



UNIVERSIDAD CENTRAL DE VENEZUELA  
FACULTAD DE CIENCIAS  
ESCUELA DE MATEMÁTICA

# Valor en Riesgo y Pérdida Esperada para Algunos Modelos con Volatilidad Constante

Trabajo Especial de Grado presentado ante la ilustre Universidad Central de Venezuela por el **Br. Yefirson Arteaga** para optar al título de Licenciado en Matemática.

**Tutor: Dra. Carenne Ludeña.**

Caracas, Venezuela

Octubre de 2013

Nosotros, los abajo firmantes, designados por la Universidad Central de Venezuela como integrantes del Jurado Examinador del Trabajo Especial de Grado titulado “**Valor en Riesgo y Pérdida Esperada para Algunos Modelos con Volatilidad Constante**”, presentado por el **Br. Yefirson Arteaga**, titular de la Cédula de Identidad **18739437**, certificamos que este trabajo cumple con los requisitos exigidos por nuestra Magna Casa de Estudios para optar al título de **Licenciado en Matemática**.

---

**Dra. Carenne Ludeña**  
**Tutor**

---

**Dr. José Benito Hernández**  
**Jurado**

---

**Dra. Mairene Colina**  
**Jurado**

## Agradecimiento

Ante todo a Dios por darme salud y vida, a mi Madre Idabelis Flores, mi Tia Maglenis Flores, Mi Padre Carlos Nadales por todo el apoyo incondicional y la paciencia que me han tenido a lo largo de toda mi vida y por quienes hoy puedo estar escribiendo estas palabras, a mis primos Genesis y Leandro Nadales y a mi novia Yaneht Marquez quien ha sido gran parte de mi inspiración y fortaleza, con quien espero poder formar una familia en un futuro reciente.

Al licenciado Maicol Ochoa quien planteo el tema a trabajar e hizo llegar a mí la publicación en la cual se baso el presente trabajo especial de grado y fue mi tutor hasta que sus otras ocupaciones fueron más importantes y decidió abandonar su responsabilidad tanto conmigo como con la Universidad Central de Venezuela por medio de un correo electrónico luego de que estuve más de dos meses seguidos intentando comunicarme con él por todos los medios posibles, sin su desplante el presente trabajo no hubiera llegado a ser ni remotamente el que se presenta a continuación.

A todas esas personas, profesores y amigos que por una u otra razón forman parte de esa gran familia que es la Facultad de Ciencias y con quienes he compartido tan buenos momentos, en particular los aspirantes a Físicos y Matemáticos con quienes me siento más identificado. A la Doctora y Profesora Carenne Ludeña quien fue mi salvación en un momento que pensaba todo estaba perdido y mi guía a lo largo del refinamiento del presente documento.

## Índice general

Introducción	1
Capítulo 1. Valor en Riesgo	2
1. Soporte teórico	2
Capítulo 2. Cálculo del Valor en Riesgo y Pérdida Esperada	22
1. Cálculo del VaR no paramétrico en $R$ .	25
2. Cálculo del VaR en $R$ con el paquete “ $VaR$ ”.	26
3. Backtesting de las estimaciones	28
4. Pérdida Esperada	30
5. Comentarios finales	31
Bibliografía	32

## Introducción

En un mundo donde la realidad está caracterizada por la incertidumbre sobre el comportamiento futuro, las instituciones financieras y en general todos los agentes de los mercados financieros se encuentran expuestos a desviarse del resultado deseado, lo cual es conocido como riesgo. La literatura distingue tres tipos de riesgo, el de crédito, el de liquidez y el de mercado. Este trabajo se enfocará en la medida más empleada por reguladores y agentes del mercado para cuantificar el riesgo del mercado: el Valor en Riesgo, VaR por sus siglas en inglés (Value at Risk).

Contar con estimaciones del VaR precisas es de crucial importancia para mantener la estabilidad del sistema financiero sin perjudicar la rentabilidad de estas instituciones. Si el VaR no es medido adecuadamente se puede presentar deterioro en la rentabilidad de las instituciones o inestabilidad financiera en las mismas.

No obstante, el VaR no es una medida coherente de riesgo en general, por lo que podría conllevar a resultados contradictorios y en algunos casos adversos para el manejo del riesgo, por lo que también se propondrá otra alternativa de medición de riesgo conocida como Pérdida Esperada o ES por sus siglas en inglés (Expected Shortfall).

En este trabajo se describirá paso a paso la estimación del VaR empleando la simulación histórica, un método paramétrico que supone una distribución normal con varianza constante y un método semi-paramétrico que modela la cola de la distribución de los retornos usando la distribución generalizada de Pareto. Así mismo, se detalla como calcular el ES y realizar el backtesting al interior de la muestra (“in sample”). Para la ilustración práctica de los conceptos, se empleará el programa R y el paquete VaR.

## Valor en Riesgo

### 1. Soporte teórico

El *Valor en Riesgo* o *VaR* por sus siglas en inglés se define como la máxima pérdida posible de un portafolio de inversión durante determinado horizonte de tiempo y un nivel de significancia  $\alpha$ , bajo circunstancias normales de mercado, no obstante, aún no se ha dicho que es un portafolio de inversión o a que se refiere el término mercado, cuyas definiciones son las siguientes.

**Definición 1.1.** Un *mercado* es un mecanismo por medio del cual los consumidores y productores interactúan para intercambiar bienes y servicios.

Existen diferentes tipos de mercado, pero en el que se enmarca el *VaR*, el *mercado financiero* es donde se comercian instrumentos financieros entre el inversionista y la institución emisora o un intermediario, nuevamente surge un término no definido, el de instrumento financiero, el cual se define:

**Definición 1.2.** Un *instrumento financiero* es un activo comerciable de cualquier tipo, ya sea dinero en efectivo, evidencia de propiedad en alguna entidad o cualquier otro documento financiero con fuerza legal.

**Definición 1.3.** Un *activo* es el conjunto de todos los bienes y derechos con valor monetario que son propiedad de una empresa.

Al momento de invertir las instituciones seleccionan según su criterio el grupo de instrumentos financieros en los cuales destinará el capital excedente de su patrimonio, esperando que el valor de estos aumente para así obtener beneficios. Formando así un *portafolio de inversión*.

**Definición 1.4.** Un *portafolio de inversión* es un conjunto de instrumentos financieros en los cuales se invierte.

Dicho conjunto será el objeto de estudio del presente trabajo, se le analizarán sus beneficios y posibles pérdidas, en lugar de considerar cada instrumento por separado.

En lo que sigue, se trabajará sobre un espacio de probabilidades  $(\Omega, F, \mathbf{P})$ .

Una vez dadas las definiciones básicas se puede proceder a definir formalmente el *Valor en Riesgo*.

**Definición 1.5.** Dado un nivel de confianza  $q \in (0, 1)$ , y una variable aleatoria  $Y$  sobre  $\Omega$ , el *Valor en Riesgo* de un portafolio de inversión viene dado por el menor valor  $x$  tal que la probabilidad de que  $Y$  exceda a  $x$  no sea mayor que  $(1 - q)$ . Formalmente:

$$(1.1) \quad VaR_q(Y) = \inf\{x \in \mathbb{R} : \mathbb{P}(Y > x) \leq 1 - q\} = \inf\{x \in \mathbb{R} : F_Y(x) \geq q\},$$

donde  $F_Y = \mathbb{P}(Y \leq x)$  es la función de distribución de la variable aleatoria  $Y$ .

El  $VaR_q(Y)$  no es más que el cuantil de orden  $q$  de la distribución de pérdidas, es decir, es el valor de la variable aleatoria  $Y$  que marca un corte de modo que una proporción  $q$  de valores de la población es menor o igual que  $Y$ .

En particular para  $F_Y$  continua y estrictamente creciente se tiene que  $VaR_q(Y) = F^{-1}(q)$ . Por ejemplo, si  $Y \sim N(\mu, \sigma^2)$ , es decir,  $Y$  se distribuye normalmente con media  $\mu$  y varianza  $\sigma^2$  tenemos que:

$$P(Y < x) = P\left(\frac{Y - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right) = P\left(Z < \frac{x - \mu}{\sigma}\right),$$

donde  $Z \sim N(0, 1)$ , luego si  $P(Y < x) = q$ , entonces:

$$P\left(Z < \frac{x - \mu}{\sigma}\right) = q$$

pero por definición,  $P(Z < \Phi^{-1}(q)) = q$ , así:

$$\frac{x - \mu}{\sigma} = \Phi^{-1}(q).$$

Por la definición del  $VaR$  y la continuidad de la distribución Normal se tiene que  $x = -VaR_q$ , y dado que  $\Phi^{-1}(q) = -\Phi^{-1}(1 - q)$  se tiene que:

$$(1.2) \quad VaR_q = \Phi^{-1}(1 - q)\sigma - \mu.$$

Las características generales de  $VaR$  se resumen en el siguiente lema:

**Lema 1.6.** *VaR cumple con las siguientes propiedades:*

- (a) Si  $Y \geq 0$  entonces  $VaR_\alpha(Y) \leq 0$ .
- (b) Si  $X \leq Y$  entonces  $VaR_\alpha(X) \leq VaR_\alpha(Y)$ .
- (c)  $VaR_\alpha(tY) = tVaR_\alpha(Y)$ ,  $\forall t \geq 0$ .
- (d)  $VaR_\alpha(Y + t) = VaR_\alpha(Y) - t$ ,  $\forall t \in \mathbb{R}$ .

DEMOSTRACIÓN.

- (a) Sea  $A = \{y : \mathbf{P}(Y \leq y) > \alpha\}$ . Como  $Y$  es una variable aleatoria no negativa se tiene que  $A \subseteq [0, \infty)$ , y así, necesariamente, se cumple que  $\inf(A) \geq 0$ . Además, de la definición del  $VaR$  se tiene que  $\inf(A) = -VaR_\alpha(Y)$ , de donde se sigue el resultado.
- (b) Por hipótesis se cumple que  $\{X \leq y\} \subseteq \{Y \leq y\}$ . Como  $\mathbf{P}$  es una medida de probabilidad, entonces  $\{y : \mathbf{P}(X \leq y) > \alpha\} \subseteq \{y : \mathbf{P}(Y \leq y) > \alpha\}$ , y por tanto, al tomar ínfimos se concluye que  $-VaR_\alpha(Y) \leq -VaR_\alpha(X)$ , de donde se sigue que  $VaR_\alpha(X) \leq VaR_\alpha(Y)$ .
- (c) Si  $t \geq 0$  entonces  $\{Y \leq y\} = \{tY \leq ty\}$  y a su vez  $\mathbf{P}(Y \leq y) = \mathbf{P}(tY \leq ty)$ .  
Luego

$$\begin{aligned} \{y : \mathbf{P}(Y \leq y) > \alpha\} &= \{t \cdot y : \mathbf{P}(t \cdot Y \leq t \cdot y) > \alpha\} \\ &= t \cdot \{y : \mathbf{P}(Y \leq y) > \alpha\} \end{aligned}$$

de donde  $\inf \{ty : \mathbf{P}(tY \leq ty) > \alpha\} = t \cdot \inf \{y : \mathbf{P}(Y \leq y) > \alpha\}$ . Es decir que  $-VaR_\alpha(tY) = t \cdot (-VaR_\alpha(Y))$  como se quería demostrar.

- (d) Dado que  $\{Y + t \leq y\} = \{Y \leq y - t\}$ ,  $\forall t \in \mathbb{R}$ , se tiene entonces:

$$\begin{aligned} \inf \{y : \mathbf{P}(Y + t \leq y) > \alpha\} &= \inf \{y - t : \mathbf{P}(Y \leq y - t) > \alpha\} \\ &= \inf \{y_0 : \mathbf{P}(Y \leq y_0) > \alpha\} + t, \end{aligned}$$

entonces, al cambiar lo signos se obtiene que

$$VaR_\alpha(Y + t) = VaR_\alpha(Y) - t, \quad \forall t \in \mathbb{R}.$$

A pesar de que  $VaR$  es ampliamente aceptado por parte de los agentes financieros debido a cualidades como su aplicación a cualquier tipo de activo, su simplicidad y su capacidad de simplificar en una sola cifra todas las posibles fuentes de riesgo de mercado existente en



un portafolio, el  $VaR$  tiene una característica que representa serios inconvenientes cuando se trabaja con portafolios muy complejos. Esta característica es la no subaditividad, es decir, dadas dos posiciones  $X, Y$  en un mercado financiero, en general no ocurre que:

$$(1.3) \quad VaR_q(X + Y) \leq VaR_q(X) + VaR_q(Y)$$

La condición anterior origina que el  $VaR$  no sea consistente con el principio de Markowitz, ampliamente aceptado: LA DIVERSIFICACIÓN DE UN PORTAFOLIO REDUCE EL RIESGO, propiedad que tiene gran importancia en la administración del riesgo ya que es útil conocer que para portafolios grandes, en los cuales es difícil conocer el riesgo global, el máximo riesgo que podría tener el portafolio corresponde a la suma de los riesgos individuales.

Para entender la importancia de la subaditividad, es suficiente notar que dicha condición expresa el hecho de que un portafolio formado por varios subportafolios arriesgará a lo sumo la suma de los riesgos individuales de cada uno de los subportafolios.

**Ejemplo 1.7.** Considérese un espacio de probabilidad  $(\Omega, \mathbf{F}, \mathbf{P})$  con  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ ,  $\mathbf{F}$  la colección de todos los subconjuntos de  $\Omega$  y la medida  $\mathbf{P}$  dada por:

$$\mathbf{P}[\{\omega_1\}] = \mathbf{P}[\omega_2] = p, \quad \mathbf{P}[\{\omega_3\}] = 1 - 2p,$$

con  $0 < p < \frac{1}{3}$ . Fijado  $N > 0$ , sean las posiciones  $X_1, X_2$  como sigue

$$X_i(\omega_j) = \begin{cases} -N & \text{si } i = j \\ 0 & \text{en caso contrario} \end{cases}$$

La función de distribución de las  $X_i$  está dada por

$$F_{X_i}(x) = \begin{cases} 0 & \text{si } x < -N \\ p & \text{si } -N \leq x < 0 \\ 1 & \text{para } x \geq 0 \end{cases}$$

Considerando  $p < q < 2p$  entonces es claro que  $VaR_q(X_i) = 0$  para  $i = 1, 2$ .

Ahora bien, suponiendo además que  $X_1$  y  $X_2$  son independientes, se puede calcular explícitamente estas mismas medidas para la posición diversificada  $X = X_1 + X_2$ , esto es:

$$X = \begin{cases} 0 & \text{con probabilidad } (1-p)^2 \\ -N & \text{con probabilidad } 2p(1-p) \\ -2N & \text{con probabilidad } p^2 \end{cases}$$

En este caso, y para el mismo  $q$ , se tiene que:  $VaR_q(X) = N$ , lo cual indica que la posición  $X$  es "más riesgosa" que las posiciones individuales. Es decir, se ve claramente que:

$$VaR_q(X_1) + VaR_q(X_2) < VaR_q(X_1 + X_2)$$

Nótese además que el  $VaR$  no da información sobre la severidad de las pérdidas que ocurren con una probabilidad menor a  $1 - q$ . Si la distribución de pérdidas tiene una cola pesada esto resultaría muy problemático.

En vista del inconveniente antes expuesto se tiene otra alternativa a la medición de riesgo conocida como *Pérdida Esperada* o *ES* (*Expected Shortfall*) por sus siglas en inglés.

**Definición 1.8.** Dado un nivel de confianza  $q \in (0, 1)$ , una variable aleatoria  $X$  con  $\mathbb{E}[|X|] < \infty$  y una función de distribución  $F_X$ , se define como *Pérdida Esperada*:

$$(1.4) \quad ES_q(X) = \frac{1}{1-q} \int_q^1 Var_u(X) du.$$

Si  $X \sim N(\mu, \sigma^2)$ , para un  $q$  dado mediante un simple cálculo se puede ver que:

$$(1.5) \quad ES_q = \sigma \frac{\phi(\Phi^{-1}(1-q))}{1-q} - \mu.$$

Donde  $\phi$  es la función de densidad de la distribución normal estándar.

La propiedad más importante del *ES* es la de ser una medida coherente de riesgo, lo cual se demostrará con el siguiente lema:

**Lema 1.9.** Sea  $q \in (0, 1)$  fijo. Considérese un conjunto de variables aleatorias a valor real  $V$  en un espacio de probabilidad  $(\Omega, A, \mathbf{P})$  tal que  $E[X] < \infty$  para todo  $X \in V$ . Entonces  $\rho : V \rightarrow \mathbb{R}$  con  $\rho(X) = ES_q(X)$  para todo  $X \in V$  es una medida de riesgo coherente, es decir:

- (a) *Monotonía:* Para todo  $X, Y \in V$  con  $X \leq Y$ , entonces  $\rho(X) \leq \rho(Y)$ .
- (b) *Homogeneidad positiva:*  $\rho(\lambda X) = \lambda \rho(X)$ ,  $\forall \lambda \geq 0$ .
- (c) *Invariante ante traslaciones:*  $\rho(X + t) = \rho(X) - t$ ,  $\forall X \in V, y t \in \mathbb{R}$ .
- (d) *Subaditividad:*  $\rho(X + Y) \leq \rho(X) + \rho(Y)$ ,  $\forall X, Y \in V$ .

DEMOSTRACIÓN.

(a) Por el item b del lema 1.6 se tiene que para  $X \leq Y$   $VaR_\alpha(X) \leq VaR_\alpha(Y)$ .

Por definición:

$$\rho(X) = ES_q(X) = \frac{1}{1-q} \int_q^1 Var_u(X) du \leq \frac{1}{1-q} \int_q^1 Var_u(Y) du = \rho(Y),$$

se tiene el resultado.

(b) Por el item c del lema 1.6 se tiene que  $VaR_\alpha(tY) = tVaR_\alpha(Y)$ ,  $\forall t \geq 0$ .

Por definición:

$$\rho(\lambda X) = ES_q(\lambda X) = \frac{1}{1-q} \int_q^1 Var_u(\lambda X) du = \frac{1}{1-q} \int_q^1 \lambda Var_u(X) du,$$

y se tiene lo buscado.

(c) Por el item d del lema 1.6 se tiene que  $VaR_\alpha(Y + t) = VaR_\alpha(Y) - t$ ,  $\forall t \in \mathbb{R}$ .

Por definición:

$$\rho(X + t) = \frac{1}{1-q} \int_q^1 Var_u(X + t) du = \frac{1}{1-q} \int_q^1 Var_u(X) - t du = \rho(X) - t.$$

(d) Para realizar esta prueba se tomara una representación alternativa de la Pérdida Esperada dada por:

$$ES_q(X) = -\frac{1}{1-q} E[X 1_{X \leq X_q}]$$

Donde  $1_{X \leq X_q}$  es la función indicatriz. Sea  $Z=X+Y$  entonces:

$$\begin{aligned} (1-q)(ES_q(X) + ES_q(Y) - ES_q(Z)) &= (E[Z 1_{Z \leq Z_q}] - E[X 1_{X \leq X_q}] - E[Y 1_{Y \leq Y_q}]) \\ &= E[Z 1_{Z \leq Z_q} - X 1_{X \leq X_q} - Y 1_{Y \leq Y_q}] \\ &= E[(X + Y) 1_{Z \leq Z_q} - X 1_{X \leq X_q} - Y 1_{Y \leq Y_q}] \\ &= E[X(1_{Z \leq Z_q} - 1_{X \leq X_q}) + Y(1_{Z \leq Z_q} - 1_{Y \leq Y_q})] \\ &\geq X_\alpha E[1_{Z \leq Z_q} - 1_{X \leq X_q}] + Y_\alpha E[1_{Z \leq Z_q} - 1_{Y \leq Y_q}] \\ &= X_\alpha(\alpha - \alpha) + Y_\alpha(\alpha - \alpha) \\ &= 0. \end{aligned}$$

En la desigualdad se uso que:  $E[(X - X_\alpha)(1_{Z \leq Z_\alpha} - 1_{X \leq X_\alpha})] \geq 0$

y  $E[(Y - Y_\alpha)(1_{Z \leq Z_\alpha} - 1_{Y \leq Y_\alpha})] \geq 0$ .

□

Existen diferentes metodologías para el cálculo del  $VaR$ , las más básicas vienen dadas por:

- Modelo Varianza-Covarianza.
- Modelo Simulación Histórica.
- Modelo Monte Carlo.

El modelo Varianza-Covarianza solo es aplicable a una cartera cuyos rendimientos se distribuyan con una curva de densidad normal. La suposición más básica del modelo es que los rendimientos de los factores de riesgo se distribuyen normalmente, y que su distribución conjunta es normal multivariante, por lo que la matriz de covarianza de los rendimientos de los factores de riesgo es todo lo que se requiere para observar la dependencia entre los rendimientos de los factores de riesgo.

El modelo Simulación Histórica asume que todas las variaciones posibles en el futuro se han experimentado en el pasado, y que la distribución histórica simulada es idéntica a la distribución de los rendimientos en el horizonte de riesgo con visión de futuro. Este modelo se puede usar para carteras con rendimientos lineales y no lineales. La aplicación de este enfoque parte de la identificación de los diferentes instrumentos que forman la cartera y de la elección de una muestra de los rendimientos históricos de los activos de la cartera. A partir de los cambios observados en los rendimientos históricos de los activos, se obtendrán valores alternativos que permitan estimar el percentil adecuado para obtener el VaR de la cartera.

El modelo Monte Carlo estima el VaR simulando todos los escenarios posibles y reevalúa sus posiciones en el portafolio. Al igual que la simulación histórica se usa para carteras con rendimientos lineales y no lineales, la diferencia es que cambia la metodología de como se generan los escenarios. Un importante punto para tener presente es que si se tiene una significativa cantidad de exposiciones no lineales en el portafolio, una simulación que reevalúa posiciones generalmente será más exacta que una aproximación paramétrica para estimar el VaR, sin embargo, esto obviamente tiene un costo por mayor complejidad.

Es decir, los tres métodos buscan estimar un valor crítico para las pérdidas potenciales, aunque cada método realiza distintos supuestos acerca de que valores son representativos sobre las futuras pérdidas potenciales y como estas se distribuyen estadísticamente, sin embargo, todos estos métodos presentan deficiencias; en el caso del método de Simulación

Históricamente el problema más conocido a este enfoque es la determinación de un modelo adecuado para la dependencia de los diferentes activos que componen el portafolio, además, existe una dependencia extrema de la información histórica utilizada, no se pueden extrapolar resultados más allá de la longitud de la muestra y se pueden presentar errores en el cálculo de los cuantiles pequeños. En el método de Varianza-Covarianza, la suposición de la normalidad de los retornos no refleja su característica de colas pesadas y no se puede aplicar este método en productos no lineales. Por último se tiene que el método de Montecarlo es intenso en términos computacionales y requiere hipótesis subjetivas de cálculo.

Es por este motivo que se plantea un método alternativo para el cálculo del VaR, denominado Método de Valores Extremos.

El objetivo de la teoría de valores extremos es ajustar distribuciones límites a las fluctuaciones de los máximos normalizados  $M_1 = X_1$ ,  $M_n = \max(X_1, \dots, X_n)$ ,  $n \geq 2$ , donde las  $(X_i)_{i \in \mathbb{N}}$ , es una sucesión de variables aleatorias independientes e idénticamente distribuidas con función de distribución común  $F$ .

**Definición 1.10.** (Distribución de Valores Extremos Generalizados (GEV))

La función de distribución GEV está dada por:

$$H_\xi(x) = \begin{cases} \exp(-(1 + \xi x)^{-\frac{1}{\xi}}), & \xi \neq 0 \\ \exp(-e^{-x}), & \xi = 0 \end{cases}$$

Donde  $1 + \xi x > 0$ . Al definir  $H_{\xi, \mu, \sigma} := H_\xi((x - \mu)/\sigma)$  se obtiene una familia de tres parámetros, un parámetro de localización  $\mu \in \mathbb{R}$  y un parámetro de escala  $\sigma > 0$ .

El parámetro  $\xi$  es conocido como parámetro de forma de la distribución GEV.

Esta definición contiene tres tipos de distribución, las cuales son conocidas por otros nombres de acuerdo al valor de  $\xi$ .

Cuando  $\xi > 0$  corresponde a una distribución Fréchet, dada por:

$$(1.6) \quad \Phi_{\frac{1}{\xi}}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\frac{1}{\xi}}) & \text{si } x > 0 \text{ y } \xi > 0. \end{cases}$$

Cuando  $\xi = 0$  es una distribución Gumbel definida por:

$$(1.7) \quad \Lambda(x) = \exp(-e^{-x}), x \in \mathbb{R}.$$

Por último, cuando  $\xi < 0$  se tiene una distribución Weibull dada por:

$$(1.8) \quad \Psi_{-\frac{1}{\xi}}(x) = \begin{cases} \exp(-(-x^{-\frac{1}{\xi}})) & \text{si } x \leq 0 \text{ y } \xi > 0 \\ 1 & \text{si } x > 0. \end{cases}$$

Supóngase que los máximos  $M_n$  de una sucesión de variables aleatorias independientes e idénticamente distribuidas (iid) convergen en distribución bajo una apropiada normalización. Recordando que  $P(M_n \leq x) = F^n(x)$ , se puede observar que esta convergencia significa que existe una sucesión de constantes reales  $(b_n)$  y  $(a_n)$  donde  $a_n > 0$  para todo  $n$  tal que:

$$(1.9) \quad \lim_{n \rightarrow \infty} P((M_n - b_n)/a_n \leq x) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x).$$

Para alguna función de distribución  $H(x)$  no degenerada.

**Definición 1.11.** Se dice que una función de distribución es degenerada si

$$\lim_{n \rightarrow \infty} P((M_n - b_n)/a_n \leq x) \in \{0, 1\}.$$

**Definición 1.12.** Dos distribuciones  $F$  y  $G$  son del mismo tipo o pertenecen a la misma familia si para algunas constantes  $A > 0, B \in \mathbf{R}$ ,

$$G(x) = F(Ax + B), x \in \mathbf{R}$$

En términos de variables aleatorias, si  $X \sim F$  y  $Y \sim G$  entonces

$$Y(x) \stackrel{d}{=} \frac{X - B}{A}$$

Los siguientes teoremas serán usados en la demostración del teorema de Fisher-Tippett, Gnedenko los cuales no serán demostrados en este documento. La demostración del Teorema de Convergencia a Familias puede ser consultada en [7] y la del Teorema de la Ecuación Funcional de Hamel en [9].

**Teorema 1.13.** *(Teorema de convergencia a Familias)*

Sean  $U(x)$  y  $V(x)$  dos funciones de distribución, ninguna de las cuales está concentrada en un punto.

- (a) Supóngase que para  $n \geq 1$ ,  $X_n$  son v.a. con funciones de distribución  $F_n$  y sean, además, constantes  $a_n > 0$ ,  $b_n \in \mathbf{R}$ ,  $\alpha_n > 0$ ,  $\beta_n \in \mathbf{R}$  y

$$(1.10) \quad F_n(a_n x + b) \rightarrow U(x), F_n(\alpha_n x + \beta) \rightarrow V(x)$$

entonces existen constantes  $A > 0$  y  $B \in \mathbf{R}$  tales que, cuando  $n \rightarrow \infty$ ,

$$(1.11) \quad \frac{\alpha_n}{a_n} \rightarrow A, \quad \frac{(\beta_n - b_n)}{a_n} \rightarrow B$$

y

$$(1.12) \quad V(x) = U(Ax + B).$$

(b) Recíprocamente, si (1.11) vale, entonces cualquiera de las relaciones en (1.10) implica la otra y (1.12) vale.

**Teorema 1.14.** (*Ecuación funcional de Hamel*)

Una función  $\phi(\lambda)$  finita, medible, positiva, y que satisface  $\phi(\lambda)\phi(\gamma) = \phi(\lambda\gamma)$  para  $\lambda > 0$  es necesariamente de la forma  $\lambda^\rho$ ,  $-\infty < \rho < \infty$ .

Mediante la siguiente definición y el siguiente teorema se englobará el papel de la distribución GEV en el estudio de los máximos.

**Definición 1.15.** (Dominio Máximo de Atracción)

Si (1.9) se cumple para alguna función de distribución no degenerada  $H$ , entonces se dice que  $F$  pertenece al dominio máximo de atracción de  $H$ , Notación  $F \in MDA(H)$ .

**Teorema 1.16.** (*Fisher-Tippett, Gnedenko*)

Si  $F \in MDA(H)$  para alguna función de distribución  $H$  no degenerada, entonces  $H$  debe ser una distribución de tipo  $H_\xi$ ; es decir, una distribución GEV.

DEMOSTRACIÓN.

Sea  $t \in \mathbf{R}$  y  $[t]$  la parte entera de  $t$ .

Como  $F \in MDA(H)$ , se tiene que para todo  $t > 0$

$$F^{[nt]}(a_{[nt]}x + b_{[nt]}) \rightarrow H(x)$$

de donde

$$F^{[nt]}(a_n x + b_n) = (F^n(a_n x + b_n))^{\frac{[nt]}{n}} \rightarrow H^t(x)$$

Luego por el teorema de convergencia a familias existe  $\alpha(t) > 0$ ,  $\beta(t) \in \mathbf{R}$ ,  $t > 0$  tales que:

$$\lim_{n \rightarrow \infty} \frac{a_n}{a_{[nt]}} = \alpha(t), \quad \lim_{n \rightarrow \infty} \frac{b_n - b_{[nt]}}{a_{[nt]}} = \beta(t)$$

y

$$(1.13) \quad H^t(x) = H(\alpha(t)x + \beta(t))$$

Para  $t > 0$  y  $s > 0$  por un lado se tiene:

$$H^{ts}(x) = H(\alpha(ts)x + \beta(ts))$$

y por el otro:

$$\begin{aligned} H^{ts}(x) &= (H^s(x))^t = (H^t(\alpha(s)x + \beta(s))) \\ &= H(\alpha(t)\{\alpha(s)x + \beta(s)\} + \beta(t)) \\ &= H(\alpha(t)\alpha(s)x + \alpha(t)\beta(s) + \beta(t)). \end{aligned}$$

Como  $H$  es no degenerada se concluye que para  $t > 0$ ,  $s > 0$ :

$$(1.14) \quad \alpha(ts) = \alpha(t)\alpha(s)$$

$$(1.15) \quad \beta(ts) = \alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s).$$

La ecuación (1.14) es la ecuación funcional de Hamel, cuya única solución no negativa es de la forma:

$$\alpha(t) = t^{-\theta}, \quad \theta \in \mathbf{R}$$

Considérense los casos: (a)  $\theta = 0$ , (b)  $\theta > 0$ , (c)  $\theta < 0$ .

**Caso (a):**  $\theta = 0$ . En este caso  $\alpha(t) = 1$  y (1.15) queda:

$$\beta(st) = \beta(t) + \beta(s),$$

La cual es variante de la ecuación de Hamel, cuya solución es de la forma:

$$\beta(t) = -c \log(t), \quad t > 0, \quad c \in \mathbf{R}$$

Con esta expresión en (1.13) queda:

$$(1.16) \quad H(t) = H(x - c \log(t))$$

Si  $c = 0$   $H$  sería degenerada. Para  $x$  fijo,  $H^t(x)$  es decreciente en  $t$  por lo tanto se concluye que  $c > 0$ .



Si para algún  $x_0$ ,  $H(x_0) = 1$  entonces de (1.16) se obtiene:

$$1 = H(x_0 - c \log(t)) \forall t$$

Cambiando de variables, haciendo  $H(u) = 1 \forall u$  se llega a una contradicción.

Luego  $H(x) < 1, \forall x$ . De igual forma se observa que no puede pasar que  $H(y) = 0$  para cualquier  $y$ .

Haciendo  $x = 0$  en (1.16) se obtiene para  $t > 0$ :

$$(1.17) \quad H^t(0) = H(-c \log(t))$$

Dado que  $H(0) = \exp(-\exp(-x)) \in (0, 1)$ , haciendo cambio de variables  $u = -c \log(t)$  con  $t \in (0, \infty)$ ,  $u \in (-\infty, \infty)$  en (1.17) queda:

$$\begin{aligned} H(u) &= \exp\{-\exp(-p)t\} = \exp\{-\exp(-(c^{-1}u + p))\} \\ &= \Lambda(c^{-1}u + p). \end{aligned}$$

Caso (b) :  $\theta > 0$ . De (1.15)

$$\alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s)$$

Para  $t \neq 1, s \neq 1$

$$\frac{\beta(s)}{1 - \alpha(s)} = \frac{\beta(t)}{1 - \alpha(t)},$$

es decir, la función  $\beta(\cdot)(1 - \alpha(\cdot))^{-1}$  es constantemente igual a  $c \in \mathbf{R}$ . Luego para  $t \neq 1$

$$\begin{aligned} \beta(t) &= \beta(s)(1 - \alpha(s))^{-1}(1 - \alpha(t)) \\ &= c(1 - t^{-\theta}) \end{aligned}$$

con esta expresión en (1.13) queda:

$$\begin{aligned} H^t(x) &= H(t^{-\theta}x + c(1 - t^{-\theta})) \\ &= H(t^{-\theta}(x - c) + c) \end{aligned}$$

De donde:

$$H^t(x + c) = H(t^{-\theta}x + c).$$

Sea  $G(x) = H(x + c)$ . Entonces  $H$  y  $G$  son del mismo tipo, por lo que resulta suficiente resolver  $G$ . La función  $G$  satisface

$$(1.18) \quad G^t(x) = G(t^{-\theta}x)$$

y  $G$  es no degenerada. Supóngase que  $x = 0$  en (1.18),  $t \log(G(0)) = \log(G(0))$  para  $t > 0$ , de donde o  $\log(G(0)) = 0$  o  $\log(G(0)) = -\infty$ , es decir, o  $G(0) = 0$  o  $1$ . Sin embargo,  $G(0) = 1$  es imposible puesto que implicaría la existencia de  $x < 0$  tal que el lado izquierdo de (1.18) es decreciente respecto a  $t$  mientras que el lado derecho es creciente respecto de  $t$ . Por tanto  $G(0) = 0$ .

De nuevo de (1.18) se obtiene que  $G^t(1) = G(t^{-\theta})$ . Si  $G(1) = 0$  entonces  $G \equiv 0$  y si  $G(1) = 1$  entonces  $G \equiv 1$ , pero ambos aspectos están en contradicción con el hecho de que  $G$  es no degenerada. Luego,  $G(1) \in (0, 1)$ . Sea  $\theta^{-1} = \alpha$ ,  $G(1) = \exp\{-p^{-\alpha}\}$ ,  $u = t^{-\theta}$ , tal que  $u^{-\alpha} = t$ .

De (1.18) con  $x = 1$  se tiene para  $u > 0$

$$\begin{aligned} G(u) &= \exp\{-p^{-\alpha}t\} = \exp\{-(pu)^{-\alpha}\} \\ &= \Phi_{\alpha}(pu). \end{aligned}$$

**Caso (c):**  $\theta < 0$ . Partiendo de la ecuación (1.18) y usando el caso (b) se tiene:

$$\begin{aligned} G(u) &= \exp\{-p^{-\alpha}(-t)\} = \exp\{-(-pu)^{-\alpha}\} \\ &= \Psi_{\alpha}(pu). \end{aligned}$$

□

Las distribuciones a las cuales hace referencia el teorema de Fisher-Tippett, Gnedenko, se denominan distribuciones generalizadas de valores extremos, el principal aporte de dicho teorema es que facilita una especificación funcional aproximada con la que estudiar la distribución del máximo de un número lo suficientemente alto de rendimientos haciendo innecesario conocer, al menos en principio, la distribución completa de la variable aleatoria original. Ahora bien, este resultado sobre la distribución asintótica del máximo parece que tiene poco que ver con el original: estimar el VaR, un cuantil extremo de una distribución de rendimientos, no de su máximo. Sin embargo existe una estrecha relación entre la distribución asintótica del máximo y el comportamiento en la cola de la distribución original.

Intuitivamente parece claro, el máximo es un suceso que, por definición, pertenece a la cola derecha de la distribución y, por tanto, la distribución asintótica del máximo debería estar estrechamente relacionada con la distribución de la variable aleatoria original cuando se está en la cola. El siguiente Teorema y su Corolario permiten centrar definitivamente esta cuestión.

**Teorema 1.17.** (*Aproximación de Poisson*)

Dado  $\tau \in [0, \infty]$ , una sucesión  $(u_n)$  de números reales y  $\bar{F}(u_n) = (1 - F(u_n))$ , es decir la cola de la función de distribución  $F$ , las siguientes dos relaciones son equivalentes

$$(1.19) \quad \lim_{n \rightarrow \infty} n \bar{F}(u_n) = \tau,$$

$$(1.20) \quad \lim_{n \rightarrow \infty} P(M_n \leq u_n) = \exp(-\tau).$$

DEMOSTRACIÓN.

Considérese  $0 \leq \tau < \infty$ . Si (1.19) se cumple, entonces:

$$P(M_n \leq u_n) = F^n(u_n) = (1 - \bar{F}(u_n))^n = \left(1 - \frac{\tau}{n}\right)^n$$

de donde se obtiene (1.20). Recíprocamente, si (1.20) se cumple, entonces  $\bar{F}(u_n) \rightarrow 0$ , porque de suponer que esto no se cumple, para alguna subsucesión  $(u_{n_k})$  y algún  $\epsilon > 0$ ,  $\bar{F}(u_{n_k}) > 0$  para todo  $k$  y  $P(M_{n_k} \leq u_{n_k}) = (1 - \bar{F}(u_{n_k}))^{n_k} \rightarrow 0$  cuando  $n_k \rightarrow \infty$ , lo que contradice (1.20). Tomando logaritmo en (1.20) se tiene:

$$-n \ln(1 - \bar{F}(u_n)) \rightarrow \tau$$

Como  $\ln(1 - x) \sim -x$  para  $x \rightarrow 0$  y como  $\bar{F}(u_n) \rightarrow 0$ , esto implica que  $n \bar{F}(u_n) = \tau$ , y por tanto se tiene (1.19).

□

Nótese que, a partir de la aproximación de Poisson, particularizando adecuadamente el valor de  $\tau$  para cada  $x$ , se puede llegar a la demostración del Teorema Fisher-Tippett, Gnedenko.

La aproximación de Poisson justifica la intuición anterior sobre la relación entre la cola y el máximo, ya que si se cumple la ecuación (1.19),  $1 - \bar{F}(u_n)$  ha de ser extremadamente pequeño, lo cual implica que se está en la cola de la distribución original y, por tanto, en ese caso y solo en ese caso, se puede pasar a (1.20). Precisamente esta relación entre la cola y el

máximo, permite caracterizar los dominios de atracción de los tres tipos de valores extremos a partir del comportamiento en la cola de la distribución original.

**Corolario 1.18.** *La función de distribución  $F$  pertenece al dominio máximo de atracción de la distribución de Valores Extremos  $H$  con constantes de normalización  $c_n > 0$ ,  $d_n \in \mathbb{R}$  si y sólo si*

$$\lim_{n \rightarrow \infty} n\bar{F}(c_n x + d_n) = -\ln H(x)$$

donde  $x \in \mathbb{R}$  y  $\bar{F}_X(u) = 1 - F_X(u)$ .

Cuando  $H(x) = 0$ , el límite es interpretado como  $\infty$ .

DEMOSTRACIÓN.

Como  $F \in MDA(H)$ ,  $F^n(c_n x + d_n) \rightarrow H(x)$  o lo que es lo mismo,  $(1 - \bar{F}(c_n x + d_n))^n \rightarrow H(x)$

Tomando logaritmo neperiano y usando el teorema de la aproximación de Poisson se tiene lo pedido, es decir:

$$n\bar{F}(c_n x + d_n) \rightarrow -\ln H(x).$$

□

Así, dicho al contrario, si se estima el índice de cola de la distribución asintótica del máximo se tendrá absolutamente determinada la parte derecha de la expresión anterior y, por tanto, se podrá utilizar, bajo ciertas condiciones, para realizar una aproximación a la cola de la distribución de la variable aleatoria original, que era el objetivo buscado.

Una vez obtenida una especificación funcional aproximada con la que estudiar la distribución del máximo de un número lo suficientemente alto de rendimientos y ante una muestra de datos financieros concreta, asumiendo que las premisas del Teorema de Fisher-Tippett, Gnedenko se cumplen, surge ahora una nueva pregunta ¿Qué distribución de valores extremos generalizados es la relevante para dicha muestra? Esta pregunta se responderá con la aplicación de los siguientes teoremas, pero antes se requieren las siguientes definiciones:

**Definición 1.19.** Se dice que la cola  $\bar{F}$  de una distribución varía regularmente con índice  $-\alpha$ ,  $\alpha \geq 0$ , si

$$\lim_{n \rightarrow \infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = t^{-\alpha}, \quad t > 0$$

NOTACIÓN:  $\bar{F} \in \mathcal{R}_{-\alpha}$ .

**Definición 1.20.** La función cuantila (o inversa generalizada) de una distribución  $F$  se define, para  $0 < p < 1$ , como:

$$Q_F(p) = \inf\{x \in \mathbb{R} \mid F(x) \geq p\} = x_p.$$

Si  $F$  es continua y estrictamente creciente,  $x_p$  es la solución de  $F(x) = p$  (función inversa de  $F$ ).

**Teorema 1.21.** (*Dominio Máximo de Atracción de la Distribución Frechet*)

La función de distribución  $F$  pertenece al dominio máximo de atracción de  $\Phi_\alpha$ ,  $\alpha > 0$ , si y sólo si  $\bar{F}(x) = x^{-\alpha}L(x)$ , para alguna función  $L$  que satisface  $\lim_{y \rightarrow \infty} \frac{L(yx)}{L(y)} = 1$ . Si  $F \in MDA(\Phi_\alpha)$ , entonces:

$$\frac{1}{c_n} M_n \xrightarrow{d} \Phi_\alpha,$$

donde  $c_n = Q_F(1 - \frac{1}{n})$  y  $(\xrightarrow{d})$  significa convergencia en distribución.

**Teorema 1.22.** (*Dominio Máximo de Atracción de la Distribución Weibull*)

La función de distribución  $F$  pertenece al dominio máximo de atracción de  $\Psi_\alpha$ ,  $\alpha > 0$ , si y sólo si  $x_f < \infty$  y  $\bar{F}(x_f - x^{-1}) = x^{-\alpha}L(x)$ , para alguna función  $L$  que satisface  $\lim_{y \rightarrow \infty} \frac{L(yx)}{L(y)} = 1$ . Si  $F \in DMA(\Psi_\alpha)$ , entonces

$$\frac{M_n - x_f}{c_n} \xrightarrow{d} \Psi_\alpha$$

donde  $c_n = x_f - Q_F(1 - \frac{1}{n})$  y  $d_n = x_f = \sup\{x \in \mathbf{R} \mid F(x) < 1\}$ .

**Teorema 1.23.** (*Dominio Máximo de Atracción de la Distribución Gumbel*)

La función de distribución  $F$  pertenece al dominio máximo de atracción de  $\Lambda_\alpha$ ,  $\alpha > 0$ , si y sólo si existe una función positiva  $\tilde{a}$  tal que se satisface

$$\lim_{x \uparrow x_f} \frac{\bar{F}(x + t\tilde{a}(x))}{\bar{F}(x)} = \exp(-t), \quad t \in \mathbb{R}$$

Se puede elegir  $\tilde{a} = a(x) = \int_x^{x_f} \frac{\bar{F}(t)}{\bar{F}(x)} dt$ ,  $x < x_f$ .

En la Teoría de Valores Extremos hay dos tipos de enfoques que generalmente se aplican, el más tradicional es el método Block-Máxima, el cual se aplica para grandes observaciones recolectadas a partir de grandes muestras de observaciones idénticamente distribuidas y consiste fundamentalmente en partir las observaciones por bloques y en estos encontrar el máximo, el principal problema que se enfrenta al utilizar este método es que se pierden

muchos datos que pueden ser de gran utilidad al momento de ajustar una distribución a la data. Una alternativa para solucionar dicho problema es considerar un método basado en las excedencias sobre un umbral preestablecido.

La distribución utilizada para modelar las excedencias sobre un umbral es la distribución de Pareto Generalizada (GPD).

**Definición 1.24.** La función de distribución de Pareto Generalizada (GPD) ésta dada por:

$$(1.21) \quad G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp(-x/\beta), & \xi = 0 \end{cases}$$

Donde  $\beta > 0$  y  $x \geq 0$  cuando  $\xi \geq 0$  y  $0 \leq x \leq -\beta/\xi$  cuando  $\xi < 0$ . Los parámetros  $\xi$  y  $\beta$  se denominan parámetros de forma y escala respectivamente.

Al igual que la distribución Generalizada de Valores Extremos (GEV) la distribución de Pareto Generalizada contiene otras distribuciones. Cuando  $\xi > 0$  la distribución corresponde a una distribución de Pareto Ordinaria con  $\alpha = 1/\xi$  y  $\kappa = \beta/\xi$ ; cuando  $\xi = 0$  se tiene la distribución exponencial y cuando  $\xi < 0$  corresponde a una distribución Pareto tipo II o de cola corta.

Entre las distribuciones que juegan un papel fundamental en el modelo para ajustar las excedencias a una distribución de Pareto Generalizada se tienen la función de distribución de excesos y la función de excedencia media.

**Definición 1.25.** (Función de distribución de los excesos)

Sea  $X$  una variable aleatoria con función de distribución  $F$ , la distribución de los excesos sobre un umbral  $u$  tiene función de distribución

$$(1.22) \quad F_u(x) = P(X - u \leq x \mid X > u) = \frac{F(x + u) - F(u)}{1 - F(u)},$$

para  $0 \leq x < x_F - u$  donde  $x_F \leq \infty$  es el punto final de la cola derecha de  $F$ .

**Definición 1.26.** (Función de excedencia media)

La función de Excedencia Media de una variable aleatoria  $X$  con media finita esta dada por:

$$(1.23) \quad e(u) = E(X - u \mid X > u).$$

A continuación se presentará un teorema que proporciona una caracterización para el dominio de atracción de una distribución de valores extremos generalizados.

**Teorema 1.27.** (*Balkema-de Hann*)

Para cada  $\xi \in \mathbb{R}$ ,  $F \in MDA(H_\xi)$  si y sólo si

$$(1.24) \quad \lim_{u \rightarrow x_f} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0,$$

para alguna función positiva  $\beta(u)$ , donde  $x_f = \sup\{x \in \mathbb{R} : F(x) < 1\}$ .

Dado que se tiene la siguiente cadena de desigualdades:

$$F_u(x) = F(X \leq u + x | X > u) = \frac{P(u \leq X \leq x + u)}{1 - P(X \leq u)} = \frac{\bar{F}(x + u)}{\bar{F}(u)}.$$

Demostrar el teorema de Balkema-de Hann es equivalente a demostrar que:

$$\lim_{u \rightarrow x_f} \frac{\bar{F}(u + x\beta(u))}{\bar{F}(u)} = \begin{cases} (1 + \xi x)^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ \exp(-x) & \text{si } \xi = 0 \end{cases}$$

para alguna función  $\beta(\cdot)$  positiva y medible.

Dicha demostración se basa en el teorema de representación de funciones de variación regular y no se incluye en el documento.

Este resultado dice que siempre tiene sentido, para  $u$  suficientemente grande ajustar una distribución de Pareto Generalizado a las excedencias.

Luego de presentar las definiciones y teoremas que proporcionan las caracterizaciones de las distribuciones Generalizadas de Valores Extremos y la distribución de Pareto Generalizada, se estudiarán algunos resultados teóricos para el cálculo de los estimadores de los parámetros.

Como bien se recordará, de la distribución de Valores Extremos Generalizados se obtiene una familia de tres parámetros:

$\theta = (\xi, \mu, \psi) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$ , la cual consiste en un parámetro de forma  $\xi$ , un parámetro de localización  $\mu$  y un parámetro de escala  $\psi$ .

Para la estimación de los parámetros existen diferentes métodos. Uno de los más usados es la estimación por Máxima Verosimilitud, en el cual se considera que los datos disponibles consisten en una muestra IID,  $X_1, \dots, X_n$  de la función de distribución  $H_\theta$ . Se supone además que  $H_\theta$  tiene función de densidad  $h_\theta$ ; entonces la función de verosimilitud de los datos  $X_1, \dots, X_n$  está dada por:

$$(1.25) \quad L(\theta, X) = \prod_{I=1}^n h_\theta(X_i) I_{\{1+\xi(X_i-\mu)/\psi > 0\}}.$$

Se denota por  $l(\theta, X) = \ln L(\theta, X)$  al logaritmo de la verosimilitud, luego el estimador de máxima verosimilitud de  $\theta$  se obtiene calculando:

$$(1.26) \quad \hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} l(\theta, X).$$

Es decir, el estimador de máxima verosimilitud se obtiene al calcular los parámetros que maximizan  $l(\theta, X)$ .

Entre las propiedades más importantes de los estimadores de máxima verosimilitud se tienen:

- Son generalmente consistentes, es decir, el valor del estimador tiende a ser el valor del parámetro a medida que el tamaño de la muestra crece.
- Son asintóticamente eficientes, es decir, su varianza tiende al límite dado por la desigualdad de Cramer-Rao al aumentar el tamaño de la muestra.
- Tienden a distribuirse normalmente cuando el tamaño de la muestra tiende a infinito.
- Son insesgados, es decir, su sesgo tiende a cero al aumentar el tamaño muestral.
- Con relación a la suficiencia, si existen estimadores suficientes para los parámetros, los estimadores de máxima verosimilitud son función de estos estimadores.
- Son estimadores invariantes, esto es, si  $\hat{\theta}$  es un estimador máximo verosimil de  $\theta$  entonces  $g(\hat{\theta})$  es un estimador máximo verosimil de  $g(\theta)$ , siendo  $g(\theta)$  una función inyectiva.

En el caso de la función de densidad de la distribución de valores extremos generalizada no existe una solución explícita para (1.26), por lo cual es necesario recurrir a procedimientos numéricos tales como variaciones del algoritmo de Newton-Raphson y el método de Nelder-Mead simplex.

Dichos procedimientos numéricos son métodos de optimización que permiten calcular la mejor estimación posible a los parámetros buscados.

El método de Newton-Raphson es eficiente para encontrar aproximaciones de los ceros o raíces de una función real. También puede ser usado para encontrar el máximo o mínimo de una función, encontrando los ceros de su primera derivada.

El método Nelder-Mead busca de modo aproximado una solución óptima local a un problema con  $N$  variables cuando la función a minimizar varía suavemente, utiliza el concepto de un simplex, que es un politopo de  $N+1$  vértices en  $N$  dimensiones: un segmento de



línea en una línea, un triángulo en un plano, un tetraedro en un espacio tridimensional y así sucesivamente.

De igual manera de la Distribución de Pareto Generalizada, se obtiene una familia de dos parámetros  $\theta = (\xi, \beta) \in \mathbb{R} \times \mathbb{R}_+$ , la cual consiste en un parámetro de forma  $\xi$  y un parámetro de escala  $\beta$ . Para estimación de los parámetros se calcula el logaritmo de la función de verosimilitud dada por

$$(1.27) \quad L(\theta, X) = \prod_{I=1}^n g_{\theta}(X_i) I_{\{1+\xi(X_i-\mu)/\beta > 0\}}.$$

Luego el estimador de máxima verosimilitud de  $\Theta$  se obtiene calculando

$$(1.28) \quad \hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} l(\theta, X).$$

De nuevo se utilizan procedimientos numéricos tales como el algoritmo de Newton-Raphson y el método de Nelder-Mead simplex para estimar los parámetros.

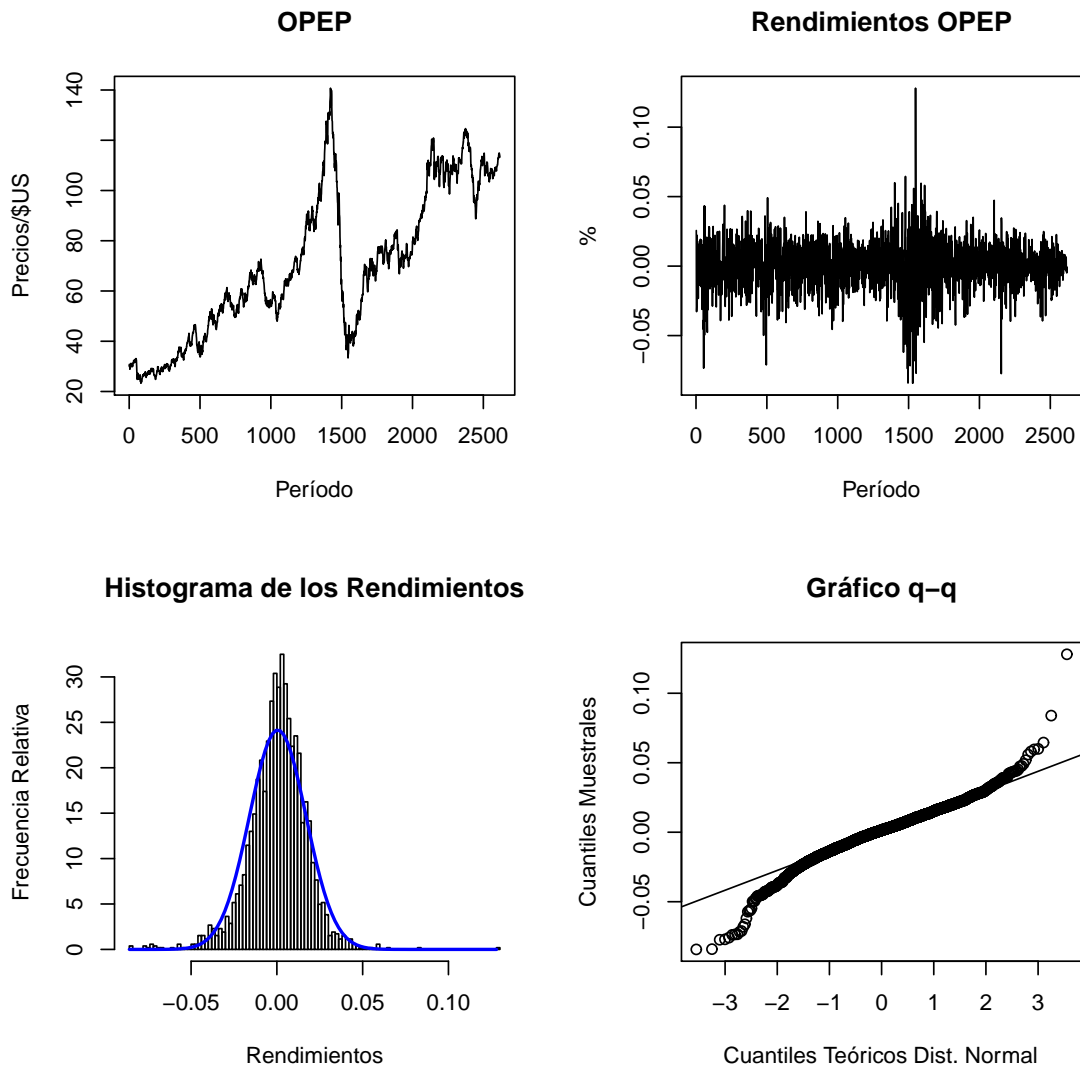
## Cálculo del Valor en Riesgo y Pérdida Esperada

En el presente capítulo se medirá, por medio de modelos *VaR*, el riesgo implícito en la tasa de cambio OPEP/WTI, es decir, como afectan las fluctuaciones de un índice sobre otro. Para esto se adoptarán tres diferentes formas de calcular el VaR y un método para evaluar el comportamiento de las diferentes aproximaciones.

Para la ilustración práctica de los conceptos se empleará el programa *R* y los datos de los precios del petróleo WTI y OPEP desde el 2 de Enero de 2003 hasta 20 de Febrero de 2013, dichos datos pueden ser obtenidos en [10] y [6] respectivamente.

Para entender la muestras se iniciará graficando la serie de los precios OPEP, sus rendimientos, su histograma y el gráfico “q-q” para la distribución normal, lo mismo se realiza para los precios WTI. Los gráficos se obtienen de la siguiente manera:

```
> opep<-read.table(file="Cesta-OPEP1.txt",header=T)#lectura de los datos
> opep<-ts(opep)#convierte los datos en serie de tiempo
> ropep<-diff(log(opep))#Cálculo de los retornos
> layout(matrix(c(1,2,3,4), 2, 2, byrow=TRUE))
> plot(opep, xlab = "Período", ylab = "Precios/$US", main = "OPEP", type = "l")
> plot(ropep, xlab = "Período", ylab = "%", main = "Rendimientos OPEP", type = "l")
> hist(ropep, xlab = "Rendimientos", ylab = "Frecuencia Relativa", nclass = 100 ,
freq = FALSE, main = "Histograma de los Rendimientos")
> curve(dnorm(x, mean=mean(ropep), sd=sqrt(var(ropep))), to=max(ropep), col = "blue",
lwd = 2, add = TRUE)
> qqnorm(ropep, main = "Gráfico q-q", xlab = "Cuantiles Teóricos Dist. Normal",
ylab= "Cuantiles Muestrales")
> qqline(ropep)
```

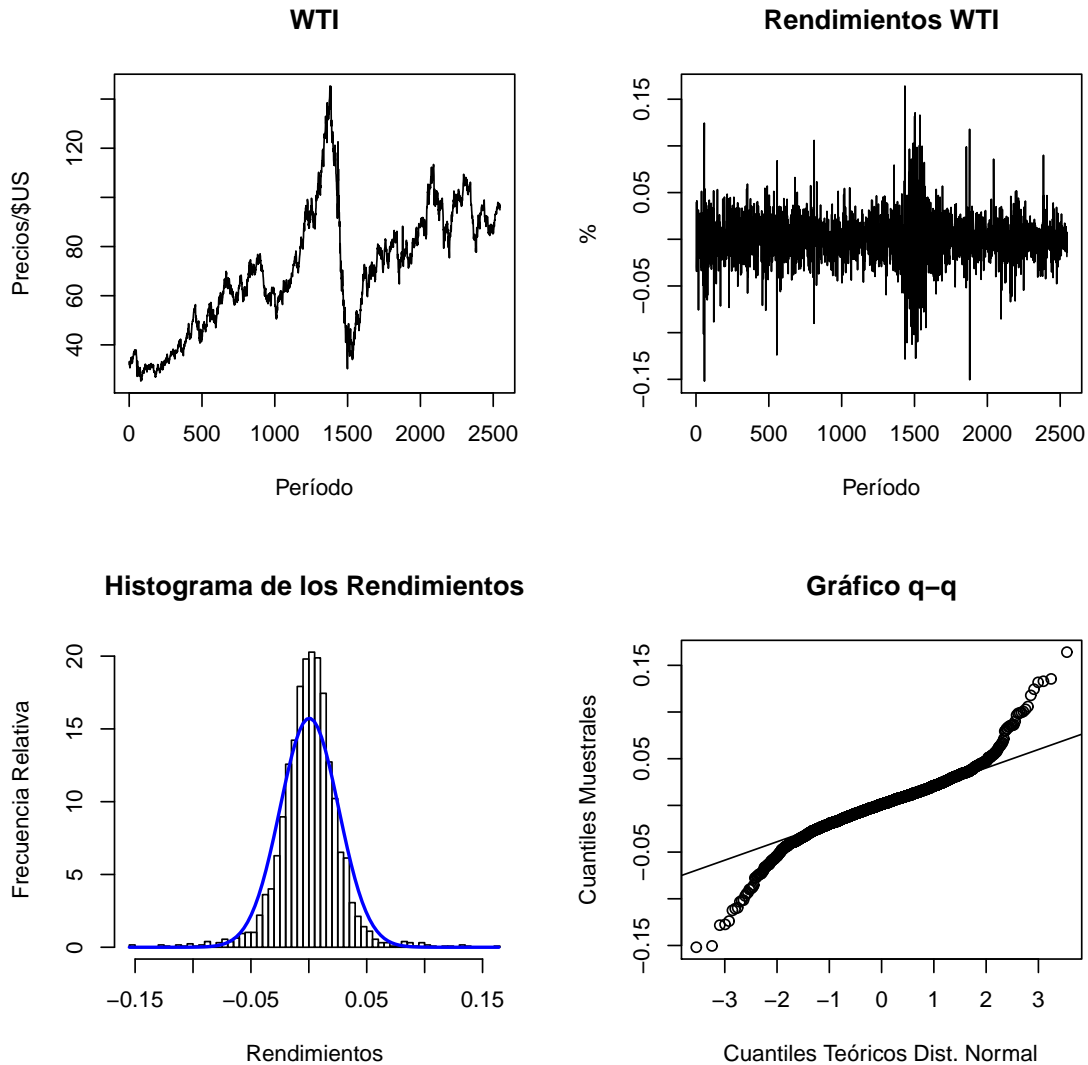


```

> wti<-read.table(file="Cesta-WTI.txt",header=T)#lectura de los datos
> wti<-ts(wti)#convierte los datos en serie de tiempo
> rwti<-diff(log(wti))#Cálculo de los retornos
> layout(matrix(c(1,2,3,4), 2, 2, byrow=TRUE))
> plot(wti, xlab = "Período", ylab = "Precios/$US", main = "WTI", type = "l")
> plot(rwti, xlab = "Período", ylab = "%", main = "Rendimientos WTI", type = "l")
> hist(rwti, xlab = "Rendimientos", ylab = "Frecuencia Relativa", nclass = 100 ,
freq = FALSE, main = "Histograma de los Rendimientos")
> curve(dnorm(x, mean=mean(rwti), sd=sqrt(var(rwti))), to=max(rwti), col = "blue",
lwd = 2, add = TRUE)

```

```
> qqnorm(rwti, main = "Gráfico q-q", xlab = "Cuantiles Teóricos Dist. Normal",
ylab= "Cuantiles Muestrales")
> qqline(rwti)
```



De las figuras anteriores se observa que los rendimientos muestran una gran variabilidad (volatilidad). En otras palabras, la desviación que presentan los rendimientos respecto a su media es muy cambiante. Pero no solo eso, la volatilidad busca “agruparse” o presentar clusters (volatility cluster), pues episodios de gran volatilidad tienden a estar seguidos de períodos de baja volatilidad y viceversa.

Nótese además, que la función de densidad  $f(x) = F'(x)$  de los rendimientos tiene características similares a la normal, como la simetría y la forma “acampanada”, pero existe una

gran diferencia entre las distribuciones empíricas de los rendimientos (representada por los histogramas) y la distribución normal; la distribución de los rendimientos tiende a ser más “picuda” o leptocúrtica que la distribución normal y posee colas más pesadas. En la práctica las colas pesadas (“fat tails”) implican que existe mayor probabilidad de obtener valores extremos (observaciones muy alejadas de la media) que la que existiría en una distribución normal. Este hecho también se observa mediante el diagrama de probabilidad normal (normal probability plot o QQ plot), el cual representa la relación entre los cuantiles muestrales y los teóricos de una distribución normal, si la muestra proviene de una distribución normal entonces, los puntos muestrales deberían estar sobre una línea recta, pero en las figuras se observa como las colas de las distribuciones son más pesadas (“fat tails”) que la distribución normal. En otras palabras, la probabilidad de obtener valores extremos es mucho mayor en la distribución empírica de los rendimientos que lo que predice una distribución normal.

A continuación se ilustrarán tres diferentes maneras de encontrar el VaR empleando la tasa de cambio OPEP/WTI usando un nivel de confianza del 99 %.

### 1. Cálculo del VaR no paramétrico en *R*.

Una de las formas más sencillas de calcular el *VaR* es utilizando el método no paramétrico conocido como la simulación histórica. Este método implica, tomar cada uno de los datos como un posible escenario, de tal forma que el *VaR* corresponderá al percentil  $\alpha$  de la distribución empírica de los datos. Así el cálculo del *VaR* al usar esta aproximación requiere las siguientes líneas de código:

```
> wti.real<-diff(wti)
> VaRw.nopara<- (-1) * quantile(wti.real, prob = 0.01)
> opep.real<-diff(opep)
> VaRo.nopara<- (-1) * quantile(opep.real, prob = 0.01)
```

En este caso, el VaR para los precios del petróleo WTI corresponde a 4,78, para los precios del petróleo OPEP corresponde a 3,22, es decir, la máxima pérdida para el siguiente día de transacción en el precio del barril WTI con un 99 % de confianza será de 4,78 dólares, mientras que para el precio del barril OPEP será de 3,22 dólares.

## 2. Cálculo del VaR en R con el paquete “VaR”.

Los métodos paramétricos implican suponer una distribución o modelo al cual se ajustan los diferentes valores del portafolio. Esta distribución puede ser muy diferente dependiendo de cada caso y aunque la más usada es la normal, tal vez no es la más adecuada para describir el comportamiento de la distribución de los rendimientos de un activo o portafolio. El paquete “VaR” de R permite calcular el VaR bajo dos aproximaciones:

- Aproximación paramétrica bajo el supuesto de que los rendimientos siguen una distribución normal y una varianza constante.
- Aproximación semi-paramétrica en la cual se modela la cola inferior de la distribución de los rendimientos con una distribución generalizada de Pareto.

### 2.1. Aproximación paramétrica: Función VaR.norm.

La función “VaR.norm” del paquete “VaR” estima el VaR para un determinado nivel de confianza, bajo el supuesto de que los rendimientos siguen una distribución normal. Esta función requiere como argumentos más importantes la serie del valor del portafolio o precio del activo, el nivel de confianza ( $p$ ) y determinar si se debe realizar una aproximación lineal de la distribución lognormal o no (`lin.appx`). Para el caso en estudio, se tiene:

```
> library(VaR)# Carga del paquete VaR en R
> VaRw.para<-VaR.norm(wti, p = 0.99, lin.appx = FALSE)
> VaRw.para$VaR
> VaRo.para<-VaR.norm(opep, p = 0.99, lin.appx = FALSE)
> VaRo.para$VaR
```

En este caso se encontró que el VaR para el próximo día de negociación concerniente a los datos WTI es de 5,54, y para los datos OPEP de 4.28, es decir, la máxima pérdida para el siguiente día de transacción en el precio del barril WTI con un 99 % de confianza será de 5,54 dólares, mientras que para el precio del barril OPEP será de 4.28 dólares.

### 2.2. Aproximación semi-paramétrica: Función VaR.gpd.

La función “VaR.gpd” estima el VaR para un determinado nivel de confianza, ajustando la distribución generalizada de Pareto a la parte de los retornos que exceden un determinado umbral. Es decir, solo emplea los datos de la cola inferior de la distribución para modelar

la cola por medio de la distribución generalizada de Pareto. En general, esta función transforma los datos suministrados como argumento a sus correspondientes rendimientos, luego son organizados y solamente conserva aquellos que sobrepasan el umbral pre—establecido. Finalmente, el logaritmo de la función de máxima verosimilitud se emplea para calcular el VaR. Los argumentos de esta función son: el vector numérico de los datos para el cual se calculará el VaR,  $p$  el nivel de confianza,  $p.tr$  el umbral preestablecido, e  $init$  el valor inicial para el cálculo del logaritmo de la función de máxima verosimilitud. En el caso de esta investigación, para obtener el VaR con dicha aproximación y un nivel de significancia del 1 %, se emplea el siguiente código:

```
> VaRw1.gdp<-VaR.gpd(wti, p = 0.01,p.tr=0.97,init=c(1,0.1))
> VaRw1.gdp$VaR
> VaRo1.gdp<-VaR.gpd(opep, p = 0.01,p.tr=0.97,init=c(1,0.1))
> VaRo1.gpd$VaR
```

En este caso se encontró que el VaR para el próximo día de negociación concerniente a los datos WTI es de 7,23, y para los datos OPEP de 4,72, es decir, la máxima pérdida para el siguiente día de transacción en el precio del barril WTI con un 99 % de confianza será de 7,23 dólares, mientras que para el precio del barril OPEP será de 4,72 dólares.

Es importante notar que este resultado es sensible al umbral que se adopte, por ejemplo si en vez de adoptar el umbral de 0.97 se adopta un umbral de 0.95 o de 0.99, los resultados cambiarán. Véase esto empleando el siguiente código:

```
> VaRw2.gdp<-VaR.gpd(wti, p = 0.01,p.tr=0.95,init=c(1,0.1))
> VaRo2.gdp<-VaR.gpd(opep, p = 0.01,p.tr=0.95,init=c(1,0.1))
> VaRo2.gpd$VaR
> VaRw3.gdp<-VaR.gpd(wti, p = 0.01,p.tr=0.99,init=c(1,0.1))
> VaRw3.gdp$VaR
> VaRo3.gdp<-VaR.gpd(opep, p = 0.01,p.tr=0.99,init=c(1,0.1))
> VaRo3.gpd$VaR
```

Nótese que con los umbrales de 0.95 y 0.99 se obtiene un VaR para WTI de 7,11 y 7,26 respectivamente y para OPEP de 4,72 y 4,54. En la práctica es imposible conocer cuál es el umbral adecuado, por lo que es recomendable estimar el VaR con esta aproximación para diferentes umbrales y emplear las medidas de ajuste (o backtesting), que se discutirán seguidamente, para determinar cual umbral es el más adecuado a una determinada muestra.

### 3. Backtesting de las estimaciones

En la práctica es necesario decidir cual es la mejor aproximación para calcular el VaR, pues como se puede constatar, cada aproximación y supuesto que se realice implicará una estimación diferente del VaR. Uno de los modelos más utilizados para verificar el ajuste realizado es el desarrollado por Kupiec en 1995. Consiste en probar cuál es la proporción en que se observa una pérdida superior a la predicción del modelo VaR. Esta proporción debe ser en promedio no mayor al nivel de significancia.

Para este fin se empleará la prueba de Kupiec (1995), quien define la proporción de las excepciones como:

$$(2.1) \quad \hat{p} = \frac{\# \text{ de excepciones}}{H}$$

Donde H es el número total de observaciones empleadas en la evaluación; en otras palabras, el número de observaciones que fueron empleadas para calcular la proporción de excepciones  $\hat{p}$ . Para evaluar la hipótesis nula que la proporción de excepciones es igual a la esperada teóricamente ( $\alpha$ )( $H_0 : \hat{p} = \alpha$ ), se empleará el siguiente estadístico t de Kupiec (1995):

$$(2.2) \quad t_U = \frac{\hat{p} - \alpha}{\sqrt{\hat{p}(1 - \hat{p})/H}}$$

Kupiec (1995) demostró que  $t_U$  sigue una distribución  $t$  con  $H - 1$  grados de libertad. En la práctica se emplean dos aproximaciones para evaluar el estadístico de Kupiec:

- Emplear datos en la muestra (in sample).
- Emplear datos por fuera de la muestra (out of sample).

La aproximación “in sample”, implica calcular el VaR empleando la muestra completa y comparar ese VaR con todas las pérdidas del portafolio o activo de la muestra que excedieron el VaR. Para el caso en estudio, se cuenta con 2546 observaciones del precio diario del barril



WTI y 2617 del precio diario del barril OPEP, esta aproximación implica observar que tantos de esos precios originan una pérdida mayor a la del VaR calculado. En otras palabras, esta aproximación mide que tan bien se ajustó el modelo a la muestra propuesta.

La aproximación “out of sample”, implica considerar una muestra recursiva de un tamaño inicial  $G$ ; para esa muestra se calcula el VaR del siguiente periodo ( $G+1$ ) y se compara con el rendimiento o pérdida observada para el periodo  $G+1$ . Después se adiciona una observación más a la muestra y se calcula el VaR para el periodo  $G+2$  y se compara con lo que realmente ocurrió en el periodo  $G+2$ . Este procedimiento se repite hasta agotar la muestra. Por tanto, esta aproximación mide que tan buena es la predicción del VaR en los diferentes modelos. A continuación se presenta como realizar la aproximación “in sample” dado que la “out of sample” no se encuentra disponible en el paquete *VaR* de *R*.

### 3.1. Evaluación en la muestra “in sample”.

La evaluación en la muestra por medio del paquete “VaR” se realiza a través de la función “VaR.backtest”. Esta función calcula el valor  $p$  de la prueba de Kupiec expresada en (2.2). Los argumentos de la función son tres:

- (i) Los datos de los rendimientos observados,
- (ii) El VaR calculado para el siguiente período,
- (iii) La proporción teórica esperada.

Para el caso del VaR paramétrico, con el supuesto de una distribución normal, se emplea el siguiente código para encontrar el valor  $p$  del estadístico de Kupiec:

```
>VaR.backtest(VaRw.para$cddata, VaRw.para$VaR, p = 0.01)
>VaR.backtest(VaRo.para$cddata, VaRo.para$VaR, p = 0.01)
```

En este caso el valor  $p$  encontrado para WTI es de 6.674506e-07, y para OPEP 4.602066e-07, por tanto, existe suficiente evidencia para rechazar la hipótesis nula de que la proporción de excepciones observada es igual a la teórica para ambas datas.

Para el caso de los cálculos realizados de la aproximación semi-paramétrica con umbrales 0.97, 0.95 y 0.99 se emplea el siguiente código:

```
> VaR.backtest(VaRw1.gpd$cddata, VaRw1.gpd$VaR, p = 0.01)
> VaR.backtest(VaRw2.gpd$cddata, VaRw2.gpd$VaR, p = 0.01)
```

```

> VaR.backtest(VaRw3.gpd$cddata, VaRw3.gpd$VaR, p = 0.01)
> VaR.backtest(VaRo1.gpd$cddata, VaRo1.gpd$VaR, p = 0.01)
> VaR.backtest(VaRo2.gpd$cddata, VaRo2.gpd$VaR, p = 0.01)
> VaR.backtest(VaRo3.gpd$cddata, VaRo3.gpd$VaR, p = 0.01)

```

En este caso el valor  $p$  encontrado para WTI con umbrales 0.97, 0.95 y 0.99 es de  $3.712959e-151$ ,  $5.920956e-121$ ,  $8.184049e-251$  respectivamente y para OPEP  $2.028299e-84$ ,  $1.437272e-61$ ,  $9.944529e-181$  correspondientemente, por tanto, en todos los casos existe suficiente evidencia para rechazar la hipótesis nula de que la proporción de excepciones observada es igual a la teórica para ambos conjuntos de datos y todos los umbrales.

Finalmente, para la simulación histórica (aproximación no paramétrica) se puede encontrar el valor  $p$  de la prueba de kupiec empleando el siguiente código:

```

>VaR.backtest(VaRw3.gpd$cddata, VaRw.nopara, p = 0.01)
>VaR.backtest(VaRo3.gpd$cddata, VaRo.nopara, p = 0.01)

```

Para este caso el valor de  $p$  es 0 en ambos casos por lo que existe suficiente evidencia para rechazar la hipótesis nula de que la proporción de excepciones observada es igual a la teórica en ambos conjuntos de datos.

En conclusión para todos los casos se puede rechazar la hipótesis nula de que el porcentaje de excepciones es igual al teóricamente esperado.

#### 4. Pérdida Esperada

Recuérdese que a pesar de todas las cualidades del  $VaR$  como medida de riesgo, esta presenta el inconveniente de no ser una medida coherente de riesgo, además de que solo tiene en cuenta el percentil de pérdidas que exceden al  $VaR$  y no las magnitudes de estas. Sin embargo, es la magnitud de la caída una de las principales preocupaciones para el correcto manejo del riesgo. Por lo tanto es necesario emplear una medida de riesgo tan simple como el  $VaR$  que permita no solo conocer cuál es la pérdida que se espera para el día de mañana tal que sólo el 1% de las posibles pérdidas sean mayores a esta, sino que también aporte información acerca de la cola de la distribución. De esta forma sería posible conocer el valor esperado de la pérdida dado que el  $VaR$  fue excedido, esto es conocido como Pérdida Esperada o Expected Shortfall (ES), formalmente:

$$ES_{\alpha} = E(r | r > VaR_{\alpha})$$

Donde  $r$  es el rendimiento. El  $ES$  es por lo tanto un promedio de los peores  $100(1 - \alpha)\%$  casos que excedan el umbral con un nivel de confianza de  $\alpha$ . El  $ES$  se calcula fácilmente para el caso de la aproximación semi-paramétrica con el paquete “VaR” de R. De hecho ya fue calculado al momento de emplear la función “VaR.gpd”, dado que esta función calcula tanto el VaR como el  $ES$ , que almacena en las posiciones “\$VaR” y “\$ES”.

Para obtener el  $ES$  solo se necesitan las siguientes líneas de código:

```
> VaRw1.gdp$ES
> VaRw2.gdp$ES
> VaRw3.gdp$ES
> VaRo1.gdp$ES
> VaRo2.gdp$ES
> VaRo3.gdp$ES
```

Así, la pérdida esperada, en caso de que se presente una pérdida mayor al VaR para los datos WTI, es de 9.37 para el VaR semi-paramétrico que emplea la distribución generalizada de Pareto y un umbral de 0.97, para OPEP 5.94. Si se consideran los umbrales de 0.95 y 0.99, los  $ES$  correspondientes son 9.34 y 9.42, para WTI, y OPEP 5.95, 6.03 respectivamente.

## 5. Comentarios finales

Al observar los resultados en el cálculo del  $VaR$  a través de las diferentes metodologías, obtenidos por medio del software estadístico  $R$  y el paquete del mismo denominado VaR se concluye que bajo circunstancias normales de mercado la máxima pérdida posible para el siguiente día de transacción siempre será superior para el activo WTI que para OPEP con un 99% de confianza, sin importar que metodología sea empleada en el cálculo, además se puede ver la robustez del modelo dada la cercanía de los resultados obtenidos por cada metodología, aunque al realizar la prueba de Kupiet se encontró que al evaluar el comportamiento dentro de la muestra, ninguna de las aproximaciones brinda la cobertura deseada. Es importante tener en cuenta que estos resultados solo se aplican a los datos trabajados en el presente documento y el resultado puede cambiar para otras muestras de activos.

## Bibliografía

- [1] ALEXANDER, CAROL.(2008) Market Risk Analysis, Value-at-Risk Models. John Wiley and Sons, Ltd.
- [2] ALONSO, J. C. SEMAÁN, P.(2009) Cálculo del Valor en Riesgo y Pérdida Esperada mediante R empleando modelos de volatilidad constante. Apuntes de Economía. ISSN 1794-029X. N° 21.
- [3] EMBRECHTS, P. KLÜPPELBERG, C. MIKOSCH, T.(1997) Modelling Extremal Events for Insurance and Finance. Springer-Verlag.
- [4] ESTEVES, M.(2010) Análisis de Riesgo de Precios de Títulos Públicos del Mercado Venezolano. Trabajo de maestría no publicado, Instituto Venezolano de Investigaciones Científicas.
- [5] MCNEIL, A. FREY, R. EMBRECHTS, P(2005) Quantitative Risk Management. Princeton University Press.
- [6] ORGANIZATION OF THE PETROLEUM EXPORTING COUNTRIES.(2013, Febrero 21). OPEC Basket Price. [Datos en Línea]. Disponible: <http://www.quandl.com/DOE-US-Department-of-Energy/RWTC-WTI-Crude-Oil-Spot-Price-Cushing-OK-FOB> 22
- [7] RESNICK, SIDNEY(1987) Extreme Values, Regular Variation, and Point Processes. Springer, New York. 10
- [8] ROSS, SHELDON, A first course in probability. Prentice Hall.
- [9] SENETA, E(1976) Regularly Varying Function. Lecture Notes in Mathematic, 508, Springer,Berlin. 10
- [10] US DEPARTMENT OF ENERGY.(2013, Febrero 21). WTI Crude Oil Spot Price Cushing, OK FOB. [Datos en Línea]. Disponible: [http://www.opec.org/opec\\_web/en/data\\_graphs/40.htm](http://www.opec.org/opec_web/en/data_graphs/40.htm) 22