbf{x}, \zeta) - \alpha}{1 - \alpha} \\ &= (1 - \alpha)^{-1} [\Psi(\mathbf{x}, \zeta) - \alpha]. \end{aligned}$$

De esta manera, los valores de ζ que proporcionan el mínimo de $F_\alpha(\mathbf{x}, \zeta)$ con respecto a ζ son aquellos que cumplen

$$\Psi(\mathbf{x}, \zeta) - \alpha = 0. \quad (2.7)$$

Dado que $F_\alpha(\mathbf{x}, \zeta)$ es convexo con respecto a ζ y además que $\Psi(\mathbf{x}, \zeta)$ es continuo, no decreciente en ζ y por la definición del VaR entonces

$$\Psi(\mathbf{x}, \zeta_\alpha(\mathbf{x})) = \alpha.$$

Luego se tiene que

$$\min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta) = F_\alpha(\mathbf{x}, \zeta_\alpha(\mathbf{x})).$$

Recordemos además de (2.5) que

$$F_\alpha(\mathbf{x}, \zeta_\alpha(\mathbf{x})) = \phi_\alpha(\mathbf{x})$$

y consecuentemente

$$\phi_\alpha(\mathbf{x}) = \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta). \quad \square$$

(Demostrando que $A_\alpha(\mathbf{x})$ es un intervalo cerrado, acotado y no vacío)

De (2.7), $\zeta \in A_\alpha(\mathbf{x}) = \arg \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta)$ cuando $\Psi(\mathbf{x}, \zeta) = \alpha$. Es decir,

$$A_\alpha(\mathbf{x}) = \{\zeta \in \mathbb{R} : \Psi(\mathbf{x}, \zeta) = \alpha\}.$$

Como $\Psi(\mathbf{x}, \zeta_\alpha(\mathbf{x})) = \alpha$, entonces $\zeta_\alpha(\mathbf{x}) \in A_\alpha(\mathbf{x})$. De esta manera, $A_\alpha(\mathbf{x})$ es no vacío. Por otro lado, dado que $A_\alpha(\mathbf{x})$ es el conjunto de minimizantes de una función convexa, entonces éste es un conjunto convexo en \mathbb{R} . Por lo tanto, $A_\alpha(\mathbf{x})$ es un intervalo. Además, como la función $\Psi(\mathbf{x}, \zeta)$ es continua con respecto a ζ y no decreciente, entonces la imagen inversa de un conjunto cerrado (α),

$$A_\alpha(\mathbf{x}) = \Psi(\mathbf{x}, \cdot)^{-1}(\alpha)$$

es cerrado. Así, $A_\alpha(\mathbf{x})$ es un intervalo cerrado. Por último, como $\lim_{\zeta \rightarrow +\infty} \Psi(\mathbf{x}, \zeta) = 1$ y $\alpha < 1$, entonces existe un ζ' suficientemente grande tal que $\Psi(\mathbf{x}, \zeta') > \alpha$. Análogamente, $\lim_{\zeta \rightarrow -\infty} \Psi(\mathbf{x}, \zeta) = 0$ y $\alpha > 0$, entonces existe un $\zeta'' < \zeta'$ tal que $\Psi(\mathbf{x}, \zeta'') < \alpha$. En consecuencia, como $\Psi(\mathbf{x}, \zeta)$ es no decreciente, $A_\alpha(\mathbf{x}) \subset [\zeta'', \zeta']$ es acotado. \square

(Demostrando que $\zeta_\alpha(\mathbf{x})$ = extremo izquierdo de $A_\alpha(\mathbf{x})$) Dado que $A_\alpha(\mathbf{x}) = \{\zeta \in \mathbb{R} : \Psi(\mathbf{x}, \zeta) = \alpha\}$ y por la definición del VaR,

$$\zeta_\alpha(\mathbf{x}) = \min \underbrace{\{\zeta \in \mathbb{R} : \Psi(\mathbf{x}, \zeta) \geq \alpha\}}_B,$$

entonces $A_\alpha(\mathbf{x}) \subset B$. De esta manera, $\zeta_\alpha(\mathbf{x}) = \min B \leq \min A_\alpha(\mathbf{x})$. Pero como se vio anteriormente, $\zeta_\alpha(\mathbf{x}) \in A_\alpha(\mathbf{x})$. Por lo tanto, $\zeta_\alpha(\mathbf{x})$ es el extremo izquierdo de $A_\alpha(\mathbf{x})$. \square

De acuerdo al siguiente teorema y como se mencionó anteriormente, no es necesario trabajar directamente con el CVaR debido a que en su formulación aparece el VaR. En su lugar uno puede trabajar directamente con la expresión más simple $F_\alpha(\mathbf{x}, \zeta)$ ayudado por su convexidad con respecto a (\mathbf{x}, ζ) , lo cual se verá en el siguiente teorema.

Teorema 14. *Minimizar el CVaR de las pérdidas asociadas a \mathbf{x} sobre todo $\mathbf{x} \in X$ es equivalente a minimizar $F_\alpha(\mathbf{x}, \zeta)$ sobre todo $(\mathbf{x}, \zeta) \in X \times \mathbb{R}$ en el siguiente sentido:*

$$\min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x}) = \min_{(\mathbf{x}, \zeta) \in X \times \mathbb{R}} F_\alpha(\mathbf{x}, \zeta)$$

donde además un par (\mathbf{x}^*, ζ^*) alcanza el segundo mínimo si y solo si \mathbf{x}^* alcanza el primer mínimo y $\zeta^* \in A_\alpha(\mathbf{x}^*)$. Además, $F_\alpha(\mathbf{x}, \zeta)$ es convexo con respecto a (\mathbf{x}, ζ) y $\phi_\alpha(\mathbf{x})$ es convexo con respecto a \mathbf{x} cuando $f(\mathbf{x}, \mathbf{y})$ es convexo con respecto a \mathbf{x} .

En particular, en circunstancias donde el intervalo $A_\alpha(\mathbf{x}^*)$ se reduce a un solo punto (como es usual), la minimización de $F_\alpha(\mathbf{x}, \zeta)$ sobre $(\mathbf{x}, \zeta) \in X \times \mathbb{R}$ produce un par, no necesariamente único, tal que \mathbf{x}^* minimiza el CVaR y ζ^* da el correspondiente VaR.

(Demostrando que $\min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x}) = \min_{(\mathbf{x}, \zeta) \in X \times \mathbb{R}} F_\alpha(\mathbf{x}, \zeta)$) El teorema previo determinó que

$$\phi_\alpha(\mathbf{x}) = \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta).$$

Minimizando ambos lados de la ecuación con respecto a $\mathbf{x} \in X$ tenemos lo siguiente:

$$\begin{aligned} \min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x}) &= \min_{\mathbf{x} \in X} \left(\min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta) \right) \\ &= \min_{(\mathbf{x}, \zeta) \in X \times \mathbb{R}} F_\alpha(\mathbf{x}, \zeta). \end{aligned}$$

Es decir, la minimización de $F_\alpha(\mathbf{x}, \zeta)$ con respecto a $(\mathbf{x}, \zeta) \in X \times \mathbb{R}$ puede ser llevada a cabo primero minimizando con respecto a $\zeta \in \mathbb{R}$ y luego el resultado minimizarlo con respecto a $\mathbf{x} \in X$. \square

(Demostrando que $F_\alpha(\mathbf{x}, \zeta)$ es convexo con respecto a (\mathbf{x}, ζ)) De acuerdo a la hipótesis, para cada $\mathbf{y} \in \mathbb{R}^m$, $f(\mathbf{x}, \mathbf{y})$ es convexo con respecto a \mathbf{x} . En ese caso, para cada $\mathbf{y} \in \mathbb{R}^m$, la función $f(\mathbf{x}, \mathbf{y}) - \zeta$ también es convexa en (\mathbf{x}, ζ) . Además, como la función $[\cdot]^+$ es convexa y no decreciente, entonces, para cada $\mathbf{y} \in \mathbb{R}^m$, la función $[f(\mathbf{x}, \mathbf{y}) - \zeta]^+$ también es convexa en (\mathbf{x}, ζ) . \square

(Demostrando que $\phi_\alpha(\mathbf{x})$ es convexo con respecto a \mathbf{x}) Sean \mathbf{x}_1 y $\mathbf{x}_2 \in \mathbb{R}^n$, ζ_1 y $\zeta_2 \in \mathbb{R}$ y $0 < t < 1$. Por la convexidad de F_α , tenemos lo siguiente:

$$\begin{aligned} (1-t)F_\alpha(\mathbf{x}_1, \zeta_1) + tF_\alpha(\mathbf{x}_2, \zeta_2) &\geq F_\alpha((1-t)\mathbf{x}_1 + t\mathbf{x}_2, (1-t)\zeta_1 + t\zeta_2) \\ &\geq \min_{\zeta \in \mathbb{R}} F_\alpha((1-t)\mathbf{x}_1 + t\mathbf{x}_2, \zeta) \\ &= \phi_\alpha((1-t)\mathbf{x}_1 + t\mathbf{x}_2). \end{aligned}$$

Entonces $(1-t)F_\alpha(\mathbf{x}_1, \zeta_1) + tF_\alpha(\mathbf{x}_2, \zeta_2) \geq \phi_\alpha((1-t)\mathbf{x}_1 + t\mathbf{x}_2)$, $\forall \zeta_1, \zeta_2 \in \mathbb{R}$. Minimizando en ζ_1 :

$$\begin{aligned} (1-t) \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}_1, \zeta) + tF_\alpha(\mathbf{x}_2, \zeta_2) &\geq \phi_\alpha((1-t)\mathbf{x}_1 + t\mathbf{x}_2) \\ (1-t)\phi_\alpha(\mathbf{x}_1) + tF_\alpha(\mathbf{x}_2, \zeta_2) &\geq \phi_\alpha((1-t)\mathbf{x}_1 + t\mathbf{x}_2), \forall \zeta_2 \in \mathbb{R}. \end{aligned}$$

Minimizando en ζ_2 :

$$\begin{aligned} (1-t)\phi_\alpha(\mathbf{x}_1) + t \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}_2, \zeta) &\geq \phi_\alpha((1-t)\mathbf{x}_1 + t\mathbf{x}_2) \\ (1-t)\phi_\alpha(\mathbf{x}_1) + t\phi_\alpha(\mathbf{x}_2) &\geq \phi_\alpha((1-t)\mathbf{x}_1 + t\mathbf{x}_2). \end{aligned}$$

Con lo cual concluimos que $\phi_\alpha(\mathbf{x})$ es convexo con respecto a \mathbf{x} . \square

De esta manera, debido a la convexidad y continuidad de F_α , se puede optimizar el CVaR y calcular el VaR del portafolio óptimo.

2.3.2 Optimización del CVaR para distribuciones de pérdidas más generales

En esta sección, siguiendo a Rockafellar y Uryasev (2002), se muestra que las propiedades fundamentales del CVaR vistas en la sección anterior se mantienen en un contexto de posibles discontinuidades e incluso saltos en la función de distribución acumulada de pérdidas Ψ . En la sección anterior los autores asumieron por simplicidad que la distribución acumulada de pérdidas era continua. En esta sección se levanta ese supuesto y se mostrará que los teoremas mostrados presentan versiones y resultados análogos. Además, se muestra la estabilidad del CVaR con respecto a α , es decir, éste presenta un comportamiento continuo con respecto a α ,

lo cual no es necesariamente cierto para el VaR^4 . Por último, se muestra que el CVaR es una medida de riesgo coherente.

Empecemos redefiniendo los conceptos de la sección anterior para el caso de funciones de pérdidas más generales.

Definición 15. Se asume que el vector aleatorio \mathbf{y} está gobernado por una medida de probabilidad P sobre Y independiente de \mathbf{x} . Para cada \mathbf{x} , denotamos a $\Psi(\mathbf{x}, \cdot)$ sobre \mathbb{R} la resultante función de distribución de las pérdidas $z = f(\mathbf{x}, \mathbf{y})$ es decir

$$\Psi(\mathbf{x}, \zeta) = P\{\mathbf{y} : f(\mathbf{x}, \mathbf{y}) \leq \zeta\}.$$

Con el supuesto que $f(\mathbf{x}, \mathbf{y})$ es continuo en \mathbf{x} y medible en \mathbf{y} . Además, $E\{|f(\mathbf{x}, \mathbf{y})|\} < +\infty$ para cada $\mathbf{x} \in X$.

Definición 16. Denotamos por $\Psi(\mathbf{x}, \zeta^-)$ al límite por izquierda de $\Psi(\mathbf{x}, \cdot)$ en ζ . Así,

$$\Psi(\mathbf{x}, \zeta^-) = \lim_{\eta \rightarrow \zeta^-} \Psi(\mathbf{x}, \eta) = P\{\mathbf{y} : f(\mathbf{x}, \mathbf{y}) < \zeta\}.$$

Definición 17. Cuando la diferencia $\Psi(\mathbf{x}, \zeta) - \Psi(\mathbf{x}, \zeta^-) = P\{\mathbf{y} : f(\mathbf{x}, \mathbf{y}) = \zeta\}$ es positiva, se dice que existe una “probabilidad átomo” en ζ .

En este contexto, el VaR estará definido como sigue:

Definición 18. El VaR de las pérdidas asociadas con respecto a \mathbf{x} es el valor

$$\zeta_\alpha(\mathbf{x}) = \min\{\zeta : \Psi(\mathbf{x}, \zeta) \geq \alpha\}.$$

El valor mínimo del VaR es alcanzado pues $\Psi(\mathbf{x}, \zeta)$ es no decreciente y continuo por la derecha en ζ . Cuando $\Psi(\mathbf{x}, \cdot)$ es continuo y estrictamente creciente, $\zeta_\alpha(\mathbf{x})$ es simplemente el único punto que satisface $\Psi(\mathbf{x}, \zeta) = \alpha$. En otro caso, esta ecuación puede no tener solución o puede tener un rango de soluciones. El caso en el cual no hay soluciones corresponde a una salto vertical de $\Psi(\mathbf{x}, \cdot)$ con α perteneciendo a un intervalo de niveles de confianza que pertenecen al mismo VaR. La Figura 2.1 muestra este caso. Los puntos extremos en ese intervalo son los siguientes:

$$\begin{aligned} \alpha^-(\mathbf{x}) &= \Psi(\mathbf{x}, \zeta_\alpha(\mathbf{x})^-) \\ \alpha^+(\mathbf{x}) &= \Psi(\mathbf{x}, \zeta_\alpha(\mathbf{x})). \end{aligned}$$

Por otro lado, el caso de un rango de soluciones corresponde a un segmento constante de la gráfica de $\Psi(\mathbf{x}, \zeta)$. La Figura 2.2 muestra este caso.

Las soluciones forman un intervalo teniendo a $\zeta_\alpha(\mathbf{x})$ como punto mínimo. El extremo superior de este intervalo es el valor $\zeta_\alpha^+(\mathbf{x})$ definido a continuación.

Definición 19. El VaR^+ (VaR superior) de las pérdidas asociadas con respecto a la decisión \mathbf{x} es el valor

$$\zeta_\alpha^+(\mathbf{x}) = \inf\{\zeta : \Psi(\mathbf{x}, \zeta) > \alpha\}.$$

Definamos ahora al CVaR de la siguiente manera:

Definición 20. El CVaR de las pérdidas asociadas a un vector decisión \mathbf{x} es el valor $\phi_\alpha(\mathbf{x})$ de la media de la distribución de α -cola de $z = f(\mathbf{x}, \mathbf{y})$. De manera más precisa,

$$\phi_\alpha(\mathbf{x}) = \int_{\mathbb{R}} \zeta d\Psi_\alpha(\mathbf{x}, \zeta),$$

donde $\Psi_\alpha(\mathbf{x}, \zeta)$ está dada por

$$\Psi_\alpha(\mathbf{x}, \zeta) = \begin{cases} 0 & ; \text{ para } \zeta < \zeta_\alpha(\mathbf{x}) \\ [\Psi(\mathbf{x}, \zeta) - \alpha] / [1 - \alpha] & ; \text{ para } \zeta \geq \zeta_\alpha(\mathbf{x}). \end{cases}$$

⁴Para esto último, ver Martín, M. (2005).

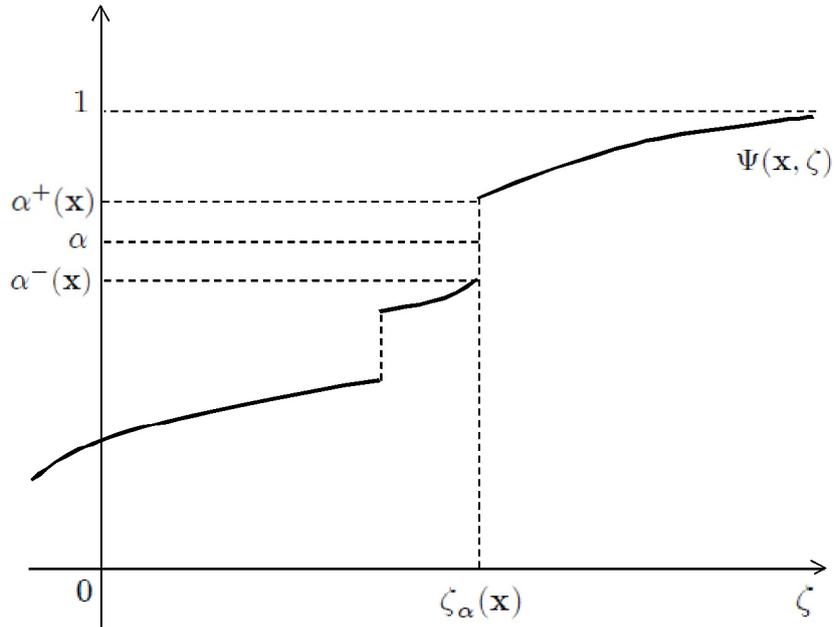


Figura 2.1: La ecuación $\Psi(\mathbf{x}, \zeta) = \alpha$ no tiene solución en ζ

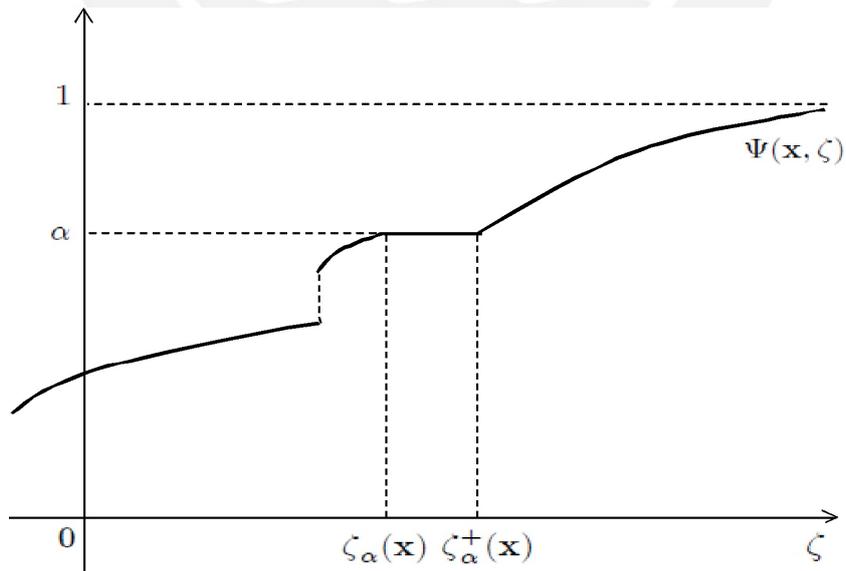


Figura 2.2: La ecuación $\Psi(\mathbf{x}, \zeta) = \alpha$ tiene infinitas soluciones en ζ

Notemos que $\Psi_\alpha(\mathbf{x}, \cdot)$ es otra función de distribución acumulada, como lo es $\Psi(\mathbf{x}, \cdot)$, ya que aquélla es no decreciente y continua por la derecha con $\Psi_\alpha(\mathbf{x}, \zeta) \rightarrow 1$ conforme $\zeta \rightarrow +\infty$. Ahora se mostrará que el VaR y el CVaR pueden ser calculados como en la sección anterior. Para ello, los autores utilizaron la versión análoga de la función $F_\alpha(\mathbf{x}, \zeta)$ de la siguiente manera:

$$F_\alpha(\mathbf{x}, \zeta) = \zeta + \frac{1}{1-\alpha} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta]^+ \} \quad (2.8)$$

Note que, con un procedimiento análogo al visto para obtener (2.5), podemos obtener $F_\alpha(\mathbf{x}, \zeta_\alpha(\mathbf{x})) = \phi_\alpha(\mathbf{x})$, de la siguiente manera

$$\begin{aligned} \phi_\alpha(\mathbf{x}) &= \int_{\mathbb{R}} \zeta d\Psi_\alpha(\mathbf{x}, \zeta) \\ &= \int_{\mathbb{R}} \zeta_\alpha(\mathbf{x}) d\Psi_\alpha(\mathbf{x}, \zeta) + \int_{\mathbb{R}} [\zeta - \zeta_\alpha(\mathbf{x})] d\Psi_\alpha(\mathbf{x}, \zeta) \\ &= \zeta_\alpha(\mathbf{x}) + \int_{(-\infty, \zeta_\alpha(\mathbf{x}))} [\zeta - \zeta_\alpha(\mathbf{x})] d\Psi_\alpha(\mathbf{x}, \zeta) + \int_{[\zeta_\alpha(\mathbf{x}), +\infty)} [\zeta - \zeta_\alpha(\mathbf{x})] d\Psi_\alpha(\mathbf{x}, \zeta) \\ &= \zeta_\alpha(\mathbf{x}) + \int_{\mathbb{R}} [\zeta - \zeta_\alpha(\mathbf{x})]^+ d\Psi_\alpha(\mathbf{x}, \zeta) \\ &= \zeta_\alpha(\mathbf{x}) + \frac{1}{1-\alpha} \int_{\mathbb{R}} [\zeta - \zeta_\alpha(\mathbf{x})]^+ d\Psi(\mathbf{x}, \zeta) \\ &= \zeta_\alpha(\mathbf{x}) + \frac{1}{1-\alpha} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta_\alpha(\mathbf{x})]^+ \} \\ &= F_\alpha(\mathbf{x}, \zeta_\alpha(\mathbf{x})). \end{aligned}$$

Los siguientes teoremas muestran que, así como en el caso de la sección anterior de supuestos especiales de la función de distribución acumulada de pérdidas, la minimización del CVaR se mantiene en el caso de discontinuidades o incluso saltos en la función de distribución acumulada de pérdidas. De esta manera, como en la sección anterior, uno de los aportes de la tesis será completar las demostraciones de los teoremas más importantes vistos en Rockafellar y Uryasev (2002).

Teorema 21. Como función de $\zeta \in \mathbb{R}$, $F_\alpha(\mathbf{x}, \zeta)$ es finita y convexa (por lo tanto continua) con

$$\phi_\alpha(\mathbf{x}) = \min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$$

además

$$\zeta_\alpha(\mathbf{x}) = \text{extremo inferior de } \arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta),$$

$$\zeta_\alpha^+(\mathbf{x}) = \text{extremo superior de } \arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta),$$

donde el $\arg\min$ se refiere al conjunto de los ζ para los cuales el mínimo es alcanzado y en este caso es un intervalo no vacío, cerrado y acotado (quizá un único punto). En particular uno siempre tiene que

$$\zeta_\alpha(\mathbf{x}) \in \arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta),$$

$$\phi_\alpha(\mathbf{x}) = F_\alpha(\mathbf{x}, \zeta_\alpha(\mathbf{x})).$$

(Demostrando que $F_\alpha(\mathbf{x}, \zeta)$ es finita y convexa con respecto a ζ) Para cada $\mathbf{x} \in \mathbb{R}^n$

$$\begin{aligned} F_\alpha(\mathbf{x}, \zeta) &= \zeta + \frac{1}{1-\alpha} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta]^+ \} \\ &\leq \zeta + \frac{1}{1-\alpha} E \{ |f(\mathbf{x}, \mathbf{y})| + |\zeta| \} \end{aligned}$$

ya que $[a - b]^+ \leq |a - b| \leq |a| + |b|$. Entonces

$$F_\alpha(\mathbf{x}, \zeta) \leq \zeta + \frac{1}{1-\alpha} E\{|f(\mathbf{x}, \mathbf{y})| + |\zeta|\} < +\infty,$$

pues $E\{|f(\mathbf{x}, \mathbf{y})|\} < +\infty$. Así, $\zeta \rightarrow F_\alpha(\mathbf{x}, \zeta)$ es finita. Por otro lado, para cada $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in Y$, la función $\zeta \rightarrow h_{\mathbf{x}, \mathbf{y}}(\zeta) = [f(\mathbf{x}, \mathbf{y}) - \zeta]^+$ es convexa, es decir, para todo $\zeta_1, \zeta_2 \in \mathbb{R}$ y $0 < t < 1$:

$$(1-t)h_{\mathbf{x}, \mathbf{y}}(\zeta_1) + th_{\mathbf{x}, \mathbf{y}}(\zeta_2) \geq h_{\mathbf{x}, \mathbf{y}}((1-t)\zeta_1 + t\zeta_2)$$

tomando esperanza

$$(1-t)E\{h_{\mathbf{x}, \mathbf{y}}(\zeta_1)\} + tE\{h_{\mathbf{x}, \mathbf{y}}(\zeta_2)\} \geq E\{h_{\mathbf{x}, \mathbf{y}}((1-t)\zeta_1 + t\zeta_2)\}.$$

Así, para cada $\mathbf{x} \in \mathbb{R}^n$, la función $\zeta \rightarrow E\{[f(\mathbf{x}, \mathbf{y}) - \zeta]^+\}$ es convexa y, en consecuencia, $\zeta \rightarrow F_\alpha(\mathbf{x}, \zeta)$ es también convexa. De esta manera, la función $\zeta \rightarrow F_\alpha(\mathbf{x}, \zeta)$ es continua. \square

(Demostrando que $\phi_\alpha(\mathbf{x}) = \min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$) Sea

$$\frac{F_\alpha(\mathbf{x}, \zeta') - F_\alpha(\mathbf{x}, \zeta)}{\zeta' - \zeta} = 1 + \frac{1}{1-\alpha} E\left\{ \frac{[f(\mathbf{x}, \mathbf{y}) - \zeta']^+ - [f(\mathbf{x}, \mathbf{y}) - \zeta]^+}{\zeta' - \zeta} \right\}. \quad (2.9)$$

Si $\zeta' > \zeta$,

$$\frac{[f(\mathbf{x}, \mathbf{y}) - \zeta']^+ - [f(\mathbf{x}, \mathbf{y}) - \zeta]^+}{\zeta' - \zeta} \begin{cases} = -1 & ; \text{si } f(\mathbf{x}, \mathbf{y}) > \zeta' \\ = 0 & ; \text{si } f(\mathbf{x}, \mathbf{y}) \leq \zeta \\ \in (-1, 0) & ; \text{si } \zeta < f(\mathbf{x}, \mathbf{y}) \leq \zeta' \end{cases}$$

entonces tenemos

$$-P(\mathbf{y} : f(\mathbf{x}, \mathbf{y}) > \zeta') - P(\mathbf{y} : \zeta < f(\mathbf{x}, \mathbf{y}) \leq \zeta') \leq E\left\{ \frac{[f(\mathbf{x}, \mathbf{y}) - \zeta']^+ - [f(\mathbf{x}, \mathbf{y}) - \zeta]^+}{\zeta' - \zeta} \right\} \leq -P(\mathbf{y} : f(\mathbf{x}, \mathbf{y}) > \zeta'). \quad (2.10)$$

Notemos que $\lim_{\zeta' \rightarrow \zeta^+} P(\mathbf{y} : \zeta < f(\mathbf{x}, \mathbf{y}) \leq \zeta') = \lim_{\zeta' \rightarrow \zeta^+} \Psi(\mathbf{x}, \zeta') - \Psi(\mathbf{x}, \zeta) = 0$ y que $\lim_{\zeta' \rightarrow \zeta^+} P(\mathbf{y} : f(\mathbf{x}, \mathbf{y}) > \zeta') = \lim_{\zeta' \rightarrow \zeta^+} (1 - \Psi(\mathbf{x}, \zeta')) = 1 - \Psi(\mathbf{x}, \zeta)$, ya que $\Psi(\mathbf{x}, \cdot)$ es continua por la derecha. Así por (2.10) y por el teorema del sandwich

$$\lim_{\zeta' \rightarrow \zeta^+} E\left\{ \frac{[f(\mathbf{x}, \mathbf{y}) - \zeta']^+ - [f(\mathbf{x}, \mathbf{y}) - \zeta]^+}{\zeta' - \zeta} \right\} = \Psi(\mathbf{x}, \zeta) - 1.$$

Aplicando esto a (2.9)

$$\lim_{\zeta' \rightarrow \zeta^+} \frac{F_\alpha(\mathbf{x}, \zeta') - F_\alpha(\mathbf{x}, \zeta)}{\zeta' - \zeta} = 1 + \frac{1}{1-\alpha} (\Psi(\mathbf{x}, \zeta) - 1) = \frac{\Psi(\mathbf{x}, \zeta) - \alpha}{1-\alpha}.$$

Si $\zeta' < \zeta$,

$$\frac{[f(\mathbf{x}, \mathbf{y}) - \zeta']^+ - [f(\mathbf{x}, \mathbf{y}) - \zeta]^+}{\zeta' - \zeta} \begin{cases} = -1 & ; \text{si } f(\mathbf{x}, \mathbf{y}) \geq \zeta \\ = 0 & ; \text{si } f(\mathbf{x}, \mathbf{y}) \leq \zeta' \\ \in (-1, 0) & ; \text{si } \zeta' < f(\mathbf{x}, \mathbf{y}) \leq \zeta \end{cases}$$

entonces tenemos

$$-P(\mathbf{y} : f(\mathbf{x}, \mathbf{y}) \geq \zeta) - P(\mathbf{y} : \zeta' < f(\mathbf{x}, \mathbf{y}) < \zeta) \leq E\left\{ \frac{[f(\mathbf{x}, \mathbf{y}) - \zeta']^+ - [f(\mathbf{x}, \mathbf{y}) - \zeta]^+}{\zeta' - \zeta} \right\} \leq -P(\mathbf{y} : f(\mathbf{x}, \mathbf{y}) \geq \zeta) \quad (2.11)$$

Pero $\lim_{\zeta' \rightarrow \zeta^-} P(\mathbf{y} : \zeta' < f(\mathbf{x}, \mathbf{y}) < \zeta) = \lim_{\zeta' \rightarrow \zeta^-} \Psi(\mathbf{x}, \zeta') - \Psi(\mathbf{x}, \zeta) = 0$ y que $\lim_{\zeta' \rightarrow \zeta^-} P(\mathbf{y} : f(\mathbf{x}, \mathbf{y}) \geq \zeta) = 1 - \Psi(\mathbf{x}, \zeta^-)$. Así por (2.11) y por el teorema del sandwich

$$\lim_{\zeta' \rightarrow \zeta^-} E\left\{ \frac{[f(\mathbf{x}, \mathbf{y}) - \zeta']^+ - [f(\mathbf{x}, \mathbf{y}) - \zeta]^+}{\zeta' - \zeta} \right\} = \Psi(\mathbf{x}, \zeta^-) - 1.$$

Aplicando esto a (2.9)

$$\lim_{\zeta' \rightarrow \zeta^-} \frac{F_\alpha(\mathbf{x}, \zeta') - F_\alpha(\mathbf{x}, \zeta)}{\zeta' - \zeta} = 1 + \frac{1}{1 - \alpha} (\Psi(\mathbf{x}, \zeta^-) - 1) = \frac{\Psi(\mathbf{x}, \zeta^-) - \alpha}{1 - \alpha}.$$

De este modo

$$\frac{\partial^+ F_\alpha(\mathbf{x}, \zeta)}{\partial \zeta} = \frac{\Psi(\mathbf{x}, \zeta) - \alpha}{1 - \alpha} \quad (2.12)$$

$$\frac{\partial^- F_\alpha(\mathbf{x}, \zeta)}{\partial \zeta} = \frac{\Psi(\mathbf{x}, \zeta^-) - \alpha}{1 - \alpha}. \quad (2.13)$$

Como $F_\alpha(\mathbf{x}, \zeta)$ es convexa con respecto a ζ , entonces los valores de ζ donde $F_\alpha(\mathbf{x}, \zeta)$ alcanza el mínimo son aquellos para los cuales se tiene que

$$\frac{\partial^- F_\alpha(\mathbf{x}, \zeta)}{\partial \zeta} \leq 0 \leq \frac{\partial^+ F_\alpha(\mathbf{x}, \zeta)}{\partial \zeta}.$$

De (2.12) y (2.13), esta condición equivale a

$$\Psi(\mathbf{x}, \zeta^-) \leq \alpha \leq \Psi(\mathbf{x}, \zeta). \quad (2.14)$$

Veamos ahora que $\zeta_\alpha(\mathbf{x})$ satisface la última condición. Por la definición del VaR, $\Psi(\mathbf{x}, \zeta_\alpha(\mathbf{x})) \geq \alpha$ y $\Psi(\mathbf{x}, \zeta) < \alpha$ para todo $\zeta < \zeta_\alpha(\mathbf{x})$. Entonces

$$\Psi(\mathbf{x}, \zeta_\alpha(\mathbf{x})^-) = \lim_{\zeta \rightarrow \zeta_\alpha(\mathbf{x})^-} \Psi(\mathbf{x}, \zeta) \leq \alpha.$$

Esto prueba que $\zeta_\alpha(\mathbf{x})$ satisface (2.14) y por lo tanto

$$\min_{\zeta} F_\alpha(\mathbf{x}, \zeta) = F_\alpha(\mathbf{x}, \zeta_\alpha(\mathbf{x})) = \phi_\alpha(\mathbf{x}). \quad \square$$

(Demostrando que $\zeta_\alpha(\mathbf{x}) = \text{extremo inferior de } \arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$ y $\zeta_\alpha^+(\mathbf{x}) = \text{extremo superior de } \arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$). Por definición, $\zeta_\alpha(\mathbf{x})$ es el menor número que cumple la desigualdad en (2.14) y como ya fue probado que $\zeta_\alpha(\mathbf{x}) \in \arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$, entonces $\zeta_\alpha(\mathbf{x})$ es el extremo inferior de $\arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$. Por otro lado, por la definición de $\zeta_\alpha^+(\mathbf{x})$ y ya que $\Psi(\mathbf{x}, \zeta)$ es no decreciente y continuo por la derecha, entonces

$$\Psi(\mathbf{x}, \zeta_\alpha^+(\mathbf{x})) \geq \alpha.$$

Además, si $\zeta < \zeta_\alpha^+(\mathbf{x})$ entonces $\Psi(\mathbf{x}, \zeta) \leq \alpha$. Entonces

$$\Psi(\mathbf{x}, \zeta_\alpha^+(\mathbf{x})^-) = \lim_{\zeta \rightarrow \zeta_\alpha^+(\mathbf{x})^-} \Psi(\mathbf{x}, \zeta) \leq \alpha.$$

Esto prueba que $\zeta_\alpha^+(\mathbf{x})$ cumple (2.14) y por lo tanto $\zeta_\alpha^+(\mathbf{x}) \in \arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$. Además, como $\Psi(\mathbf{x}, \cdot)$ es no decreciente y continua por la derecha, por la definición de $\zeta_\alpha^+(\mathbf{x})$, $\Psi(\mathbf{x}, \zeta) > \alpha$ para todo $\zeta > \zeta_\alpha^+(\mathbf{x})$. Es más, como $\Psi(\mathbf{x}, \cdot)$ es no decreciente, entonces $\Psi(\mathbf{x}, \zeta) > \alpha$ para todo $\zeta > \zeta_\alpha^+(\mathbf{x})$. En particular, si $\zeta > \zeta_\alpha^+(\mathbf{x})$ entonces ζ no cumple (2.14) y por tanto $\zeta \notin \arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$. Esto muestra que $\zeta_\alpha^+(\mathbf{x})$ es el extremo superior de $\arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$. \square

(Demostrando que $\arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$ es un conjunto no vacío, cerrado y acotado) Como se vio anteriormente, $\zeta_\alpha(\mathbf{x}) \in \arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$ implica que $\arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$ es un conjunto no vacío. Por otro lado, tomando límites en (2.12) y (2.13) tenemos

$$\begin{aligned} \lim_{\zeta \rightarrow +\infty} \frac{\partial^+ F_\alpha(\mathbf{x}, \zeta)}{\partial \zeta} &= \lim_{\zeta \rightarrow +\infty} \frac{\partial^- F_\alpha(\mathbf{x}, \zeta)}{\partial \zeta} = 1 \\ \lim_{\zeta \rightarrow -\infty} \frac{\partial^+ F_\alpha(\mathbf{x}, \zeta)}{\partial \zeta} &= \lim_{\zeta \rightarrow -\infty} \frac{\partial^- F_\alpha(\mathbf{x}, \zeta)}{\partial \zeta} = -\frac{\alpha}{1 - \alpha}, \end{aligned}$$

y como $F_\alpha(\mathbf{x}, \zeta)$ es convexo con respecto a ζ , entonces los conjuntos de nivel de $F_\alpha(\mathbf{x}, \cdot)$ son cerrados y acotados y por lo tanto el conjunto $\arg\min_{\zeta} F_\alpha(\mathbf{x}, \zeta)$ es también cerrado y acotado. \square

Teorema 22. Minimizar $\phi_\alpha(\mathbf{x})$ con respecto a $\mathbf{x} \in X$, es equivalente a minimizar $F_\alpha(\mathbf{x}, \zeta)$ sobre todo $(\mathbf{x}, \zeta) \in X \times \mathbb{R}$ en el sentido que

$$\min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x}) = \min_{(\mathbf{x}, \zeta) \in X \times \mathbb{R}} F_\alpha(\mathbf{x}, \zeta)$$

además

$$(\mathbf{x}^*, \zeta^*) \in \arg \min_{(\mathbf{x}, \zeta) \in X \times \mathbb{R}} F_\alpha(\mathbf{x}, \zeta) \iff \mathbf{x}^* \in \arg \min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x}) \text{ y } \zeta^* \in \arg \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}^*, \zeta).$$

(Demostrando que $\min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x}) = \min_{(\mathbf{x}, \zeta) \in X \times \mathbb{R}} F_\alpha(\mathbf{x}, \zeta)$) Por el teorema anterior tenemos lo siguiente:

$$\phi_\alpha(\mathbf{x}) = \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta)$$

minimizando ambos lados de la igualdad con respecto a $\mathbf{x} \in X$ tenemos lo siguiente:

$$\min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x}) = \min_{\mathbf{x} \in X} \left\{ \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta) \right\} = \min_{(\mathbf{x}, \zeta) \in X \times \mathbb{R}} F_\alpha(\mathbf{x}, \zeta).$$

Para la ida de la segunda parte del teorema tenemos por hipótesis:

$$F_\alpha(\mathbf{x}^*, \zeta^*) \leq F_\alpha(\mathbf{x}, \zeta), \quad \forall \mathbf{x} \in X \text{ y } \forall \zeta \in \mathbb{R}.$$

En particular, si $\mathbf{x} = \mathbf{x}^*$, $\forall \zeta \in \mathbb{R}$, entonces

$$F_\alpha(\mathbf{x}^*, \zeta^*) \leq F_\alpha(\mathbf{x}^*, \zeta), \quad \forall \zeta \in \mathbb{R}.$$

Así $\zeta^* \in \arg \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}^*, \zeta)$. Además, por el teorema previo,

$$\phi_\alpha(\mathbf{x}^*) = \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}^*, \zeta) \leq F_\alpha(\mathbf{x}^*, \zeta^*) \leq F_\alpha(\mathbf{x}, \zeta), \quad \forall \zeta \in \mathbb{R}.$$

Minimizando en $\zeta \in \mathbb{R}$,

$$\phi_\alpha(\mathbf{x}^*) \leq \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}, \zeta) = \phi_\alpha(\mathbf{x}).$$

Así, $\mathbf{x}^* \in \arg \min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x})$.

Para la parte de regreso, tenemos:

$$\min_{(\mathbf{x}, \zeta) \in X \times \mathbb{R}} F_\alpha(\mathbf{x}, \zeta) = \min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x}).$$

Por hipótesis

$$\min_{\mathbf{x} \in X} \phi_\alpha(\mathbf{x}) = \phi_\alpha(\mathbf{x}^*).$$

Por el teorema previo

$$\phi_\alpha(\mathbf{x}^*) = \min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}^*, \zeta).$$

Nuevamente por la hipótesis

$$\min_{\zeta \in \mathbb{R}} F_\alpha(\mathbf{x}^*, \zeta) = F_\alpha(\mathbf{x}^*, \zeta^*).$$

Por lo tanto

$$(\mathbf{x}^*, \zeta^*) \in \arg \min_{(\mathbf{x}, \zeta) \in X \times \mathbb{R}} F_\alpha(\mathbf{x}, \zeta). \quad \square$$

Otra característica del CVaR es el hecho que se comporta de forma continua con respecto al nivel de probabilidad involucrado α . La continuidad del CVaR con respecto al nivel de confianza α se presenta en la siguiente proposición.

Proposición 23 (estabilidad del CVaR). El valor $\phi_\alpha(\mathbf{x})$ se comporta continuamente con respecto a la elección $\alpha \in (0, 1)$ e incluso tiene derivadas izquierda y derecha dadas por las siguientes expresiones:

$$\begin{aligned}\frac{\partial^-}{\partial \alpha} \phi_\alpha(\mathbf{x}) &= \frac{1}{(1-\alpha)^2} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta_\alpha(\mathbf{x})]^+ \} \\ \frac{\partial^+}{\partial \alpha} \phi_\alpha(\mathbf{x}) &= \frac{1}{(1-\alpha)^2} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta_\alpha^+(\mathbf{x})]^+ \}\end{aligned}$$

(Demostrando la estabilidad del CVaR) Para cada \mathbf{x} , $\zeta \in \mathbb{R}$, sea la función $\theta_\zeta : \mathbb{R} \rightarrow \mathbb{R}$ dada por

$$\theta_\zeta(\gamma) = \zeta + \gamma E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta]^+ \}$$

y sea $\theta : \mathbb{R} \rightarrow \mathbb{R}$ la función dada por

$$\theta(\gamma) = \min_{\zeta \in \mathbb{R}} \theta_\zeta(\gamma).$$

Por el Teorema 21, tenemos que

$$\phi_\alpha(x) = \theta \left(\frac{1}{1-\alpha} \right)$$

Como $\theta = \min_{\zeta \in \mathbb{R}} \theta_\zeta$ es el mínimo punto a punto de la colección de funciones θ_ζ , las cuales son afines para cada $\zeta \in \mathbb{R}$, entonces θ es cóncava. En particular, θ es continua y tiene derivadas a la derecha e izquierda en cada punto. De la definición de θ , $\frac{\partial^+}{\partial \gamma} \theta(\tilde{\gamma})$ es la menor de las pendientes de las funciones afines θ_ζ tales que $\theta_\zeta(\tilde{\gamma}) = \theta(\tilde{\gamma})$. De manera análoga, $\frac{\partial^-}{\partial \gamma} \theta(\tilde{\gamma})$ es la mayor de las pendientes de las funciones afines θ_ζ tales que $\theta_\zeta(\tilde{\gamma}) = \theta(\tilde{\gamma})$. Además, para cada $\gamma = \frac{1}{1-\alpha}$, y por el Teorema 21,

$$\theta_{\zeta^*}(\gamma) = \min_{\zeta \in \mathbb{R}} \theta_\zeta(\gamma) \iff \zeta^* \in [\zeta_\alpha(\mathbf{x}), \zeta_\alpha^+(\mathbf{x})].$$

Es decir,

$$\begin{aligned}\frac{\partial^+}{\partial \gamma} \theta(\gamma) &= \min_{\zeta \in [\zeta_\alpha(\mathbf{x}), \zeta_\alpha^+(\mathbf{x})]} \frac{\partial}{\partial \gamma} \theta_\zeta(\gamma) = \min_{\zeta \in [\zeta_\alpha(\mathbf{x}), \zeta_\alpha^+(\mathbf{x})]} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta]^+ \} = E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta_\alpha^+(\mathbf{x})]^+ \} \\ \frac{\partial^-}{\partial \gamma} \theta(\gamma) &= \max_{\zeta \in [\zeta_\alpha(\mathbf{x}), \zeta_\alpha^+(\mathbf{x})]} \frac{\partial}{\partial \gamma} \theta_\zeta(\gamma) = \max_{\zeta \in [\zeta_\alpha(\mathbf{x}), \zeta_\alpha^+(\mathbf{x})]} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta]^+ \} = E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta_\alpha(\mathbf{x})]^+ \}.\end{aligned}$$

La función $\alpha \rightarrow \phi_\alpha(\mathbf{x})$ es la composición de θ con la función

$$\alpha \xrightarrow{\varphi} \frac{1}{1-\alpha}.$$

Como φ es continua y θ es cóncava, entonces la función $\alpha \rightarrow \phi_\alpha(\mathbf{x})$ es continua. Finalmente, usando la regla de la cadena para obtener las derivadas laterales de $\phi_\alpha(\mathbf{x})$ con respecto a α , tenemos lo siguiente:

$$\begin{aligned}\frac{\partial^-}{\partial \alpha} \phi_\alpha(\mathbf{x}) &= \frac{\partial^-}{\partial \gamma} \theta(\varphi(\alpha)) \cdot \varphi'(\alpha) = \frac{1}{(1-\alpha)^2} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta_\alpha(\mathbf{x})]^+ \} \\ \frac{\partial^+}{\partial \alpha} \phi_\alpha(\mathbf{x}) &= \frac{\partial^+}{\partial \gamma} \theta(\varphi(\alpha)) \cdot \varphi'(\alpha) = \frac{1}{(1-\alpha)^2} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta_\alpha^+(\mathbf{x})]^+ \}.\end{aligned} \quad \square$$

Por último, siguiendo a Rockafellar y Uryasev (2002), se mostrará que si $f(\mathbf{x}, \mathbf{y})$ es sublineal, entonces el CVaR es una medida de riesgo coherente en un sentido similar al enunciado en la sección (2.2.3). Más precisamente, el mapeo $\phi_\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ satisface las siguientes propiedades:

1. Monotonicidad: para todo $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ con $f(\mathbf{x}_1, \mathbf{y}) \leq f(\mathbf{x}_2, \mathbf{y}), \forall \mathbf{y} \in Y$ tenemos $\phi_\alpha(\mathbf{x}_1) \leq \phi_\alpha(\mathbf{x}_2)$.
2. Homogeneidad positiva: para todo $\lambda \geq 0$ y para todo $\mathbf{x} \in \mathbb{R}^n$, $\phi_\alpha(\lambda \mathbf{x}) = \lambda \phi_\alpha(\mathbf{x})$.
3. Subaditividad: para todo \mathbf{x}_1 y $\mathbf{x}_2 \in \mathbb{R}^n$, $\phi_\alpha(\mathbf{x}_1 + \mathbf{x}_2) \leq \phi_\alpha(\mathbf{x}_1) + \phi_\alpha(\mathbf{x}_2)$.

Antes de la demostración definamos lo que es una función *sublineal*.

Definición 24 (Sublineal). Una función es llamada sublineal si cumple lo siguiente: $h(\mathbf{x} + \mathbf{x}') \leq h(\mathbf{x}) + h(\mathbf{x}')$ y además $h(\lambda \mathbf{x}) = \lambda h(\mathbf{x})$, para $\lambda > 0$. La segunda propiedad se llama homogeneidad positiva.

Teorema 25 (Coherencia del CVaR). El CVaR es una medida de riesgo coherente. Cuando $f(\mathbf{x}, \mathbf{y})$ es sublineal con respecto a \mathbf{x} , no solo $\phi_\alpha(\mathbf{x})$ es sublineal con respecto a \mathbf{x} , sino que cuando $f(\mathbf{x}, \mathbf{y}) = c$ (reflejando la falta de riesgo), $\phi_\alpha(\mathbf{x}) = c$ y además cuando $f(\mathbf{x}, \mathbf{y}) \leq f(\mathbf{x}', \mathbf{y})$, $\phi_\alpha(\mathbf{x}) \leq \phi_\alpha(\mathbf{x}')$, es decir, $\phi_\alpha(\mathbf{x})$ es monótona.

Lema 26. $F_\alpha(\mathbf{x}, \zeta)$ es sublineal en (\mathbf{x}, ζ) .

(Demostrando que $F_\alpha(\mathbf{x}, \zeta)$ es sublineal en (\mathbf{x}, ζ)) Sean $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ y $\zeta_1, \zeta_2 \in \mathbb{R}$, entonces queremos probar que $F_\alpha(\mathbf{x}_1 + \mathbf{x}_2, \zeta_1 + \zeta_2) \leq F_\alpha(\mathbf{x}_1, \zeta_1) + F_\alpha(\mathbf{x}_2, \zeta_2)$. Como f es sublineal con respecto a \mathbf{x} tenemos lo siguiente

$$\begin{aligned} f(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) &\leq f(\mathbf{x}_1, \mathbf{y}) + f(\mathbf{x}_2, \mathbf{y}) \\ f(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) - (\zeta_1 + \zeta_2) &\leq (f(\mathbf{x}_1, \mathbf{y}) - \zeta_1) + (f(\mathbf{x}_2, \mathbf{y}) - \zeta_2). \end{aligned}$$

Como la función $\mathbf{x} \rightarrow [\mathbf{x}]^+ = \max\{\mathbf{x}, 0\}$ es no decreciente y $[a + b]^+ \leq [a]^+ + [b]^+$,

$$\begin{aligned} [f(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) - (\zeta_1 + \zeta_2)]^+ &\leq [(f(\mathbf{x}_1, \mathbf{y}) - \zeta_1) + (f(\mathbf{x}_2, \mathbf{y}) - \zeta_2)]^+ \\ &\leq [f(\mathbf{x}_1, \mathbf{y}) - \zeta_1]^+ + [f(\mathbf{x}_2, \mathbf{y}) - \zeta_2]^+. \end{aligned}$$

Tomando esperanza en \mathbf{y}

$$E([f(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) - (\zeta_1 + \zeta_2)]^+) \leq E([f(\mathbf{x}_1, \mathbf{y}) - \zeta_1]^+) + E([f(\mathbf{x}_2, \mathbf{y}) - \zeta_2]^+).$$

Finalmente

$$\begin{aligned} F_\alpha(\mathbf{x}_1 + \mathbf{x}_2, \zeta_1 + \zeta_2) &\leq \zeta_1 + \frac{1}{1 - \alpha} E\{[f(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) - \zeta_1]^+\} + \zeta_2 + \frac{1}{1 - \alpha} E\{[f(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) - \zeta_2]^+\} \\ &\leq F_\alpha(\mathbf{x}_1, \zeta_1) + F_\alpha(\mathbf{x}_2, \zeta_2). \end{aligned}$$

Por otro lado, sean $\mathbf{x} \in \mathbb{R}^n$, $\zeta \in \mathbb{R}$, y $\lambda > 0$, entonces queremos probar que $F_\alpha(\lambda \mathbf{x}, \lambda \zeta) = \lambda F_\alpha(\mathbf{x}, \zeta)$.

$$\begin{aligned} F_\alpha(\lambda \mathbf{x}, \lambda \zeta) &= \lambda \zeta + \frac{1}{1 - \alpha} E\{[f(\lambda \mathbf{x}, \mathbf{y}) - \lambda \zeta]^+\} \\ &= \lambda \zeta + \frac{1}{1 - \alpha} E\{[\lambda f(\mathbf{x}, \mathbf{y}) - \lambda \zeta]^+\} \end{aligned}$$

pues f es sublineal, luego

$$\begin{aligned} F_\alpha(\lambda \mathbf{x}, \lambda \zeta) &= \lambda \zeta + \frac{1}{1 - \alpha} E\{\lambda [f(\mathbf{x}, \mathbf{y}) - \zeta]^+\} \\ &= \lambda \zeta + \frac{1}{1 - \alpha} \lambda E\{[f(\mathbf{x}, \mathbf{y}) - \zeta]^+\} \\ &= \lambda \left(\zeta + \frac{1}{1 - \alpha} E\{[f(\mathbf{x}, \mathbf{y}) - \zeta]^+\} \right) \\ &= \lambda F_\alpha(\mathbf{x}, \zeta). \end{aligned}$$

De esta manera, $F_\alpha(\mathbf{x}, \zeta)$ es sublineal en (\mathbf{x}, ζ) . □

Este resultado junto con el Teorema 21 nos ayudarán a demostrar que el CVaR es una medida de riesgo coherente según Rockafellar y Uryasev (2002).

(Demostrando que $\phi_\alpha(\mathbf{x})$ es sublineal con respecto a \mathbf{x}) Primero probaremos que si $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$, entonces $\phi_\alpha(\mathbf{x}_1 + \mathbf{x}_2) \leq \phi_\alpha(\mathbf{x}_1) + \phi_\alpha(\mathbf{x}_2)$. Como $F_\alpha(\mathbf{x}, \zeta)$ es sublineal con respecto a (\mathbf{x}, ζ) , para $\zeta_1, \zeta_2 \in \mathbb{R}$ y por el Teorema 21

$$\begin{aligned}\phi_\alpha(\mathbf{x}_1 + \mathbf{x}_2) &= \min_{\zeta} F_\alpha(\mathbf{x}_1 + \mathbf{x}_2, \zeta) \\ &\leq F_\alpha(\mathbf{x}_1 + \mathbf{x}_2, \zeta_1 + \zeta_2) \\ &\leq F_\alpha(\mathbf{x}_1, \zeta_1) + F_\alpha(\mathbf{x}_2, \zeta_2)\end{aligned}$$

Minimizando sobre ζ_1

$$\begin{aligned}\phi_\alpha(\mathbf{x}_1 + \mathbf{x}_2) &\leq \min_{\zeta} F_\alpha(\mathbf{x}_1, \zeta) + F_\alpha(\mathbf{x}_2, \zeta_2) \\ &= \phi_\alpha(\mathbf{x}_1) + F_\alpha(\mathbf{x}_2, \zeta_2).\end{aligned}$$

Minimizando sobre ζ_2

$$\begin{aligned}\phi_\alpha(\mathbf{x}_1 + \mathbf{x}_2) &\leq \phi_\alpha(\mathbf{x}_1) + \min_{\zeta} F_\alpha(\mathbf{x}_2, \zeta) \\ &= \phi_\alpha(\mathbf{x}_1) + \phi_\alpha(\mathbf{x}_2).\end{aligned}$$

De esta manera, el CVaR es subaditivo. Ahora si $\mathbf{x} \in \mathbb{R}^n$ y $\lambda > 0$, entonces queremos probar que $\phi_\alpha(\lambda \mathbf{x}) = \lambda \phi_\alpha(\mathbf{x})$. Por el Teorema 21 sabemos que

$$\begin{aligned}\phi_\alpha(\lambda \mathbf{x}) &= \min_{\zeta} F_\alpha(\lambda \mathbf{x}, \zeta) \\ &= \min_{\zeta} \lambda F_\alpha\left(\mathbf{x}, \frac{\zeta}{\lambda}\right).\end{aligned}$$

Ya que $F_\alpha(\mathbf{x}, \zeta)$ es sublineal en (\mathbf{x}, ζ)

$$\begin{aligned}\phi_\alpha(\lambda \mathbf{x}) &= \min_{\zeta} \lambda F_\alpha\left(\mathbf{x}, \frac{\zeta}{\lambda}\right) \\ &= \lambda \min_{\zeta} F_\alpha\left(\mathbf{x}, \frac{\zeta}{\lambda}\right) \\ &= \lambda \min_{\tilde{\zeta}} F_\alpha(\mathbf{x}, \tilde{\zeta}) \\ &= \lambda \phi_\alpha(\mathbf{x})\end{aligned}$$

con $\tilde{\zeta} = \frac{\zeta}{\lambda}$. De esta manera, el CVaR es homogénea positiva. Por lo tanto, $\phi_\alpha(\mathbf{x})$ es sublineal con respecto a \mathbf{x} . \square

(Demostrando que $\phi_\alpha(\mathbf{x})$ es monótona) Sean $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$, tales que

$$\begin{aligned}f(\mathbf{x}, \mathbf{y}) &\leq f(\mathbf{x}', \mathbf{y}), \forall \mathbf{y} \in Y \\ f(\mathbf{x}, \mathbf{y}) - \zeta &\leq f(\mathbf{x}', \mathbf{y}) - \zeta, \forall \mathbf{y} \in Y.\end{aligned}$$

Entonces, como $[\cdot]^+$ es no decreciente

$$[f(\mathbf{x}, \mathbf{y}) - \zeta]^+ \leq [f(\mathbf{x}', \mathbf{y}) - \zeta]^+, \forall \mathbf{y} \in Y.$$

Tomando esperanza en \mathbf{y}

$$\begin{aligned}E([f(\mathbf{x}, \mathbf{y}) - \zeta]^+) &\leq E([f(\mathbf{x}', \mathbf{y}) - \zeta]^+) \\ F_\alpha(\mathbf{x}, \zeta) &\leq F_\alpha(\mathbf{x}', \zeta).\end{aligned}$$

Minimizando en ζ y por el Teorema 21

$$\phi_\alpha(\mathbf{x}) = \min_{\zeta} F_\alpha(\mathbf{x}, \zeta) \leq \min_{\zeta} F_\alpha(\mathbf{x}', \zeta) = \phi_\alpha(\mathbf{x}').$$

De esta manera, el CVaR es monótona. \square

(Demostrando que cuando $f(\mathbf{x}, \mathbf{y}) = c, \phi_\alpha(\mathbf{x}) = c$) Por el Teorema 21 sabemos que

$$\begin{aligned}\phi_\alpha(\mathbf{x}) &= \min_{\zeta} F_\alpha(\mathbf{x}, \zeta) \\ &= \min_{\zeta} \left\{ \zeta + \frac{1}{1-\alpha} E \{ [f(\mathbf{x}, \mathbf{y}) - \zeta]^+ \} \right\} \\ &= \min_{\zeta} \left\{ \zeta + \frac{1}{1-\alpha} [c - \zeta]^+ \right\}.\end{aligned}$$

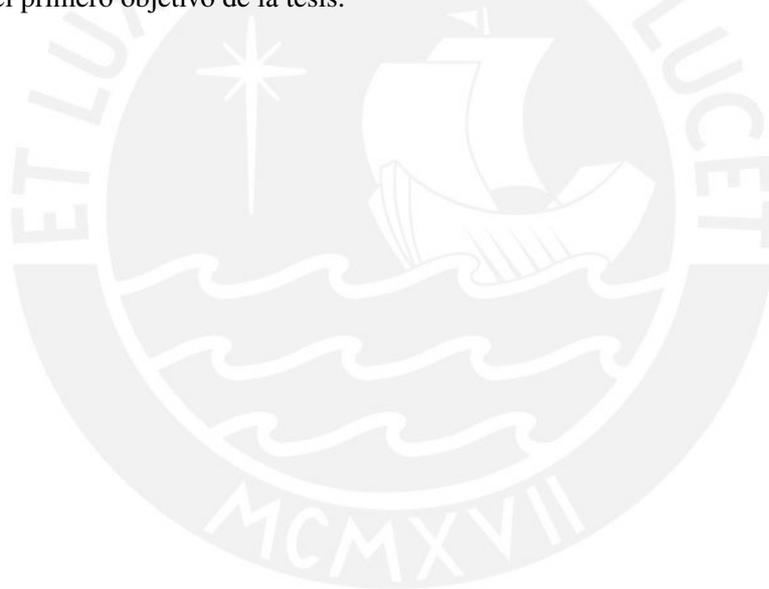
Sea $g(\zeta) = \zeta + \frac{1}{1-\alpha} [c - \zeta]^+$ y luego

$$g(\zeta) = \begin{cases} \zeta & ; \zeta \geq c \\ \frac{c-\alpha\zeta}{1-\alpha} & ; \zeta < c \end{cases}$$

entonces tenemos que

$$\phi_\alpha(\mathbf{x}) = \min_{\zeta} g(\zeta) = g(c) = c. \quad \square$$

En este capítulo se mostró de manera exhaustiva las propiedades del CVaR cuando la función de distribución de pérdidas es continua. En la tercera sección se mostró que estas propiedades tienen versiones análogas cuando la función de distribución es discontinua y presenta saltos. Además, se demostró la continuidad del CVaR con respecto al nivel de confianza α y se demostró que es una medida de riesgo coherente. Las demostraciones vistas forman parte del primero objetivo de la tesis.



Capítulo 3

Cóputas

En este capítulo, se derrallora brevemente la teoría de cóputas, la cual tiene creciente uso en diferentes campos, en particular, en finanzas. Se presentarán las principales definiciones y las cóputas más utilizadas en trabajos aplicados. Además, se mostrará las formas de modelar los datos mediante la utilización de herramientas gráficas. Luego, se presentarán las diferentes formas de estimar los parámetros de las cóputas así como los diferentes criterios de selección de una determinada cóputa.

Se utilizará esta teoría debido a que en los casos en los cuales las distribuciones marginales no se distribuyen como una normal (como se verá más adelante), es imposible definir una distribución conjunta. Siguiendo a Jondeau, Poon y Rockinger (2006), cuando se quiere relacionar dos variables que tienen diferentes distribuciones marginales o cuando la extensión multivariada de alguna distribución marginal no existe, una solución es utilizar cóputas. Por ello, para la simulación de los escenarios se utilizará esta teoría.

El punto de partida de la aplicación de la teoría de cóputas en finanzas se da con el teorema de Sklar (1959), el cual establece la relación entre la función de distribución multivariada y sus marginales univariadas. Sin embargo, antes de mostrar formalmente el teorema y en general mostrar la teoría de cóputas, se presentarán definiciones previas que servirán más adelante.

3.1 Conceptos básicos

Siguiendo a Nelsen (2005), de manera informal diremos que las cóputas son funciones que *juntan* funciones de distribución multivariadas a sus funciones de distribución marginales univariadas, las cuales son uniformes. Como ejemplo, consideremos dos variables aleatorias X e Y con funciones de distribución $F(x) = P[X \leq x]$ y $G(y) = P[Y \leq y]$, respectivamente y una función de distribución conjunta $H(x, y) = P[X \leq x, Y \leq y]$. Así, para cada par de números reales (x, y) podemos asociar tres números: $F(x)$, $G(y)$ y $H(x, y)$. Dado que estos números son valores de funciones de distribución que pertenecen al intervalo $[0, 1]$, entonces cada par (x, y) de números reales lleva a un punto $(F(x), G(y))$ en el cuadrado unitario $[0, 1] \times [0, 1]$, el cual a su vez corresponde a un número $H(x, y)$ en $[0, 1]$. De esta manera, el elemento que asigna el valor de la función de distribución a cada par ordenado de valores de las funciones de distribución individuales es una función, la cual es llamada una *cóputa*.

Para presentar formalmente la definición de cóputa así como sus propiedades, necesitamos previamente desarrollar ciertos conceptos.

Definición 27. Sean S_1, \dots, S_n subconjuntos no vacíos de $\bar{\mathbb{R}} = [-\infty, +\infty]$ y sea H una función real n -dimensional tal que $DomH = S_1 \times \dots \times S_n$. Sea $B = [\mathbf{a}, \mathbf{b}]$ un punto n -dimensional cuyos vértices están en $DomH$. Entonces el H -volumen de B está dado por

$$V_H(B) = \sum \text{sign}(\mathbf{c})H(\mathbf{c}),$$

donde la suma es tomada sobre todos los vértices \mathbf{c} de B y $\text{sign}(\mathbf{c})$ está dado por

$$\text{sign}(\mathbf{c}) = \begin{cases} 1, & \text{si } c_k = a_k \text{ para un número par de } k \\ -1, & \text{si } c_k = a_k \text{ para un número impar de } k. \end{cases}$$

Definición 28. Una función real H es n -creciente si $V_H(B) \geq 0$ para todos los n -puntos de B cuyos vértices se encuentran en el $DomH$.

Ejemplo 29. Para el caso $n = 2$, sea H una función definida sobre \mathbf{I}^2 tal que $H(x, y) = \max(x, y)$. Entonces, H es una función no decreciente de x e y . Sin embargo, $V_H(\mathbf{I}^2) = -1$, de modo que H no es 2-creciente.

3.2 Cópulas

Primero presentaremos la definición de una cópula bidimensional y luego la definición de una cópula n -dimensional.

Definición 30 (Cópula bivariada). Una cópula bidimensional es una función $C : [0, 1]^2 \rightarrow [0, 1]$ con las siguientes propiedades:

(i) Para cada $u_1, u_2 \in [0, 1]$,

$$C(u_1, 0) = 0 = C(0, u_2)$$

y

$$C(u_1, 1) = u_1 \text{ y } C(1, u_2) = u_2.$$

(ii) C es 2-creciente, es decir, para cada $u_{11}, u_{21}, u_{12}, u_{22} \in [0, 1]$ tal que $u_{11} \leq u_{21}$ y $u_{12} \leq u_{22}$,

$$C(u_{21}, u_{22}) - C(u_{21}, u_{12}) - C(u_{11}, u_{22}) + C(u_{11}, u_{12}) \geq 0.$$

Definición 31 (Cópula). Una cópula n -dimensional es una función $C : [0, 1]^n \rightarrow [0, 1]$ con las siguientes propiedades:

(i) Para cada $\mathbf{u} \in [0, 1]^n$,

$$C(\mathbf{u}) = 0 \text{ si al menos una coordenada de } \mathbf{u} \text{ es } 0$$

y

$$\text{si todas las coordenadas de } \mathbf{u} \text{ son } 1 \text{ excepto } u_k, \text{ entonces } C(\mathbf{u}) = u_k.$$

(ii) C es n -creciente, es decir, para cada \mathbf{a} y $\mathbf{b} \in [0, 1]^n$ tal que $\mathbf{a} \leq \mathbf{b}$,

$$V_C([\mathbf{a}, \mathbf{b}]) \geq 0.$$

Presentemos ahora al teorema de Sklar.

3.3 El teorema de Sklar

El teorema de Sklar es de suma importancia en la teoría de cópulas. Como se mencionó anteriormente, el teorema pone de manifiesto el rol que juegan las cópulas para relacionar las funciones de distribución multivariadas y sus marginales univariadas. Antes de mostrar al teorema de Sklar, se debe definir lo que son las funciones de distribución.

Definición 32. Una función de distribución es una función F con $DomF = \overline{\mathbb{R}}$ tal que

(i) F es no decreciente;

(ii) $F(-\infty) = 0$ y $F(\infty) = 1$.

Definición 33. Una función de distribución n -dimensional es una función H con dominio en $\overline{\mathbb{R}}^n$ tal que

(i) H es n -creciente;

(ii) $H(\mathbf{t}) = 0$ para todo \mathbf{t} en $\overline{\mathbb{R}}^n$ tal que $t_k = -\infty$ para al menos un k y $H(\infty, \dots, \infty) = 1$.

Teorema 34 (Teorema de Sklar). Sea H una función de distribución n -dimensional con marginales F_1, \dots, F_n . Entonces existe una n -cópula C tal que $\forall \mathbf{x} \in \overline{\mathbb{R}}^n$

$$H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)).$$

C es única, si F_1, \dots, F_n son continuas. De manera inversa, si C es una n -cópula y F_1, \dots, F_n son funciones de distribución, entonces la función H es una función de distribución conjunta n -dimensional con marginales F_1, \dots, F_n .

(Demostración) Ver More y Spruill (1975), Deheuvels (1978), Sklar (1959) y Sklar (1996). □

De esta manera, a través del teorema de Sklar, la probabilidad conjunta de los datos se puede separar entre sus marginales y la cópula, la cual representa la estructura de dependencia de los datos.

Lo interesante del teorema radica en que la expresión anterior puede ser invertida para expresar la cópula en términos de la función de distribución conjunta y las “inversas” de las marginales. Sin embargo, si una marginal no es estrictamente creciente, entonces ésta no posee inversa en el sentido usual. Así, debemos definir la cuasi-inversa de una función de distribución.

Definición 35. Sea F una función de distribución. Entonces una función cuasi-inversa de F es cualquier función $F^{(-1)}$ con dominio en \mathbf{I} tal que

1. si t está en $\text{Ran}F$, entonces $F^{(-1)}(t)$ es cualquier número x en $\overline{\mathbb{R}}$ tal que $F(x) = t$, es decir, $\forall t \in \text{Ran}F$,

$$F(F^{(-1)}(t)) = t;$$

2. si t no está en $\text{Ran}F$, entonces

$$F^{(-1)}(t) = \inf\{x \mid F(x) \geq t\} = \sup\{x \mid F(x) \leq t\}.$$

Si F es estrictamente creciente, entonces F tiene una sola cuasi-inversa, la cual es la inversa ordinaria, para la cual se usará la notación usual F^{-1} .

Corolario 36. Sean H, C, F_1, \dots, F_n definidos en el teorema anterior y sean $F_1^{(-1)}, \dots, F_n^{(-1)}$ las cuasi-inversas de F_1, \dots, F_n , respectivamente. Entonces para cualquier $\mathbf{u} \in \mathbf{I}^n$,

$$C(u_1, \dots, u_n) = H(F_1^{(-1)}(u_1), \dots, F_n^{(-1)}(u_n)).$$

Definición 37 (Cópula densidad). Sea C una cópula n -dimensional n -diferenciable, entonces la función $c : [0, 1]^n \rightarrow [0, 1]$ con

$$c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n}$$

es llamada la cópula densidad de la cópula C .

3.3.1 Cópulas de sobrevivencia

En ciertas aplicaciones (como en el control de riesgos) resulta de interés el tiempo de sobrevivencia de una variable más allá de cierto umbral. En ese sentido, la probabilidad de sobrevivir más allá de un tiempo dado x se da a través de la función de sobrevivencia $\overline{F}(x) = P[X > x] = 1 - F(x)$, donde F es la función de distribución de X . Así, para un par de variables aleatorias (X, Y) con función de distribución conjunta H , la función de sobrevivencia está dada por $\overline{H}(x, y) = P[X > x, Y > y]$.

Existe una forma análoga del teorema de Sklar para las funciones de sobrevivencia. Suponga que la cópula de X e Y es C . Entonces tenemos lo siguiente:

$$\overline{H}(x, y) = 1 - F(x) - G(y) + H(x, y)$$

$$\begin{aligned}
 &= \overline{F}(x) + \overline{G}(y) - 1 + C(F(x), G(y)) \\
 &= \overline{F}(x) + \overline{G}(y) - 1 + C(1 - \overline{F}(x), 1 - \overline{G}(y))
 \end{aligned}$$

de tal forma que si definimos $\widehat{C} : \mathbf{I}^2 \rightarrow \mathbf{I}$ mediante

$$\widehat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$$

tendremos

$$\overline{H}(x, y) = \widehat{C}(\overline{F}(x), \overline{G}(y)).$$

La función \widehat{C} recibe el nombre de cópula de sobrevivencia de X e Y .

Más adelante se mostrarán algunas cópulas de sobrevivencia llamadas cópulas rotadas 180°.

3.4 Medidas de dependencia

Existen una variedad de formas de medir la dependencia entre dos variables. Estas medidas son importantes no solo porque nos permiten determinar la dependencia entre dos variables, sino porque permitirán estimar, si se cumplen ciertos supuestos, los parámetros de ciertas familias de cópulas.

La medida de dependencia más conocida es el coeficiente de correlación de Pearson. Sin embargo, ésta es una medida de dependencia “lineal” y, lo más importante, es una medida no invariante bajo transformaciones estrictamente crecientes de las variables. Por ello, se mostrarán las medidas de dependencia “invariantes por traslación” más conocidas como el tau de Kendall y el rho de Spearman, las cuales son las medidas de dependencia más conocida dentro de las medidas de asociación. Estas medidas de asociación se basan en los conceptos de *concordancia* y *discordancia*.

Siguiendo a Nelsen (2005), de manera informal diremos que dos variables aleatorias son concordantes si grandes valores de una de ellas tienden a estar asociados con grandes valores de la otra y viceversa. En forma más formal, sean (x_i, y_i) y (x_j, y_j) dos observaciones de un vector aleatorio (X, Y) de variables aleatorias continuas. Diremos que (x_i, y_i) y (x_j, y_j) son *concordantes* si $x_i < x_j$ y $y_i < y_j$ o si $x_i > x_j$ y $y_i > y_j$. De manera similar, diremos que (x_i, y_i) y (x_j, y_j) dos *discordantes* si $x_i < x_j$ y $y_i > y_j$ o si $x_i > x_j$ y $y_i < y_j$.

Definición 38. Una medida numérica de asociación δ entre dos variables aleatorias X e Y cuya cópula es C es una medida de dependencia si satisface las siguientes propiedades:

1. δ está definida para cualquier par de variables aleatorias X e Y .
2. $\delta_{X,Y} = \delta_{Y,X}$.
3. $0 \leq \delta_{X,Y} \leq 1$.
4. $\delta_{X,Y} = 0$ si y solo si X e Y son independendientes.
5. $\delta_{X,Y} = 1$ si y solo si X e Y son función estrictamente monótona no decreciente c.s. una de la otra.
6. Si α y β son funciones monótonas no decrecientes c.s. sobre $RanX$ y $RanY$, respectivamente, entonces $\delta_{\alpha(X),\beta(Y)} = \delta_{X,Y}$.
7. Si $\{(X_n, Y_n)\}$ es una sucesión de variables aleatorias con cópulas C_n y si $\{C_n\}$ converge a C , entonces $\lim_{n \rightarrow \infty} \delta_{C_n} = \delta_C$.

Como mencionamos anteriormente, los ejemplos más conocidos de medidas de asociación son el tau de Kendall y el rho de Spearman.

3.4.1 Tau de Kendall

Sean (X_1, Y_1) y (X_2, Y_2) vectores aleatorios independientes e idénticamente distribuidos (iid) con funciones de distribución conjunta H_1 y H_2 . El tau de Kendall es definido como la probabilidad de concordancia menos la probabilidad de discordancia

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Nelsen (2005) demostró que el tau de Kendall se puede expresar en términos de cópulas mediante el siguiente teorema.

Teorema 39. Sean X e Y dos variables aleatorias continuas cuya cópula es C . Entonces el tau de Kendall de X e Y está dado por

$$\tau(X, Y) = 4 \int \int_{\mathbb{I}^2} C(u, v) dC(u, v) - 1.$$

(Demostración) Nelsen (2005). □

3.4.2 Rho de Spearman

Así como el tau de Kendall, el rho de Spearman es otra medida de asociación basada en la discordancia y concordancia. La representación de esta medida en función de la cópula se muestra en el siguiente teorema.

Teorema 40. Sean X e Y variables aleatorias continuas cuya cópula es C . Entonces el rho de Spearman de X e Y está dado por

$$\rho(X, Y) = 12 \int \int_{\mathbb{I}^2} C(u, v) dudv - 3 = 12 \int \int_{\mathbb{I}^2} uv dC(u, v) - 3.$$

(Demostración) Nelsen (2005). □

3.4.3 Coeficiente de dependencia de colas

Las medidas de dependencia introducidas hasta ahora (tau de Kendall y rho de Spearman) están diseñadas para describir cuán grande o pequeño los valores de una variable aleatoria están asociados con valores grandes o pequeños de otra variable aleatoria. No obstante, otras medidas están centradas en determinar la asociación entre las variables cuando estos toman valores extremos. A estas medidas se les conocen como medidas de dependencia de las colas.

Definición 41. Sean X e Y variables aleatorias continuas con funciones de distribución F y G , respectivamente. El parámetro de dependencia de cola superior λ_U es el límite (si éste existe) de la probabilidad condicional que Y sea mayor que el percentil 100 de G dado que X es mayor que el percentil 100 de F cuando t se aproxima a 1, es decir,

$$\lambda_U = \lim_{t \rightarrow 1^-} P[Y > G^{(-1)}(t) | X > F^{(-1)}(t)].$$

De forma similar, el parámetro de dependencia de cola inferior λ_L es el límite (si éste existe) de la probabilidad condicional que Y sea menor o igual al percentil 100 de G dado que X es menor o igual al percentil 100 de F cuando t se aproxima a 0, es decir,

$$\lambda_L = \lim_{t \rightarrow 0^+} P[Y \leq G^{(-1)}(t) | X \leq F^{(-1)}(t)].$$

Según Joe (1997), las definiciones del parámetro de dependencia de cola superior e inferior en términos de cópulas son las siguientes:

Definición 42. Sean X e Y variables aleatorias continuas con funciones de distribución F y G , respectivamente y cópula C . Si los límites de

$$\lambda_U = \lim_{t \rightarrow 1^-} \frac{(1 - 2t + C(t, t))}{(1 - t)}$$

y

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t}$$

existen, entonces si $\lambda_U \in (0, 1]$, diremos que C tiene dependencia de cola superior. Si $\lambda_U = 0$, diremos que C no tiene dependencia de cola superior. De manera similar, si $\lambda_L \in (0, 1]$, diremos que C tiene dependencia de cola inferior. Si $\lambda_L = 0$, diremos que C no tiene dependencia de cola inferior.

3.5 Cópulas multivariadas

3.5.1 Cópulas elípticas

Una de las principales clases de cópulas son las elípticas, las cuales son generadas por distribuciones elípticas. Además, este tipo de cópulas son obtenidas directamente al invertir el teorema de Sklar.

Definición 43 (Distribución elíptica). El vector aleatorio n -dimensional \mathbf{X} tiene una distribución elíptica si y solo si la función de densidad $f_{\mathbf{X}}(\mathbf{x})$ tiene la siguiente representación

$$f_{\mathbf{X}}(\mathbf{x}) = c_n |\Sigma|^{-\frac{1}{2}} g((\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu)),$$

donde $c_n \in \mathbb{R}^n$, el vector de media $\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n \times n}$ es una matriz simétrica positivo definida y g es alguna función independiente de n .

Los ejemplos más conocidos de distribuciones elípticas son la distribución normal multivariada y la t -student multivariada. Dado que estas cópulas se obtienen al invertir el teorema de Sklar, la cópula normal multivariada es la siguiente:

$$C(u_1, \dots, u_n) = \Phi_{\rho}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)), \quad (3.1)$$

donde Φ_{ρ} representa la función de distribución normal estándar con parámetro de correlación ρ (matriz de correlación simétrica es positivo definida) y Φ^{-1} es la inversa de la función de distribución normal estándar. La cópula densidad de la cópula normal multivariada es la siguiente:

$$c(u_1, \dots, u_n) = |\rho|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{x}'(\rho^{-1} - I)\mathbf{x}\right),$$

donde $\mathbf{x} = (x_1, \dots, x_n)' \in \mathbb{R}^n$ con $x_i = \Phi^{-1}(u_i)$, $i = 1, \dots, n$.

Por otro lado, la t -cópula multivariada es la siguiente:

$$C(u_1, \dots, u_n) = t_{\rho, v}(t_v^{-1}(u_1), \dots, t_v^{-1}(u_n)),$$

donde $t_{\rho, v}$ es la distribución t -student multivariada con matriz de correlaciones ρ y v grados de libertad. t_v^{-1} representa la inversa de la función de distribución univariada. La cópula densidad de la t -cópula multivariada es la siguiente:

$$c(u_1, \dots, u_n) = \frac{\Gamma\left(\frac{v+n}{2}\right)}{\Gamma\left(\frac{v}{2}\right) |\rho|^{1/2}} \left(\frac{\Gamma\left(\frac{v}{2}\right)}{\Gamma\left(\frac{v+1}{2}\right)}\right)^n \left(\frac{\left(1 + \frac{1}{v} \mathbf{x}' \rho^{-1} \mathbf{x}\right)^{-\frac{v+n}{2}}}{\prod_{j=1}^n \left(1 + \frac{x_j^2}{v}\right)^{-\frac{v+1}{2}}}\right),$$

donde $\mathbf{x} = (x_1, \dots, x_n)' \in \mathbb{R}^n$ con $x_i = t_v^{-1}(u_i)$, $i = 1, \dots, n$.

3.5.2 Cópulas arquimedianas

Otra importante clase de cópulas son las arquimedianas, las cuales tienen una gran variedad de aplicaciones, debido a la facilidad en su construcción, la gran variedad de familias y sus propiedades, tales como su simetría y asociatividad.

Para definir a las cópulas arquimedianas es necesario definir a la pseudo inversa de una función.

Definición 44 (pseudo inversa). Sea $\varphi : [0, 1] \rightarrow [0, \infty]$ una función continua y estrictamente decreciente tal que $\varphi(0) = \infty$ y $\varphi(1) = 0$. La pseudo inversa de la función φ es la función $\varphi^{[-1]}$ con $\text{Dom}\varphi^{[-1]} = [0, \infty]$ y $\text{Ran}\varphi^{[-1]} = [0, 1]$ dado por

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0), \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases}$$

Teorema 45 (Cópula arquimediana). Sea $\varphi : [0, 1] \rightarrow [0, \infty]$ una función continua y estrictamente decreciente tal que $\varphi(0) = \infty$ y $\varphi(1) = 0$ y sea $\varphi^{[-1]}$ la pseudo inversa de φ . Entonces

$$C(u_1, \dots, u_n) = \varphi^{[-1]}(\varphi(u_1) + \dots + \varphi(u_n))$$

es una cópula.

(Demostración) Ver Nelsen (2005). □

φ es llamada la función generadora de la cópula C .

Teorema 46 (Propiedades de las cópulas arquimedianas). Sea C una cópula arquimediana con generador φ . Entonces:

1. C es simétrica, es decir, $C(u, v) = C(v, u)$ para todo u, v en $[0, 1]$.
2. C es asociativa, es decir, $C(C(u, v), w) = C(u, C(v, w))$ para todo u, v, w en $[0, 1]$.
3. Si $c > 0$ es cualquier constante, entonces $c\varphi$ es también un generador de C .

(Demostración) Ver Nelsen (2005). □

3.6 Cópulas bivariadas

3.6.1 Familia de cópulas bivariadas elípticas

Dentro de las cópulas bivariadas elípticas, las más conocidas son la gaussiana y la t-student.

Cópula gaussiana

De acuerdo a la versión multivariada definida en la sección 3.5, la cópula gaussiana bivariada con parámetro de correlación $\rho \in (-1, 1)$ es la siguiente:

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

donde Φ_ρ representa la función de distribución normal estándar bivariada con parámetro de correlación ρ y Φ^{-1} representa la inversa de la función de distribución normal estándar univariada. La correspondiente cópula densidad es la siguiente:

$$c(u_1, u_2) = \frac{1}{\sqrt{1-\rho^2}} \exp \left[-\frac{\rho^2(x_1^2 + x_2^2) - 2\rho x_1 x_2}{2(1-\rho^2)} \right],$$

donde $x_1 = \Phi^{-1}(u_1)$ y $x_2 = \Phi^{-1}(u_2)$. Finalmente, el tau de Kendall de la cópula gaussiana es $\frac{2}{\pi} \arcsin(\rho)$. La Figura 3.1 muestra dos gráficos de dispersión del rango de una cópula normal bivariada (1,000 simulaciones) para diferentes valores de ρ .

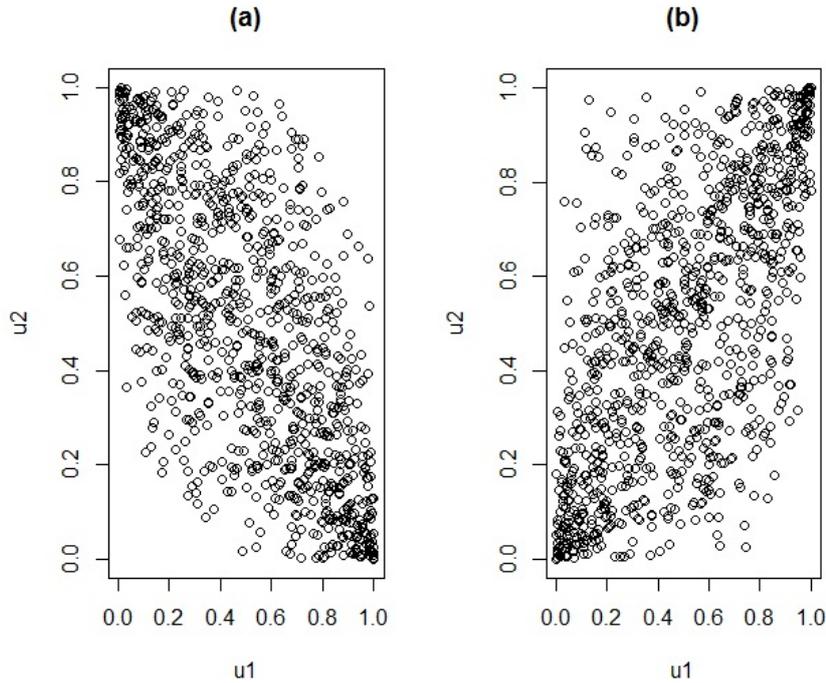


Figura 3.1: Gráficos de dispersión del rango de una cópula bivariada gaussiana. (a): $\rho = -0.7$; (b): $\rho = 0.7$

t-cópula

La t-cópula bivariada es la siguiente:

$$C(u_1, u_2) = t_{\rho, v}(t_v^{-1}(u_1), t_v^{-1}(u_2)),$$

donde $t_{\rho, v}$ representa la función de distribución t-student bivariada con parámetro de correlación ρ y v grados de libertad. Además, t_v^{-1} es la inversa de la función de distribución univariada t-student con v grados de libertad. La correspondiente cópula densidad es la siguiente:

$$c(u_1, u_2) = \frac{\Gamma\left(\frac{v+2}{2}\right) / \Gamma\left(\frac{v}{2}\right)}{v\pi dt_v(x_1) dt_v(x_2) \sqrt{1-\rho^2}} \left[1 + \frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{v(1-\rho^2)} \right]^{-\frac{v+2}{2}},$$

donde $x_1 = t_v^{-1}(u_1)$, $x_2 = t_v^{-1}(u_2)$ y dt_v es la densidad de probabilidad de la distribución estándar t con v grados de libertad. El tau de Kendal de la t-cópula es $\frac{2}{\pi} \arcsin(\rho)$. La Figura 3.2 muestra los gráficos de dispersión del rango de una t-cópula bivariada (1,000 simulaciones) para diferentes valores de ρ y v .

3.6.2 Familias de cópulas bivariadas arquimedianas

Como en el caso multivariado, las cópulas arquimedianas bivariadas están definidas de la siguiente manera:

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2)),$$

donde $\varphi : [0, 1] \rightarrow [0, \infty]$ es una función convexa estrictamente decreciente y continua tal que $\varphi(1) = 0$ y $\varphi^{[-1]}$ es la pseudo inversa

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0), \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases}$$

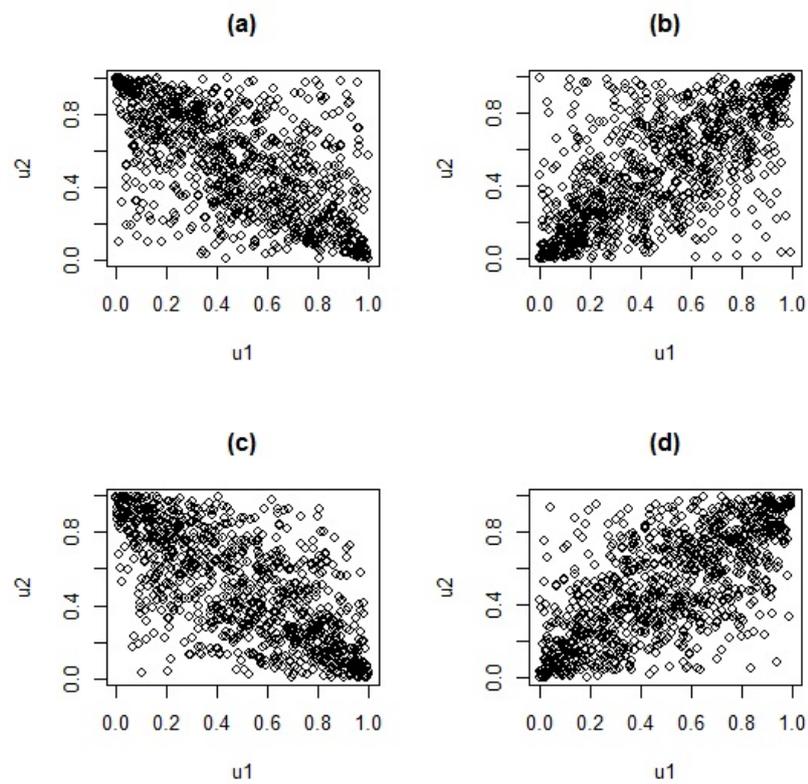


Figura 3.2: Gráficos de dispersión del rango de una t-cópula bivariada. (a) $\rho = -0.7, v = 3$; (b) $\rho = 0.7, v = 3$; (c) $\rho = -0.7, v = 10$; (d) $\rho = 0.7, v = 10$

φ es llamada la función generadora de la cópula C . Por otro lado, para las cópulas arquimedianas, el tau de Kendall se expresa en función de la función generadora de la siguiente forma:

$$\tau(X, Y) = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt,$$

donde X e Y son variables aleatorias y φ es la función generadora definida anteriormente.

A continuación se presentarán algunas de la cópulas bivariadas arquimedianas más conocidas¹.

Cópula Clayton

Con función generadora $\varphi(t) = \frac{1}{\theta} (t^{-\theta} - 1)$ y cópula dada por

$$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$$

para $\theta > 0$, la cópula Clayton es una cópula arquimediana asimétrica que muestra una mayor dependencia en la cola negativa. La cópula densidad de la cópula Clayton es la siguiente:

$$c(u_1, u_2) = (1 + \theta)(u_1 u_2)^{-1-\theta} (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}-2}.$$

El tau de Kendall de la cópula Clayton es el siguiente:

$$\tau(\theta) = \frac{\theta}{\theta + 2}.$$

La Figura 3.3 muestra los gráficos de dispersión del rango de una cópula Clayton bivariada (1,000 simulaciones) para diferentes valores de θ .

Cópula Gumbel

Con función generadora $\varphi(t) = (-\log t)^\theta$ y cópula dada por

$$C(u_1, u_2) = \exp \left[- \left((-\log u_1)^\theta + (-\log u_2)^\theta \right)^{\frac{1}{\theta}} \right],$$

para $\theta \geq 1$, la cópula Gumbel es una cópula arquimediana asimétrica que muestra una mayor dependencia en la cola positiva. La cópula densidad de la cópula Gumbel es la siguiente:

$$c(u_1, u_2) = \frac{C(u_1, u_2)}{u_1 u_2} \frac{(\log u_1 \log u_2)^{\theta-1}}{\left((-\log u_1)^\theta + (-\log u_2)^\theta \right)^{2-\frac{1}{\theta}}} \left[\left((-\log u_1)^\theta + (-\log u_2)^\theta \right)^{\frac{1}{\theta}} + \theta - 1 \right].$$

El tau de Kendall de la cópula Gumbel es el siguiente:

$$\tau(\theta) = 1 - \frac{1}{\theta}.$$

La Figura 3.4 muestra los gráficos de dispersión del rango de una cópula Gumbel bivariada (1,000 simulaciones) para diferentes valores de θ .

¹Para mayor detalle acerca de los funciones generadoras así como las cópulas densidad, ver Nelsen (2005) y Joe (1997).

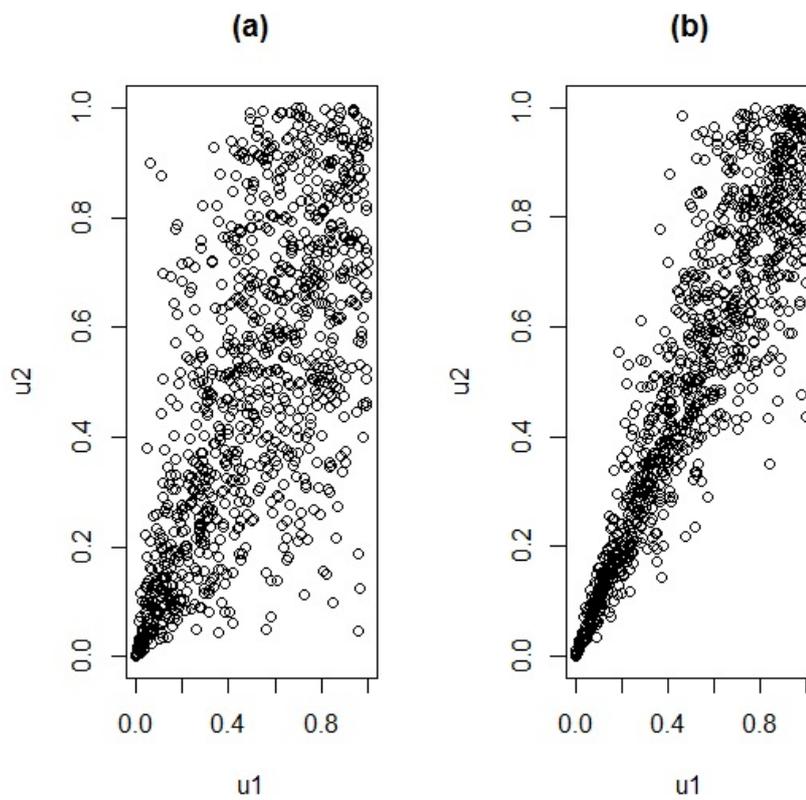


Figura 3.3: Gráficos de dispersión del rango de una cópula Clayton bivariada. (a): $\theta = 2$; (b): $\theta = 6$

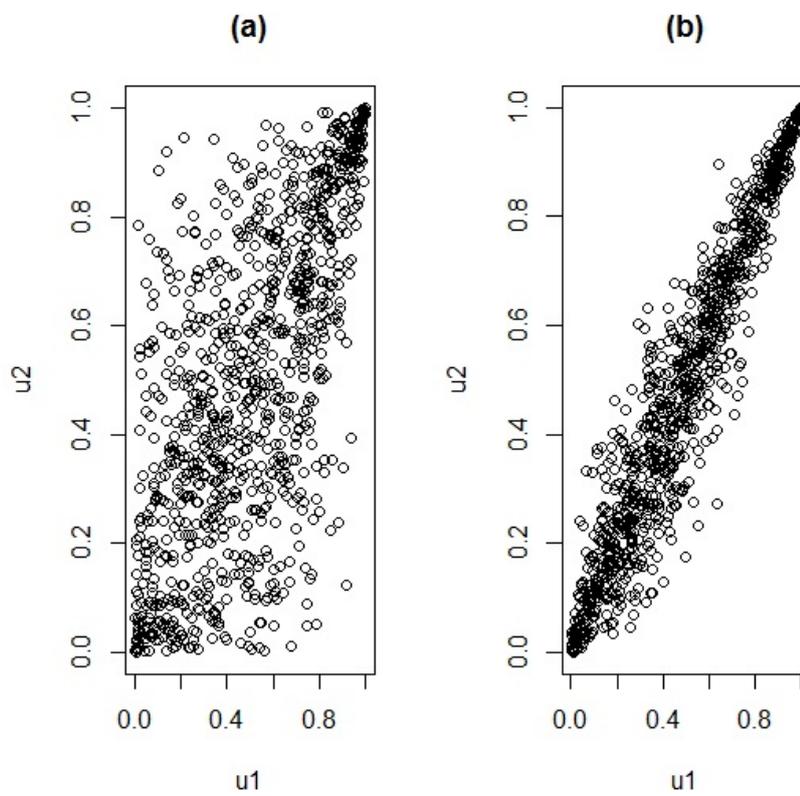


Figura 3.4: Gráficos de dispersión del rango de una cópula Gumbel bivariada. (a): $\theta = 2$; (b): $\theta = 6$

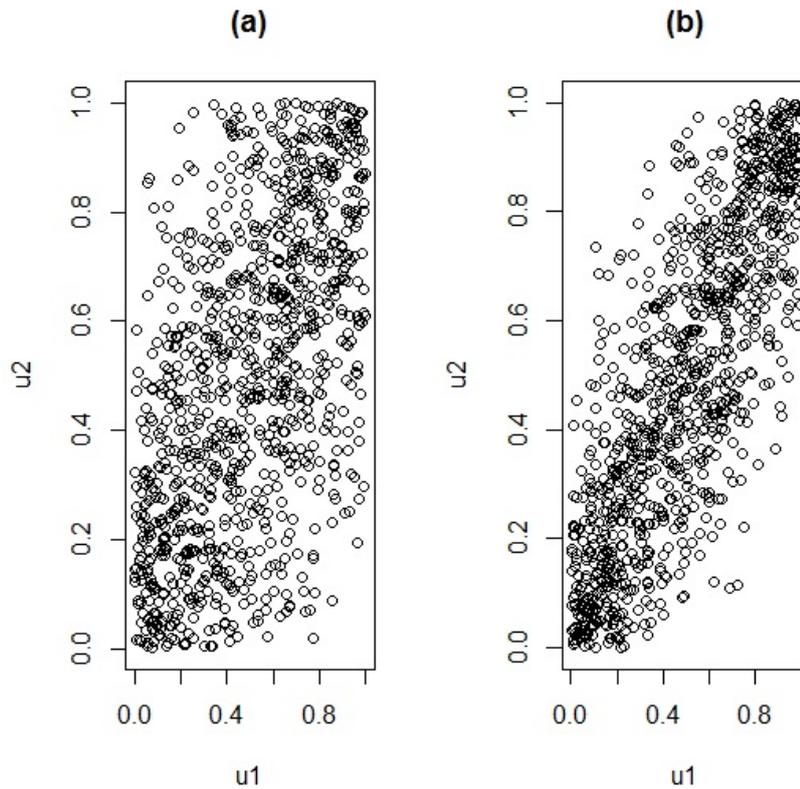


Figura 3.5: Gráficos de dispersión del rango de una cópula Frank bivariada. (a): $\theta = 4$; (b): $\theta = 8$

Cópula Frank

Con función generadora $\varphi(t) = -\log \left[\frac{\exp(\theta t) - 1}{\exp(\theta) - 1} \right]$ y cópula dada por

$$C(u_1, u_2) = -\frac{1}{\theta} \log \left[1 + \frac{(\exp(\theta u_1) - 1)(\exp(\theta u_2) - 1)}{\exp(\theta) - 1} \right],$$

donde $\theta \in \mathbb{R} \setminus \{0\}$, la cópula Frank es una cópula arquimediana simétrica. La cópula densidad de la cópula Frank es la siguiente:

$$c(u_1, u_2) = \frac{\theta(\exp(-\theta) - 1)\exp((-\theta)(u_1 + u_2))}{[\exp(-\theta) - 1 + (\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)]}$$

El tau de Kendall de la cópula Frank es el siguiente:

$$\tau(\theta) = 1 - \frac{4}{\theta} + 4 \frac{1}{\theta} \int_0^\theta \frac{c/\theta}{\exp(x) - 1} dx.$$

La Figura 3.5 muestra los gráficos de dispersión del rango de una cópula Frank bivariada (1,000 simulaciones) para diferentes valores de θ .

Cópula Joe

Con función generadora $\varphi(t) = -\log [1 - (1 + t)^\theta]$ y cópula dada por

$$C(u_1, u_2) = 1 - [(1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta (1 - u_2)^\theta]^{1/\theta},$$

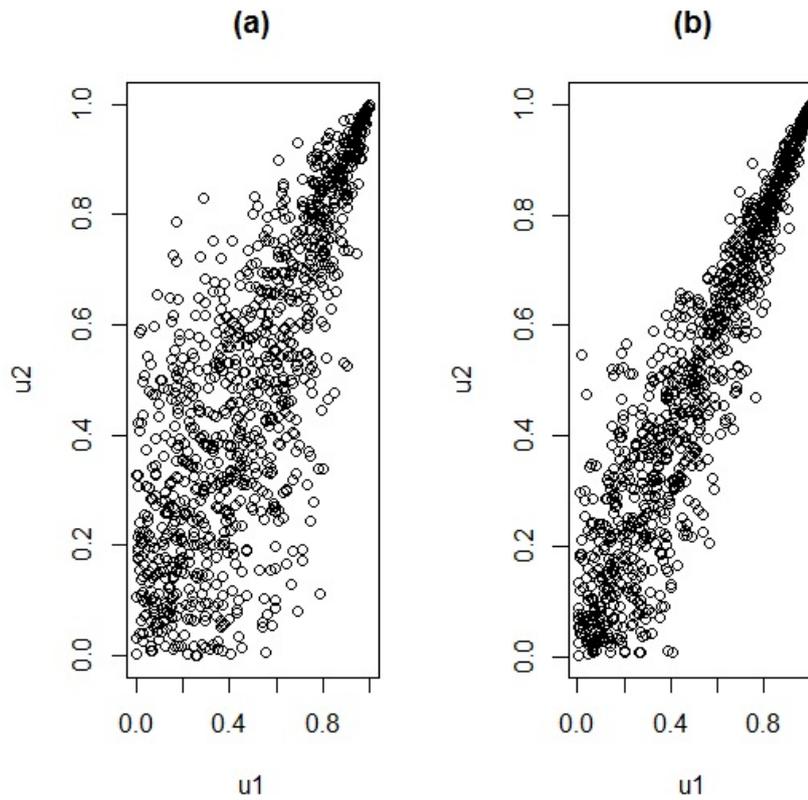


Figura 3.6: Gráficos de dispersión del rango de una cópula Joe bivariada. (a): $\theta = 4$; (b): $\theta = 8$

para $\theta > 1$, la cópula Joe es una cópula arquimediana asimétrica. La cópula densidad de la cópula Joe es la siguiente:

$$c(u_1, u_2) = [(1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta (1 - u_2)^\theta]^{\frac{1}{\theta} - 2} \times (1 - u_1)^{\theta - 1} (1 - u_2)^{\theta - 1} \times [\theta - 1 + (1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta (1 - u_2)^\theta].$$

El tau de Kendall de la cópula Joe es el siguiente:

$$\tau(\theta) = 1 + \frac{4}{\theta^2} \int_0^1 t \log(t) (1 - t)^{\frac{2(1-\theta)}{\theta}} dt.$$

La Figura 3.6 muestra los gráficos de dispersión del rango de una cópula Joe bivariada (1,000 simulaciones) para diferentes valores de θ .

Cópula Clayton-Gumbel (BB1)

Introducida por Joe (1997), la cópula BB1 es una cópula arquimediana de dos parámetros, la cual puede ser considerada como la generalización de las familias Clayton y Gumbel. Con función generadora $\varphi(t) = [t^{-\theta} - 1]^\delta$ y cópula dada por

$$C(u_1, u_2) = [1 + [(u_1^{-\theta} - 1)^\delta + (u_2^{-\theta} - 1)^\delta]^\frac{1}{\delta}]^{-\frac{1}{\theta}},$$

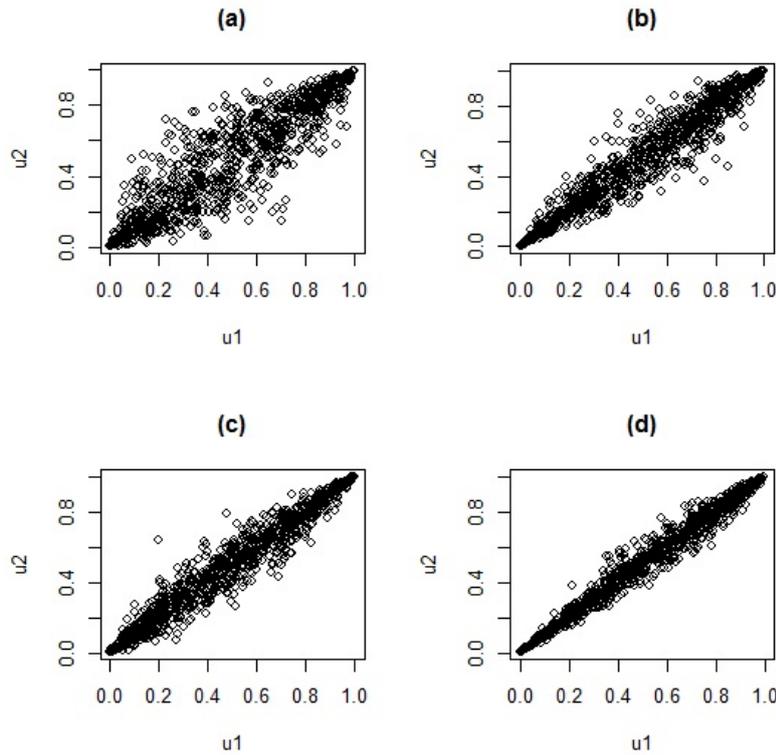


Figura 3.7: Gráficos de dispersión del rango de una cópula Clayton-Gumbel (BB1). (a) $\theta = 0.5$, $\delta = 3$; (b) $\theta = 2$, $\delta = 3$; (c) $\theta = 0.5$, $\delta = 6$; (d) $\theta = 2$, $\delta = 6$

para $\theta > 0$ y $\delta \geq 1$. La cópula densidad de la cópula BB1 es la siguiente:

$$c(u_1, u_2) = [1 + [(u_1^{-\theta} - 1)^\delta + (u_2^{-\theta} - 1)^\delta]^{\frac{1}{\delta}}]^{-\frac{1}{\theta}} \\ \times [(u_1^{-\theta} - 1)^\delta + (u_2^{-\theta} - 1)^\delta]^{\frac{2}{\delta} - 2} \\ \times [\theta\delta + 1 + \theta(\delta - 1)[(u_1^{-\theta} - 1)^\delta + (u_2^{-\theta} - 1)^\delta]^{\frac{1}{\delta}}] \\ \times (u_1^{-\theta} - 1)^{\delta-1} u_1^{-\theta-1} (u_2^{-\theta} - 1)^{\delta-1} u_2^{-\theta-1}.$$

El tau de Kendall de la cópula BB1 es el siguiente:

$$\tau(\theta, \delta) = 1 - \frac{2}{\delta(\theta + 2)}.$$

La Figura 3.7 muestra los gráficos de dispersión del rango de una cópula Clayton-Gumbel (BB1) (1,000 simulaciones) para diferentes valores de θ y δ .

Cópula Joe-Gumbel (BB6)

Introducida por Joe (1997), la cópula BB6 es una cópula arquimediana de dos parámetros, la cual puede ser considerada como la generalización de las familias Joe y Gumbel. Con función generadora $\varphi(t) = (-\log[1 - (1-t)^\theta])^\delta$ sus parámetros son $\theta > 0$ y $\delta \geq 1$. El tau de Kendall de la cópula BB6 es el siguiente:

$$\tau(\theta, \delta) = 1 + 4 \int_0^1 \left(-\log(-(1-t)^\theta + 1) \frac{(1-t - (1-t)^\theta + t(1-t)^{-\theta})}{\delta\theta} \right) dt$$

La Figura 3.8 muestra los gráficos de dispersión del rango de una cópula Joe-Gumbel (BB6) (1,000 simulaciones) para diferentes valores de θ y δ .

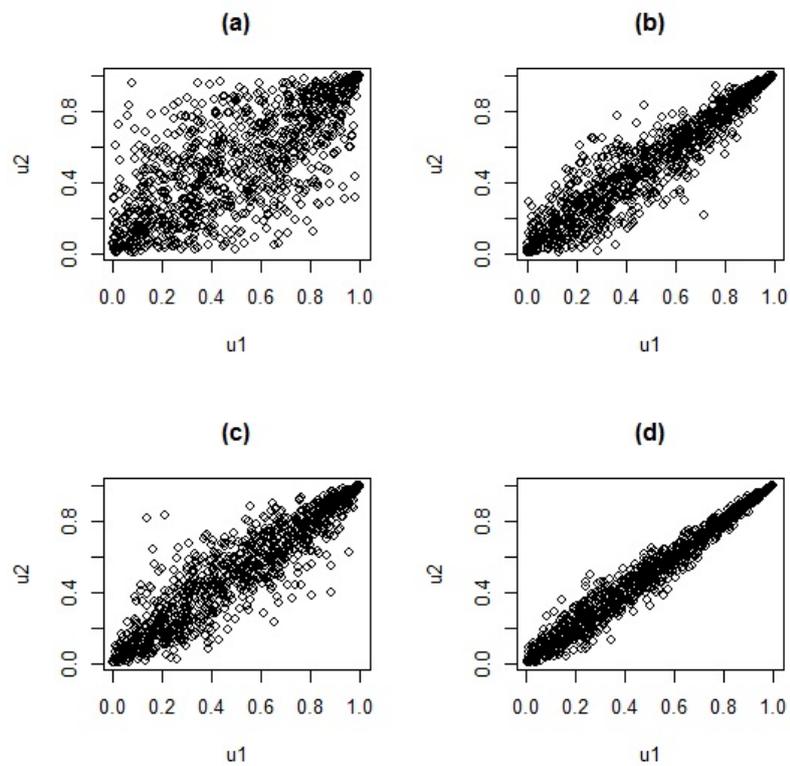


Figura 3.8: Gráficos de dispersión del rango de una cópula Joe-Gumbel (BB6). (a) $\theta = 0.5$, $\delta = 3$; (b) $\theta = 2$, $\delta = 3$; (c) $\theta = 0.5$, $\delta = 6$; (d) $\theta = 2$, $\delta = 6$

Cópula Joe-Clayton (BB7)

Introducida también por Joe (1997), la cópula BB7 es una cópula arquimediana de dos parámetros, la cual puede ser considerada como la generalización de las familias Joe y Clayton. Con función generadora $\varphi(t) = [1 - (1 - t)^\theta]^{-\delta}$ y cópula dada por

$$C(u_1, u_2) = 1 - [1 - [(1 - (1 - u_1)^\theta)^{-\delta} + (1 - (1 - u_2)^\theta)^{-\delta}]^{-\frac{1}{\delta}}]^{-\frac{1}{\theta}},$$

para $\theta \geq 1$ y $\delta > 0$. La cópula densidad de la cópula BB7 es la siguiente:

$$\begin{aligned} c(u_1, u_2) &= \left(-\frac{1}{\theta}\right) \left(\frac{1}{\delta} - 1\right) \cdot h^{\frac{1}{\theta}-2} du_2 h \cdot du_1 h - \frac{1}{\theta} \cdot h^{\frac{1}{\theta}-1} du_1 u_2 h \\ h &= 1 - ((1 - (1 - u_1)^\theta)^{-\delta} + (1 - (1 - u_2)^\theta)^{-\delta} - 1)^{\frac{1}{\delta}} \\ du_1 h &= -\theta((1 - (1 - u_1)^\theta)^{-\delta} - (1 - (1 - u_2)^\theta)^{-\delta} - 1)^{\frac{1}{\delta}-1} \\ &\quad \times (1 - (1 - u_1)^\theta)^{-\delta-1} (1 - u_1)^{\theta-1} \\ du_2 h &= -\theta((1 - (1 - u_1)^\theta)^{-\delta} - (1 - (1 - u_2)^\theta)^{-\delta} - 1)^{\frac{1}{\delta}-1} \\ &\quad \times (1 - (1 - u_2)^\theta)^{-\delta-1} (1 - u_2) \\ du_1 u_2 h &= \frac{1}{\delta} \left(-\frac{1}{\delta} - 1\right) ((1 - (1 - u_1)^\theta)^{-\delta} - (1 - (1 - u_2)^\theta)^{-\delta} - 1)^{\frac{1}{\delta}-2} \cdot du_1 S \cdot du_2 S \\ du_1 S &= -\theta\delta(1 - (1 - u_1)^\theta)^{-\delta-1} (1 - u_1)^{\theta-1} \\ du_2 S &= -\theta\delta(1 - (1 - u_2)^\theta)^{-\delta-1} (1 - u_2)^{\theta-1} \end{aligned}$$

El tau de Kendall de la cópula BB7 es el siguiente:

$$\tau(\theta, \delta) = 1 - \frac{2}{\delta(2 - \theta)} + \frac{4}{\theta^2 \delta} B\left(\frac{2 - 2\theta}{\theta} + 1, \delta + 2\right),$$

donde

$$B(x, y) = \int_0^1 t^{x+1} (1 - t)^{y-1} dt.$$

La Figura 3.9 muestra los gráficos de dispersión del rango de una cópula Joe-Clayton (BB7) (1,000 simulaciones) para diferentes valores de θ y δ .

Cópula Joe-Frank (BB8)

Introducida por Joe (1997), la cópula BB8 es una cópula arquimediana de dos parámetros, la cual puede ser considerada como la generalización de las familias Joe y Frank. Con función generadora $\varphi(t) = -\log \left[\frac{1 - (1 - \delta t)^\theta}{1 - (1 - \delta)^\theta} \right]$ para $\theta \geq 1$ y $0 < \delta \leq 1$, el tau de Kendall de la cópula BB8 es el siguiente:

$$\tau(\theta, \delta) = 1 + 4 \int_0^1 \left(-\log \left(\frac{(1 - t\delta)^\theta - 1}{(1 - \delta)^\theta - 1} \right) \frac{1 - t\delta - (1 - t\delta)^{-\theta} + t\delta(1 - t\delta)^{-\theta}}{\theta\delta} \right) dt.$$

La Figura 3.10 muestra los gráficos de dispersión del rango de una cópula Joe-Frank (BB8) (1,000 simulaciones) para diferentes valores de θ y δ .

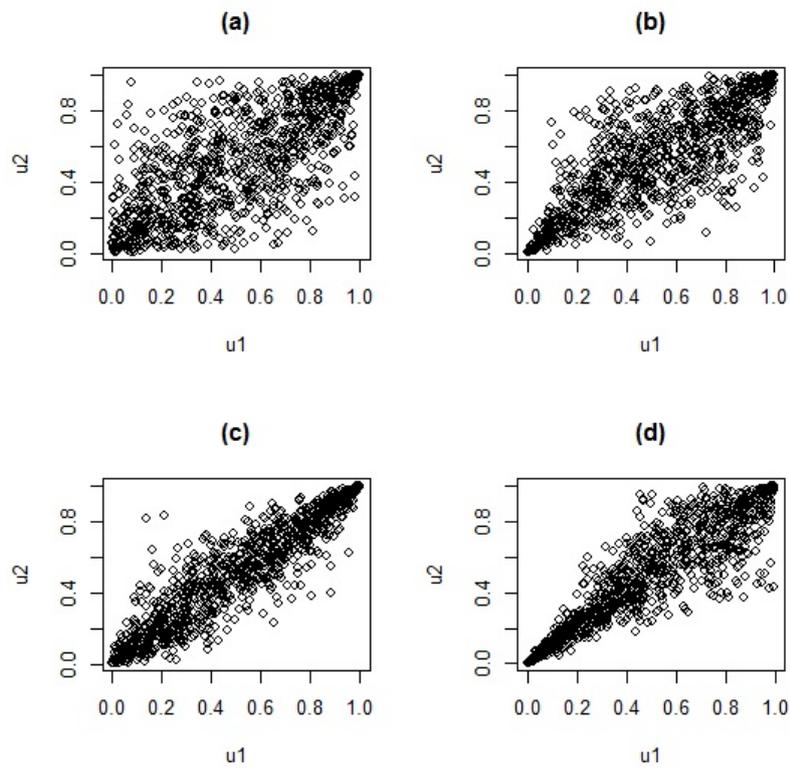


Figura 3.9: Gráficos de dispersión del rango de una cópula Joe-Clayton (BB7). (a) $\theta = 0.5$, $\delta = 3$; (b) $\theta = 2$, $\delta = 3$; (c) $\theta = 0.5$, $\delta = 6$; (d) $\theta = 2$, $\delta = 6$

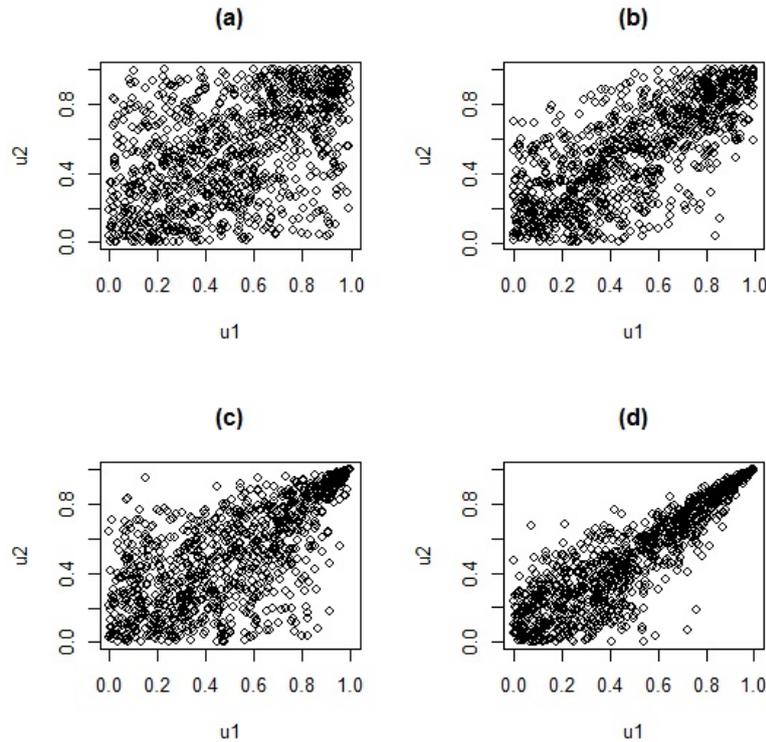


Figura 3.10: Gráficos de dispersión del rango de una cópula Joe-Frank (BB8). (a) $\theta = 3$, $\delta = 0.7$; (b) $\theta = 6$, $\delta = 0.7$; (c) $\theta = 3$, $\delta = 1$; (d) $\theta = 6$, $\delta = 1$

Cóputas rotadas

Las funciones de distribución C_{90} , C_{180} y C_{270} de una cópula rotada 90, 180 y 270 grados respectivamente son, de acuerdo al teorema 2.4.4 en Nelsen (2005), las siguientes:

$$C_{90}(u_1, u_2) = u_2 - C(1 - u_1, u_2),$$

$$C_{180}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2),$$

$$C_{270}(u_1, u_2) = u_1 - C(u_1, 1 - u_2).$$

Cuando la cópula es rotada 180° se obtiene la llamada cópula de sobrevivencia (vista en la sección 3.3); mientras que para modelar la dependencia negativa se pueden utilizar las cópulas rotadas 90° o 270° .

De esta forma, existen una gran familia de cópulas bivariadas arquimedianas que pueden ser utilizadas para modelar los datos.

La Figura 3.11 muestra los gráficos de dispersión del rango de una cópula Joe (1,000 simulaciones) rotar y rotada 0, 90, 180 y 270 grados.

3.7 Cóputas en pares, vine regular y simulación

Como se mencionó anteriormente, los recientes trabajos empíricos acerca de cópulas se han centrado en capturar la estructura de interdependencia bivariada. La razón es que la complejidad de las cópulas multivariadas se incrementa rápidamente a medida que aumenta la dimensión del problema. Las cópulas estándar multivariadas como la multivariada Gaussiana, la t-Student o las cópulas multivariadas arquimedianas carecen de flexibilidad para modelar adecuadamente la dependencia entre grandes número de variables. Las vine no sufren de ese problema pues modelan los datos usando una sucesión de cópulas bivariadas, llamadas cópulas en pares.

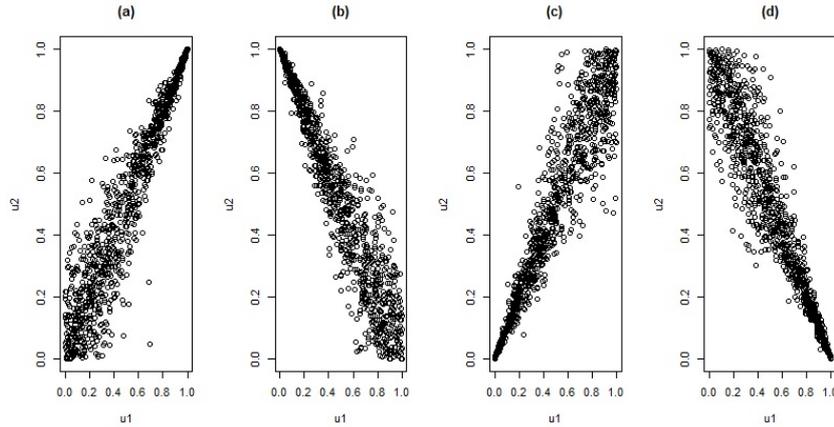


Figura 3.11: Gráficos de dispersión del rango de una cópula Joe rotada 0, 90, 180 y 270 grados. Parámetros correspondientes a los valores del tau de Kendall de 0.8 para la dependencia positiva y -0.8 para la dependencia negativa

3.7.1 Cópulas en pares

Al usar cópulas bivariadas se puede construir funciones de distribución multivariadas. Para ello, se debe especificar la dependencia y dependencia condicional de las variables aleatorias así como las funciones de distribución de todas las marginales. A esta construcción en bloque de cópulas bivariadas Bedford y Cooke (2001, 2002) las llamaron cópulas en pares.

Siguiendo a Aas, Czado, Frigessi y Bakken (2009), considere n variables aleatorias $\mathbf{X} = (X_1, \dots, X_n)$ con función de densidad conjunta $f(x_1, \dots, x_n)$. Sabemos que esta densidad puede ser factorizada como

$$f(x_1, \dots, x_n) = f(x_n) \cdot f(x_{n-1} | x_n) \cdots f(x_1 | x_2, \dots, x_n),$$

donde $f(x_{n-1} | x_n)$ es la función de densidad de x_{n-1} dado x_n .

Por otro lado, según el teorema de Sklar, cada distribución multivariada F con marginales F_1, \dots, F_n puede ser escrito como

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

para alguna cópula n -dimensional apropiada. Además, para alguna distribución absolutamente continua F con densidades marginales continuas y crecientes F_1, \dots, F_n se tiene

$$f(x_1, \dots, x_n) = c_{12 \dots n}(F_1(x_1), \dots, F_n(x_n)) \cdot f_1(x_1) \cdots f_n(x_n)$$

para alguna cópula densidad n -variada $c_{12 \dots n}(\cdot)$. En el caso bivariado la expresión anterior será la siguiente

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2),$$

donde $c_{12}(\cdot, \cdot)$ es la densidad de una apropiada cópula en par para las variables $F_1(x_1)$ y $F_2(x_2)$. Para la densidad condicional tenemos

$$f(x_1 | x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1)$$

para la misma cópula par c_{12} .

Extendiendo esta argumentación para el caso de tres variables aleatorias X_1, X_2 y X_3 tenemos

$$f(x_1 | x_2, x_3) = c_{12|3}(F_{1|3}(x_1 | x_3), F_{2|3}(x_2 | x_3)) \cdot f(x_1 | x_3),$$

para la apropiada cópula en par $c_{12|3}$ aplicado a las variables transformadas $F(x_1 | x_3)$ y $F(x_2 | x_3)$. No obstante, una descomposición alternativa es la siguiente

$$f(x_1 | x_2, x_3) = c_{13|2}(F_{1|2}(x_1 | x_2), F_{3|2}(x_3 | x_2)) \cdot f(x_1 | x_2),$$

donde $c_{13|2}$ es diferente de la cópula en par $c_{12|3}$. Descomponiendo $f(x_1 | x_2)$ en la ecuación anterior tenemos

$$f(x_1 | x_2, x_3) = c_{13|2}(F_{1|2}(x_1 | x_2), F_{3|2}(x_3 | x_2)) \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1),$$

donde ahora aparecen dos cópulas en pares.

De esta manera, cada término en

$$f(x_1, \dots, x_n) = f(x_n) \cdot f(x_{n-1} | x_n) \cdots f(x_1 | x_2, \dots, x_n)$$

puede ser descompuesto en marginales y adecuadas cópulas bivariadas usando la fórmula general

$$f(x | v) = c_{xv|v_{-j}}(F(x | v_{-j}), F(v_j | v_{-j})) \cdot f(x | v_{-j})$$

para un vector d -dimensional. v_j es un componente arbitrario de v y v_{-j} es el vector v excluyendo el componente v_j . De esta manera, bajo las condiciones adecuadas de regularidad, una densidad multivariada puede ser expresada como el producto de cópulas en pares.

La construcción de las cópulas en pares involucra distribuciones marginales condicionales de la forma $F(x | v)$. Para cada j , Joe (1996) mostró que

$$F(x | v) = \frac{\partial C_{xv|v_{-j}}(F(x | v_{-j}), F(v_j | v_{-j}))}{\partial F(v_j | v_{-j})},$$

donde $C_{ij|k}$ es una función de distribución de una cópula bivariada.

Aas, Czado, Frigessi y Bakken (2009) sostienen que hay un número significativo de posibles construcciones de cópula en pares cuando el número de variables aumenta. Por ejemplo, los autores mencionan que existen 240 diferentes construcciones para una densidad 5-dimensional. Por ello, se concentran en dos casos especiales de vines regulares que reducen significativamente el número de posibles construcciones: C-vine y D-vine.

3.7.2 Vines regular

Bedford y Cooke (2001, 2002) y Kurowicka y Cooke (2006) introdujeron a los llamados vines regulares como un método gráfico para representar la multitud de posibles descomposiciones de cópulas en pares. Un vine regular sobre d variables es un conjunto anidado de árboles, donde dos bordes en el árbol j están unidos por un borde en el árbol $j + 1$ solo si esos bordes comparten un nodo en común. La definición formal de una vine regular es la siguiente:

Definición 47 (Vine regular). \mathcal{V} es una vine regular de d elementos si

- (i) $\mathcal{V} = (T_1, \dots, T_{d-1})$.
- (ii) $T_1 = (N_1, E_1)$ es un árbol con nodos $N_1 = \{1, \dots, d\}$. Para $i = 2, \dots, n - 1$, $T_i = (N_i, E_i)$ es un árbol con nodos $N_i = E_{i-1}$.
- (iii) Para $i = 2, \dots, d - 1$, si $\{a, b\} \in E_i$, donde $a = \{a_1, a_2\}$ y $b = \{b_1, b_2\}$, entonces exactamente uno de los a_j es igual a uno de los b_j

La propiedad (iii) es conocida como la condición de proximidad pues expresa el hecho que dos nodos son adyacentes en el árbol T_j solo si los correspondientes bordes en el árbol T_{j-1} son adyacentes, es decir, comparten el mismo nodo.

Definición 48 (C-vine y D-vine). Una vine regular \mathcal{V} de d elementos es llamada un

- (i) C-vine si cada árbol T_i , $i = 1, \dots, d - 1$, es un inicio, es decir, si cada árbol T_i tiene un único nodo de grado $d - i$, llamado el nodo raíz.
- (ii) D-vine si T_1 es una senda, es decir, si cada nodo en T_1 tiene un grado de como máximo 2.

De esta manera, en el caso de un C-vine, en el primer árbol, la dependencia con respecto a una variable en particular, la primera raíz, es modelada usando cópulas bivariadas para cada par de variables. Luego, condicionada sobre esta variable, las dependencias en pares con respecto a la segunda variable es modelada, el segundo nodo. En general, un nodo es escogido en cada árbol y todas las dependencias en pares con respecto a ese nodo son modeladas condicionadas a todos los nodos previos.

Por otro lado, las D-vine son construidas escogiendo un orden específico de las variables. De este modo, en el primer árbol, la dependencia de la primera y la segunda variable, de la segunda y la tercera, de la tercera y la cuarta, de la $n-1$ -ésima y la n -ésima variable son modeladas usando cópulas en pares. En el segundo árbol, la dependencia condicionada de la primera y la tercera dada la segunda variable es modelada. Luego, es modelada la segunda y la cuarta dada la tercera y así sucesivamente.

Bedford y Cooke (2001) mostraron la forma de la densidad de una distribución d -dimensional en términos de una vine regular. Posteriormente, Aas, Czado, Frigessi y Bakken (2009) mostraron la forma de esta densidad en el caso particular de una C-vine y D-vine.

Así, la densidad de una distribución d -dimensional $f(x_1, \dots, x_d)$ correspondiente a una D-vine puede ser escrita de la siguiente manera:

$$f(\mathbf{x}) = \prod_{k=1}^d f_k(x_k) \times \prod_{i=1}^{d-1} \prod_{j=1}^{d-i} c_{j,j+i|(j+1):(j+i-1)},$$

donde

$$c_{j,j+i|(j+1):(j+i-1)} := c_{j,j+i|(j+1):(j+i-1)}(F(x_j | x_{j+1}, \dots, x_{j+i-1}), F(x_{j+i} | x_{j+1}, \dots, x_{j+i-1}) | \theta_{j,j+i|(j+1):(j+i-1)}),$$

donde f_k , $k = 1, \dots, d$, representa las densidades marginales y $c_{j,j+i|(j+1):(j+i-1)}$ las densidades de las cópulas bivariadas con parámetros $\theta_{j,j+i|(j+1):(j+i-1)}$.

Por otro lado, la densidad de una distribución d -dimensional $f(x_1, \dots, x_d)$ correspondiente a una C-vine puede ser escrita de la siguiente manera:

$$f(\mathbf{x}) = \prod_{k=1}^d f_k(x_k) \times \prod_{i=1}^{d-1} \prod_{j=1}^{d-i} c_{i,i+j|1:(i-1)},$$

donde

$$c_{i,i+j|1:(i-1)} := c_{i,i+j|1:(i-1)}(F(x_i | x_1, \dots, x_{i-1}), F(x_{i+j} | x_1, \dots, x_{i-1}) | \theta_{i,i+j|1:(i-1)}),$$

donde f_k , $k = 1, \dots, d$, representa las densidades marginales y $c_{i,i+j|1:(i-1)}$ las densidades de las cópulas bivariadas con parámetros $\theta_{i,i+j|1:(i-1)}$.

Ejemplo 49. Si deseamos trabajar con 4 variables, entonces será necesario estimar $4(4-1)/2=6$ cópulas bivariadas en $4-1=3$ árboles. Esto se puede ver en la Figura 3.12, donde se esquematiza el proceso de estimación de las cópulas bivariadas para las cuatro variables en el ejemplo.

De esta manera, la densidad de una C-vine puede ser escrita de la siguiente manera:

$$\begin{aligned} f(x_1, x_2, x_3, x_4) &= f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot f(x_4) \\ &\cdot c_{12}(F(x_1), F(x_2)) \cdot c_{13}(F(x_1), F(x_3)) \cdot c_{14}(F(x_1), F(x_4)) \\ &\cdot c_{23|1}(F(x_2|x_1), F(x_3|x_1)) \cdot c_{24|1}(F(x_2|x_1), F(x_4|x_1)) \\ &\cdot c_{34|12}(F(x_3|x_1, x_2), F(x_4|x_1, x_2)) \end{aligned}$$

Por otro lado, la densidad de la D-vine puede ser escrita de la siguiente manera:

$$\begin{aligned} f(x_1, x_2, x_3, x_4) &= f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot f(x_4) \\ &\cdot c_{12}(F(x_1), F(x_2)) \cdot c_{23}(F(x_2), F(x_3)) \cdot c_{34}(F(x_3), F(x_4)) \\ &\cdot c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot c_{24|3}(F(x_2|x_3), F(x_4|x_3)) \\ &\cdot c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)) \end{aligned}$$

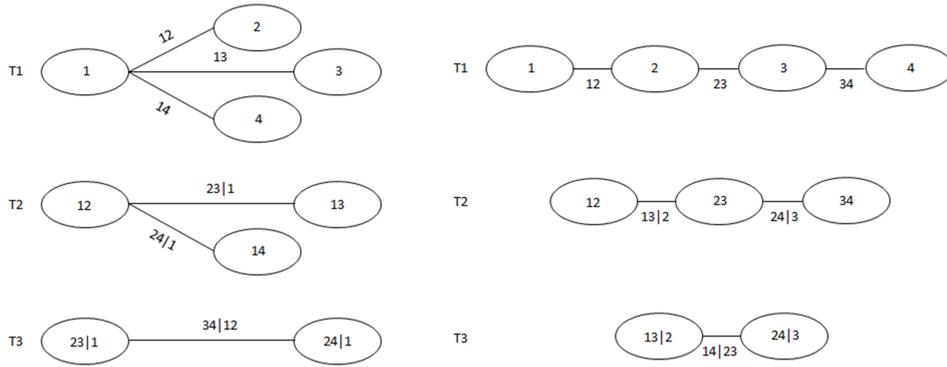


Figura 3.12: Ejemplo de árboles C-vine (izquierda) y D-vine (derecha) de dimensión 4

3.7.3 Simulación

En esta sección se muestra el procedimiento para realizar las simulaciones de muestras utilizando las cópulas en pares mostradas en Aas, Czado, Frigessi y Bakken (2009). Primero se muestra el caso bivariado, luego se muestran las simulaciones para el C-vine y D-vine. Se asume por simplicidad que las marginales de las variables de interés son uniformes.

Simulando una muestra bivariada

Primero, se crea una muestra $w_i; i = 1, 2$ de variables aleatorias independientes con distribución uniforme en $[0, 1]$. Entonces,

$$\begin{aligned} x_1 &= w_1 \\ x_2 &= F_{2|1}^{-1}(w_2 | x_1) = h^{-1}(w_2, x_1, \Theta_{11}), \end{aligned}$$

donde Θ_{11} es el parámetro estimado de una determinada cópula bivariada, $h(x_2, x_1, \Theta_{11})$ representa la función de distribución condicional de x_2 dado x_1 y $h^{-1}(w_2, x_1, \Theta_{11})$ es el valor de x_2 .

Simulando una C-vine y D-vine

Primero, se crea una muestra $w_i; i = 1, \dots, n$ de variables aleatorias independientes con distribución uniforme en $[0, 1]$. Entonces,

$$\begin{aligned} x_1 &= w_1 \\ x_2 &= F_{2|1}^{-1}(w_2 | x_1) \\ x_3 &= F_{3|1,2}^{-1}(w_3 | x_1, x_2) \\ &\dots = \dots \\ x_n &= F_{n|1,2,\dots,n-1}^{-1}(w_n | x_1, \dots, x_{n-1}). \end{aligned}$$

La determinación de la función $F_{n|1,2,\dots,n-1}^{-1}$ es diferente dependiendo si se trata de una C-vine o D-vine. Para el caso de un C-vine, tenemos lo siguiente

$$F_{n|1,2,\dots,n-1}(x_j | x_1, \dots, x_{j-1}) = \frac{\partial C_{j,j-1|1,\dots,j-2}(F(x_j | x_1, \dots, x_{j-2}), F(x_{j-1} | x_1, \dots, x_{j-2}))}{\partial F(x_{j-1} | x_1, \dots, x_{j-2})},$$

mientras que para un D-vine tenemos

$$F_{n|1,2,\dots,n-1}(x_j | x_1, \dots, x_{j-1}) = \frac{\partial C_{j,1|2,\dots,j-1}(F(x_j | x_2, \dots, x_{j-1}), F(x_1 | x_2, \dots, x_{j-1}))}{\partial F(x_1 | x_2, \dots, x_{j-1})}.$$

3.8 Herramientas para el análisis exploratorio de los datos

3.8.1 Test de independencia

Como paso previo al análisis exploratorio de los datos, un inicio útil es ver si los datos son independientes desde el punto de vista estadístico. Para ello, Genest y Favre (2007) proponen el uso de un test de independencia bivariada basado en el tau de Kendall. El test explota la normalidad asintótica de la siguiente expresión:

$$T = \sqrt{\frac{9n(n-1)}{2(2n+5)}} \times |\hat{\tau}|,$$

donde n es el número de observaciones y $\hat{\tau}$ el tau de empírico de Kendall. El p -value de la hipótesis nula de independencia bivariada es asintóticamente igual a p -value = $2 \times (1 - \Phi(T))$, donde Φ es la función de distribución normal estándar.

3.8.2 Chi plot

Propuesto por Fisher y Switzer (1985, 2001), el chi plot es un diagrama de dispersión que sirve para investigar posibles asociaciones de datos bivariados. El gráfico está diseñado de tal forma que éste es aproximadamente horizontal cuando los datos son independientes.

Sean $(x_1, y_1), \dots, (x_n, y_n)$ una muestra aleatoria de la función de distribución conjunta H de un par de variables aleatorias (X, Y) y sea $I(A)$ la función indicador del evento A . Para cada par de puntos (x_i, y_i) sean las siguientes expresiones que dependen de los rangos de las observaciones

$$H_i = \sum_{j \neq i} I(x_j \leq x_i, y_j \leq y_i) / (n-1),$$

$$F_i = \sum_{j \neq i} I(x_j \leq x_i) / (n-1),$$

$$G_i = \sum_{j \neq i} I(y_j \leq y_i) / (n-1)$$

y

$$S_i = \text{sign} \left\{ \left(F_i - \frac{1}{2} \right), \left(G_i - \frac{1}{2} \right) \right\}.$$

Cos estas expresiones, Fisher y Switzer (2001) proponen un gráfico de dispersión de los pares (λ_i, χ_i) , donde:

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}}$$

$$\lambda_i = 4 \text{sign}(\tilde{F}_i \tilde{G}_i) \cdot \text{máx}(\tilde{F}_i^2, \tilde{G}_i^2),$$

donde $\tilde{F}_i = F_i - 0.5$ y $\tilde{G}_i = G_i - 0.5$ para $i = 1, \dots, n$. λ_i puede ser interpretado como la distancia de un punto (x_i, y_i) al centro del conjunto de datos; mientras que χ_i puede ser visto como el test estadístico χ^2 para comprobar la independencia de una tabla 2×2 . De esta manera, el par (λ_i, χ_i) tenderá a estar localizado por encima de cero para datos bivariados con dependencia positiva y viceversa para dependencia negativa. Finalmente, dado que uno podría esperar que $H_i \approx F_i \times G_i$ para todo i cuando F_i y G_i son independientes, los valores de χ_i que caen lejos de cero, son indicios desviaciones de esta hipótesis. Para ello, Fisher y Switzer (1985, 2001) sugirieron adicionar al gráfico las líneas horizontales de control $\pm c_p/n^{\frac{1}{2}}$, que representan la desviación estándar asintótica de χ_i bajo independencia. c_p es seleccionado de modo que aproximadamente 100p% de los pares (λ_i, χ_i) caen entre las líneas.

La Figura 3.13 muestra en la primera columna los valores simulados de un vector aleatorio normal con diferentes parámetros de correlación ρ y en la segunda columna, los respectivos gráficos χ -plot.

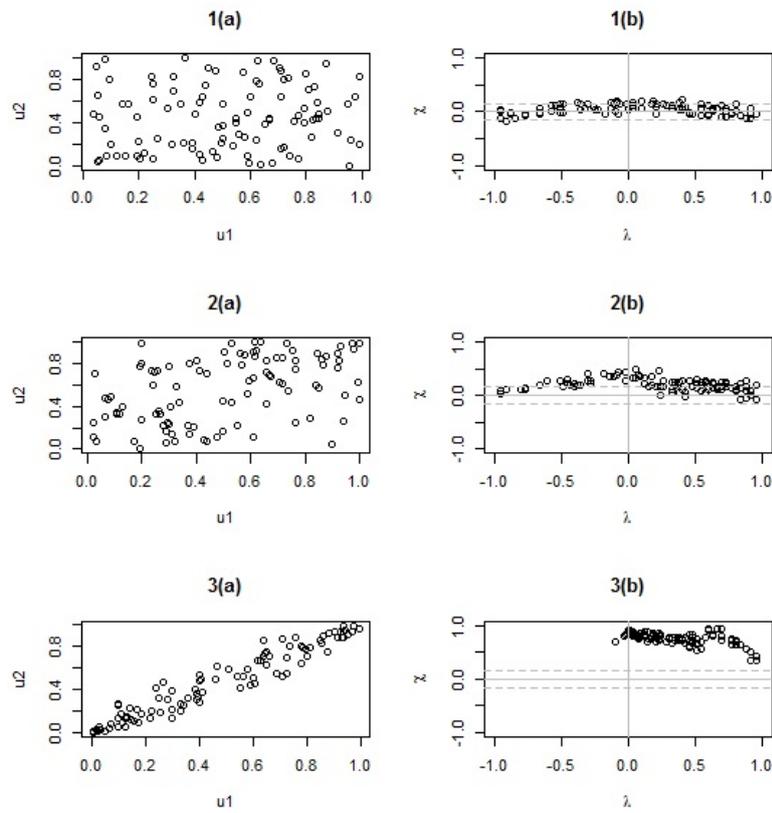


Figura 3.13: χ -plot basados en muestras aleatorias de tamaño $n = 100$ de un vector normal bivariado. 1 (a) $\rho = 0.05$; 1 (b) χ -plot. 2 (a) $\rho = 0.50$; 2 (b) χ -plot. 3 (a) $\rho = 0.95$; 3 (b) χ -plot

3.8.3 Kendall plot (K-plot)

Otra herramienta gráfica que depende de los rangos de las observaciones es el propuesto por Genest y Boies (2003) llamado Kendall plot o en forma abreviada K-plot. Sea $(x_1, y_1), \dots, (x_n, y_n)$ una muestra aleatoria de la función de distribución conjunta H de un par de variables aleatorias (X, Y) . Para la construcción del K-plot se deben seguir los siguientes pasos:

1. Para cada $1 \leq i \leq n$, calcular $H_i = \sum_{j \neq i} I(x_j \leq x_i, y_j \leq y_i) / (n - 1)$.
2. Ordenar los H_i para obtener $H_1 \leq \dots \leq H_n$.
3. Dibujar los pares $(W_{i:n}, H_i)$, $1 \leq i \leq n$, donde $W_{i:n}$ representa el valor esperado del estadístico de orden i -ésimo en una muestra aleatoria de tamaño n . Esto viene de la distribución K_0 de los H_i bajo la hipótesis nula de independencia.

Para todo $1 \leq i \leq n$, $W_{i:n}$ puede ser calculado de la siguiente manera:

$$W_{i:n} = n \binom{n-1}{i-1} \int_0^1 \omega (K_0(\omega))^{i-1} (1 - K_0(\omega))^{n-i} dK_0(\omega),$$

donde $K_0(\omega) = \omega - \omega \log(\omega)$.

Si los puntos del K-plot caen aproximadamente sobre la diagonal, entonces X e Y son aproximadamente independientes. Cualquier desviación de la diagonal implica dependencia. Habría dependencia positiva o negativa en los datos si los puntos del K-plot están localizados por encima o debajo de la diagonal $y = x$. En particular, mientras mayor sea la desviación de la diagonal, mayor será el grado de dependencia. Finalmente, si existe una perfecta dependencia positiva, entonces los puntos $(W_{i:n}, H_i)$ estarán en la curva $K_0(\omega)$ localizada por encima de la diagonal. Por otro lado, si los puntos se ubican encima del eje x , esto indica que existe una perfecta dependencia negativa entre X e Y .

La Figura 3.14 muestra en la primera columna los valores simulados de un vector aleatorio normal con diferentes parámetros de correlación ρ y en la segunda columna, los respectivos gráficos K-plot.

3.8.4 λ - función

Propuesto por Genest y Rivest (1993) como un medio para seleccionar una cópula que represente la estructura de dependencia entre dos variables, la λ - función está definida de la siguiente manera:

$$\lambda(v, \theta) = v - K(v, \theta)$$

$$K(v, \theta) = P(C_\theta(u_1, u_2) \leq v), \quad v \in [0, 1].$$

Por otro lado, para las cópulas arquimedianas, Genest y Rivest (1993) determinaron explícitamente la λ - función en términos de su función generadora y su derivada dada por

$$\lambda(v, \theta) = \frac{\varphi(v)}{\varphi'(v)}.$$

De esta manera, la λ - función servirá para determinar de manera gráfica el tipo de cópula que mejor ajusta los datos. Como ejemplo, ver la Figura 3.15.

3.9 Elección de la vine regular

Un primer criterio para la elección de la estructura de la vine (C-vine o D-vine) dependerá si existe dependencia con respecto a una variable en particular, es decir, cuando una variable actúe como una que gobierne las interacciones con las demás dentro de la base de datos, llamada pivote. En ese caso, será conveniente utilizar una C-vine; mientras que si no existe una variable pivote en la base de datos, será conveniente utilizar D-vine.

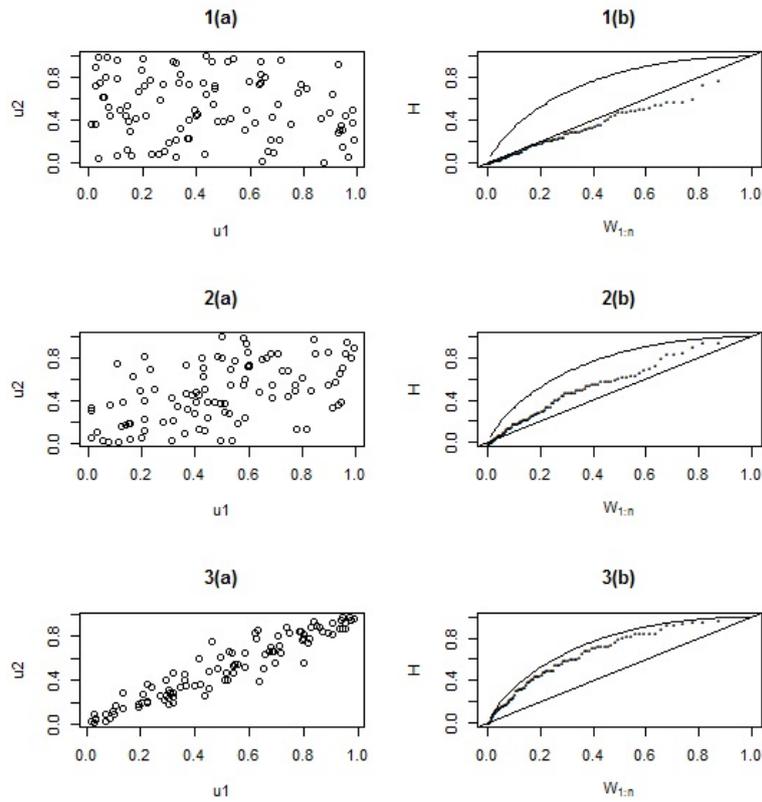


Figura 3.14: K-plot basados en muestras aleatorias de tamaño $n = 100$ de un vector normal bivariado. 1 (a) $\rho = 0.05$; 1 (b) k-plot. 2 (a) $\rho = 0.50$; 2 (b) k-plot. 3 (a) $\rho = 0.95$; 3 (b) k-plot

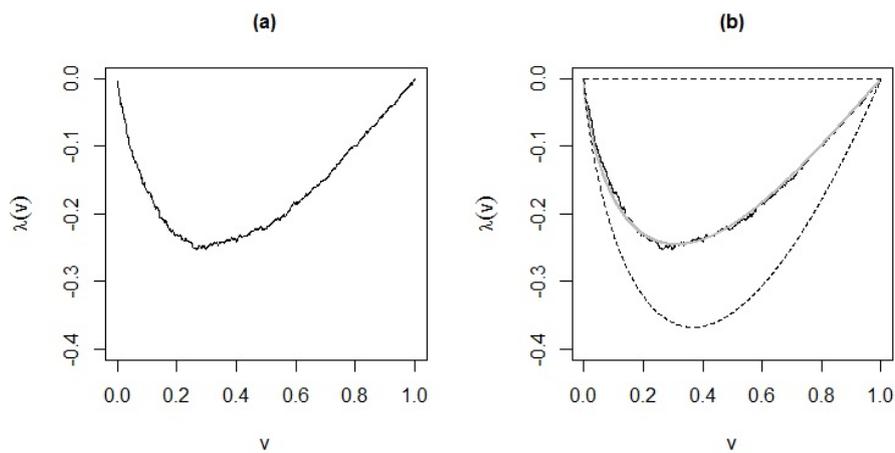


Figura 3.15: (a) λ - función empírica de una cópula Joe simulada. (b) λ - función teórica de una cópula Joe simulada y la λ - función empírica

Aunque el primer criterio es una cuestión más empírica que analítica, Aas, Czado, Frigessi y Bakken (2009) sugieren un criterio adicional para determinar si un conjunto de datos debe ser modelado como una C-vine o D-vine. Ellos sugieren estimar una t-cópula bivariada a cada par de variables para extraer los grados de libertad de cada estimación y ordenar las variables de acuerdo con estos valores. Un menor número de grados de libertad indican mayor dependencia entre las variables, lo cual ayudaría a elegir la estructura de la vine.

Aunque estos dos procedimientos (juicio crítico y estimación de los grados de libertad) son herramientas que ayudan a elegir la vine adecuada anterior a la estimación, existen pruebas estadísticas (posteriores a la estimación de los parámetros de las cópulas en pares) para saber si haber elegido una determinada estructura de vine ajusta mejor los datos. Estas pruebas serán vistas en la sección 3.11.

3.10 Estimación de parámetros de las cópulas

Los métodos de estimación de parámetros de una cópula han sido discutidos extensamente en Genest (1995), Joe (1997) y Genest y Favre (2007). Aquí solo se van a presentar las principales diferencias de esos métodos, los cuales dependerán si se han hecho supuestos acerca de la marginal.

- (i) *Inversión de la tau de Kendall, τ , o el rho de Spearman, ρ* . Para aquellas familias de cópulas bivariadas que tienen una forma cerrada de tau de Kendall, τ , o el rho de Spearman, ρ , se pueden invertir estos parámetros para obtener un estimado de parámetro de la cópula, las cuales son consistentes bajo ciertas condiciones (ver Genest y Favre (2007)). Además, para las cópulas elípticas de mayor dimensión, el tau de Kendall, τ , o el rho de Spearman, ρ , puede ser usados para estimar los parámetros de la matriz de correlación.
- (ii) *Máxima verosimilitud (MV)*. Cuando se conoce que las marginales pertenecen a una específica familia paramétrica, se puede utilizar la función de verosimilitud

$$l_{MV}(\theta, \delta) = \sum_{i=1}^n \log [c(F_{\delta,1}(\mathbf{X}_{i,1}), \dots, F_{\delta,1}(\mathbf{X}_{i,d}))],$$

donde $\delta = (\delta_1, \dots, \delta_d)'$ y $\mathbf{X}_i = (\mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,d})$, $i = 1, \dots, n$ son muestras del vector aleatorio $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_d)$.

- (iii) *Inferencia de las marginales (IM)*. Dado que no se suele conocer las marginales, Joe (1997) sugirió realizar la maximización en dos etapas. En la primera, estimar separadamente las marginales y, en la segunda, introducir esos estimados en el logaritmo de la función de verosimilitud

$$l_{IFM}(\theta, \delta) = \sum_{i=1}^n \log [c(F_{\hat{\delta},1}(\mathbf{X}_{i,1}), \dots, F_{\hat{\delta},1}(\mathbf{X}_{i,d}))].$$

No obstante, a pesar de ser computacionalmente más eficiente, Joe (2005) mostró que este procedimiento en dos etapas pierde eficiencia en algunos casos. Más aún, Kim (2007) mostró que los dos últimos métodos no son robustos frente a malas especificaciones de las marginales.

- (iv) *Pseudo máxima verosimilitud (PMV)*. Debido a que las distribuciones de las marginales son casi siempre desconocidas, el método de estimación natural de las marginales sería la estimación no paramétrica (ver Genest (1995)). Con este método se reemplaza la función de distribución desconocida por su versión empírica

$$\hat{F}_j(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[X_{i,j} \leq t]}.$$

Esta versión empírica de la función de distribución se introduce en el logaritmo de la función de verosimilitud

$$l_{PMV}(\theta) = \sum_{i=1}^n \log [c_{\theta}(\hat{F}_{n,1}(X_{i,1}), \dots, \hat{F}_{n,d}(X_{i,d}))].$$

Kim (2007) comparó los estimados obtenidos a través de las metodoloías MV, IM y PMV y mostró que el estimado obtenido a través de PMV es la mejor alternativa en la mayoría de casos. En particular, sobre la base de simulaciones, los métodos MV y IM no son robustos frente a malas especificaciones de las marginales. En cambio, este autor mostró que cuando las marginales no son conocidas, los parámetros estimados con el método PMV es la mejor alternativa.

De esta manera, tomando en cuenta lo visto en las cópulas en pares así como la mejor forma de estimar los parámetros de las cópulas, el logaritmo de la función a maximizar es la siguiente: sea $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_N)'$ observaciones d -dimensional con $\mathbf{u}'_i = (u'_{i,1}, \dots, u'_{i,d})' \in [0, 1]^d$, $i = 1, \dots, N$. Entonces el logaritmo de la función de verosimilitud de una C-vine es el siguiente:

$$\text{loglik} := l_{CVine}(\theta | \mathbf{u}) = \sum_{i=1}^N \sum_{j=1}^{d-1} \sum_{k=1}^{d-j} \ln[c_{j,j+k|1,\dots,j-1}],$$

donde

$$c_{j,j+k|1,\dots,j-1} := c_{j,j+k|1:(j-1)}(F(u_{i,j} | u_{i,1}, \dots, u_{i,j-1}), F(u_{i,j+k} | u_{i,1}, \dots, u_{i,j-1}) | \theta_{j,j+k|1,\dots,j-1})$$

denota las cópulas en pares con parámetros $\theta_{j,j+k|1,\dots,j-1}$.

De manera similar, el logaritmo de la función de verosimilitud de una D-vine es el siguiente:

$$\text{loglik} := l_{DVine}(\theta | \mathbf{u}) = \sum_{i=1}^N \sum_{j=1}^{d-1} \sum_{k=1}^{d-j} \ln[c_{k,k+j|k+1,\dots,k+j-1}],$$

donde

$$c_{k,k+j|k+1,\dots,k+j-1} :=$$

$$c_{k,k+j|k+1,\dots,k+j-1}(F(u_{i,k} | u_{i,k+1}, \dots, u_{i,k+j-1}), F(u_{i,k+j} | u_{i,k+1}, \dots, u_{i,k+j-1}) | \theta_{k,k+j|k+1,\dots,k+j-1}).$$

Las funciones de distribución condicional en ambas expresiones son obtenidas recursivamente usando la función $h()$

$$h(u | \mathbf{v}, \theta) := F(u | \mathbf{v}) = \frac{\partial C_{uv_j | \mathbf{v}_{-j}}(F(u_j | \mathbf{v}_{-j}), F(v_j | \mathbf{v}_{-j}))}{\partial F(v_j | \mathbf{v}_{-j})},$$

donde $C_{uv_j | \mathbf{v}_{-j}}$ es la función de distribución de una cópula bivariada arbitraria con parámetros θ y \mathbf{v}_{-j} representa un vector con el j -ésimo componente v_j , ver Aas, Czado, Frigessi, y Bakken (2009).

3.11 Selección del modelo

Una vez estimado los parámetros de varias cópulas para un mismo par de variables el siguiente paso será determinar cuál de las cópulas candidatas será la elegida. Para ello, existen algunos criterios de selección de modelos. Uno de ellos es el criterio de información Kullback-Leiber (KLIC, por sus siglas en inglés), desarrollado por Kullback y Leiber (1951), el cual mide la distancia entre dos modelos estadísticos. El KLIC está definido de la siguiente manera:

$$KLIC := E_0[\log h_0(Y_i | x_i)] - E_0[\log f(Y_i | x_i, \hat{\beta})],$$

donde $h_0(\cdot | \cdot)$ es la verdadero, pero desconocida, función de probabilidad condicional de Y_i dado x_i . E_0 es el valor esperado bajo el verdadero modelo y $\hat{\beta}$ es el estimador del parámetro β en el modelo $f(Y_i | x_i, \hat{\beta})$. De esta manera, el modelo con la menor distancia KLIC será el mejor. Los test de Vuong y Clarke se basan en el KLIC. No obstante, antes de presentar los test de Vuong y Clarke, se presentarán los conocidos criterios de información del Akaike y bayesiano de Schwarz.

3.11.1 AIC

El criterio de información del Akaike (AIC, por sus siglas en inglés), desarrollado por Akaike (1974), es uno de los criterios de selección más utilizados. Dado un conjunto de datos, varios modelos son comparados de acuerdo al AIC siendo el mejor el que tenga el menor AIC. El AIC es definido de la siguiente manera:

$$AIC = -2l(\hat{\theta} | x) + 2k,$$

donde $l(\cdot)$ es el logaritmo de la función de verosimilitud, $\hat{\theta}$ es el vector de parámetros estimados del modelo, x son los datos y k el número de parámetros del modelo. La primera parte del criterio es una medida de la bondad de ajuste del modelo; mientras que la segunda parte es un término de penalidad. Sin embargo, una desventaja del término de penalidad es que no considera el número de observaciones.

3.11.2 BIC

A diferencia del AIC, el criterio de información bayesiano (BIC) o el criterio de Schwarz, desarrollado por Schwarz (1978), toma en cuenta el número de observaciones en el término de penalidad. El BIC es definido de la siguiente manera:

$$BIC = -2l(\hat{\theta} | x) + k \log(n),$$

donde $l(\cdot)$ es el logaritmo de la función de verosimilitud, $\hat{\theta}$ es el vector de parámetros estimados del modelo, x son los datos, k el número de parámetros del modelo y n es el número de observaciones. Así como en el caso del AIC, la primera parte del criterio es una medida de la bondad de ajuste del modelo; mientras que la segunda parte es un término de penalidad.

3.11.3 El test de Vuong

Sean dos modelos $f_1(\cdot | \hat{\theta}_1)$ y $f_2(\cdot | \hat{\theta}_2)$ con parámetros estimados $\hat{\theta}_1$ y $\hat{\theta}_2$, respectivamente. El test desarrollado por Vuong (1989) utiliza el criterio de información Kullback-Leiber (KLIC) para estimar la hipótesis nula

$$\begin{aligned} KLIC(h_0, f_1, \hat{\theta}_1) &= KLIC(h_0, f_2, \hat{\theta}_2) \\ E_0 [\log f_1(X | \hat{\theta}_1)] &= E_0 [\log f_2(X | \hat{\theta}_2)]. \end{aligned}$$

Si $E_0 [\log f_1(X | \hat{\theta}_1)] > E_0 [\log f_2(X | \hat{\theta}_2)]$, el modelo 1 es el preferido. De esta manera, para observaciones x_i , $i = 1, \dots, n$, se define las diferencias en logaritmos de las funciones de verosimilitud $m_i := \log \left[\frac{f_1(x_i | \hat{\theta}_1)}{f_2(x_i | \hat{\theta}_2)} \right]$. Bajo la verdadera distribución h_0 , $\mathbf{m} = (m_1, \dots, m_n)'$ es un vector aleatorio con media $\mu_0^m = (\mu_1^m, \dots, \mu_n^m)'$. Dado que μ_0^m no es conocido, no es posible someter a prueba directa que ambos modelos sean igualmente buenos en términos de sus aproximaciones a la verdadera distribución. No obstante, Vuong (1989) mostró que

$$\lim_{n \rightarrow +\infty} \left(\frac{1}{n} \sum_{i=1}^n m_i \right) = \mu_0^m.$$

Además, Vuong (1989) mostró que la distribución asintótica de $\frac{1}{n} \sum_{i=1}^n m_i$ se distribuye como una normal estándar. Así, la hipótesis nula del test de Vuong es la siguiente:

$$H_0 : E_0[m_i] = 0 \text{ versus } H_1 : E_0[m_i] \neq 0, \forall i = 1, \dots, n.$$

Podemos rechazar la hipótesis nula a un nivel de significancia α si

$$|v| > \Phi^{-1} \left(1 - \frac{\alpha}{2} \right),$$

donde Φ^{-1} es la inversa de la función de distribución normal estándar. Si $v < \Phi^{-1} \left(1 - \frac{\alpha}{2} \right)$, elegimos el modelo 2. Si $v > \Phi^{-1} \left(1 - \frac{\alpha}{2} \right)$, preferimos el modelo 1 frente al 2.

3.11.4 El test de Clarke

El test propuesto por Clarke (2007), así como el test de Vuong, también utiliza el criterio de información Kullback-Leiber (KLIC). La diferencia entre ellos es la hipótesis nula. Ésta es la siguiente:

$$H_0 : P(m_i > 0) = 0.5, \forall i = 1, \dots, n,$$

donde $m_i = \log \left[\frac{f_1(x_i | \hat{\theta}_1)}{f_2(x_i | \hat{\theta}_2)} \right]$. La hipótesis nula establece que los logaritmos de los ratios de verosimilitud deben de estar igualmente distribuidos alrededor del cero. De esta manera, la mitad de los ratios de verosimilitud deberían ser mayores a cero y la mitad menores a cero.

El estadístico es el siguiente

$$B = \sum_{i=1}^n I_{(0,+\infty)}(d_i),$$

donde I es la función indicadora y

$$d_i = \log f_1(x_i | \hat{\theta}_1) - \log f_2(x_i | \hat{\theta}_2).$$

Clarke (2007) demostró que B tiene una distribución binomial con parámetros n y $p = 0.5$. De esta manera, si el modelo 1 es mejor que el modelo 2, B será significativamente mayor que su valor esperado bajo la hipótesis nula $\left(\frac{n}{2}\right)$.

3.12 Test de bondad de ajuste

El test de bondad de ajuste de un modelo mide el nivel de ajuste de éste con respecto a los datos observados, es decir, se centran en la discrepancia entre el valor observado y esperado bajo el modelo en cuestión. Algunos de los test estadísticos más conocidos son el de Kolmogorov-Smirnov (K-S) y el de Anderson-Darling (A-D). No obstante, antes de presentarlos, una herramienta gráfica que da información acerca del nivel de ajuste de los datos de un modelo es el conocido gráfico cuantil cuantil o qq-plot.

3.12.1 Análisis gráfico

Una herramienta gráfica de diagnóstico alternativa son los gráficos cuantil cuantil o qq-plot propuesto por Wilk y Gnanadesikan (1968). Estos gráficos permiten comparar las funciones de distribución acumulada empíricas de los datos observados y simulados. Si estas dos funciones son iguales, entonces el gráfico es similar a una línea de 45° .

3.12.2 Análisis estadístico

El test de Kolmogorov-Smirnov (K-S), desarrollado por Kolmogorov y Smirnov (1933), es una prueba no paramétrica de bondad de ajuste de dos distribuciones de probabilidad. El test compara la máxima diferencia entre la distribución de los retornos simulados y los retornos históricos, donde la hipótesis nula es que las dos variables tienen la misma distribución.

$$\text{K-S} = \sup_{x \in \mathbb{R}} |F(x) - \hat{F}(x)|,$$

donde $F(x)$ es la función de distribución de los datos y $\hat{F}(x)$ es la función de distribución estimada.

No obstante, el estadístico K-S otorga la misma importancia a las colas que al centro de las distribuciones que está comparando, es decir, mide la distancia máxima entre las dos distribuciones, pero sin importar dónde mide esa distancia. Una alternativa a ello es utilizar el test de Anderson Darling (A-D), desarrollado por Anderson y Darling (1952). Como en el caso del test K-S, una variante del test de A-D mide la distancia entre las dos distribuciones pero reescalado al dividir la distancia por el desvío estándar de esa distancia.

$$\text{A-D} = \sup_{x \in \mathbb{R}} \frac{|F(x) - \hat{F}(x)|}{\sqrt{\hat{F}(x)(1 - \hat{F}(x))}}.$$

De esta manera, el denominador se vuelve pequeño en los valores extremos de x . Así, las desviaciones entre F y F_n en las colas obtienen más peso que en el centro de la distribución.



Capítulo 4

Optimización del modelo M-CVaR con retornos no gaussianos

4.1 Introducción

En este capítulo se realizará un ejercicio de optimización de un portafolio de inversión utilizando al CVaR como medida de riesgo y tomando en cuenta que los retornos de un conjunto de índices no se distribuyen como una normal. Para ello, en la segunda sección, se mostrarán brevemente las diferentes formas de generar los escenarios para la optimización del modelo M-CVaR. Luego se elegirá una de ellas y se presentará la forma en la cual se generarán tales escenarios. Dado que esto involucrará el uso de varias teorías, éstas se presentarán en las siguientes secciones.

Así, en la tercera sección, se presentará la teoría de series de tiempo, en particular, los modelo autocorrelacionados y de volatilidad condicional (ARMA-GARCH), algunas de sus extensiones más utilizadas (GJR) y el test de bondad de ajuste de Ljung y Box (1978). Además, en la cuarta sección, se presentará brevemente la teoría de optimización de portafolios desarrolladas por H. Markowitz.

Finalmente, en la quinta sección, se aplican estos conceptos junto con la teoría de cópulas con el fin de simular los escenarios para luego, en la sexta sección, realizar la optimización de un portafolio de inversión. No obstante, antes se mostrará los hechos estilizados de los retornos de las series financieras y las maneras de modelarlas.

En particular, dado que las series presentan autocorrelación, se modelarán las series para extraer el componente temporal. Con los retornos filtrados se modelará de dependencia contemporánea de los datos a través de la teoría de cópulas. El proceso de obtención de los retornos simulados será el siguiente: se simularán los escenarios para la optimización del portafolios de inversión y con esto finalmente, se muestran los resultados de la optimización del modelo M-CVaR y se compararán con los obtenidos a través del modelo M-V de Markowitz.

4.2 Generación y simulación de escenarios

4.2.1 Generación de escenarios

Siguiendo a G. Guastaroba, R. Mansini, G. Speranza (2009), se revisarán algunas técnicas de generación de escenarios sin pretender cubrir todas ellas. Dentro de las técnicas de generación de escenarios, podemos encontrar las siguientes:

1. *Técnica de datos históricos*: uno de los más usados por su simplicidad. Se basa en el supuesto de que los datos históricos son posibles escenarios futuros. Esta técnica no requiere un supuesto sobre la función de distribución para los retornos ni se requiere calcular las correlaciones entre los activos, debido a que están implícitamente consideradas en los datos del mercado. Esta técnica preserva la media y varianza de los retornos. A pesar de ser uno de los más usados, la desventaja de esta técnica radica en que los datos futuros pueden ser sustancialmente diferentes a los observados en el pasado. Además, el número de escenarios que pueden ser producidos está limitado a la disponibilidad de la información.

2. *Bootstrapping*: cada escenario corresponde a la realización conjunta de los retornos de todos los activos dado un período de tiempo dado. De esta manera, cada escenario puede ser remuestrados dada una probabilidad constante repitiéndose el proceso hasta que la muestra alcance el tamaño deseado. Esta aproximación preserva la correlación entre los activos. La potencial desventaja radica en la utilización de datos históricos pues los retornos futuros pueden ser sustancialmente diferentes a los observados en el pasado.
3. *Simulación de montecarlo*: aproximación paramétrica que consiste en generar escenarios de acuerdo a una específica función de distribución. El punto más importante es elegir aquella distribución conjunta que mejor ajusta a los retornos históricos. Normalmente se utiliza la función de distribución normal estándar multivariada. Sin embargo, otras distribuciones han sido utilizadas debido a que los retornos tienden a ser leptocúrticos, siendo particularmente utilizada de distribución t-student. La principal ventaja radica en la no utilización de los datos.
4. *M-GARCH*: ésta es una de las formas de modelar la dinámica conjunta de los datos. Sin embargo, la dificultad de estos métodos radica en la cantidad de parámetros a estimar pues estos se incrementan dramáticamente con el número de variables. Otra dificultad con estos modelos radica en la elección de la distribución condicional de los retornos. Cuando las distribuciones marginales son gaussianas, la extensión al caso multivariado es trivial. Sin embargo, cuando las distribuciones no muestran este comportamiento, a veces la distribución multivariada no existe.

En resumen, las aproximaciones no paramétricas (histórico y bootstrapping) tienen la ventaja de ser simples de implementar. Las paramétricas (montecarlo y M-GARCH) dependen de la distribución asumida. Sin embargo, las no paramétricas son criticadas debido a que dependen de la muestra. Si la muestra está sesgada, entonces también lo estarán los escenarios. Por ello, teniendo en cuenta los hechos estilizados y las ventajas y desventajas de las distintas alternativas de generación de escenarios, realizaremos la generación simulando escenarios utilizando cópulas en pares para modelar la dependencia contemporánea de los datos. En la siguiente sección se detallará el procedimiento a seguir para generar los escenarios simulados.

4.2.2 Simulación de escenarios

Con el fin de generar los escenarios para la optimización del modelo M-CVaR y teniendo en cuenta que los retornos no son gaussianos, se realizarán los siguientes pasos:

1. Dado que las series presentan autocorrelación en los retornos y en las volatilidades, se removerán estos efectos fijando para cada índice un modelo ARMA(1,1)-GARCH(1,1). Se utilizó este modelo debido a que dentro de la literatura empírica es el más usado. No obstante, se escogerá el modelo que presente mejor ajuste tomando en cuenta tres criterios de selección (el valor del logaritmo de la función de verosimilitud, el criterio del Akaike y el criterio bayesiano de Schwarz). Además, se modelarán los índices con un modelo alternativo GJR para capturar el efecto apalancamiento, el cual se detallará más adelante. Finalmente, se asumirán distintas distribuciones para los errores como la distribuciones t-student.
2. Utilizando los parámetros estimados de los modelos escogidos en el paso 1, filtramos las series y obtenemos los residuos estandarizados. Esto será de especial importancia para poder luego utilizar la metodología de cópulas, pues ella requiere que los datos tengan esa característica.
3. Teniendo los residuos estandarizados de todos los índices, se modelará la interdependencia contemporánea de estos datos a través de la teoría de cópulas. En particular, se utilizarán cópulas en pares de reciente utilización para estimar los parámetros.
4. Con los parámetros estimados de todas las cópulas en pares, realizamos la simulación de residuos estandarizados.

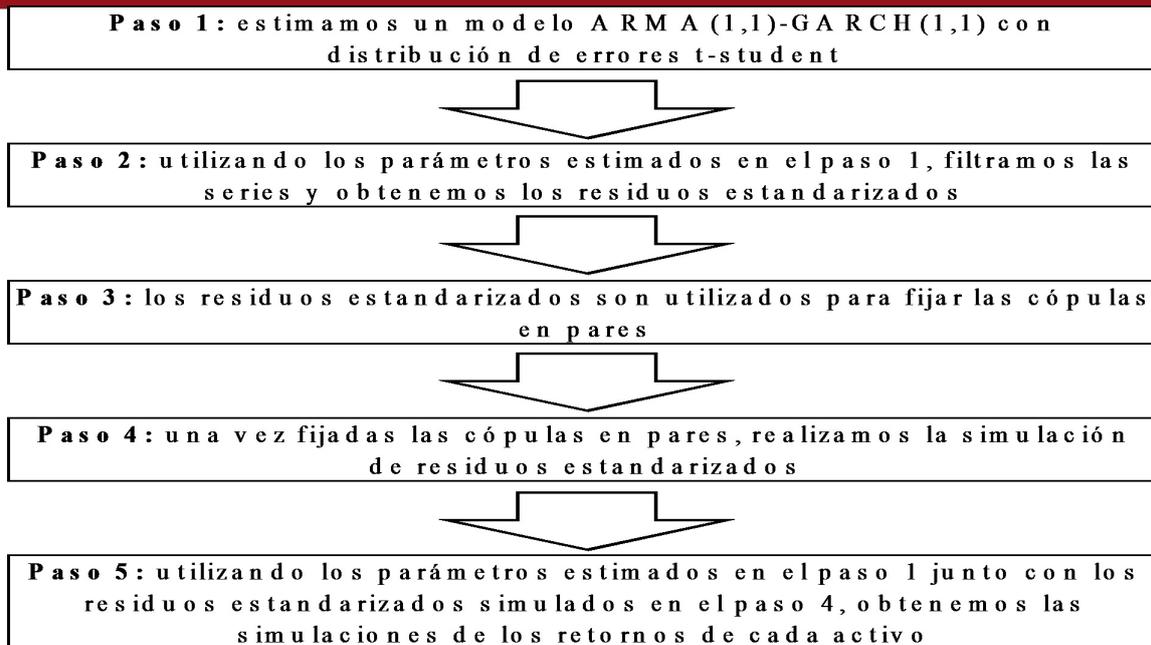


Figura 4.1: Proceso para la simulación de escenarios

- Finalmente, usando los residuos estandarizados simulados en el paso previo como insumos i.i.d., reintroducimos la autocorrelación serial de los retornos y las volatilidades utilizando los parámetros estimados de los modelos escogidos del paso 1 para, finalmente, obtener los retornos simulados con los cuales se realizará la optimización del modelo M-CVaR.

Los pasos arriba mencionados están esquematizados en la Figura 4.1.

4.3 Modelos de series de tiempo

En esta sección se introducirá brevemente algunos conceptos de la teoría de series de tiempo que serán necesarios para modelar los datos. En primer lugar, se presentará el conocido modelo ARMA-GARCH. Luego, se presentarán algunas de las extensiones del modelo GARCH más utilizadas para modelar los retornos de las series financieras. Finalmente, se presentará el test de bondad de ajuste de Ljung Box (1978). Para mayores referencias ver Enders (2010) y Hamilton (1994).

4.3.1 Modelo ARMA-GARCH

La necesidad de predecir las rentabilidades futuras de un activo así como de analizar el riesgo de mantener ese activo por un determinado período de tiempo ha llevado al desarrollo de nuevas técnicas de modelamiento dentro de la teoría de las series de tiempo. En esa línea, los Modelos Autoregresivos de Heteroscedasticidad Condicionada (ARCH), introducidos por Engle (1982), fueron específicamente diseñados para modelar y hacer pronósticos de la varianza condicionada. En estos modelos, la varianza de la variable dependiente es modelada como una función del valor pasado de la misma variable dependiente. La especificación de la varianza condicional ARCH(1,1) es la siguiente:

$$Y_t = X_t' \theta + \epsilon_t$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2,$$

donde la primera ecuación es la de la media condicional; mientras que la segunda es llamada la ecuación de la variainza condicional. Esta última ecuación es una función de tres componentes:

- Un término constante: ω .
- Noticias acerca de la volatilidad del período previo, medido como el rezago del residuo al cuadrado de la ecuación de la media ϵ_{t-1}^2 (término ARCH).
- La varianza del período anterior σ_{t-1}^2 .

En forma intuitiva, la ecuación de la varianza condicional se puede interpretar la predicción de la varianza es la media ponderada de la siguiente manera: un promedio de largo plazo (la constante), la varianza del período previo (el término GARCH) y la información acerca de la volatilidad observada en el período previo (el término ARCH).

No obstante, en aplicaciones empíricas a menudo se requiere que el modelo de la varianza condicional tenga una memoria más larga y una estructura de retardos más flexible. Esto se logra extendiendo los modelos ARCH con la clase de Modelos Generalizados Autoregresivos de Heteroscedasticidad Condicionada (GARCH) introducidos por Bollerslev (1986). Estos modelos pueden ser estimados escogiendo los parámetros p y q . La representación de la varianza condicional en el modelo GARCH(p,q) es la siguiente:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2.$$

4.3.2 Extensiones al modelo GARCH

Aunque el modelo GARCH es ampliamente utilizado en las aplicaciones financieras para modelar la volatilidad condicionada de las series de tiempo, debido a que en su formulación solamente aparecen los residuos al cuadrado, el modelo no logra capturar una interesante característica de las series financieras: el hecho que noticias negativas (*shocks* negativos) parecen tener un efecto más pronunciado sobre la volatilidad que las noticias positivas (*shocks* positivos). Esto se debería a que cuando el mercado es sorprendido negativamente (noticias negativas), los precios tienden a bajar produciéndose un efecto de "palanca", lo cual genera mayor volatilidad. Glosten, Jagathan y Runkle (1994) mostraron la manera cómo las noticias negativas y positivas tienen diferentes efectos sobre la volatilidad. En ese sentido, $\epsilon_{t-1} = 0$ representa el umbral a través del cual *shocks* mayores al umbral tienen un efecto diferente que *shocks* debajo de este umbral. Formalmente, el modelo general GJR(p,q) para la varianza condicional, el cual contempla el término de apalancamiento, es el siguiente:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \gamma_i S_{t-i} \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

donde

$$S_{t-i} = \begin{cases} 1, & \text{si } \epsilon_{t-i} < 0 \\ 0, & \text{si } \epsilon_{t-i} \geq 0 \end{cases}.$$

De esta manera, las buenas noticias, $\epsilon_{t-i} \geq 0$ y las malas noticias, $\epsilon_{t-i} < 0$ tienen diferentes efectos sobre la varianza condicional. Las buenas noticias tienen un impacto de α_i ; mientras que las malas noticias tienen un impacto de $\alpha_i + \gamma_i$.

Otro modelo que permite capturar el efecto asimétrico de las noticias es el GARCH exponencial (EGARCH) propuesto por Nelson (1991). La especificación para la varianza condicional es la siguiente:

$$\log(\sigma_t^2) = \omega + \sum_{j=1}^q \beta_j \log(\sigma_{t-j}^2) + \sum_{i=1}^p \alpha_i \left| \frac{\epsilon_{t-i}}{\sigma_{t-i}} \right| + \sum_{i=1}^p \gamma_i \frac{\epsilon_{t-i}}{\sigma_{t-i}}.$$

El hecho que la variable dependiente sea $\log(\sigma_t^2)$ implica que el efecto de palanca es exponencial en lugar de ser cuadrático. Para capturar el efecto palanca, en lugar de usar el valor de ϵ_{t-i}^2 , el modelo EGARCH usa el valor estandarizado de ϵ_{t-i} . Nelson (1991) argumenta que esta estandarización permite una mejor interpretación del tamaño y persistencia de los *shocks*. De esta manera, si $\frac{\epsilon_{t-i}}{\sigma_{t-i}}$ es positivo, el efecto de los *shocks* sobre el logaritmo de la varianza condicional es $(\alpha_i + \gamma_i)$. Por otro lado, si $\frac{\epsilon_{t-i}}{\sigma_{t-i}}$ es negativo, el efecto de los *shocks* sobre el logaritmo de la varianza condicional es $(-\alpha_i + \gamma_i)$.

No obstante, aunque estos modelos (GJR, EGARCH) permitirían tener un mejor ajuste al tomar en cuenta una característica presente en la mayoría de las series de tiempo, el modelo GARCH tiene varios beneficios:

1. El modelo GARCH representa el modelo más parsimonioso, lo cual implica que se estimen menos coeficientes. De acuerdo con Box y Jenkins (1994), a menores parámetros a estimar, menor la posibilidad de que uno se pueda equivocar.
2. El modelo GARCH captura mucha de la variabilidad de las series de tiempo. En particular, suele ser común que en aplicaciones empíricas se utilicen solamente modelos GARCH con bajos valores para p y q . De acuerdo con Bollerslev, Chou y Kroner (1992), los modelos GARCH(1,1), GARCH(2,1) o GARCH(1,2) son adecuadas representaciones de la volatilidad de las series financieras, incluso sobre largos períodos de tiempo.

4.3.3 El test de Ljung Box

Ljung y Box (1978) desarrollaron el test Ljung-Box para medir la independencia de los residuos de un modelo. Luego de especificar un modelo de serie de tiempo se observan los residuos y su autocorrelación. Si el modelo es el adecuado, el estadístico Ljung-Box es el siguiente:

$$Q = n(n + 2) \sum_{k=1}^L \frac{\hat{r}_k^2}{(n - k)},$$

donde n es el tamaño de la muestra; L es el número de rezagos incluidos en el estadístico; y r_k^2 es la autocorrelación muestral al cuadrado hasta el rezago k , se distribuye como una χ_m^2 con m grados de libertad.

4.4 Optimización del modelo M-V de Markowitz

El punto de partida de la teoría de la optimización de portafolios, desde un punto de vista matemático, es el trabajo de H. Markowitz (1952) sobre el criterio media-varianza. Debido a la facilidad de su aplicación, el modelo de Markowitz continúa siendo utilizado. No obstante, la simplicidad de su modelo es también la base de algunas desventajas. El modelo es el siguiente.

Considere un mercado donde d diferentes activos con precios $p_1, p_2, \dots, p_d > 0$ en el tiempo $t = 0$ son intercambiados. Los precios $P_1(T), P_2(T), \dots, P_d(T)$ al final del período $t = T$ no son conocidos. Así, ellos son modelados sobre un espacio de probabilidad (Ω, \mathcal{F}, P) .

Por otro lado, sean los retornos de los activos

$$R_i(T) = \frac{P_i(T)}{p_i}, i = 1, \dots, d,$$

y se asume que se conoce, o se ha estimado, las medias, varianzas y covarianzas

$$E(R_i(T)) = \mu_i \text{ para } i = 1, \dots, d,$$

$$Cov(R_i(T), R_j(T)) = \sigma_{ij} \text{ para } i, j = 1, \dots, d.$$

Se asume que cada activo es perfectamente divisible de modo que uno podría tener $\varphi_i \in \mathbb{R}$ participación del activo.

Definición 50. Se asume que un inversionista con una riqueza inicial $x > 0$ tiene una participación de $\varphi_i \geq 0$ del activo $i = 1, \dots, d$ con

$$\sum_{i=1}^d \varphi_i \cdot p_i = x.$$

Entonces, el vector del portafolio $\pi = (\pi_1, \dots, \pi_d)$ está definido de la siguiente manera

$$\pi_i = \frac{\varphi_i \cdot p_i}{x}, i = 1, \dots, d,$$

y

$$R^{\pi} = \sum_{i=1}^d \pi_i \cdot R_i(T)$$

es llamado el retorno del portafolio.

Definición 51. La media y la varianza de los retornos del portafolio son

$$E(R^{\pi}) = \sum_{i=1}^d \pi_i \cdot \mu_i,$$

$$Var(R^{\pi}) = \sum_{i=1}^d \sum_{j=1}^d \pi_i \cdot \sigma_{ij} \cdot \pi_j.$$

De esta manera, una formulación de la aproximación media-varianza será aquella que maximiza la media del retorno del portafolio $E(R^{\pi})$ dado un valor mínimo c_1 para la varianza.

$$\max_{\pi \in \mathbb{R}^d} E(R^{\pi})$$

$$\text{sujeto a } \pi_i \geq 0 \text{ para } i = 1, \dots, d, \sum_{i=1}^d \pi_i = 1, Var(R^{\pi}) \leq c_1.$$

Se está considerando el supuesto usual de no ventas en corto ($\pi_i \geq 0$). El problema de maximizar la media del retorno del portafolio implica que bajo todas los posibles portafolio $\pi \in \mathbb{R}^d$, se considera solo aquellos que satisfacen la restricción de tener una varianza menor o igual a c_1 .

Por otro lado, una formulación alternativa implica minimizar la varianza de los retornos del portafolio $Var(R^{\pi})$ dado un nivel mínimo sobre el retorno esperado del portafolio, $E(R^{\pi}) \geq c_2$

$$\min_{\pi \in \mathbb{R}^d} Var(R^{\pi})$$

$$\text{sujeto a } \pi_i \geq 0 \text{ para } i = 1, \dots, d, \sum_{i=1}^d \pi_i = 1, E(R^{\pi}) \geq c_2.$$

La formulación anterior implica que bajo todos los posibles portafolios $\pi \in \mathbb{R}^d$, considere solo aquellos que satisfacen la restricción de un retorno esperado como mínimo de c_2 .

4.5 Tratamiento de los datos

En esta sección se realizará la simulación de los retornos de un conjunto de índices financieros para ser utilizados en la optimización de un portafolio de inversiones. Se verá, en línea con los hechos estilizados, que los retornos de estos índices no se distribuyen normales, debido a la presencia de sesgo y curtosis. Además, las series muestran autocorrelación y agrupamiento de las volatilidades. Para remover estas dos últimas características, se utilizará la teoría de series de tiempo.

Luego se procederá a modelar la dependencia contemporánea de los residuos a través de cópulas en pares para, de esta manera, simular los escenarios a utilizar en la optimización del portafolio de inversión.

4.5.1 Datos

La información utilizada para la optimización del portafolio son índices representativos (*benchmarks*) de los mercados de renta fija, renta variable y *commodities*. Para el caso de los índices de renta fija, se utilizaron los bonos categoría grado de inversión de EE.UU. (USIG), bonos categoría grado de inversión de países emergentes (EMIG) y bonos de alto riesgo de países emergentes (EMHY). Para el caso de la renta variable, se utilizaron los índices de acciones que forman parte del S&P500 (EE.UU.), acciones de otros países desarrollados (EAFE: Europe, Australasia, and Far East) y de países emergentes (EM). Finalmente, se utilizan índices representativos de los mercados de materias primas (*commodities*).

La muestra utilizada para la optimización es de frecuencia mensual, los retornos se han calculado como $R_t = (P_t/P_{t-1} - 1) \times 100$, donde P_t es el valor del índice en un mes determinado y P_{t-1} es el valor del índice en el mes previo, cubre poco más de 10 años de historia (137 observaciones) y cubre dos episodios de crisis en EE.UU. (2001 y 2008). Las fuentes de información son Barclays Bank, J.P. Morgan y Bloomberg. El Cuadro 4.1 muestra un resumen de los índices utilizados

Activo	Benchmark	Fuente
Renta Fija		
USIG	Barclays U.S. Credit 1-5 years Tot Ret Idx	Barclays Bank
EMIG	EMBIG Credit IG only Cum Tot Ret Idx	J.P.Morgan
EMHY	EMBIG Credit Non-IG Cum Tot Ret Idx	J.P.Morgan
Renta Variable		
EEUU	S&P500 Net Tot Ret	Bloomberg
EAFE	MSCI Daily TR Net EAFE USD	Bloomberg
EM	MSCI Daily TR Net Emerging Market USD	Bloomberg
Otros		
Commodity	DJ UBS Commodity Index Tot Ret	Bloomberg

Cuadro 4.1: Índices utilizados para el proceso de optimización

4.5.2 Hechos estilizados de los retornos de los activos financieros

Existe amplia evidencia empírica¹ que muestra que los retornos de los activos no se distribuyen como una normal. En particular, Cont (2001) resume los siguientes hechos estilizados de los retornos de los diferentes tipos de activos:

1. *Colas anchas*: la distribución incondicional de los retornos tiene colas anchas (mayores a las determinadas por la distribución normal). Esto significa que si utilizamos la distribución normal para modelar los retornos de los activos financieros, subestimaremos el número y la magnitud de caídas bruscas o subidas rápidas de los precios de los activos. Esto se conoce como curtosis.

¹Ver, por ejemplo, Bollerslev, Chou y Kroner (1992), Pagan (1996), Shephard (1996), Campbell, Lo y MacKinlay (1997), Gouriéroux y Jasiak (2001).

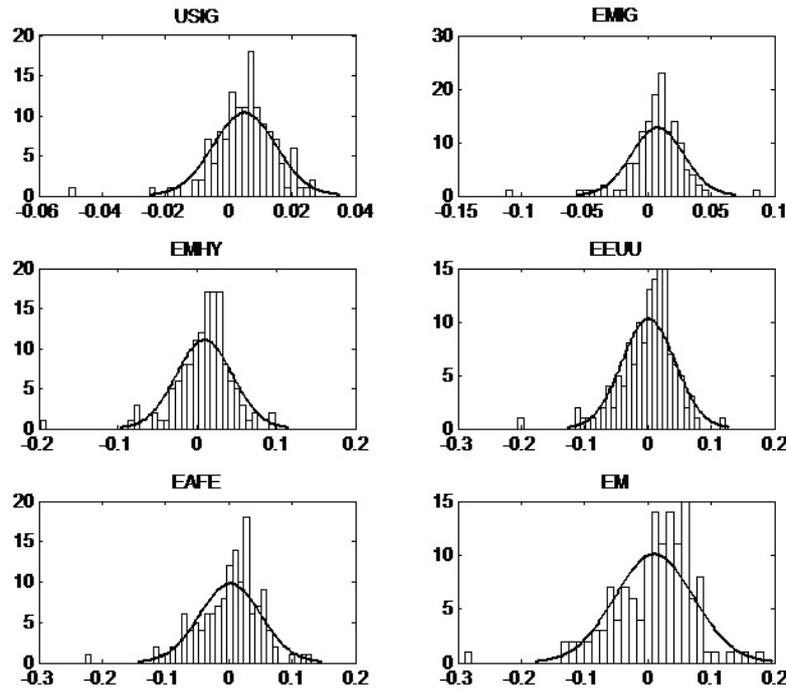


Figura 4.2: Histograma de retornos mensuales y distribución normal de USIG, EMIG, EMHY, EEUU, EAFE y EM

2. *Asimetría:* las distribuciones incondicionales de los retornos están negativamente sesgadas, lo cual sugiere que retornos negativos extremos son más frecuentes que los retornos positivos extremos.
3. *No normalidad:* bajo diferentes frecuencias (diaria, semanal, mensual), los retornos de los activos no siguen una distribución normal. Sin embargo, conforme la frecuencia de los retornos aumenta, la distribución de los retornos se acercan a la distribución normal.
4. *Ausencia de autocorrelación serial:* los retornos de los activos generalmente no muestran una significativa autocorrelación serial, excepto en los datos de alta frecuencia (diaria, semanal).
5. *Agrupamiento en la volatilidad:* la volatilidad de los retornos está serialmente autocorrelacionada, lo cual sugiere que grandes retornos (positivos o negativos) tienden a ser seguidos por otros grandes retornos (positivos o negativos).
6. *Correlación cruzada variante en el tiempo:* las correlaciones entre los retornos de los distintos activos tienden a incrementarse particularmente durante períodos de alta incertidumbre.

Ahora utilizaremos los *benchmarks* arriba mencionados para comprobar si estos muestran algunos de los hechos estilizados mencionados anteriormente. La Figura 4.2 muestra los histogramas de los retornos mensuales de los índices y se ha superpuesto las respectivas distribuciones normales asociadas utilizando la media y la varianza de cada serie. Las diferencias entre los histogramas y las distribuciones normales mostrarían, de manera gráfica, asimetría y colas anchas.

Por su parte, el Cuadro 4.2 el resumen de los estadísticos descriptivos² de los índices. Los datos muestran que todos presentan asimetría así como curtosis.

Por otro lado, el Cuadro 4.3 muestra el estadístico basado en momentos de Jarque Bera (1980) para contrastar la hipótesis nula que los retornos se distribuyen como una normal. El estadístico está definido de la siguiente

²Los datos de la media y desvío estándar están anualizados.

Activo	Media	Des. Est.	Sesgo	Curtosis
USIG	0.0604	0.0348	-1.3047	9.0811
EMIG	0.0925	0.0724	-1.3690	11.6133
EMHY	0.1117	0.1237	-1.6249	10.8439
EEUU	0.0126	0.1471	-1.1421	6.6665
EAFE	0.0289	0.1684	-0.9031	5.9544
EM	0.1205	0.2168	-0.9543	6.0957
Comm	0.0610	0.1483	-1.0936	5.9714

Cuadro 4.2: Estadísticas descriptivas de los índices

manera:

$$JB = \frac{n}{6} \left(s^2 + \frac{(k-3)^2}{4} \right),$$

donde n es el tamaño de la muestra, s es el coeficiente de asimetría muestral y k es la curtosis muestral. Bajo la hipótesis nula de que los datos se distribuyen como una normal, el test se distribuye como una χ^2_2 con dos grados de libertad. De acuerdo con los datos del Cuadro 4.3, en todos los casos se rechaza la hipótesis nula de normalidad. El rechazo en los tests se debe a la asimetría y la curtosis observadas anteriormente.

Activo	Sesgo	Curtosis	J-B	p-value
USIG	-1.3047	9.0811	249.9633	0.0000
EMIG	-1.3690	11.6133	466.2930	0.0000
EMHY	-1.6249	10.8439	411.5000	0.0000
EEUU	-1.1421	6.6665	106.5206	0.0000
EAFE	-0.9031	5.9544	68.4482	0.0000
EM	-0.9543	6.0957	75.4990	0.0000
Comm	-1.0936	5.9714	77.7063	0.0000

Cuadro 4.3: P-values de las pruebas de normalidad de Jarque Bera

Otro importante hecho estilizado de los retornos de los activos es la autocorrelación en los retornos y en las volatilidades. Para someter esto a prueba, utilizamos el test estadístico Q de Ljung-Box, cuya hipótesis nula es la ausencia de autocorrelación serial hasta el rezago k . El estadístico está definido de la siguiente manera:

		Autocorrelación en los retornos hasta el orden				
		1	5	10	15	20
USIG	Q test	7.386	12.030	16.620	21.506	23.430
	p-value	(0.007)	(0.034)	(0.083)	(0.121)	(0.268)
EMIG	Q test	1.909	16.443	19.962	26.816	29.519
	p-value	(0.167)	(0.006)	(0.030)	(0.030)	(0.078)
EMHY	Q test	2.443	3.704	5.246	9.619	16.039
	p-value	(0.118)	(0.593)	(0.874)	(0.843)	(0.714)
EEUU	Q test	8.293	15.158	18.474	20.572	23.684
	p-value	(0.004)	(0.010)	(0.047)	(0.151)	(0.256)
EAFE	Q test	16.115	27.190	34.409	33.944	36.466
	p-value	(0.000)	(0.000)	(0.000)	(0.003)	(0.014)
EM	Q test	24.016	28.958	35.428	40.271	43.045
	p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)
Comm	Q test	15.179	28.306	48.836	56.282	58.671
	p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Cuadro 4.4: P-values de las pruebas de autocorrelación de los retornos de Ljung Box

$$Q = n(n + 2) \sum_{k=1}^L \frac{r_k^2}{(n - k)},$$

donde n es el tamaño de la muestra; L es el número de rezagos incluidos en el estadístico; y r_k^2 es la autocorrelación muestral al cuadrado hasta el rezago k . El test para la dependencia serial en las volatilidades (retornos al cuadrado) puede ser ejecutado también a través del estadístico Q de Ljung-Box. Los Cuadros 4.4 y 4.5 muestran los p-value del estadístico Q de Ljung-Box entre paréntesis para los retornos y los retornos al cuadrado (para la volatilidad) de la hipótesis nula de ausencia de correlación serial hasta los órdenes 1, 5, 10, 15 y 20, respectivamente.

En línea con los hechos estilizados mencionados anteriormente, los resultados muestran que los retornos están serialmente autocorrelacionados y las correlaciones seriales son mayores para los retornos al cuadrado. De manera gráfica, la dependencia serial en los retornos y en las volatilidades se observa en los autocorrelogramas de los retornos y los retornos al cuadrado hasta el orden 20 (Figuras 4.3 y 4.4).

		Autocorrelación en los retornos al cuadrado hasta el orden				
		1	5	10	15	20
USIG	Q test	6.777	16.427	32.711	33.699	35.967
	p-value	(0.009)	(0.006)	(0.000)	(0.004)	(0.016)
EMIG	Q test	2.175	20.640	22.433	24.654	26.826
	p-value	(0.140)	(0.001)	(0.013)	(0.055)	(0.140)
EMHY	Q test	4.932	6.146	7.480	8.119	9.502
	p-value	(0.026)	(0.292)	(0.680)	(0.919)	(0.976)
EEUU	Q test	1.914	3.679	16.433	17.054	17.424
	p-value	(0.166)	(0.597)	(0.088)	(0.316)	(0.625)
EAFE	Q test	4.353	6.138	21.211	21.623	25.064
	p-value	(0.037)	(0.293)	(0.020)	(0.118)	(0.199)
EM	Q test	7.800	8.291	23.236	25.977	27.447
	p-value	(0.005)	(0.141)	(0.010)	(0.038)	(0.123)
Comm	Q test	19.066	43.549	46.522	48.676	51.990
	p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Cuadro 4.5: P-values de las pruebas de autocorrelación de los retornos al cuadrado de Ljung Box

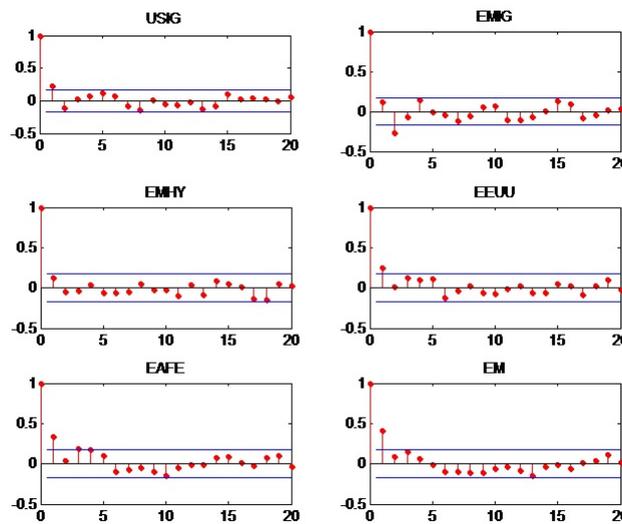
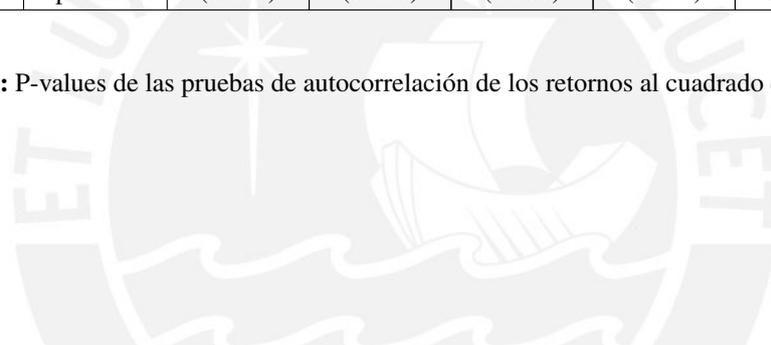


Figura 4.3: Autocorrelograma de los retornos

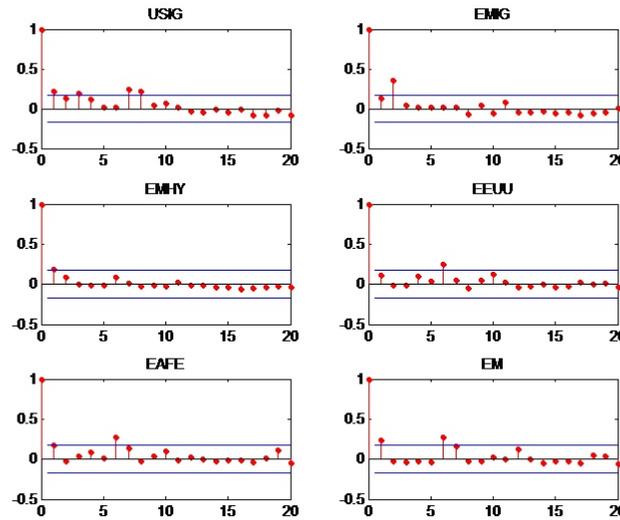


Figura 4.4: Autocorrelograma de los retornos al cuadrado

Ahora veremos, de manera gráfica, que los cuatro momentos de los retornos varían en el tiempo. La Figura 4.5 muestra la evolución de los primeros cuatro momentos del índice EEUU (para el S&P500) usando una ventana de 6 meses de rentabilidades diarias. En ella se observa que en los períodos de crisis (2001, 2008) la volatilidad (medida a través del desvío estándar) aumentó. El sesgo y la curtosis también se ven afectados.

Finalmente, los Cuadros 4.7 y 4.8 muestran la matriz de correlaciones de los retornos mensuales de todos los índices en dos períodos de tiempo. Como mencionamos, las correlaciones cambian en el tiempo, en particular, en momentos de alto estrés en los mercados. Esto implica que cuando las correlaciones se incrementan durante períodos de alta turbulencia, los beneficios de la diversificación se reducen fuertemente justo cuando ellos son más necesarios. El Cuadro 4.7 muestra las correlaciones utilizando toda la muestra; mientras que el Cuadro 4.8 muestra la matriz de correlaciones en el período más crítico de la crisis (2008). Se observa que, en la mayoría de los casos, las correlaciones entre los diferentes índices aumentó en el período más crítico de la crisis.

Índices	Estat	USIG	EMIG	EMHY	EEUU	EAFE	EM	Comm
Toda la muestra								
USIG	Corr	1.000	0.755	0.484	0.164	0.246	0.254	0.198
	p-value	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)	(0.00)
EMIG	Corr	0.755	1.000	0.757	0.310	0.361	0.368	0.136
	p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.11)
EMHY	Corr	0.484	0.757	1.000	0.504	0.511	0.525	0.251
	p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
EEUU	Corr	0.164	0.310	0.504	1.000	0.898	0.800	0.351
	p-value	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
EAFE	Corr	0.246	0.361	0.511	0.898	1.000	0.880	0.498
	p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
EM	Corr	0.254	0.368	0.525	0.800	0.880	1.000	0.492
	p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Comm	Corr	0.198	0.136	0.251	0.351	0.498	0.492	1.000
	p-value	(0.00)	(0.11)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Cuadro 4.6: Matriz de correlación y p-value de los retornos

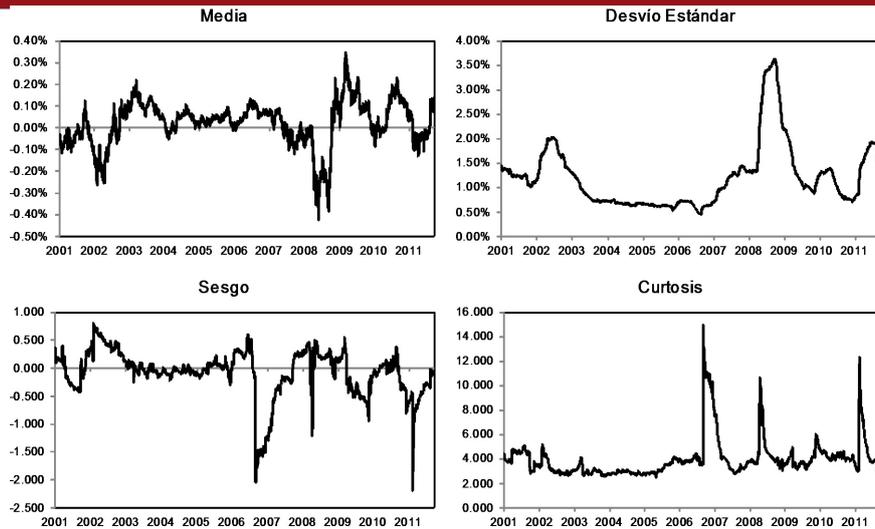


Figura 4.5: Los cuatro primeros momentos de los retornos diarios del S&P500 (EE.UU.) varían en el tiempo usando una muestra móvil de 6 meses

De manera gráfica, se puede observar que las correlaciones de los retornos del S&P500 (EEUU) frente a cuatro diferentes índices varían en el tiempo. En particular, la Figura 4.6 muestra que las correlaciones de EEUU vs EM y Comm se ha incrementado luego de la crisis y se han mantenido en esos niveles.

En conclusión, dados los resultados anteriores, asumir que los retornos de los índices tengan un comportamiento gaussiano podría generar una equivocada elección del portafolio, debido a la subestimación de pérdidas extremas. El siguiente paso será filtrar las series de la autocorrelación.

Índices	Estat	USIG	EMIG	EMHY	EEUU	EAFE	EM	Comm
Durante la crisis								
USIG	Corr	1.000	0.786	0.761	0.428	0.557	0.611	0.317
	p-value	(0.00)	(0.03)	(0.04)	(0.33)	(0.19)	(0.14)	(0.48)
EMIG	Corr	0.786	1.000	0.942	0.754	0.841	0.862	0.457
	p-value	(0.03)	(0.00)	(0.00)	(0.05)	(0.01)	(0.01)	(0.30)
EMHY	Corr	0.761	0.942	1.000	0.820	0.873	0.932	0.688
	p-value	(0.04)	(0.00)	(0.00)	(0.02)	(0.01)	(0.00)	(0.08)
EEUU	Corr	0.428	0.754	0.820	1.000	0.985	0.895	0.729
	p-value	(0.33)	(0.05)	(0.02)	(0.00)	(0.00)	(0.00)	(0.06)
EAFE	Corr	0.557	0.841	0.873	0.985	1.000	0.925	0.699
	p-value	(0.19)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.08)
EM	Corr	0.611	0.862	0.932	0.895	0.925	1.000	0.794
	p-value	(0.14)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)
Comm	Corr	0.317	0.457	0.688	0.729	0.699	0.794	1.000
	p-value	(0.48)	(0.30)	(0.08)	(0.06)	(0.08)	(0.03)	(0.00)

Cuadro 4.7: Matriz de correlación y p-value de los retornos en el 2008

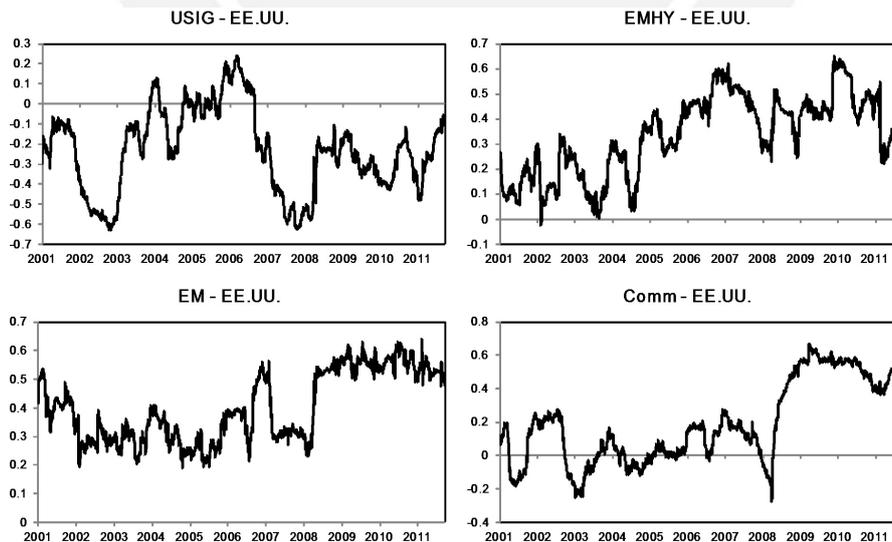


Figura 4.6: Correlaciones de retornos diarios del S&P500 (EE.UU.) y otros activos usando una ventana de 6 meses

4.5.3 Elección del modelo de serie de tiempo

En la sección anterior mostré, a través de las gráficas de autocorrelación de los retornos y los retornos al cuadrado, que los retornos mostraban indicios de autocorrelación y heteroscedasticidad. El test de Ljung Box confirmó la presencia de estas características en todas las series. Por ello, ajustaremos los modelos univariados ARMA(1,1) para la media condicional y, en primera instancia, modelos GARCH(1,1) para la varianza condicional. Luego, modelaremos la varianza a través del modelo GJR para capturar el efecto de palanca. Primero asumiremos que los errores se distribuyen como una normal; luego, dada la presencia de colas anchas, asumiremos que los errores se distribuyen como una t-student.

De esta manera, estimaremos cuatro modelos por índice. Para elegir al modelo más adecuado de entre los cuatro estimados utilizaremos el valor de la función de verosimilitud. Además, dado que se estimarán modelos con distinto número de variables, utilizaremos los criterios del Akaike (AIC) y el criterio de Schwarz (BIC).

El Cuadro 4.8 muestra los parámetros estimados de los cuatro modelos del índice USIG junto con sus respectivos p-value entre paréntesis.

Índice	USIG			
	ARMA-GARCH		ARMA-GJR	
Distribución	Normal	t-Student	Normal	t-Student
Constante	0.0050 (0.000)	0.0054 (0.000)	0.0051 (0.000)	0.0054 (0.000)
AR(1)	-0.4497 (0.080)	-0.6190 (0.000)	-0.6764 (0.000)	-0.6073 (0.000)
MA(1)	0.6861 (0.002)	0.8162 (0.000)	0.8581 (0.000)	0.8082 (0.000)
Constante	0.0000 (0.649)	0.0000 (0.478)	0.0000 (0.421)	0.0000 (0.486)
ARCH(1)	0.0270 (0.670)	0.0708 (0.484)	-0.0640 (0.396)	0.0952 (0.613)
Coef. Apalanc	- -	- -	0.1216 (0.424)	-0.0301 (0.865)
GARCH(1)	0.8098 (0.047)	0.8065 (0.001)	0.8215 (0.001)	0.8055 (0.001)
Grados de Libertad	- -	4.7520 (0.001)	- -	4.7244 (0.001)

Cuadro 4.8: Coeficientes estimados, p-values y grados de libertad del índice USIG

Para el modelo de la media condicional, existe evidencia estadística que soporta la idea de que los retornos están autocorrelacionados, lo cual se mostró en los valores de los test Q de Ljung-Box de orden 5 y 10. Por otro lado, para el modelo de la varianza condicional, no existe evidencia del efecto ARCH. Tampoco existe evidencia del efecto apalancamiento en el modelo GJR. No obstante, existe evidencia de que existe el efecto GARCH en todos los modelos. Finalmente, el valor de los grados de libertad y sus p-values asociados parecen soportar la idea de que asumir una distribución t-student para los errores parece ser adecuada.

El Cuadro 4.9 muestra el valor de los tres criterios de selección del modelo con mejor ajuste.

El mayor valor de la función de verosimilitud y los menores valores del AIC y BIC sugieren que la mejor elección sería el modelo ARMA-GARCH con distribución de errores t-student. Esto es consistente con el hecho que los coeficientes de apalancamiento no sean significativos, descartando al modelo ARMA-GJR. Además, el valor así como la significancia de los grados de libertad sugieren que asumir una distribución t-student para los errores sería adecuada.

Modelo	Distribución	\hat{l}	AIC	BIC
ARMA-GARCH	Normal	440.9311	-6.3960	-6.2675
	t-Student	454.0789	-6.5747	-6.4248
ARMA-GJR	Normal	441.6405	-6.3918	-6.2419
	t-Student	454.1047	-6.5604	-6.3890

Cuadro 4.9: Valor de la función de verosimilitud, AIC y BIC

A pesar que el modelo ARMA-GARCH con distribución de errores t-student sería la mejor elección, nos falta ver si los residuos estandarizados de esta regresión no presentan evidencia de autocorrelación ni heteroscedasticidad. Los residuos estandarizados, \hat{z}_t , se calculan a partir de la muestra (x_1, \dots, x_n) y de los parámetros estimados de la siguiente manera:

$$\hat{z}_t = \frac{(x_t - \hat{\mu}_t)}{\hat{\sigma}_t}$$

para $t = 1, 2, \dots, n$. El Cuadro 4.10 muestra los estadísticos de Ljung-Box de orden 5, 10, 15 y 20, así como sus p-values entre paréntesis, de los residuos estandarizados del modelo seleccionado.

	Autocorrelación en los residuos estandarizados de orden			
	5	10	15	20
Q test	1.258	6.314	10.636	12.227
p-value	(0.739)	(0.612)	(0.641)	(0.835)
	Autocorrelación en los residuos estandarizados al cuadrado de orden			
	5	10	15	20
Q test	0.248	1.171	2.007	3.041
p-value	(0.969)	(0.997)	(1.000)	(1.000)

Cuadro 4.10: P-values de las pruebas de autocorrelación de Ljung Box de los residuos estandarizados

En efecto, los resultados muestran que no se puede rechazar la hipótesis nula de ausencia de autocorrelación en los residuos estandarizados ni en los residuos estandarizados al cuadrado. Por ello, para el caso del índice USIG, el modelo elegido será el ARMA(1,1)-GARCH(1,1) con distribución de errores t-student.

De manera análoga, se realizará el mismo análisis para el resto de índices. El Cuadro 4.18 muestra las elecciones finales.

Por otro lado, el Cuadro 4.17 muestra el resumen de los tres criterios utilizados en la selección del modelo ARMA-GARCH de todos los índices.

Índice	EMIG			
	ARMA-GARCH		ARMA-GJR	
Modelo	Normal	t-Student	Normal	t-Student
Constante	0.0079 (0.000)	0.0098 (0.000)	0.0070 (0.001)	0.0090 (0.000)
AR(1)	-0.2383 (0.733)	-0.9740 (0.000)	-0.3730 (0.671)	-0.8039 (0.204)
MA(1)	0.3845 (0.561)	0.9975 (0.000)	0.4565 (0.599)	0.7808 (0.245)
Constante	0.0002 (0.005)	0.0001 (0.233)	0.0002 (0.014)	0.0001 (0.114)
ARCH(1)	0.2517 (0.003)	0.1978 (0.167)	-0.1070 (0.415)	-0.0682 (0.721)
Coef. Apalanc	- -	- -	0.5286 (0.006)	0.3598 (0.166)
GARCH(1)	0.2156 (0.323)	0.4687 (0.208)	0.3060 (0.232)	0.5379 (0.044)
Grados de Libertad	- -	3.5381 (0.005)	- -	3.5992 (0.010)

Cuadro 4.11: Coeficientes estimados, p-values y grados de libertad del índice EMIG

Índice	EMHY			
	ARMA-GARCH		ARMA-GJR	
Modelo	Normal	t-Student	Normal	t-Student
Constante	0.0107 (0.003)	0.0132 (0.000)	0.0085 (0.021)	0.0127 (0.000)
AR(1)	0.0146 (0.962)	-0.0010 (0.997)	0.0587 (0.868)	0.0258 (0.930)
MA(1)	0.2302 (0.474)	0.2400 (0.419)	0.1868 (0.603)	0.2174 (0.470)
Constante	0.0007 (0.036)	0.0006 (0.079)	0.0006 (0.018)	0.0006 (0.065)
ARCH(1)	0.4256 (0.001)	0.5498 (0.103)	0.1623 (0.413)	0.3817 (0.382)
Coef. Apalanc	- -	- -	0.4483 (0.134)	0.2576 (0.601)
GARCH(1)	-0.0368 (0.925)	0.0701 (0.839)	0.0291 (0.933)	0.0848 (0.798)
Grados de Libertad	- -	3.7403 (0.017)	- -	3.9273 (0.021)

Cuadro 4.12: Coeficientes estimados, p-values y grados de libertad del índice EMHY

Índice	EEUU			
	ARMA-GARCH		ARMA-GJR	
Modelo	Normal	t-Student	Normal	t-Student
Constante	0.0052 (0.182)	0.0079 (0.016)	0.0004 (0.959)	0.0066 (0.047)
AR(1)	-0.7093 (0.000)	-0.5936 (0.004)	0.8047 (0.000)	-0.5574 (0.006)
MA(1)	0.9132 (0.000)	0.7808 (0.000)	-0.6234 (0.000)	0.7570 (0.000)
Constante	0.0001 (0.377)	0.0001 (0.370)	0.0002 (0.013)	0.0001 (0.122)
ARCH(1)	0.1613 (0.049)	0.1209 (0.151)	-0.1528 (0.019)	-0.1167 (0.404)
Coef. Apalanc	- -	- -	0.7342 (0.000)	0.2830 (0.020)
GARCH(1)	0.8093 (0.000)	0.8390 (0.000)	0.7335 (0.000)	0.8607 (0.000)
Grados de Libertad	- -	5.1470 (0.026)	- -	5.8705 (0.056)

Cuadro 4.13: Coeficientes estimados, p-values y grados de libertad del índice EE.UU.

Índice	EAFE			
	ARMA-GARCH		ARMA-GJR	
Modelo	Normal	t-Student	Normal	t-Student
Constante	0.0058 (0.248)	0.0041 (0.406)	0.0024 (0.589)	0.0049 (0.262)
AR(1)	-0.6467 (0.000)	-0.1171 (0.229)	-0.6586 (0.000)	-0.3769 (0.051)
MA(1)	0.9514 (0.000)	0.6020 (0.000)	0.9528 (0.000)	0.7040 (0.000)
Constante	0.0002 (0.257)	0.0010 (0.382)	0.0002 (0.001)	0.0002 (0.018)
ARCH(1)	0.1049 (0.048)	-0.0537 (0.354)	-0.1756 (0.019)	-0.2066 (0.013)
Coef. Apalanc	- -	- -	0.3351 (0.000)	0.3230 (0.007)
GARCH(1)	0.8195 (0.000)	0.5669 (0.285)	0.9128 (0.000)	0.9423 (0.000)
Grados de Libertad	- -	4.6818 (0.021)	- -	6.4466 (0.118)

Cuadro 4.14: Coeficientes estimados, p-values y grados de libertad del índice EAFE

Índice	EM			
	ARMA-GARCH		ARMA-GJR	
Modelo	Normal	t-Student	Normal	t-Student
Constante	0.0157 (0.022)	0.0187 (0.003)	0.0114 (0.115)	0.0133 (0.048)
AR(1)	-0.1938 (0.253)	-0.2389 (0.134)	-0.1999 (0.186)	-0.1492 (0.389)
MA(1)	0.6913 (0.000)	0.6850 (0.000)	0.6956 (0.000)	0.6564 (0.000)
Constante	0.0034 (0.014)	0.0037 (0.056)	0.0007 (0.006)	0.0029 (0.002)
ARCH(1)	0.1834 (0.026)	0.1539 (0.206)	-0.0965 (0.255)	-0.0224 (0.893)
Coef. Apalanc	- -	- -	0.3321 (0.004)	0.3700 (0.107)
GARCH(1)	-0.2726 (0.511)	-0.3345 (0.466)	0.6928 (0.000)	-0.1292 (0.595)
Grados de Libertad	- -	6.9561 (0.217)	- -	14.8643 (0.507)

Cuadro 4.15: Coeficientes estimados, p-values y grados de libertad del índice EM

Índice	Comm			
	ARMA-GARCH		ARMA-GJR	
Modelo	Normal	t-Student	Normal	t-Student
Constante	0.0088 (0.020)	0.0084 (0.024)	0.0075 (0.059)	0.0075 (0.062)
AR(1)	-0.2116 (0.653)	-0.2220 (0.602)	-0.2330 (0.626)	-0.2333 (0.626)
MA(1)	0.3831 (0.402)	0.3957 (0.339)	0.4124 (0.357)	0.4126 (0.357)
Constante	0.0002 (0.349)	0.0002 (0.362)	0.0004 (0.119)	0.0004 (0.139)
ARCH(1)	0.1533 (0.020)	0.1235 (0.187)	-0.0526 (0.719)	-0.0525 (0.721)
Coef. Apalanc	- -	- -	0.3311 (0.065)	0.3305 (0.101)
GARCH(1)	0.7012 (0.001)	0.7436 (0.001)	0.6067 (0.004)	0.6073 (0.015)
Grados de Libertad	- -	18.0657 (0.569)	- -	2,116.3 (0.997)

Cuadro 4.16: Coeficientes estimados, p-values y grados de libertad del índice Comm

		l	AIC	BIC
Modelo	Distribución		USIG	
ARMA-GARCH	Normal	440.9311	-6.3960	-6.2675
	t-Student	454.0789	-6.5747	-6.4248
ARMA-GJR	Normal	441.6405	-6.3918	-6.2419
	t-Student	454.1047	-6.5604	-6.3890
			EMIG	
ARMA-GARCH	Normal	348.5577	-5.0376	-4.9091
	t-Student	360.2708	-5.1952	-5.0452
ARMA-GJR	Normal	352.1995	-5.0765	-4.9265
	t-Student	359.3537	-5.1670	-4.9956
			EMHY	
ARMA-GARCH	Normal	282.0129	-4.0590	-3.9305
	t-Student	288.9639	-4.1465	-3.9966
ARMA-GJR	Normal	283.3364	-4.0638	-3.9139
	t-Student	289.1160	-4.1341	-3.9627
			EEUU	
ARMA-GARCH	Normal	250.2158	-3.5914	-3.4629
	t-Student	254.4749	-3.6393	-3.4894
ARMA-GJR	Normal	253.6127	-3.6267	-3.4767
	t-Student	257.0533	-3.6625	-3.4912
			EAFE	
ARMA-GARCH	Normal	233.6252	-3.3474	-3.2189
	t-Student	236.7957	-3.3793	-3.2294
ARMA-GJR	Normal	239.7222	-3.4224	-3.2725
	t-Student	240.2395	-3.4153	-3.2440
			EM	
ARMA-GARCH	Normal	203.5817	-2.9056	-2.7771
	t-Student	205.1988	-2.9147	-2.7648
ARMA-GJR	Normal	205.8453	-2.9242	-2.7743
	t-Student	206.4904	-2.9190	-2.7476
			Comm	
ARMA-GARCH	Normal	250.9883	-3.6028	-3.4743
	t-Student	251.3350	-3.5932	-3.4432
ARMA-GJR	Normal	253.1844	-3.6204	-3.4704
	t-Student	253.1825	-3.6056	-3.4343

Cuadro 4.17: Valor de la función de verosimilitud, AIC y BIC de los índices

De esta manera, dados los valores de los criterios de selección de los modelos, así como del análisis de los niveles de significancia de los coeficientes estimados de todos los índices, los modelos finalmente seleccionados se presentan en el Cuadro 4.18.

Finalmente, solo queda analizar si se logró remover la autocorrelación de los índices. Para ello, nuevamente se recurrirá al estadístico Q de Ljung Box. En los Cuadros 4.19 y 4.20 se observan los resúmenes de los valores del estadístico Q de Ljung Box así como sus p-values entre paréntesis para los residuos estandarizados y residuos estandarizados al cuadrado de los modelos escogidos de acuerdo con los criterios del valor de la función de verosimilitud, el AIC y BIC.

El ejercicio muestra que los residuos estandarizados de los modelos escogidos no muestran evidencia estadística de autocorrelación en los retornos ni en las volatilidades. Estos residuos estandarizados serán utilizados como insumos para calcular los parámetros de las cópulas en pares y hacer las simulaciones respectivas.

Índice	Modelo	Distribución
USIG	ARMA-GARCH	t-Student
EMIG	ARMA-GARCH	t-Student
EMHY	ARMA-GARCH	t-Student
EEUU	ARMA-GJR	t-Student
EAFE	ARMA-GJR	Normal
EM	ARMA-GJR	Normal
Comm	ARMA-GJR	Normal

Cuadro 4.18: Modelos seleccionados

		Autocorrelación en los residuos estandarizados de orden			
		5	10	15	20
USIG	Q test	1.258	6.314	10.636	12.227
	p-value	(0.739)	(0.612)	(0.641)	(0.835)
EMIG	Q test	5.555	9.282	18.162	20.223
	p-value	(0.135)	(0.319)	(0.151)	(0.320)
EMHY	Q test	1.531	3.467	6.653	14.245
	p-value	(0.675)	(0.902)	(0.919)	(0.713)
EEUU	Q test	4.125	4.992	5.701	7.999
	p-value	(0.248)	(0.758)	(0.956)	(0.979)
EAFE	Q test	7.581	8.273	10.873	14.384
	p-value	(0.108)	(0.407)	(0.621)	(0.704)
EM	Q test	4.402	7.608	12.362	17.521
	p-value	(0.221)	(0.473)	(0.498)	(0.488)
Comm	Q test	2.153	13.508	18.690	21.046
	p-value	(0.541)	(0.141)	(0.133)	(0.277)

Cuadro 4.19: Resumen de los p-values de las pruebas de autocorrelación de los residuos estandarizados

		Autocorrelación en los residuos estandarizados al cuadrado de orden			
		5	10	15	20
USIG	Q test	0.248	1.171	2.007	3.041
	p-value	(0.969)	(0.997)	(1.000)	(1.000)
EMIG	Q test	4.761	6.941	9.183	11.328
	p-value	(0.190)	(0.543)	(0.759)	(0.880)
EMHY	Q test	1.782	3.032	7.018	9.360
	p-value	(0.619)	(0.932)	(0.901)	(0.951)
EEUU	Q test	0.473	3.543	4.179	7.202
	p-value	(0.925)	(0.896)	(0.989)	(0.988)
EAFE	Q test	5.403	9.642	10.474	27.312
	p-value	(0.248)	(0.291)	(0.655)	(0.073)
EM	Q test	1.236	5.286	7.491	11.884
	p-value	(0.744)	(0.727)	(0.875)	(0.853)
Comm	Q test	5.639	6.988	11.285	15.668
	p-value	(0.131)	(0.638)	(0.587)	(0.616)

Cuadro 4.20: Resumen de los p-values de las pruebas de autocorrelación de los residuos estandarizados al cuadrado

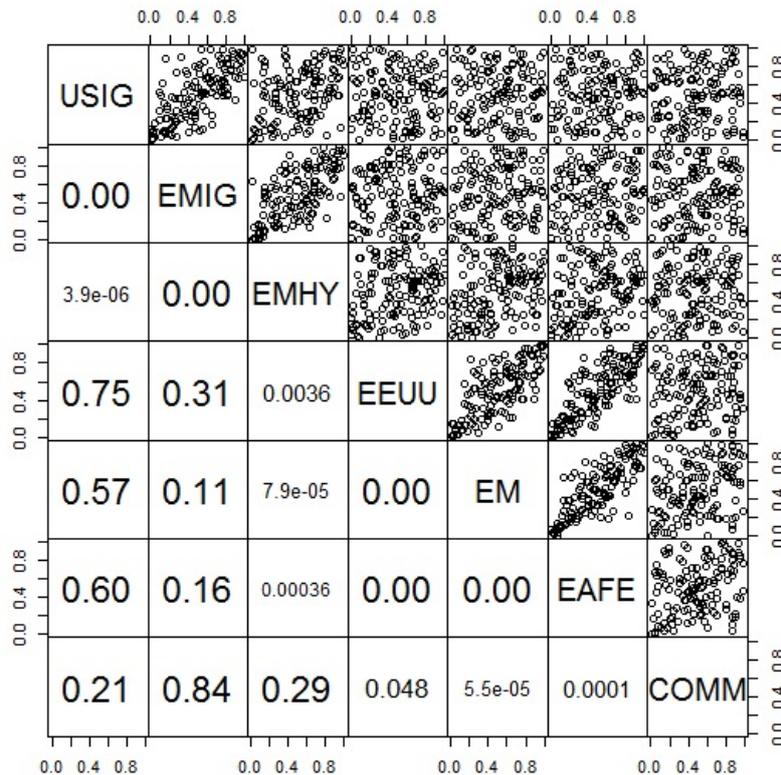


Figura 4.7: Sobre la diagonal principal se muestra el scatter plot de cada par de series mientras que debajo de la diagonal principal se muestran los p values de los test de independencia de cada par de variables

4.5.4 Elección de la vine cópula

La Figura 4.7 muestra sobre la diagonal principal los gráficos de dispersión de todos los índices y debajo de la diagonal principal los p-values de la prueba de independencia bivariada propuesto por Genest y Favre (2007) basado en el τ de Kendall. Se puede observar la baja dependencia, por ejemplo, entre USIG y el resto de activos que no forman parte de la renta fija (EE.UU., EM, EAFE, Comm) pues el p -value es mayor a 0.05.

Los p -value en la diagonal inferior de la figura anterior nos dice que no hay evidencia estadística de alguna variable pivote, pues no existe un caso en el cual el p -value de algún índice con respecto a los otros rechace la hipótesis nula de independencia bivariada.

Este resultado es coherente con el hecho que los índices, por ejemplo, de renta fija de EE.UU. (USIG) y los de la renta variable emergente (EM) o USIG y el índice de *commodities* (Comm) tengan una fuerte relación. Los determinantes del comportamiento de los índices de renta fija grado de inversión en EE.UU. (USIG) son diferentes a, por ejemplo, los de *commodities*. Si bien es cierto parte del cambio en el precio de los *commodities* se debe a factores financieros, las noticias económicas provenientes de China (mayor consumidor mundial de *commodities*) es uno de los principales determinantes de sus precios.

De esta manera, teniendo en cuenta este hecho junto con los resultados del test de independencia bivariada, se elegirá como estructura de datos a la D-Vine.

4.5.5 Estimación de parámetros de las cópulas en pares de los retornos filtrados

El siguiente paso es seleccionar las familias de cópulas bivariadas adecuadas para cada par de variables en todos los árboles. Como se mencionó en la sección 3.11, existen varios criterios de selección de una cópula. Se utilizarán estos criterios para explicar en detalle la elección de la primera cópula bivarida (USIG-EMIG) del primer árbol.

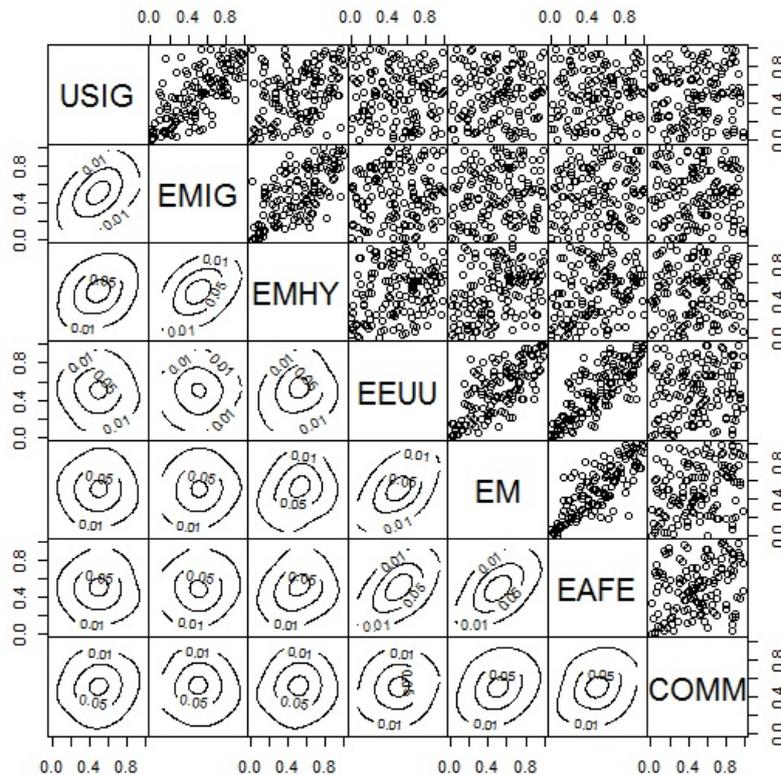


Figura 4.8: Debajo de la diagonal principal se encuentran los gráficos de contornos; mientras que encima de la diagonal principal se encuentran los gráficos de dispersión de las nueve variables

La Figura 4.8 muestra sobre la diagonal principal los gráficos de dispersión y los gráficos de contornos empíricos en la inferior para todas las variables. En general, se puede apreciar que no existe una relación negativa entre las diferentes variables. Esto implica que no se tomarían en cuenta como candidatas a las versiones de cópulas bivariadas rotadas 90° y 270° . Por otro lado, se puede apreciar que, en términos relativos, existe una mayor relación positiva entre los índices de la renta variable. En el caso de la relación USIG-EMIG, se puede apreciar que existe una relación positiva entre ellas.

Por otro lado, las Figuras 4.9 y 4.10 muestran debajo de la diagonal principal los gráficos K-plot y chi-plot, respectivamente. Estos gráficos confirman que no existen relaciones negativas entre las diferentes variables. Para el caso de la relación USIG-EMIG, ambos gráficos confirman que existe entre ellas una relación positiva.

La Figura 4.11 muestra en la esquina superior izquierda la función lambda empírica y en el resto la misma función junto con las funciones lambda de una normal, t, Clayton, Gumbel, Frank, Joe, BB1 y BB6. Debido a la cantidad de cópulas bivariadas, en la Figura 4.12 se encuentran las funciones lambda teóricas de las BB7, BB8 y las versiones rotadas 180° de la Clayton, Gumbel, Joe, BB1, BB6, BB7 y BB8.

Varias de estas cópulas son buenas candidatas para representar la relación bivariada de USIG-EMIG. En particular, se podría descartar a la cópula Joe y la rotada Clayton.

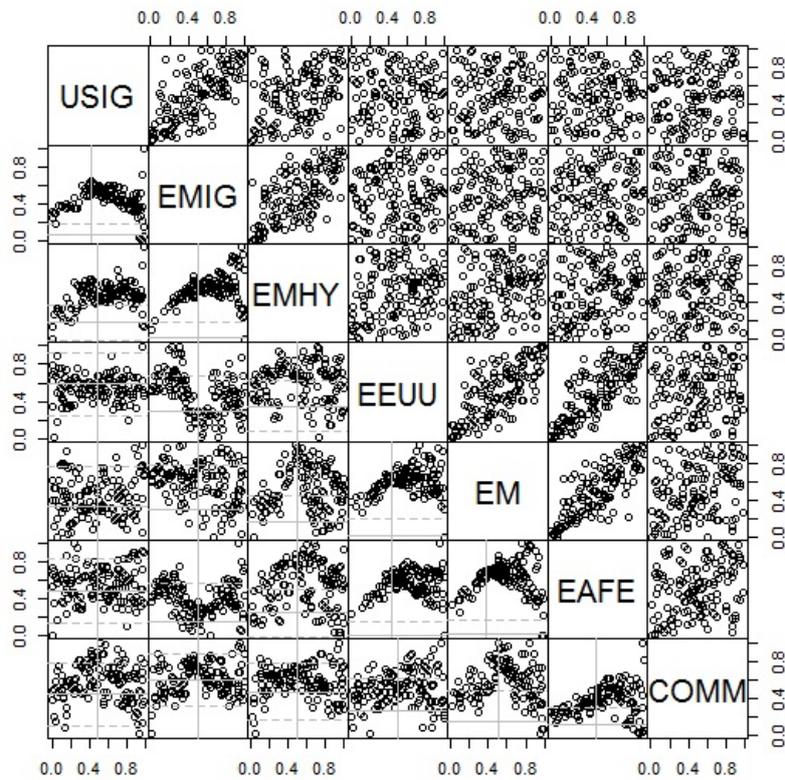


Figura 4.9: Sobre la diagonal principal se muestra el scatter plot de cada par de series mientras que debajo de la diagonal principal se muestran los Chi plots para cada par de variables

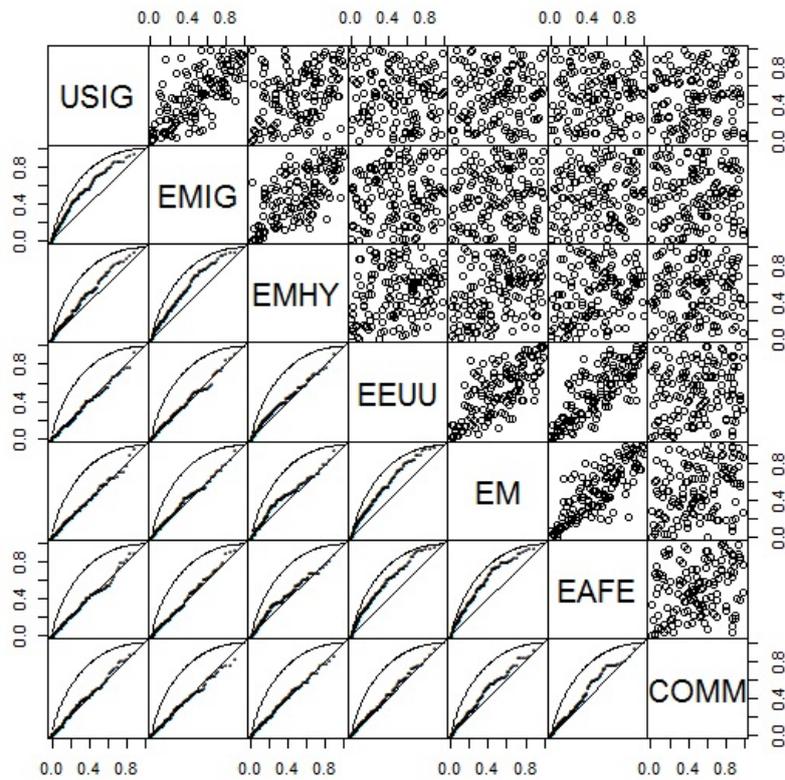


Figura 4.10: Sobre la diagonal principal se muestra el scatter plot de cada par de series mientras que debajo de la diagonal principal se muestran los K plots para cada par de variables

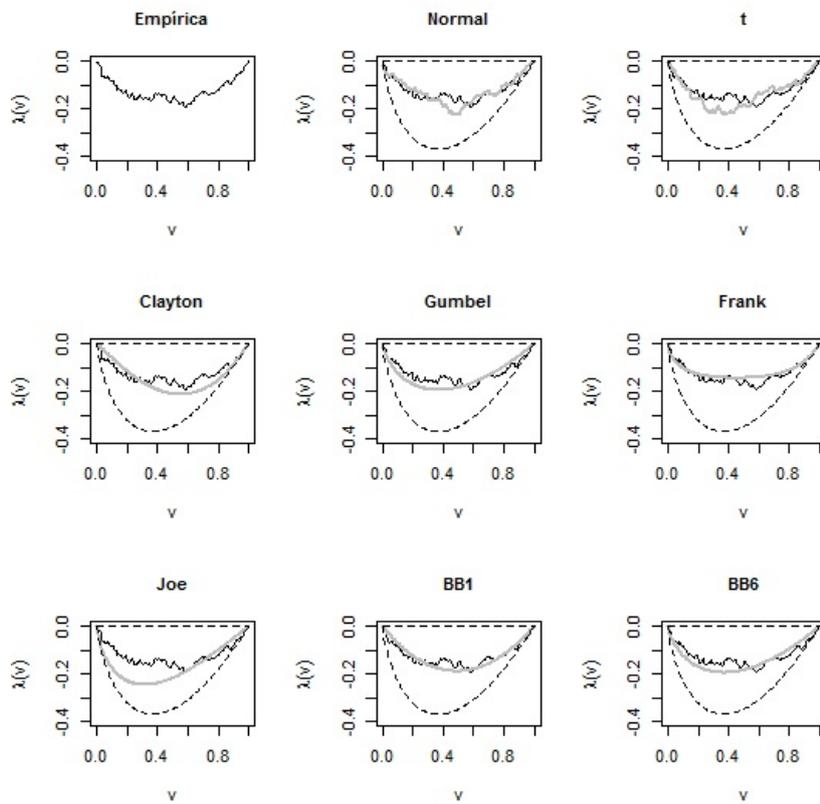


Figura 4.11: λ -función empírica de la cópula en par USIG-EMIG y de las teóricas Gaussiana, t, clayton, Gumbel, Frank, Joe, BB1 y BB6

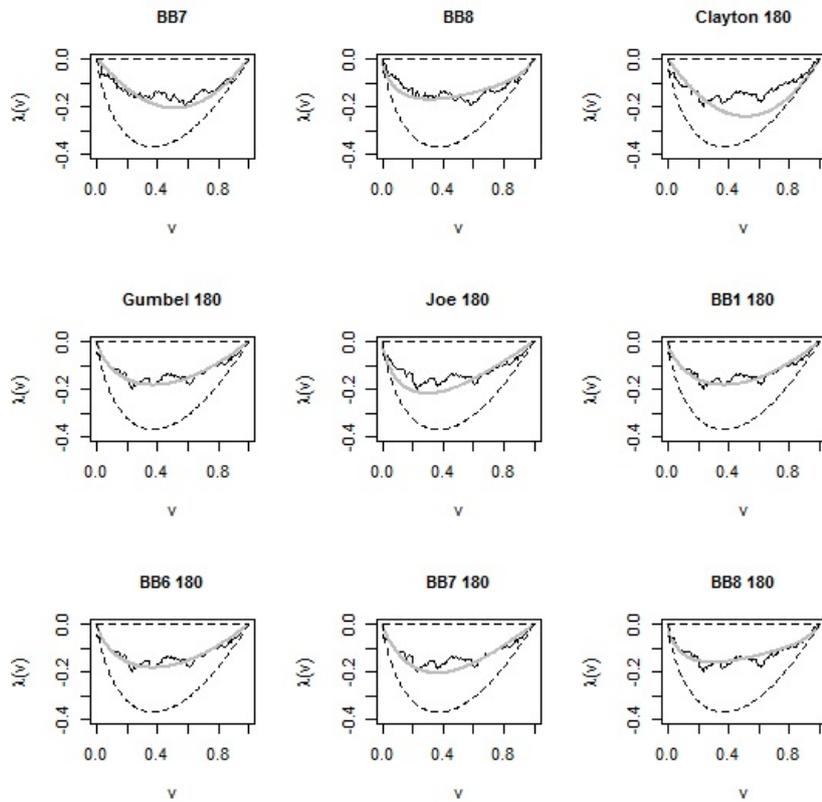


Figura 4.12: λ -función de las cópula teóricas BB7, BB8 y las rotadas 180° Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7 y BB8

Finalmente, se estimarán los parámetros de la cópula bivariada para USIG-EMIG a través de máxima verosimilitud. El Cuadro 4.21 muestra el valor del logaritmo de la función de verosimilitud, el valor del AIC, el BIC y los criterios de selección de Vuong y Clarke.

De acuerdo al valor del logaritmo de la verosimilitud la cópula elegida sería la BB1 rotada 180°. No obstante, dado que las cópulas BB1, BB6, BB7 y BB8 son de dos parámetros, los valores de AIC y BIC serían más adecuados para elegir a la cópula para USIG-EMIG. Ambos criterios indican que la cópula elegida sería la Gumbel rotada 180°. Además, el criterio de selección de Vuong, basado en darle un score positivo al modelo que mejor ajusta los datos, indica lo mismo, aunque el criterio de Clarke indica que la elegida sería la Frank.

	USIG-EMIG				
	l	AIC	BIC	Vuong	Clarke
Gaussiana	47.0802	-92.1604	-89.24774	4	7
t-Student	48.1511	-92.30226	-86.47695	4	9
Clayton	44.93877	-87.87755	-84.96489	1	-10
Gumbel	40.47271	-78.94542	-76.03277	1	-2
Frank	47.69148	-93.38297	-90.47031	3	15
Joe	27.59645	-53.19291	-50.28025	-14	-15
BB1	48.77615	-93.5523	-87.72699	4	4
BB6	40.46121	-76.92242	-71.09711	-1	-8
BB7	46.96222	-89.92443	-84.09912	0	-5
BB8	44.032974	-84.06594	-78.24064	1	4
Clayton 180	29.24313	-56.48626	-53.5736	-12	-15
Gumbel 180	49.12565	-96.2513	-93.33864	6	9
Joe 180	43.38846	-84.77693	-81.86427	-2	-9
BB1 180	49.2005	-94.40099	-88.57568	3	6
BB6 180	49.12293	-94.24585	-88.42054	1	3
BB7 180	46.45733	-88.91466	-83.08935	-1	-4
BB8 180	48.34615	-92.6923	-86.86699	2	11

Cuadro 4.21: Criterios de selección para la cópula bivariada USIG-EMIG

De esta manera, la cópula Gumbel rotada 180° es la elegida para representar la dependencia bivariada de USIG-EMIG. A modo de resumen, la Figura 4.13 muestra el gráfico de dispersión de los retornos del índice USIG frente al EMIG, así como los datos en escala de cópulas. El gráfico de contorno mostró junto con el k-plot y el Chi-plot que la relación entre ambas variables es positiva. La función lambda sugirió que la cópula Gumbel rotada 180° sería una buena elección; mientras que los criterios de elección dieron a ésta como la que mejor ajusta los datos.

El ejercicio es análogo para el resto de índices así como para el resto de árboles.

Con la elección de las cópulas en pares se obtienen los parámetros respectivos y se procede a realizar la simulación de 5,000 valores de cada uno de los residuos estandarizados. Estos valores preservan la dependencia contemporánea de los datos. Luego, se procede a reintroducir la dependencia temporal a través de los parámetros estimados de los modelos ARMA-GARCH. Finalmente, se obtienen los retornos simulados de los siete índices que se utilizarán en la optimización del portafolio de inversión.

4.5.6 Test de bondad de ajuste de las cópulas

Aunque la elección del D-vine como estructura de los datos parece la más adecuada, se sometió a prueba estadística si haber elegido un C-vine en lugar de un D-vine ajusta mejor los datos. Para ello, se estimó un modelo C-vine y se calculó el valor del logaritmo de la función de verosimilitud así como los AIC y BIC. Los resultados presentados en el Cuadro 4.22 muestran que haber optado por un D-vine ajusta mejor los datos.

Adicionalmente, se realizan las pruebas de Clarke y Vuong. Los valores de los p-values no son concluyentes entre la elección de una estructura u otra (Cuadro 4.23).

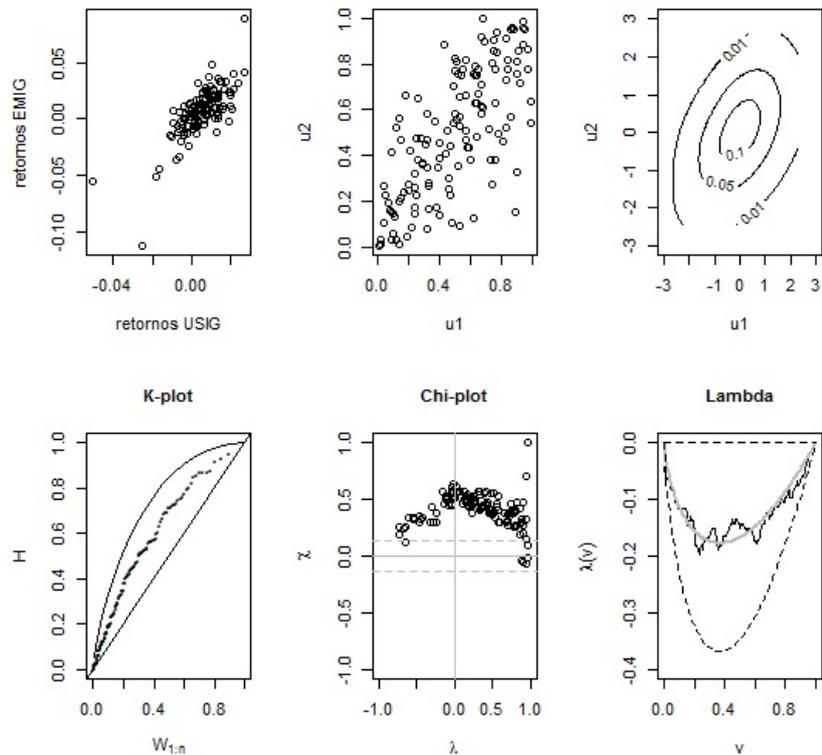


Figura 4.13: En la primera fila se observan los gráficos de dispersión de los retornos, de los datos en escala de cópulas y el gráfico de contornos. En la fila inferior, el K-plot, Chi-plot y la función lambda empírica y la teórica Gumbel rotada 180°

	C-vine	D-vine
l	290.179	299.0366
AIC	-550.358	-576.0733
BIC	-506.6682	-544.0341

Cuadro 4.22: Valor de la función de verosimilitud, AIC y BIC de la estructura de la vine

	Clarke	Vuong
Stat	64	-0.9266
p-value	(0.5485)	(0.3541)
Stat AIC	62	-1.3451
p-value	(0.3455)	(0.1785)
Stat BIC	60	-1.9545
p-value	(0.1981)	(0.0506)

Cuadro 4.23: P-value de las pruebas de Clarke y Vuong de la vine

4.5.7 Test de bondad de ajuste de los retornos simulados de los activos

Luego de la estimación de las cópulas bivariadas, la simulación de los residuos estandarizados y la posterior re-incorporación de la dependencia temporal (utilizando los coeficientes estimados del modelo ARMA-GARCH), se quiere saber si este ejercicio logró replicar el comportamiento de los retornos de los diferentes índices en un contexto donde estos no son normales. Se debe tener presente que en los casos en los cuales las distribuciones marginales no son normales (como es nuestro caso), la distribución multivariada tampoco será normal y, más aun, en los casos en los cuales las distribuciones marginales son diferentes, la distribución conjunta simplemente no existe.

Análisis gráfico

Para cada índice, se muestra a la izquierda el qq-plot de los retornos históricos asumiendo normalidad y a la derecha el qq-plot de los retornos históricos frente a los retornos simulados. Se observa que en todos los casos los retornos simulados aproximan mejor a los retornos históricos, en particular, se observa un mejor ajuste en las colas. Esto será sometido a prueba estadística a través de las pruebas de Kolmogorov Smirnov y Anderson Darling.

Análisis estadístico

De manera más formal, se realizan las pruebas de bondad de ajuste a través de la prueba de Kolmogorov-Smirnov (K-S) para dos muestras. El contraste estadístico compara la máxima diferencia entre la distribución de los retornos simulados y los retornos históricos, donde la hipótesis nula es que las dos variables tienen la misma distribución. El estadístico K-S es el siguiente:

$$K-S = \sup_{x \in \mathbb{R}} |F(x) - \hat{F}(x)|,$$

donde $F(x)$ es la función de distribución de los datos y $\hat{F}(x)$ es la función de distribución estimada.

Como se mencionó anteriormente, el test de K-S le otorga la misma importancia a las colas que al centro de las distribuciones que está comparando, es decir, mide la distancia máxima entre las dos distribuciones, pero sin importar dónde se mide esa distancia. Una alternativa a ello es utilizar el test de Anderson Darling (A-D). Como en el caso del test K-S, el test de A-D mide la distancia entre las dos distribuciones, pero está reescalado al dividir la distancia por el desvío estándar de esa distancia. De esta manera, el denominador se vuelve pequeño

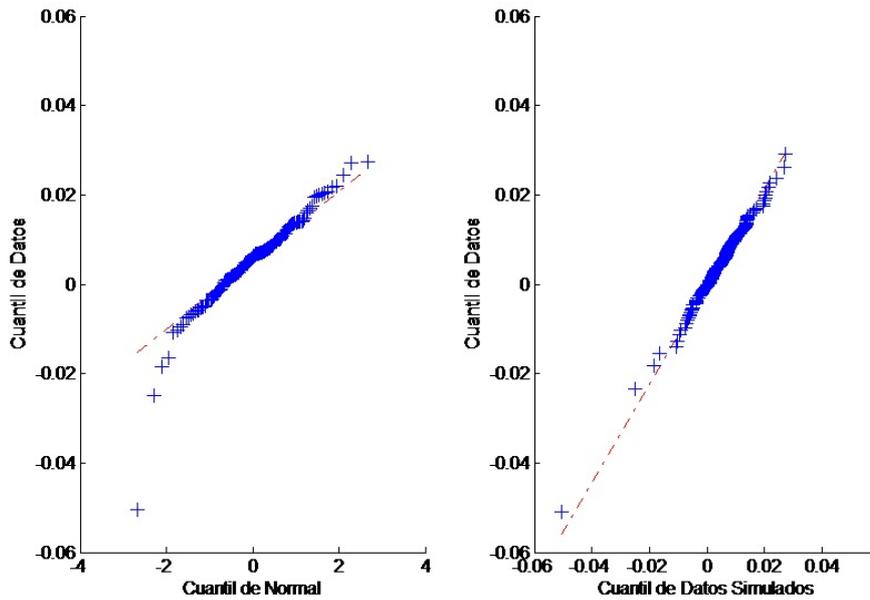


Figura 4.14: QQ-plot de USIG con distribución normal (izquierda) y con los datos simulados (derecha)

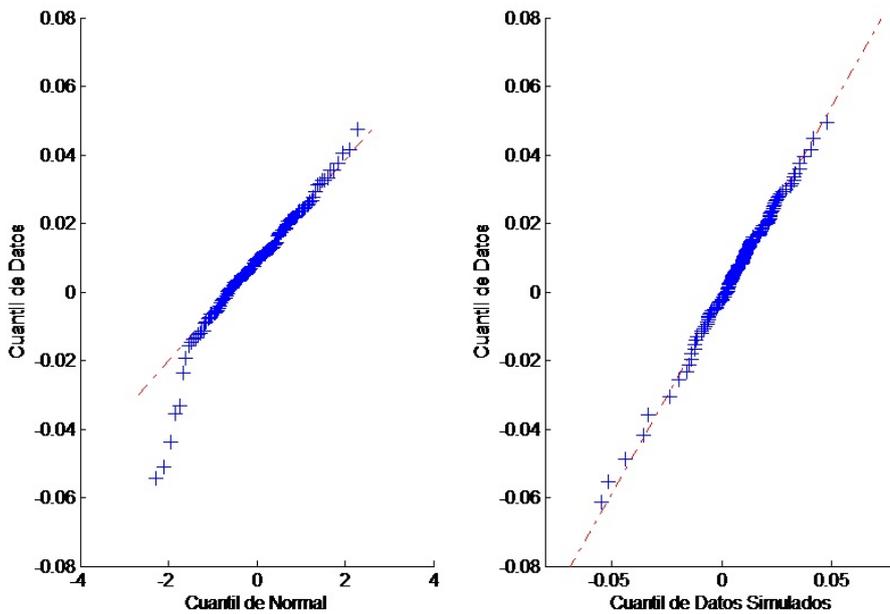


Figura 4.15: QQ-plot de EMIG con distribución normal (izquierda) y con los datos simulados (derecha)

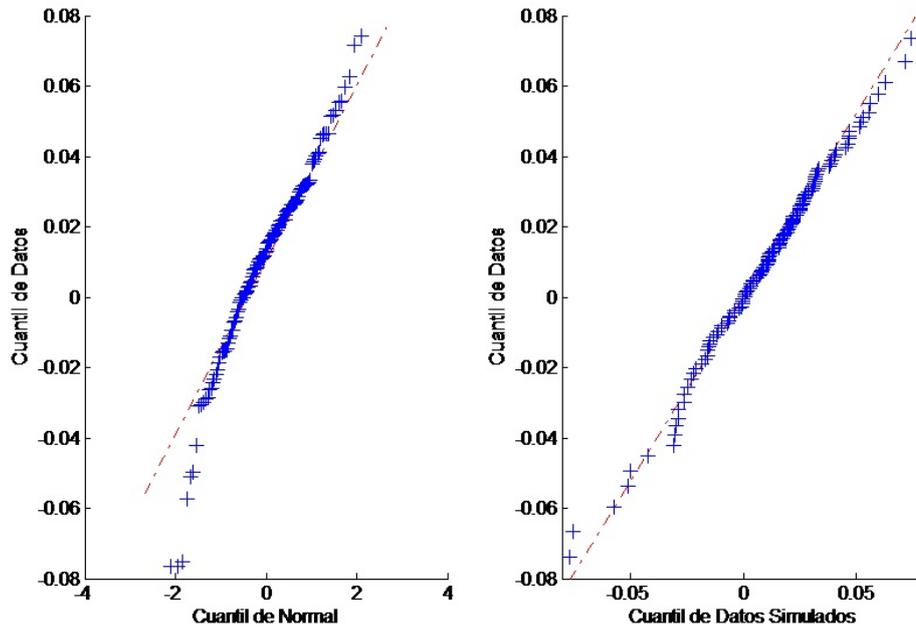


Figura 4.16: QQ-plot de EMHY con distribución normal (izquierda) y con los datos simulados (derecha)

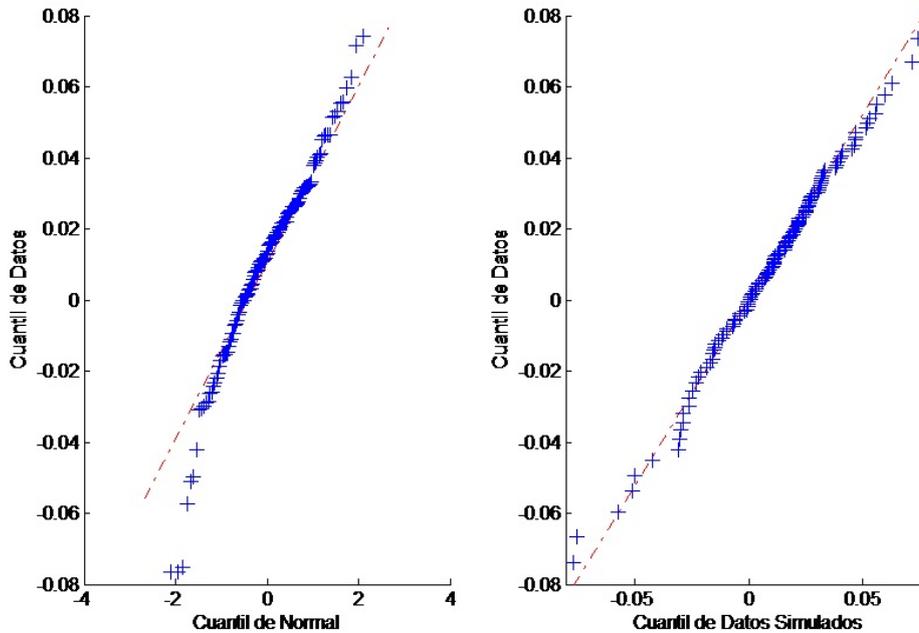


Figura 4.17: QQ-plot de EEUU con distribución normal (izquierda) y con los datos simulados (derecha)

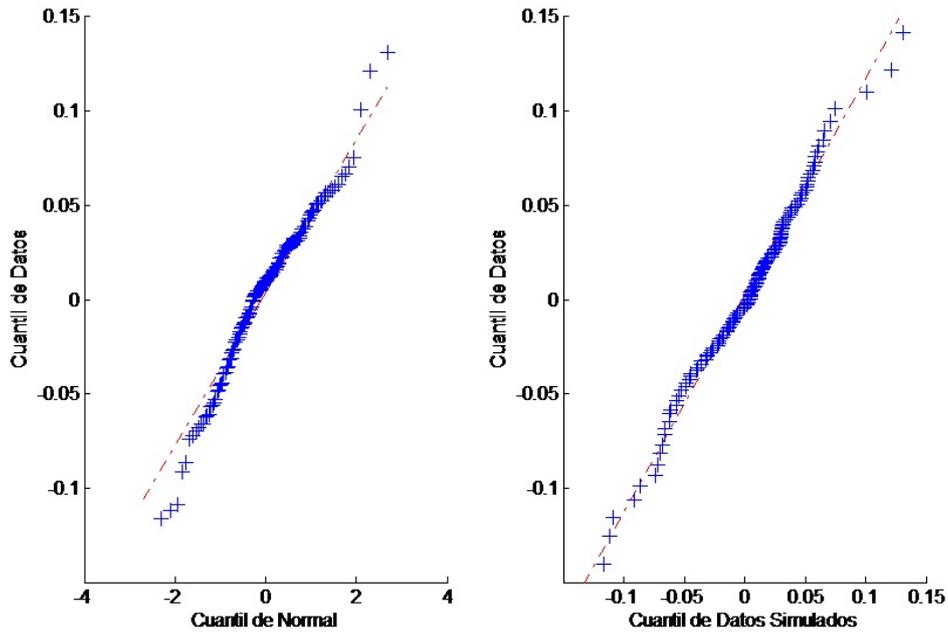


Figura 4.18: QQ-plot de EAFE con distribución normal (izquierda) y con los datos simulados (derecha)

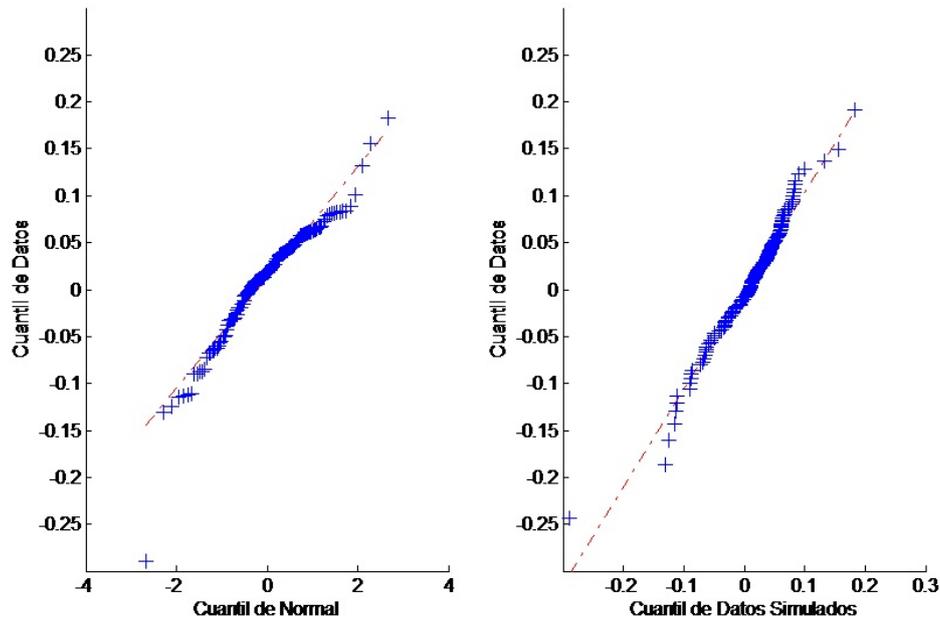


Figura 4.19: QQ-plot de EM con distribución normal (izquierda) y con los datos simulados (derecha)

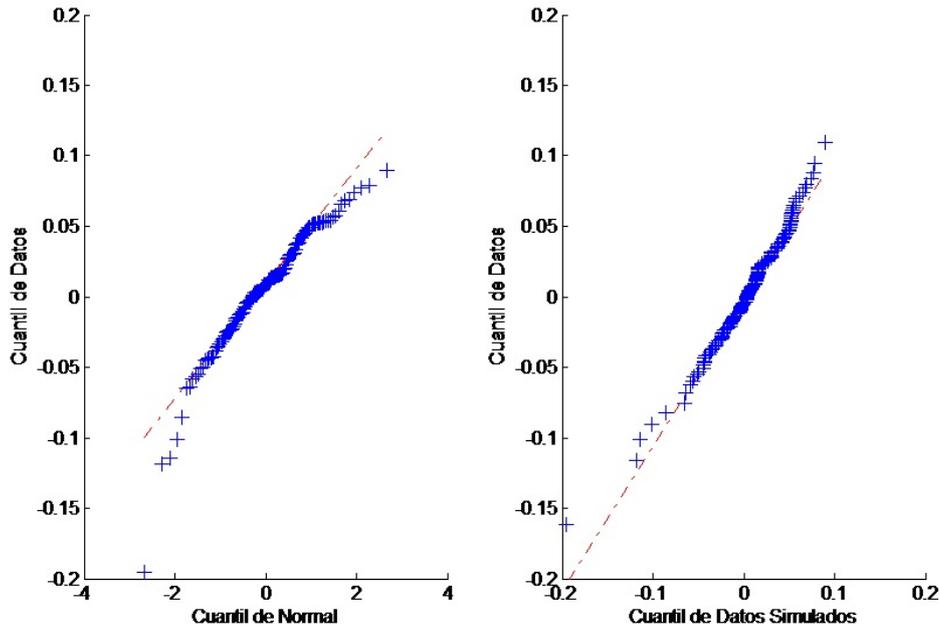


Figura 4.20: QQ-plot de Comm con distribución normal (izquierda) y con los datos simulados (derecha)

en los valores extremos de x . Así, la misma desviación entre F y F_n en las colas obtienen más peso que en el centro de la distribución. El estadístico A-D es el siguiente:

$$A-D = \sup_{x \in \mathbb{R}} \frac{|F(x) - \hat{F}(x)|}{\sqrt{\hat{F}(x)(1 - \hat{F}(x))}}$$

Los estadísticos así como los respectivos p -value de las pruebas se muestran en el Cuadro 4.24.

Índice	K-S		A-D	
	Stat	p-value	Stat	p-value
USIG	0.0344	0.9969	-1.0539	0.6428
EMIG	0.0553	0.7981	-0.7004	0.5380
EMHY	0.0372	0.9916	-1.1242	0.6623
EEUU	0.0575	0.7577	-0.4142	0.4515
EAFE	0.0709	0.4990	-0.0359	0.3414
EM	0.0781	0.3763	0.0500	0.3182
Comm	0.0663	0.5860	-0.4367	0.4583

Cuadro 4.24: P-value de las pruebas de K-S y A-D

Todos los p -value indican que los datos históricos y los simulados, para cada índice, tienen la misma distribución, con lo cual se valida lo mostrado por los gráficos qq-plot. De esta manera, se logró simular el comportamiento de los índices a través de la simulación de la dependencia contemporánea de los datos con la utilización de las cópulas en pares.

En la siguiente sección se realizará el ejercicio de optimización del portafolio de inversión con los datos simulados. Para ello, se optimizará el portafolio utilizando como medida de riesgo al CVaR.

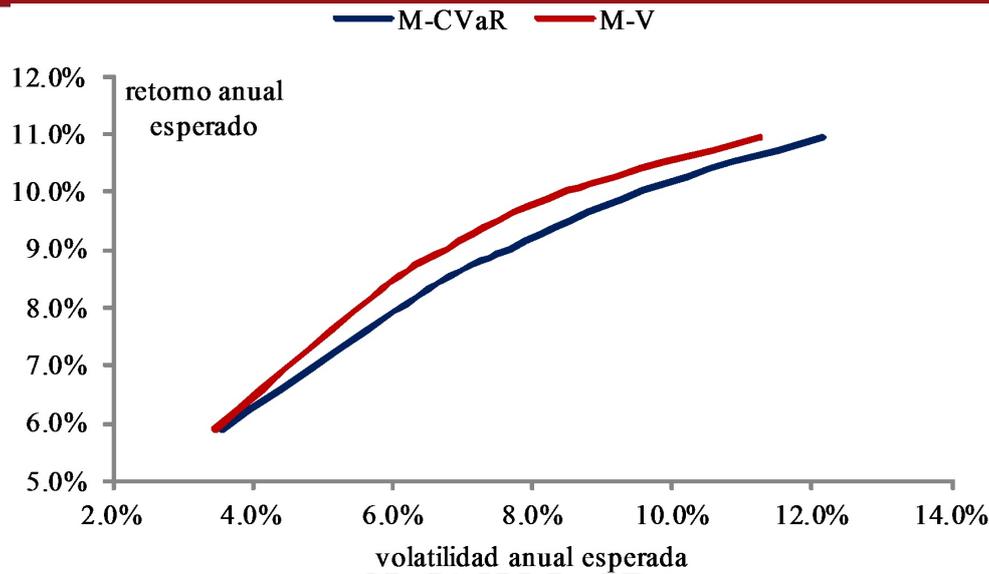


Figura 4.21: Fronteras eficientes basadas en el modelo M-V y M-CVaR

4.6 Optimización del modelo M-CVaR

En esta sección se presentan los resultados de las optimizaciones del modelo M-CVaR y del modelo M-V. En la sección anterior se simuló 5,000 escenarios para cada índice con los cuales se realizará la optimización del modelo M-CVaR. Con fines de comparación, las optimizaciones no se realizaron con restricciones de participación mínima en algún activo. Solo se realizaron con la usual restricción de ventas en corto.

4.6.1 Frontera eficiente

La frontera eficiente derivada del modelo M-CVaR difiere de la frontera eficiente derivada del modelo M-V (Figura 4.21 y Cuadro 4.25). El punto más resaltante radica en el hecho que los niveles de riesgos, medidos a través de la desviación estándar o a través del CVaR de los portafolios, son sistemáticamente menores en el modelo M-V frente al M-CVaR, dado un mismo nivel de retorno esperado. Este resultado soporta la idea de que al no tomar en cuenta la no normalidad de los retornos de los activos, el modelo M-V subestimaría los riesgos de estos portafolios, en particular, durante episodios de estrés en los mercados. Estos aparentes menores riesgos a la baja del modelo M-V se deben a la incapacidad de la varianza, como medida de riesgo, de capturar los hechos estilizados de los activos.

Retorno	Desvío estándar		CVaR	
	M-V	M-CVaR	M-V	M-CVaR
5.88%	3.57%	3.57%	2.16	2.16
8.40%	6.58%	6.65%	4.65	4.61
9.00%	6.62%	7.69%	5.38	5.33
9.50%	8.49%	8.56%	5.97	5.94
10.00%	9.56%	9.60%	6.67	6.64
10.50%	10.81%	10.88%	7.52	7.48
10.92%	11.95%	12.16%	8.33	8.29

Cuadro 4.25: Valores de las fronteras eficientes de los modelos M-V y M-CVaR

4.6.2 Portafolio óptimo

El Cuadro 4.26 muestra las diferencias en las asignaciones de ambos portafolio óptimos. Para un mismo nivel de retorno requerido de 8.4% anual, ambos modelos le dan más peso a los activos de renta fija. Para el caso de la renta variable, el modelo M-V parece ser más conservador, al darle un menor peso que el modelo M-CVaR a los mercados emergentes (EM). Sin embargo, aunque el modelo M-V parecería ser más consevador al darle menor peso a la renta variable, el modelo M-V le da más de la mitad del peso del portafolio a los bonos grado de inversión de las economías emergentes (EMIG). Por su parte, el modelo M-CVaR tiende a no concentrar los riesgos, no solo dentro de la renta fija, sino en todo el portafolio pues la asignación de los pesos está menos concentrada.

	Allocation M-V	Allocation M-CVaR
Renta Fija		
USIG	31.28 %	31.07 %
EMIG	62.71 %	37.36 %
EMHY	1.73 %	23.61 %
Renta Variable		
EEUU	0.00 %	0.00 %
EAFE	0.00 %	0.00 %
EM	4.28 %	7.96 %
Otros		
Commodity	0.00 %	0.00 %
Total	100.00 %	100.00 %
Estadísticos		
Retorno esperado	8.40 %	8.40 %
Volatilidad esperada	5.96 %	6.65 %
Ratio de Sharpe	0.73	0.66
VaR	1.54	2.76
CVaR	4.14	4.61

Cuadro 4.26: Asignación de activos de los modelos M-V y M-CVaR para un mismo nivel de retorno

De esta manera, dado que el modelo M-V minimiza la desviación estándar, expone al inversionista a peores escenarios cuando los mercados se contraen frente al modelo M-CVaR. El modelo M-CVaR serviría como una herramienta adicional a la toma de decisiones de los inversionistas. Finalmente, aunque en este ejercicio los resultados de la optimización de los portafolio a través de los modelos M-V y M-CVaR no difieren marcadamente, sí indican que si un inversionista no considera el hecho que los retornos no se distribuyen como una normal, entre otros hechos estilizados, podría estar expuesto a pérdidas ante caída en los precios de los activos que componen su cartera.

Capítulo 5

Conclusiones

La presente tesis tuvo como objetivos presentar las demostraciones de las principales propiedades del CVaR y realizar la optimización de un portafolio de inversión utilizando al CVaR como medida de riesgo considerando los hechos estilizados de los retornos de los activos.

Con respecto al primer objetivo, Rockafellar, R.T., Uryasev, S. (2000, 2002) presentaron una manera de optimizar al CVaR y a la vez calcular al VaR en el proceso de optimización. Para ello, propusieron el uso de una función auxiliar y establecieron los teoremas que muestran las características de esta función. En la primera parte de la tesis se demostró de manera exhaustiva los principales teoremas establecidos por estos autores. Luego de los aportes de estos autores se ha desarrollado teoría acerca de las medidas de riesgo, en particular, se pueden encontrar referencias a medidas de riesgo dinámicas. Éstos podrían ser futuros campos de investigación.

Con respecto al segundo objetivo, se realizó la optimización de un portafolio de inversión a través del modelo clásico M-V, del modelo M-CVaR y se compararon los resultados. La optimización del modelo M-CVaR se realizó con datos simulados. Para obtener los retornos simulados se procedió en varios pasos. En primer lugar, dada la presencia de autocorrelación en los retornos y en las volatilidades, se limpiaron las series utilizando modelos ARMA-GARCH y ARMA-GJR. Una vez limpio de la dependencia temporal, se procedió a modelar la dependencia contemporánea a través de cópulas en pares. A pesar que las diferencias en las fronteras eficientes no son muy grandes, los resultados de las optimizaciones confirman que utilizar la varianza como medida de riesgo subestimaría los riesgo en el portafolio. Además, los resultados obtenidos utilizando al CVaR generan portafolios menos concentrados y, por ende, menos expuesto a pérdidas ante eventos de estrés en los mercados. Como futuras mejoras se podría explorar los resultados de las optimizaciones durante diferentes ventanas de tiempo para tener una noción temporal y observar si, por ejemplo, los resultados cambian marcadamente durante períodos de estrés en los mercados. Dado que las correlaciones de los diferentes activos tienden a aumentar fuertemente durante esos períodos, realizar un ejercicio de esa naturaleza sería parte de un nuevo trabajo.

Por otro lado, durante el proceso de simulación de escenarios, se pudo identificar varias oportunidades de mejoras para trabajos futuros. En primer lugar, la modelación de las marginales a través de los modelos GARCH y GJR supusieron que los errores se distribuyen como una normal o t-student. Como futura mejora, se podría asumir que los errores se distribuyen como una t-student sesgada en lugar de solamente t-student, debido a la presencia de asimetría negativa en los retornos. Aunque el objetivo de la modelación de las marginales fue solamente la de remover la autocorrelación, se podría haber considerado modelarlas considerando la posibilidad de cambio de régimen. Una posible mejora en el proceso de simulación de escenarios y previo a la modelación de los datos con cópulas, podría ser utilizar Teoría del Valor Extremo para poder modelar las colas de las distribuciones de los activos.

Pasando a la teoría de cópulas, éste es un campo de gran y rápido avance. Algunas mejoras en este campo podrían ser modelar los datos a través de cópulas dinámicas o considerar cambio de régimen para modelar la dependencia contemporánea en un contexto dinámico.

Finalmente, para la realización de los diferentes cálculos a lo largo de la tesis se han utilizado principalmente EViews 5.0, Matlab 7.10 y R 2.15.1.

Bibliografía

- [1] Aas, Czado, Frigessi, y Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 2, 1, 1-25
- [2] Abberger, K. (2004) A simple graphical method to explore tail-dependence in stock-return pairs. Discussion Paper, University of Konstanz, Alemania.
- [3] Akaike, H. (1974). A New Look at Statistical Model Identification. *IEEE Trans. Automatic Control*, 19: 716-723.
- [4] Anderson, T. W.; Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics* 23: 193–212.
- [5] Artzner, P. Delbaen, F. Eber, J. Heath, D. (1999), Coherent Measures of Risk, *Mathematical Finance*, 9(1999), 203-228.
- [6] Artzner, P. Delbaen, F. Eber, J. Heath, D. (1997). Thinking coherently, *Risk*, 10(1997), 68-71.
- [7] Benati, S. (2003). The Optimal Portfolio Problem with Coherent Risk Measure Constrains. *European Journal of Operational Research* 150 (2003) 572-584.
- [8] Bedford, T.J. y Cooke, R. (2001) Probability density decomposition for conditionally dependent random variables modeled by Vines. *Annals of Mathematics and Artificial Intelligence*, 32 (1). pp. 245-268. ISSN 1012-2443
- [9] Bedford, T.J. y Cooke, R. (2002) Vines: A New Graphical Model for Dependent Random Variables. *The Annals of Statistics*, Vol. 30, No. 4 (Aug., 2002), pp. 1031-1068.
- [10] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1994
- [11] Bollerslev, T (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 31:307-327.
- [12] Bollerslev, T. Choau, R. Kroner K. (1992). ARCH modelling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* 52:5-59.
- [13] Campbell J.Y., Lo A.W., MacKinlay A.C. (1997). *The econometrics of financial markets*. Princeton University Press, Princeton, New Jersey.
- [14] Chen, S., Liu, Y. y Wang, Sh. (2004). Progress in risk measurement. *Advanced modelling and optimization*. Vol 6, N° 1.
- [15] Clarke, K. (2007). A Simple Distribution-Free Test for Nonnested Model Selection. *Political Analysis* (2007) 15 (3): 347-363.
- [16] Cont R. (2001). Empirical properties of asset return: stylized facts and statistical issues. *Quantitative Finance* 1:223-236.

- [17] Enders, W. (2010) Applied econometric time series. Hoboken, NJ : John Wiley, c2010.
- [18] Engle R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50 (4):987-1007.
- [19] Fisher, N. I., and Switzer, P. (1985). Chi-plots for assessing dependence. *Biometrika*, 72 2 , 253–265.
- [20] Fisher, N. I., and Switzer, P. (2001). Graphical assessment of dependence: Is a picture worth 100 tests? *Am. Stat.*, 55 3 , 233–239.
- [21] Genest, C. Favre, A. C. (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.* 12(4) 349-368.
- [22] Genest Ghoudi y Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* (1995) 82 (3): 543-552.
- [23] Genest, Ch. y Rivest, L. (1993). Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association* Vol. 88, No. 423 (Sep., 1993), pp. 1034-1043.
- [24] G. Guastaroba, R. Mansini, G. Speranza (2009). On the effectiveness of scenario generation techniques in single-period portfolio optimization. *European Journal of Operational Research* 192 (2009) 500–511.
- [25] Glosten, L.R. Jagannathan, R. y Runkle, D. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, American Finance Association, vol. 48(5), pages 1779-1801, December.
- [26] Gouriéroux C., Jasiak J. (2001). *Financial econometrics: Problems, models, and methods*. Princeton University Press, Princeton.
- [27] Hamilton, J. (1994) *Time series analysis*. Princeton, N.J. : Princeton University Press, 1994
- [28] Jarque C.M.; Bera, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6 (3): 255–259.
- [29] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.
- [30] Joe (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, Vol. 94, No. 2. (June 2005), pp. 401-419.
- [31] Joe, H (1996) Families of m-Variate Distributions with Given Margins and $m(m-1)/2$ Bivariate Dependence Parameters. *Lecture Notes-Monograph Series* Vol. 28, *Distributions with Fixed Marginals and Related Topics* (1996), pp. 120-141.
- [32] Jondeau, E.; Poon, S. y Rockinger, M. (2006). *Financial Modeling Under Non-Gaussian Distributions*. Springer Finance, 2006.
- [33] Kim, Silvapulle y Silvapulle (2007). Semiparametric estimation of the dependence parameter of the error terms in multivariate regression. Monash University, Department of Econometrics and Business Statistics.
- [34] Kolmogorov A. (1933) Sulla determinazione empirica di una legge di distribuzione, *Giornale dell'Istituto Italiano degli Attuari*, 4, 83-91.
- [35] Kurowicka y Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley Series in Probability and Statistics.
- [36] Kullback y Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, Vol. 22, N° 1, 79-86.
- [37] Lleo, S. (2009). Risk management: a review. *Research Foundation - CFA Institute*. Feb 2009. Vol 4, N° 1.

- [38] Ljung, G. M.; Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, Vol. 65, No. 2 (Aug., 1978), pp. 297-303.
- [39] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance* 7 (1), 47-62. 1952
- [40] Markowitz, H. (1956). The Optimización of a Quadratic Function Subject to Linear Constraints. *Naval Research Logistics Quarterly* 3(1/2), 111-133. 1956.
- [41] Martín, M. (2005). El Conditional Value at Risk en la gestión de carteras latinoamericanas. *Revista de Temas Financieros*. SBS, Volumen II, número 2.
- [42] Michal Kaut, Stein W. Wallace (mayo 2003). Evaluation of scenario-generation methods for stochastic programming. <http://edoc.hu-berlin.de/series/speps/2003-14/PDF/14.pdf>.
- [43] Munenzon, M. (2010). Risk Measurement from Theory to Practice: Is Your Risk Metric Coherent and Empirically Justified? http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1605315. 12 Mayo 2010.
- [44] Nelsen, R. (2006). *An Introduction to Copulas*. Springer Series in Statistics, New York, 2006.
- [45] Ortega, R. Pullirsch, J. Teichmann, J. Wergieluk (2009). A Dynamic Approach for Scenario Generation in Risk Management. http://arxiv.org/PS_cache/arxiv/pdf/0904/0904.0624v2.pdf.
- [46] Pagan A.R. (1996). The econometrics of financial markets. *Journal of Empirical Finance* 3:15-102.
- [47] Rockafellar, R.T., Uryasev, S. (2002). Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance* 26/7, 1443-1447.
- [48] Rockafellar, R.T., Uryasev, S. (2000). Optimization of Conditional Value-at-Risk *The Journal of Risk* 2(3), 21-41.
- [49] Shephard N. (1996). Statistical aspects of ARCH and stochastic volatility. En: Cox D.R., Hinkley D.V., Barndoff-Nielsen O.E. (editores) *Times series models in econometrics, finance and other fields*. Chapman & Hall, London, 1-17.
- [50] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, Vol. 6, No. 2 (Mar., 1978), pp. 461-464.
- [51] Siwei Cheng, Yanhui Liu y Shouyang Wang (2004). *Progress in Risk Measurement*. Advanced Modelling and Optimization, Volume 6, Number 1, 2004.
- [52] Sklar A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* 8:229-231.
- [53] Sovan Mitra (2006). Scenario Generation for Stochastic Programming. <http://www.optirisk-systems.com/papers/opt004.pdf>.
- [54] Szegö, G. (2005). Measures of Risk. *European Journal of Operational Research* 163 (2005) 5-19.
- [55] Vuong, Q. (1989). Likelihood ratio test for model selection and non-nested hypotheses. *Econometrica* Vol. 57, No. 2 (Mar., 1989), pp. 307-333
- [56] Werner Römisch. Scenario generation in stochastic programming. http://www.matheon.de/preprints/6733_wileyRoem.pdf.
- [57] Wozabal, D. A (2008). *New Methods for Value-at-Risk Constrained Optimization using the Difference of convex Algorithm (DCA)*. Department of Statistics and Decision Support Systems, University of Vienna. April 21, 2008.
- [58] Wilk, M. B. y Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*. Vol. 55, No. 1 (Mar., 1968), pp. 1-17.