



# Propuesta de desarrollo de algoritmos clasificadores clínicos para pacientes con dengue usando métodos estadísticos y de minería de datos

Andrés Felipe Palomino Montezuma

Universidad del Valle  
Facultad de Ingeniería, Escuela de Estadística  
Santiago de Cali, Colombia  
2022

# Propuesta de desarrollo de algoritmos clasificadores clínicos para pacientes con dengue usando métodos estadísticos y de minería de datos

**Andrés Felipe Palomino Montezuma**

Trabajo de grado presentado como requisito parcial para optar al título de:  
**Estadístico**

Director:

Ph.D. Jose Rafael Tovar Cuevas

Codirector:

Ph.D. Jaime Mosquera Restrepo

Universidad del Valle

Facultad de Ingeniería, Escuela de Estadística

Santiago de Cali, Colombia

2022

# Agradecimientos

*Andrés Felipe Palomino Montezuma*

# Agradecimientos

# Resumen

**Palabras clave:** *FALTA*

# Abstract

**Keywords:** *housing supply, external covariates, web scraping, regression model, generalized linear model.*



# Contenido

<b>1. Introducción</b>	<b>2</b>
1.1. Planteamiento del Problema . . . . .	4
1.2. Justificación . . . . .	4
1.3. Objetivos . . . . .	5
1.3.1. Objetivo General . . . . .	5
1.3.2. Objetivos Específicos . . . . .	5
<b>2. Antecedentes</b>	<b>6</b>
<b>3. Marco Teórico</b>	<b>7</b>
3.1. Marco Conceptual . . . . .	7
3.2. Marco Estadístico . . . . .	8
3.2.1. Distribución Normal . . . . .	8
3.2.2. Distribución de Error Genralizado: GED . . . . .	9
3.2.3. Inferencia Clásica . . . . .	10
3.2.4. Inferencia Bayesiana . . . . .	11
<b>4. Metodología</b>	<b>17</b>
4.0.1. Archivo de datos . . . . .	17
4.0.2. Análisis descriptivo . . . . .	17
4.0.3. Estimación por Maxima Verosimilitud . . . . .	17
4.0.4. Enfoque Bayesiano . . . . .	19
<b>5. Resultados</b>	<b>22</b>
5.1. Análisis Exploratorio de Datos . . . . .	22
5.2. Análisis Bayesiano . . . . .	33
5.2.1. Índice S&P500 . . . . .	33
5.2.2. Índice FSTOK10 . . . . .	34
5.2.3. Índice CRBQX . . . . .	35
<b>6. Conclusiones y recomendaciones</b>	<b>37</b>
6.1. Conclusiones . . . . .	37
6.2. Recomendaciones . . . . .	37
<b>Bibliografía</b>	<b>38</b>

<b>A. Anexo:</b>
------------------

<b>39</b>
-----------



# Lista de Figuras

1-1. Precios de cierres de Principales criptomonedas durante el período 2018 -2023	3
5-1. Serie de los precios diarios SP&500	22
5-2. Serie de los precios diarios FSTOK	23
5-3. Serie de los precios diarios CRBQX	23
5-4. Boxplot precios de cierre de los índices	24
5-5. Histograma precios de cierre de los índices	25
5-6. QQ-plot precios de cierre de los índices	25
5-7. s	26
5-8. s	26
5-9. s	27
5-10.s	27
5-11.Tabla resumen atípicos	27
5-12.s	28
5-13.s	28
5-14.Tabla resumen atípicos	28
5-15.Serie de los precios diarios normalizados SP&500	29
5-16.Serie de los precios diarios normalizados FSTOK	29
5-17.Serie de los precios diarios normalizados CRBQX	30
5-18.Log retorno S&P500	31
5-19.Log retorno FSTOK10	31
5-20.Log retorno CRBQX	32
5-21.QQ-Plot de los Log retornos	32
5-22.Grafico posterior	33
5-23.Gráfico posterior FSTOK10	34
5-24.Gráfico posterior FSTOK10	35

# Lista de Tablas

5-1. Estadísticas descriptivas del precio de cierre de cada índice . . . . .	24
5-2. Estadísticas descriptivas de los log retornos diarios del precio de cierre de los índices . . . . .	30
5-3. Prueba Shapiro-Wilk . . . . .	32
5-4. Estadísticas Descriptivas . . . . .	34
5-5. Estadísticas Descriptivas . . . . .	35
5-6. Estadísticas Descriptivas . . . . .	36

# 1. Introducción

El desarrollo y el uso extendido de la tecnología abierto una nueva dimensión a la era del mundo digital, aprovechando los diferentes beneficios de conectividad e información que nos puede brindar. Debido a esta revolución, las monedas virtuales han obtenido un gran posicionamiento en el interés de los inversionistas en el mercado. Una criptomoneda es una divisa virtual, que está protegida por criptografía que es una técnica de seguridad lo que la hace casi imposible de falsificar. El control de cada una de estas monedas funciona a través de redes descentralizadas, usualmente no son reguladas por ninguna autoridad, lo que la vuelve inmune a intermediaciones o manipulación de los gobiernos.

La descentralización es uno de los principios fundamentales de las criptomonedas. La mayoría de las criptomonedas se basan en una tecnología llamada blockchain, que es una base de datos distribuida que permite a todos los participantes de la red mantener una copia del registro de transacciones. Esto significa que no hay una autoridad central que controle las criptomonedas o la red.

La descentralización tiene varias ventajas. En primer lugar, elimina la necesidad de confiar en un ente central para realizar transacciones o almacenar fondos. En segundo lugar, la descentralización hace que sea más difícil ataques cibernéticos que puedan comprometer la seguridad de la red. Por último, la descentralización permite que las criptomonedas sean utilizadas por cualquier persona en cualquier parte del mundo, sin restricciones geográficas o gubernamentales. Esto significa que las personas que no tienen acceso a los servicios financieros tradicionales pueden utilizar criptomonedas para enviar y recibir pagos, ahorrar dinero e invertir en proyectos que les interesen.

(Nakamoto, 2008), publicó el primer artículo que describe la creación del Bitcoin la primera y más reconocida moneda digital del mercado, lo que dio paso a crear diversos activos digitales todos con características diferentes, entre lo más reconocidos están Litecoin (2011), Ethereum(2015) y Ripple(2012). La industria de las criptomonedas ha presentado el interés de muchos inversionistas lo que ha hecho que su crecimiento sea imparable, prueba de ello es que en la actualidad existen miles de criptomonedas en todo el mundo, aunque la gran mayoría tienen una capitalización de mercado muy baja y un volumen de operaciones muy limitado. Según el sitio web CoinMarketCap, que rastrea el precio, la capitalización de mercado y otros datos de las criptomonedas, hay actualmente más de 11,000 criptomonedas listadas en su plataforma a partir de marzo de 2023. Sin embargo, es importante tener

en cuenta que no todas las criptomonedas son activas o tienen un uso práctico. Algunas criptomonedas pueden ser consideradas como "muertas" porque han sido abandonadas por sus desarrolladores o porque no han tenido éxito en el mercado. Por lo tanto, el número real de criptomonedas útiles y en uso activo puede ser menor que el número total de criptomonedas existentes.



**Figura 1-1.:** Precios de cierres de Principales criptomonedas durante el período 2018 -2023

*Fuente de información:* Elaboración propia a partir de los datos de Yahoo Finance

Las criptomonedas presentan patrones de crecimiento bursátil debido a su aprovechamiento en temas de portabilidad y operaciones internacionales; estas actúan como una buena cobertura contra la inflación durante periodos de crecimiento estable. Sin embargo, su resistencia puede verse afectada cuando hay mucha incertidumbre en los mercados. La particularidad de las monedas digitales es la variabilidad que presentan en pequeños periodos de tiempo y comportamiento atípico respecto a los mercados tradicionales.

Teniendo en cuenta lo anterior, el objetivo de este trabajo es proponer una metodología estadística que permita analizar el comportamiento y distribución de este tipo activos dentro de los portafolios que contengan activos tradicionales; estimando el parámetro de volatilidad del rendimiento de las acciones desde el enfoque bayesiano, donde se considera la distribución GED o Distribución Generalizada de Errores, la cual no es considerada según la respectiva revisión de la literatura. Esta distribución se caracteriza por ser una buena candidata para modelar series financieras ya que debido a su naturaleza se espera que capture el efecto de asimetría por sus colas pesadas. La volatilidad es definida por (Tsay, 2005) como la desviación estándar de los rendimientos de un activo.

## 1.1. Planteamiento del Problema

## 1.2. Justificación

## 1.3. Objetivos

### 1.3.1. Objetivo General

Estimar la volatilidad de portafolios de inversión que incluyen criptoactivos y activos tradicionales usando métodos bayesianos

### 1.3.2. Objetivos Específicos

- Estimar la volatilidad usando métodos bayesianos cuando se asume que la serie de los retornos es una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución NORMAL.
- Estimar la volatilidad usando métodos bayesianos cuando se asume que la serie de los retornos es una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución GED.
- Comparar los resultados obtenidos después de utilizar las dos formas distribucionales.

## 2. Antecedentes

## 3. Marco Teórico

### 3.1. Marco Conceptual



## 3.2. Marco Estadístico

### 3.2.1. Distribución Normal

La Distribucion normal o Gaussiana fue presentada por primera vez en 1733 por el matemático francés Abraham de Moivre y posteriormente citada por Gaus en 1809. (Canavos, 1991) Una variable aleatoria  $X$  se encuentra normalmente distribuida si su función de densidad de probabilidad esta dada por:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right] \quad -\infty < x < \infty \quad -\infty < \mu < \infty, \sigma > 0 \quad (3-1)$$

Los parámetros que determinan la función de densidad de la distribución normal son  $\mu$  es un parámetro de forma y  $\sigma$  es un parámetro de escala. Estos parámetros se les conoce como la media y la desviación estándar de la variable aleatoria  $X$ , respectivamente.

La función de distribución acumulada (f.d.a) de la variable aleatoria  $X$  con distribución Normal esta dada por:

$$F(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (3-2)$$

Sea  $Z$  una variable aleatoria definida por la siguiente relación  $Z = (X - \mu)/\sigma$ , donde  $\mu$  y  $\sigma$  son la media y la desviación estandar de  $X$ , respectivamente. De acuerdo con esto  $Z^*$  es una variable aleatoria estandarizada con media cero y desviación estandar uno. La transformación se aplica en la formula (3-2) y esta expresada de la siguiente forma:

$$P[Z \leq (X - \mu)/\sigma] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{(x-\mu)/\sigma} \exp(-z^2/2)(\sigma dz) \quad (3-3)$$

Si  $X$  se encuentra normalmente distribuida con media  $\mu$  y desviación estandar  $\sigma$ , entonces  $Z = (X - \mu)/\sigma$  también se encuentra normalmente distribuida con media cero y desviación estándar uno. Entonces  $z = (x - \mu)/\sigma$ ,  $P(X \leq x) = P(Z \leq z)$  y:

$$F_X(x; \mu, \sigma) = F_Z(z; 0, 1) \quad (3-4)$$

### 3.2.2. Distribución de Error Genralizado: GED

La familia Generalized Error Distribution(G.E.D.) fue introducida por Subbotin (1923) y ha sido empleada por varios autores con diferentes nombres y parametrizaciones (por ejemplo, Box Tiao, 1973; Agrò, 1999; Mineo, 2007; Bottazzi Secchi, 2011)).

La Distribución del Error Generalizado no es mas que una extesión de la distribución Normal que pertenece a la familia simétrica de distribuciones que se utilizan en el modelado matemático, generalmente cuando los errores (la diferencia entre el valor esperado y los valores observados) no se distribuyen normalmente como es el caso de las series financieras donde esta distribución permite modelar el comportamiento leptocurtico y de colas pesadas. Los casos especiales de esta distribución son idénticos a la distribución normal y la distribución de Laplace.

Adicionalmente, se usa la parametrización realizada por (Box and Tiao, 2011)

La variable aleatoria  $X$  se distribuye GED (Distribución de Error Generalizado) con párametros  $\mu$ ,  $\sigma$  y  $\gamma$ , denotada por  $X \sim G(\mu, \sigma^2, \gamma)$  con función de densidad  $f(x)$  dada por:

$$f(x) = \omega(\gamma)\sigma^{-1}\exp\left[-c(\gamma)\left|\frac{x-\mu}{\sigma}\right|^{2/(1+\gamma)}\right] \quad (3-5)$$

Donde,

$$\omega(\gamma) = \frac{\{\Gamma[3(1+\gamma)/2]\}^{1/2}}{(1+\gamma)\{\Gamma[(1+\gamma)/2]\}^{3/2}} \quad (3-6)$$

y

$$c(\gamma) = \left\{ \frac{\Gamma[3(1+\gamma)/2]}{\Gamma[(1+\gamma)/2]} \right\}^{1/(1+\gamma)} \quad (3-7)$$

- $\mu \in (-\infty, \infty)$  parámetro de localización
- $\sigma \in (0, \infty)$  parámetro de escala
- $0 \leq \gamma \leq 1$  parámetro que permite controlar el grosor de la cola

Aquí  $x$  varia de  $-\infty$  a  $\infty$ ,  $\mu$  es un parámetro de ubicación y  $\sigma$  es un parámetro de escala. El  $\gamma$  es un parámetro de forma que esta relacionado con la curtosis de la distribucion que caracteriza la no normalidad de la poblacion.

La media y la varianza de la Distribucion del Error Generalizada estan definidas como:

$$E(x) = \mu$$

$$Var(x) = \sigma^2$$

La distribución es simétrica y por ende su asimetría es 0.

Su curtosis,

$$g_2 = \frac{\Gamma[5(1+\gamma)/2]\Gamma[(1+\gamma)/2]}{\Gamma[3(1+\gamma)/2]^2}$$

Notese que cuando  $\gamma=0$  la función de distribución de probabilidad (3-5) se reconoce como la función de probabilidad de una Distribución Normal Univariada, i.e.  $G(\mu, \sigma^2, 0) = N(\mu, \sigma^2)$ . Ahora, si  $\gamma=1$  la función de probabilidad se expresa como la distribución Doble Exponencial o de Laplace, es decir,  $G(\mu, \sigma^2, 1) = L(\mu, \sigma^2)$ . Cuando  $\gamma$  tiende a -1 la función de probabilidad se asemeja a una distribución Uniforme  $U(\mu - \sigma, \mu + \sigma)$ .

Ahora, si la media es igual a cero y la desviación estándar igual a uno, entonces se dice que  $X$  se distribuye GED generalizada.

### 3.2.3. Inferencia Clásica

La inferencia clásica no es más que el proceso de sacar información proporcionada por una muestra mediante un conjunto de métodos y técnicas estadísticas.

La inferencia estadística usa un número calculado a partir de los datos de la muestra para hacer inferencias sobre un número desconocido que describe la población más grande. En este sentido, un parámetro es un número que describe la población y una estadística es un número que se puede calcular a partir de los datos de la muestra sin hacer uso de ningún parámetro desconocido.

#### Método de Máxima Verosimilitud

Dadas  $n$  variables aleatorias  $X_1 \dots X_n$  la función de verosimilitud se define como la función de densidad conjunta de las  $n$  variables y se denota por  $L(x_1, \dots, x_n, \theta)$

En particular, cuando  $n$  variables aleatorias conforman una muestra aleatoria, la función de verosimilitud queda expresada como:

$$L(x_1, \dots, x_n, \theta) = f(x_1, \theta) \dots f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (3-8)$$

donde  $f$  es la función de densidad común para las  $n$  variables. Cuando solo se dispone de una variable aleatoria  $X$ , la función de verosimilitud es simplemente la función de densidad de  $X$ , esto es  $f_X(x)$ .

Dado lo anterior, el método de máxima verosimilitud para un parámetro  $\theta$  consiste en encontrar el valor de  $\theta$  que maximice esta función, éste será el estimador de máxima verosimilitud de  $\theta$  y se denota por  $\theta_{MV}$ . (Rojas, 2010)

### 3.2.4. Inferencia Bayesiana

La inferencia bayesiana proporciona métodos que permiten caracterizar las poblaciones por medio de sus parámetros, ((Hoff, 2009) menciona que el uso de probabilidades para expresar información puede hacerse en un sentido matemático por medio del Teorema de Bayes. En consecuencia, el método bayesiano se basa en establecer distribuciones a priori que explicitan el conocimiento previo que permite construir una nueva función denominada distribución a posterior la cual contiene la información sobre el parámetro a ser estimado  $\theta$  condicional a los datos observados.

#### Teorema de Bayes

Suponga que  $y = (y_1, \dots, y_n)$  es un vector  $n$  observaciones cuya distribución de probabilidad  $f(y|\theta)$  depende de los valores  $k$  parámetros  $\theta = (\theta_1, \dots, \theta_k)$ . Suponga también que  $\theta$  tiene una distribución de probabilidad  $\pi(\theta)$ . Entonces, dado los datos observados  $y$ , la distribución condicional de  $\theta$  es

$$f(\theta|y) = \frac{L(\theta|y)\pi(\theta)}{\int_{\Theta} L(\theta|y)\pi(\theta)d\theta} \quad (3-9)$$

Donde  $\Theta$  es el espacio paramétrico de  $\theta$  y

$$Ef(\theta|y) = c^{-1} = \int_{\Theta} L(\theta|y)\pi(\theta)d\theta \quad (3-10)$$

La integral esta definida sobre todo el espacio paramétrico de  $\theta$  y donde  $Ef(\theta|y)$  es el valor esperado de  $f(\theta)$  con respecto a la distribución  $\pi(\theta)$ . La cantidad de  $c$  es meramente una constante de normalización necesaria para asegurar que la distribución a posterior  $f(\theta|y)$  se integre o sume a uno. (Box and Tiao, 2011).

$$f(\theta|y) = cL(\theta|y)\pi(\theta). \quad (3-11)$$

La ecuación (3-5) o su equivalente (3-7) es usualmente referido como el Teorema de Bayes. En la expresión  $\pi(\theta)$ , dice lo que se sabe sobre  $\theta$  sin el conocimiento de los datos, es llamada distribución a priori de  $\theta$ . Correspondiente,  $f(\theta|y)$  expresa lo que se sabe de  $\theta$  dado el conocimiento de los datos, es llamada distribución a posteriori de  $\theta$ .

## Método Empírico de Bayes

El método empírico de Bayes plantea la distribución a priori  $\pi(\theta)$  a partir del conocimiento del parámetro  $\theta$ . A este método se le puede dar una interpretación frecuencial y se puede estimar usando observaciones apropiadas, así que este puede ser esencialmente no Bayesiano, en el sentido que no involucra probabilidades subjetivas.

En términos matemáticos, determinar una distribución posterior (3-5) donde  $L(\theta|x_{(n)})$  es la función de verosimilitud evaluada en  $\theta$  y  $\pi(\theta)$  es la distribución previa. En el caso de empírico de Bayes los parámetros desconocidos de  $\pi(\theta)$  que se deben estimar para trabajar la ecuación (3-5) de la misma manera que el método de Bayes

El Bayes empírico concuerda con el método de Bayes, la diferencia fundamental radica en que el método de Bayes supone el conocimiento total de la distribución previa  $\pi(\theta)$  mientras el método empírico no especifica el valor de los parámetros en la distribución a priori (Prieto et al., 1995)

## Distribución A priori

Como se mencionó anteriormente, la metodología bayesiana hace uso de la distribución a priori, la cual se puede definir como el proceso de tomar la información que tiene el especialista y expresarla en forma de una distribución de probabilidades (Tovar Cuevas, 2015) La distribución a priori cumple un papel importante en el análisis bayesiano ya que mide el grado de conocimiento inicial que se tiene de los parámetros en estudio. Si bien su influencia disminuye a medida que más información muestral está disponible, el uso de una u otra distribución a priori determinará ciertas diferencias en la distribución a posteriori.

Si se tiene un conocimiento previo sobre los parámetros, este se traducirá en una distribución a priori. Así será posible plantear tantas distribuciones a priori como estados iniciales de conocimiento existan y los diferentes resultados obtenidos en la distribución a posteriori bajo cada uno de los enfoques, adquirirán una importancia en relación con la convicción que tenga el investigador sobre cada estado inicial. Sin embargo, cuando nada es conocido sobre los parámetros, la selección de una distribución a priori adecuada adquiere una connotación especial pues será necesario elegir una distribución a priori que no influya sobre ninguno de los posibles valores de los parámetros en cuestión. Estas distribuciones a priori reciben el nombre de difusas o no informativas.

Se consideran dos interpretaciones básicas que se pueden dar a las distribuciones previas (a priori). En la interpretación de la población, la distribución a priori representa una población de posibles parámetros o valores, del que se ha sacado el  $\theta$  de interés actual. En el estado más subjetivo de la interpretación del conocimiento, el principio rector es que se debe expresar el conocimiento (e incertidumbre) sobre  $\theta$  como si su valor pudiera ser considerado como

una realización aleatoria de la distribución a priori. Para muchos problemas, tales como la estimación de la probabilidad de fallo en un nuevo proceso industrial, no hay ninguna población perfectamente relevante de  $\theta$  desde el cual el  $\theta$  actual se ha dibujado, excepto en la contemplación hipotética. Por lo general, la distribución a priori debe incluir todos los valores plausibles de  $\theta$ , pero la distribución no necesita ser realista concentrada en torno al valor verdadero, porque muchas veces la información sobre  $\theta$  está contenida en los datos que superan con creces cualquier especificación o probabilidad previa razonable (Gelman et al., 2013).

### Distribución A priori no Informativa

Para realizar inferencia estadística desde el enfoque bayesiano, se requiere contar con información externa al proceso de muestreo que dio origen a los datos observados, la cual puede tener como fuente el conocimiento subjetivo de un especialista en el tema, datos históricos recolectados sobre el tema de interés, resultados publicados sobre el mismo asunto en diferentes fuentes, etc. Una a priori no informativa o también llamada a priori de referencia las cuales son distribuciones que representan la mínima cantidad posible de información a priori disponible, generalmente útiles cuando no se cuenta con un especialista en el tema o se tiene tan poca información publicada sobre el comportamiento natural de las cantidades aleatorias de interés, generalmente, las a priori de referencia se utilizan para tener una línea de base con la cual comparar otras distribuciones con los resultados obtenidos usando métodos clásicos (Gelman et al., 2013)

### Principio de Invariancia Jeffreys

El principio de Jeffreys conduce a la definición de la densidad a priori no informativa como  $h(\theta) \propto [J(\theta)]^{1/2}$ , donde  $J(\theta)$  es la información de Fisher para (Gelman et al., 2013).

$$J(\theta) = -E \left( \frac{d^2 \ln(\theta|x)}{d\theta^2} \right) \quad (3-12)$$

La distribución a priori de jeffreys ser  $h(\theta) = \sqrt{J(\theta)}$ . Por lo tanto:

$$\sqrt{J(\theta_j)} = \sqrt{\frac{n_j}{\theta_j(1-\theta_j)}} = \sqrt{n_j} \theta_j^{-1/2} (1-\theta_j)^{-1/2} \quad (3-13)$$

Dado lo anterior se concluye que la distribución a priori Jeffrey es:

$$h(\theta_j) = \text{Beta}(1/2, 1/2) \quad (3-14)$$

### Distribución A priori Informativa

Se presenta cuando se tiene información externa del evento de interés, dicho conocimiento se puede dar por la experiencia del especialista o investigaciones previas. El hecho de expresarla en forma de una distribución de probabilidades, se ha denominado proceso de elicitación a priori el cual es un proceso complejo con el que se formalizan las creencias que se tiene acerca del comportamiento natural de un fenómeno (Tovar Cuevas, 2015)

De acuerdo con (Gelman et al., 2013), cuando no se cuenta con fuentes externas de información, es posible usar una parte de los datos o realizar procesos con los mismos que permitan tener una primera aproximación a lo que podría ser el comportamiento de los parámetros sin llegar a ser concluyentes en nuestras apreciaciones. Los autores denominaron a dicho proceder como métodos empíricos de Bayes o Bayes Empíricos. (Gelman et al., 2013)

### Deviance Information Criterion (DIC)

Es un criterio de selección de modelos que consiste en definir una desviación de 2 veces la log-verosimilitud, corregida por una constante aditiva.(Spiegelhalter et al., 2002).

$$D(\theta) \equiv -2\text{Log}[L(\theta)] + C \quad (3-15)$$

Donde  $C$  es cualquier constante que no depende de  $\theta$ .

$$E[D(\theta)|y] \quad (3-16)$$

Lo anterior se utiliza como una medida de ajuste de un modelo con valores pequeños. También para definir un numero efectivo de parámetros. Esto se utiliza una medida de ajuste del modelo con valores pequeños que son buenos. También definir un número efectivo de parámetros. (Christensen R et al., 2012)

$$P_D \equiv E[D(\theta)|y] - D[\hat{\theta}] \quad (3-17)$$

Donde  $\hat{\theta}$  es una estimación bayesiana de , digamos la media, mediana posteriori. Finalmente, DIC se define como la suma de la medida del ajuste del modelo y una penalización para el número efectivo de parámetros.

$$DIC \equiv E[D(\theta|y)] + P_D = 2E[D(\theta|y)] - D[\hat{\theta}] \quad (3-18)$$

El DIC se calcula fácilmente utilizando una muestra de la distribución a posteriori (Christensen R et al., 2012)

### Cadenas de Markov y Método Montecarlo (MCMC)

Se han ideado muchos métodos inteligentes para construir y muestrear distribuciones posteriores arbitrariamente. La simulación de una cadena de Markov (también llamada cadena de Markov de Monte Carlo, o MCMC) es un método general basado en obtener valores de  $\theta$  a partir de distribuciones aproximadas y luego corrigiendo dichos valores para aproximar mejor la distribución posterior objetivo,  $p(\theta|y)$ . (Gelman et al., 2013)

Se realiza un muestreo de forma secuencial, con la distribución de los valores muestreados en función del último valor obtenido; Por lo tanto, dichos valores forman una cadena de Markov. (Como se define en la teoría de la probabilidad, una cadena de Markov es una secuencia de variables aleatorias  $\theta_1, \theta_2, \dots$ , para la cual, para cualquier  $t$ , la distribución de  $\theta^t$  dados todos los  $\theta'$ s previos depende solo del valor más reciente,  $\theta^{t-1}$ ). La clave del éxito del método, sin embargo, no es la propiedad de Markov sino que las distribuciones aproximadas se mejoran en cada paso de la simulación, en el sentido de que convergen a la distribución objetivo. (Gelman et al., 2013)

### Muestreo de Gibbs

El muestreo de Gibbs es un esquema MCMC donde el kernel de transición es formado con la distribución condicional completa. Asuma que la distribución de interés es  $\pi(\theta)$  donde  $\theta = (\theta_1, \dots, \theta_d)'$ . Cada uno de los componentes  $\theta_i$  puede ser un escalar, un vector o una matriz. Considere también que la distribución condicional completa  $\pi_i(\theta_i) = \pi(\theta_i|\theta_{1-i})$ ,  $i = 1, \dots, d$  están disponibles. Esto significa que son completamente conocidos y pueden ser muestreados (Gamerman and Lopes, 2006)

El problema a resolver es obtener valores de  $\pi$  cuando los esquemas de generación directa son costosos, complicados o simplemente no están disponibles pero cuando la generación de los  $\pi_i$  es posible. El Muestreo de Gibbs provee un esquema de generación alternativo basado en generaciones sucesivas a partir de la distribución condicional completa. Este puede ser descrito de la siguiente manera (Gamerman and Lopes, 2006):

1. Inicializar el contador de iteraciones de la cadena.  $j = 1$  y un conjunto de valores iniciales  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})'$
2. Obtenga el nuevo valor  $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})'$  a partir de  $\theta^{j-1}$  mediante generación sucesiva de valores

$$\theta_1^j \sim \pi(\theta_1|\theta_2^{j-1}, \dots, \theta_d^{j-1}), \quad (3-19)$$

$$\theta_2^j \sim \pi(\theta_2|\theta_1^j, \theta_3^{j-1}, \dots, \theta_d^{j-1}),$$

$$\theta_d^j \sim \pi(\theta_d|\theta_1^j, \dots, \theta_d^{j-1}),$$



3. Cambiar el contador  $j$  a  $j + 1$  y retornar al paso 2 hasta alcanzar la convergencia

Cuando se alcanza la convergencia, el valor resultante  $\theta^j$

es obtenido a partir de  $\pi$ . Como el número de iteraciones incrementa, la cadena se acerca a su condición de equilibrio. (Gamerman and Lopes, 2006)

### **Algoritmo Metropolis-Hasting**

## 4. Metodología

Se presenta la metodología determinada para el estudio de Estimación de la volatilidad de los índices los cuales representan activos tradicionales y criptoactivos usando métodos bayesianos.

### 4.0.1. Archivo de datos

Inicialmente se detalla la base de datos extraída del aplicativo REFINITIV y la API, la cual contiene el precio de cierre de los índices **NCI**, **CRBQX** y **SP500** con datos diarios del 18 de noviembre del 2021 hasta el 17 de diciembre del 2021, a los cuales se les calculó el retorno diario excluyendo los días en los que no se obtuvo la cuantificación del precio de cierre de los índices en la siguiente forma:

Sea  $R_t$  el retorno de un activo en un tiempo  $t$ ,

$$R_t = \ln(X_t) - \ln(X_{t-1}) = \ln\left(\frac{X_t}{X_{t-1}}\right) \quad (4-1)$$

donde  $X_{t-1}$  es el precio del activo en el tiempo  $t-1$  y  $X_t$  es el precio en el tiempo  $t$ .

### 4.0.2. Análisis descriptivo

El análisis Estadístico inicia con un análisis exploratorio de los tres índices de interés como para sus respectivos retornos, se emplearon tablas con medidas de tendencia central tales como la media, mediana, cuartiles, máximos y mínimos; medidas de dispersión como la desviación estándar y la varianza, curtosis y la asimetría, por último; se emplearon gráficos que permitieran evaluar las particularidades distribucionales y de correlación entre los retornos de los activos apoyados con cálculos como la medida de correlación a través del coeficiente de Spearman.

### 4.0.3. Estimación por Máxima Verosimilitud

Para obtener la estimación del parámetro de volatilidad  $\sigma^2$  se usa como instrumento el método de Máxima Verosimilitud que consiste en encontrar  $\hat{\sigma}_{MV}^2$  de tal forma que maximice la función de verosimilitud.

Generando los resultados de los objetivos planteados donde se determina la estimación del parámetro de interés a través de la distribución Normal y Distribución del Error Generalizada, se procede a evaluar la máxima verosimilitud de ambas distribuciones.

### Distribución Normal

Se detalla el método de máxima verosimilitud de la distribución Normal

$$L(\hat{\sigma}^2, \mu | r_1, r_2, \dots, r_n) = \prod_{i=1}^n (2\pi\hat{\sigma}^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}(r_i - \mu)^2\right] \quad (4-2)$$

$$= (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}(r_i - \mu)^2\right]$$

donde  $(r_1, r_2, \dots, r_n)$  es una muestra de tamaño  $n$ . Considerando la función

$$\ln L(\hat{\sigma}^2, \mu | r_1, r_2, \dots, r_n) = -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (r_i - \mu)^2 \quad (4-3)$$

Seguidamente el estimador de Máxima Verosimilitud se expresa por:

$$\hat{\sigma}_{MV}^2 = \frac{1}{n} \sum_{i=1}^n (r_i - \mu)^2 \Leftrightarrow \hat{\sigma}_{mv}^{\frac{1}{2}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_i - \mu)^2} \quad (4-4)$$

se tiene que  $(E(\hat{\sigma}^2) \neq \sigma^2)$  evidenciando sesgo lo cual es una peculiaridad no deseada en un estimador, es por esto que se busca utilizar la forma insesgada del mismo:

$$\hat{\sigma}_{MV}^2 = \frac{\sum_{j=1}^n (r_i - \bar{r})^2}{n-1} \quad (4-5)$$

donde  $\bar{r}$  es el promedio muestral de los retornos.

### Distribución del Error Generalizado (GED)

Supongamos que tenemos una muestra de  $n$  observaciones i.i.d. extraídas de la función de densidad, entonces la función de verosimilitud es:

$$L(x; \mu, \sigma_p, p) = [2p^{\frac{1}{p}} \sigma_p \Gamma(1 + \frac{1}{p})]^{-n} \exp\left(-\frac{\sum_{i=1}^n |x_i - \mu|^p}{p\sigma_p^p}\right) \quad (4-6)$$

y la función de log-verosimilitud esta dada por,

$$\iota(x; \mu, \sigma_p, p) = \log L(x; \mu, \sigma_p, p) = -n \log[2p^{\frac{1}{p}} \sigma_p \Gamma(1 + \frac{1}{p})] - \frac{\sum_{i=1}^n |x_i - \mu|^p}{p \sigma_p^p} \quad (4-7)$$

Derivando la función de verosimilitud con respecto a  $\mu$  y  $\sigma_p$  e igualando a 0 las expresiones obtenidas,

$$\frac{\partial \iota(x)}{\partial \mu} = -\frac{1}{\sigma_p^p} \sum_{i=1}^n |x_i - \mu|^{p-1} \text{sign}(x_i - \mu) = 0 \quad (4-8)$$

$$\frac{\partial \iota(x)}{\partial \sigma_p} = -\frac{1}{\sigma_p} + \frac{1}{\sigma_p^{p+1}} \sum_{i=1}^n |x_i - \mu|^p = 0 \quad (4-9)$$

La  $\frac{\partial \iota(x)}{\partial \mu}$  no tiene solución explícita y se resuelve mediante métodos numéricos, en tanto la  $\frac{\partial \iota(x)}{\partial \sigma_p}$  se obtiene el estimador de máxima verosimilitud de  $\sigma_p$

$$\hat{\sigma}_p = \left( \frac{\sum_{i=1}^n |x_i - \mu|^p}{n} \right)^{\frac{1}{p}} \quad (4-10)$$

#### 4.0.4. Enfoque Bayesiano

Dando respuesta al objetivo principal de esta investigación, centrada en la estimación del parámetro de volatilidad, la cual es una medida de incertidumbre y altamente estocástica, se inicia la estimación desde el enfoque bayesiano proponiendo distribuciones previas y apartir de estas hallar las distribuciones posteriores con el fin de obtener la estimación.

##### Distribución apriori no informativa

Una de las principales fortalezas del método de Jeffreys es su invarianza bajo transformaciones inyectivas, basándose en la medida de información de Fisher sobre  $\sigma^2 > 0$ , es decir,

$$I(\sigma^2) = E\left[\left(\frac{\partial \ln L(\sigma^2 | r_1, r_2, \dots, r_n)}{\partial \sigma^2}\right)^2 | \sigma^2\right] = -E\left[\frac{\partial^2 \ln L(\sigma^2 | r_1, r_2, \dots, r_n)}{\partial (\sigma^2)^2} | \sigma^2\right] \quad (4-11)$$

De este modo, la distribución previa propuesta por Jeffreys para el caso uniparamétrico,

$$h(\sigma^2) \propto \sqrt{I(\sigma^2)} \quad (4-12)$$

se tiene que,

$$i(\sigma^2|r_1, r_2, \dots, r_n) = \ln L(\hat{\sigma}^2, \mu|r_1, r_2, \dots, r_n) = -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (r_i)^2 \quad (4-13)$$

Posteriormente, las derivadas parciales de primer y segundo orden quedan expresadas como

$$\frac{\partial i}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n r_i^2 \Rightarrow \frac{\partial^2 i}{\partial (\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^n r_i^2 \quad (4-14)$$

Observe,

$$I(\sigma^2) = -E\left[\frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^n r_i^2\right] = \frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3} \sum_{i=1}^n E[r_i^2] = \frac{n}{2(\sigma^2)^2} + \frac{n}{(\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} \quad (4-15)$$

en otras palabras, la distribución previa No informativa obtenida usando el método de Jeffreys se demarca como

$$h(\sigma^2) \propto \frac{\sqrt{n}}{\sigma^2} \quad (4-16)$$

Para la **Distribucion del Error Generalizada** la información de Fisher queda expresada de la forma

$$I(\sigma_p) = -E\left[\frac{\partial^2 \ln i(x)}{\partial (\sigma^2)^2}\right] = \frac{p}{(\sigma^2)^2} \quad (4-17)$$

llegando a la siguiente expresión como la distribución previa no informativa

$$h(\sigma^2) = \left(\frac{p}{\sigma^2}\right)^{\frac{1}{2}} = \frac{\sqrt{p}}{\sigma^2} \quad (4-18)$$

### Distribución Posteriori

Para la distribución normal, partiendo de la distribución previa

$$h(\sigma^2|x_1, \dots, x_n) = \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right] \frac{\sqrt{n}}{\sigma^2}}{\int_0^\infty (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right] \frac{\sqrt{n}}{\sigma^2} d\sigma^2} \quad (4-19)$$

$$\propto \frac{(\sigma^2)^{-(\frac{n}{2}+1)} \exp[-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2]}{\int_0^\infty (\sigma^2)^{-(\frac{n}{2}+1)} \exp[-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2] d\sigma^2}$$

$$= \frac{(\sum_{i=1}^n \frac{x_i^2}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} (\sigma^2)^{-(\frac{n}{2}+1)} \exp[-\frac{1}{\sigma^2} \sum_{i=1}^n \frac{x_i^2}{2}]$$

distribución posterior obtenida, es una distribución Gamma Inversa con parámetros  $(\alpha_1^*, \beta_1^*)$  definiendo:

$$(\alpha_1^* = \frac{n}{2}, \beta_1^* = \sum_{i=1}^n \frac{x_i^2}{2})$$

Para obtener las estimaciones del parametro, se observa que las distribuciones a posteriori obtenidas pertenecen a la familia Beta, la cual es una distribución bastante conocida facilmente programable en la mayoría de los software estadísticos existentes en el mercado. Es posible entonces utilizar un método para encontrar los valores de las probabilidades y las estimaciones a posteriori usando cadenas de Markov. Para obtener las estimaciones, en este trabajo se simularon en el software R, cadenas de 10000 valores de la distribución Beta a posteriori, se observo la forma de la distribución para tener información acerca de la simetría.

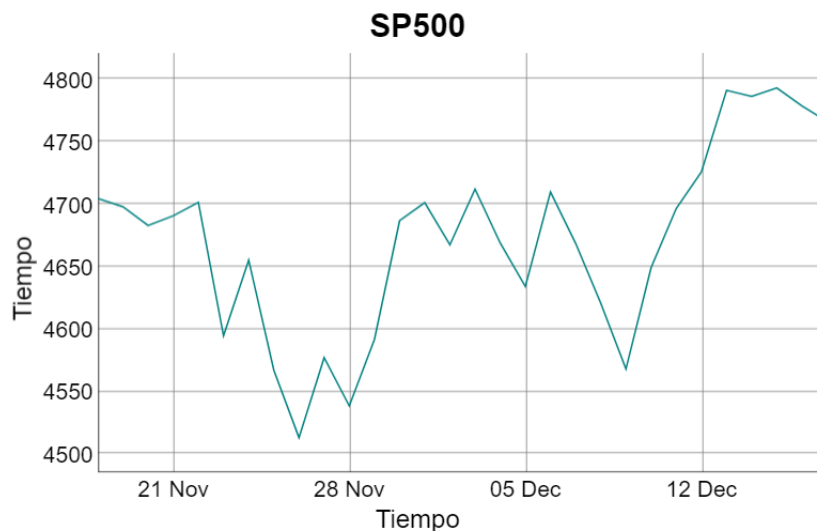
## 5. Resultados

En esta sección se presentan los resultados obtenidos, de acuerdo a la metodología planteada para cumplir con los objetivos propuestos.

### 5.1. Análisis Exploratorio de Datos

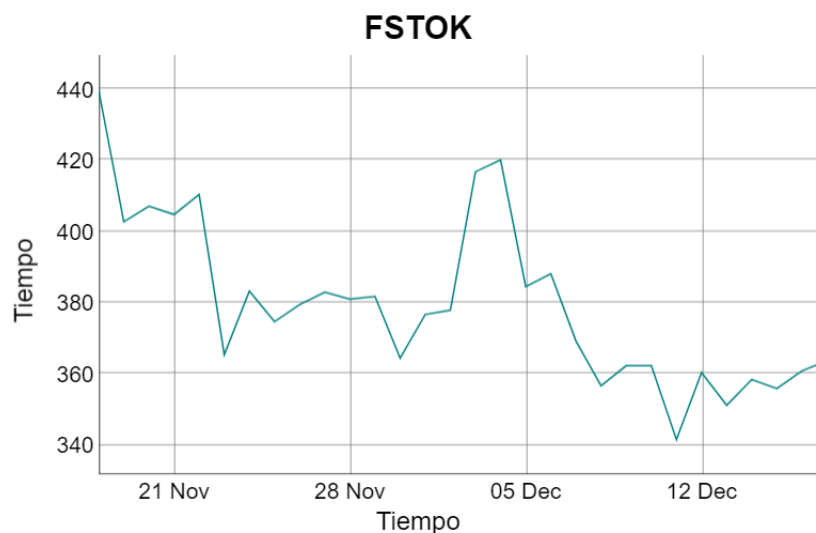
En las figuras 5.1, 5.2, y 5.3 se observa el comportamiento del precio de cierre para cada uno de los índices estudiados tomando como periodo tiempo desde el 18/11/2021 hasta el 17/12/2021.

Para el índice S&P500 se evidencia cierta tendencia creciente para los últimos cinco días, adicionalmente, de manera general se puede indagar que a lo largo del tiempo observado no se observa una tendencia de alza o a la baja.

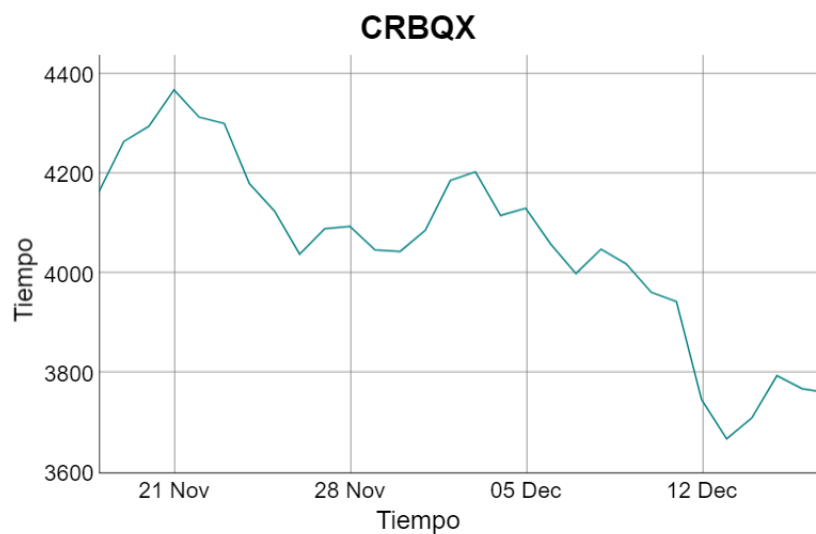


**Figura 5-1.:** Serie de los precios diarios SP&500

La Figura 5.2 se observa una tendencia mas variante en el tiempo, con picos altos y bajos drásticamente en días muy próximos, este índice es muy importante, puesto que es el que incluye los principales criptoactivos del mercado y es la base del problema del trabajo propuesto.



En la figura 5.3, muestra cierta tendencia decreciente para el índice CRBQX con unos cambios no tan bruscos como los que se observaron para el índice de las criptomonedas pero significativos para el análisis del comportamiento de la variabilidad.



Para contextualizar los gráficos anteriores, la tabla 5.1 expone algunas medidas de posición y dispersión del precio de los índices. Se aprecia que los índices de acciones y commodities presentan un promedio similar en su precio de cierre, por otro lado se encuentra que el índice FSTOK de las criptomonedas fue creado con un valor base más bajo, pero algo importante



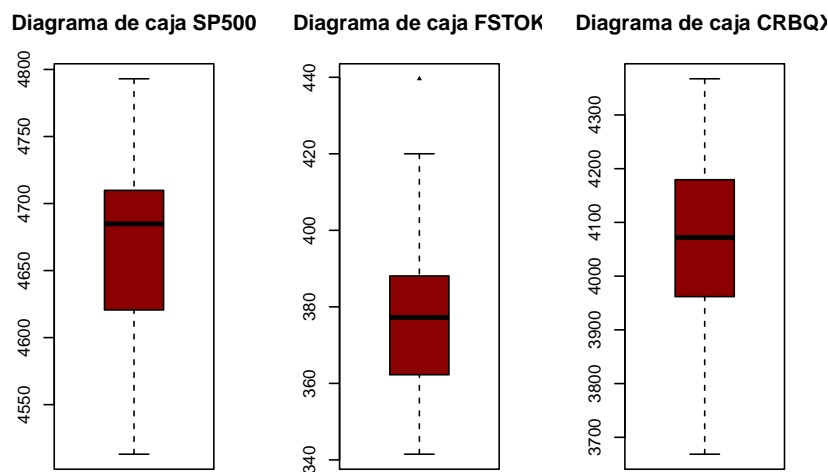
de destacar es que el precio en si mismo no tienen ningún interés y sus diferencias en el valor base no representan alguna dificultad, puesto que la información que aporta al problema base se extrae a partir de los log retornos.

Como medidas relevantes de la tabla, se tiene el coeficiente de variación para el índice de las criptomonedas, el cual presenta un valor mayor que los dos índices restantes, indicando una mayor volatilidad en los precios de cierre.

**Tabla 5-1.:** Estadísticas descriptivas del precio de cierre de cada índice

Estadístico	SP500	FSTOK	CRBQX
n	30	30	30
Media	4671	379.4	4051
Desv. Estándar	75.23917	23.15564	188.7198
Coef. Variación	0,0161077	0,06103226	0,04658597
Mínimo	4513	341.5	3669
Q1 (25 %)	4624	362.3	3971
Mediana (Q2)	4685	377.2	4072
Q3 (75 %)	4709	387.2	4175
Máximo	4793	439.6	4367

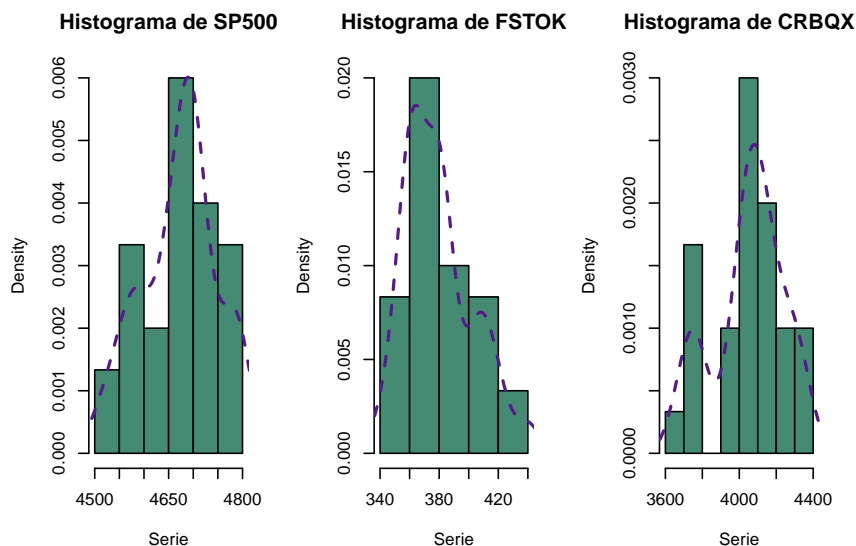
De forma visual, la figura 5.1 evidencia lo expuesto anteriormente, el precio del índice de las criptomonedas es mucho mas bajo que los precios de los otros índices.



**Figura 5-4.:** Boxplot precios de cierre de los índices

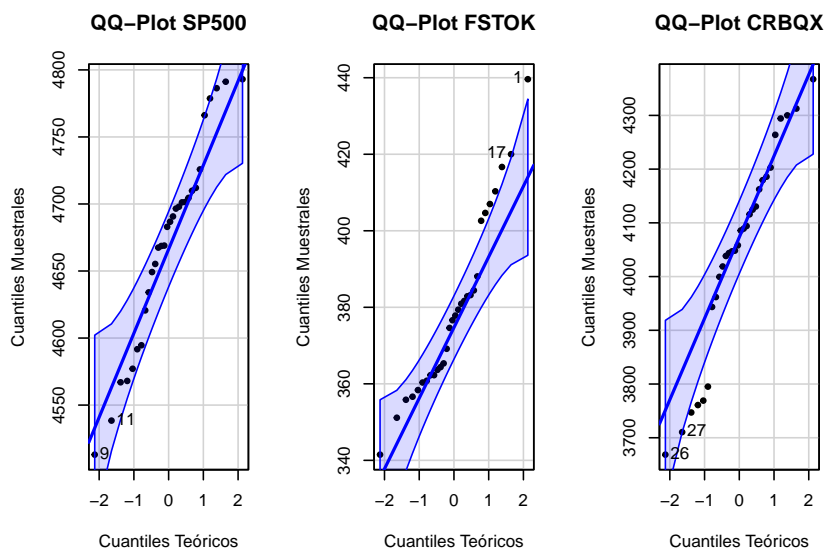
En la figura 5.5, el histograma de los índices de CRBQX y S&P500 muestra cierta asimetría negativa, en caso contrario, se encuentra el índice FSTOK con una asimetría positiva. De

manera general, los tres índices presentan forma leptocurticas, puesto que su concentración esta alrededor de los valores centrales de la variable, indicando a su vez la presencia de valores atipicos.



**Figura 5-5.:** Histograma precios de cierre de los índices

La figura 5.6, el QQ-plot del índice FSTOK se observan varios datos por fuera de las bandas de confianza, lo que podría indicar un comportamiento de "no normalidad", al igual que el índice CRBQX.



**Figura 5-6.:** QQ-plot precios de cierre de los índices

### Puntos atípicos:

A continuación se presenta la detección de puntos atípicos y las pruebas realizadas para la influencia de estos.

**S&P500** Para el índice S&P500 se observa que no presenta puntos que se puedan considerar atípicos.

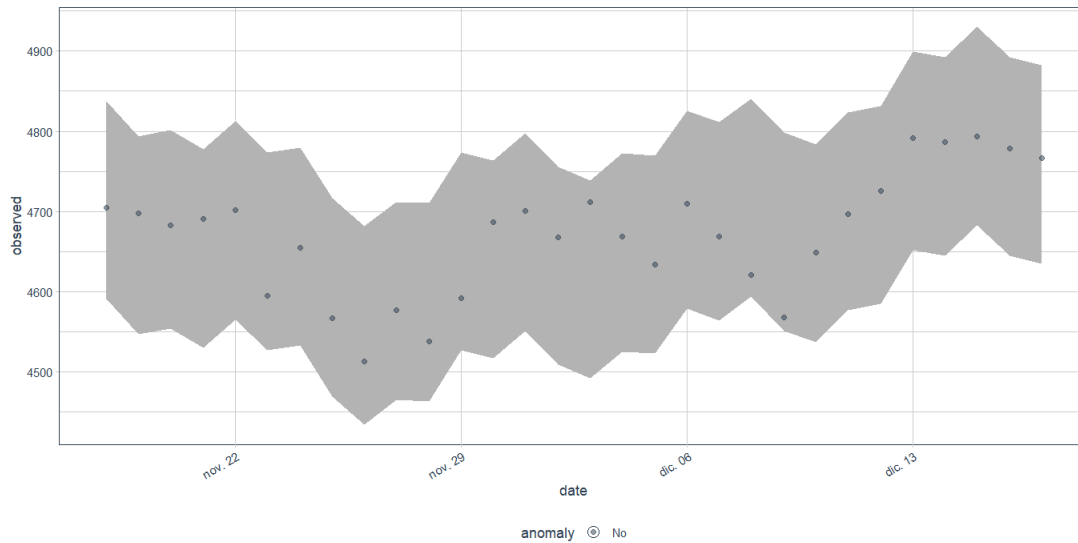


Figura 5-7.: s

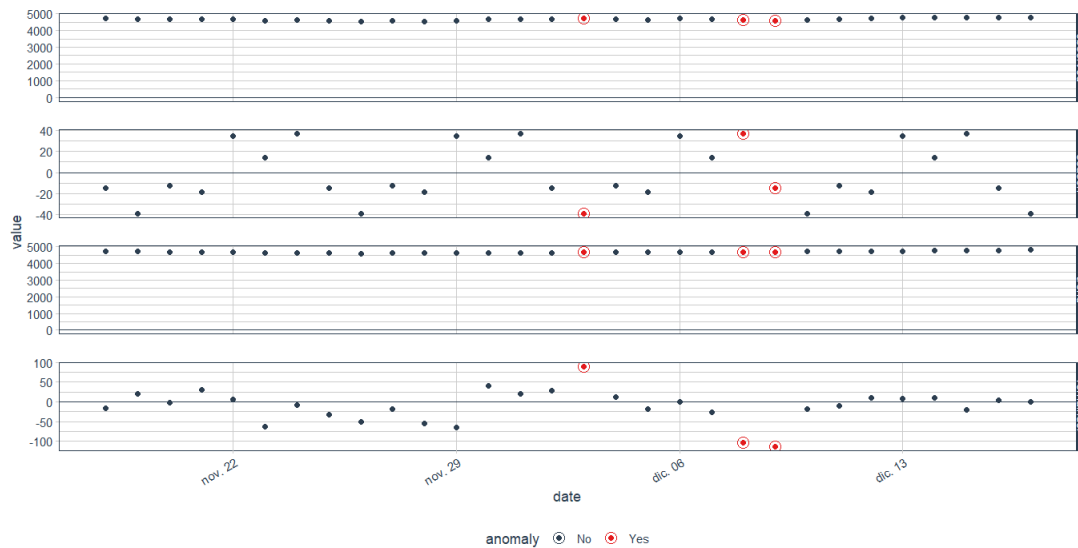


Figura 5-8.: s

En las figuras 5.9 y 5.10 evidencian la presencia de puntos atípicos, esto quizás se deba a los picos observados en las series anteriores.

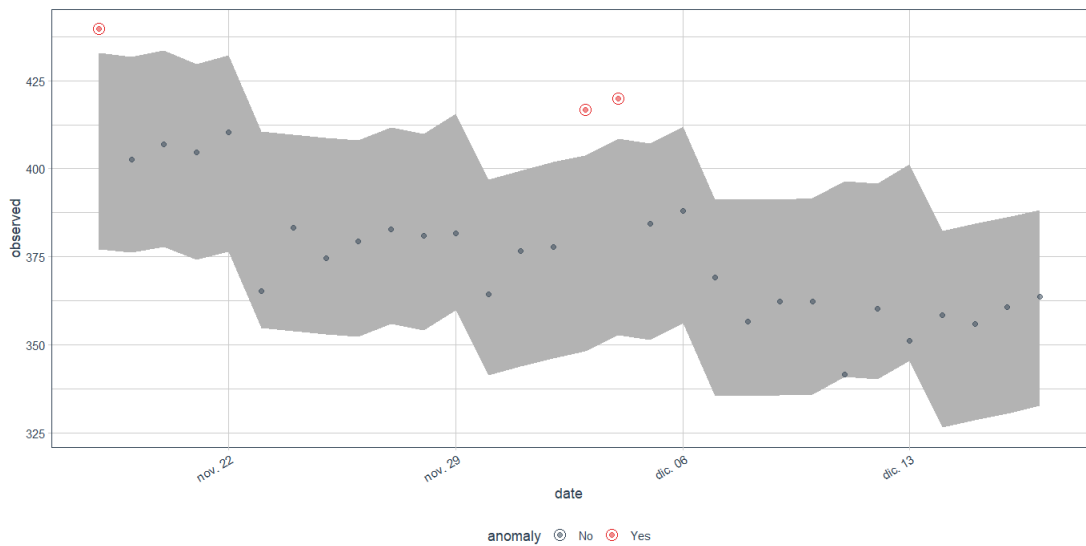


Figura 5-9.: s

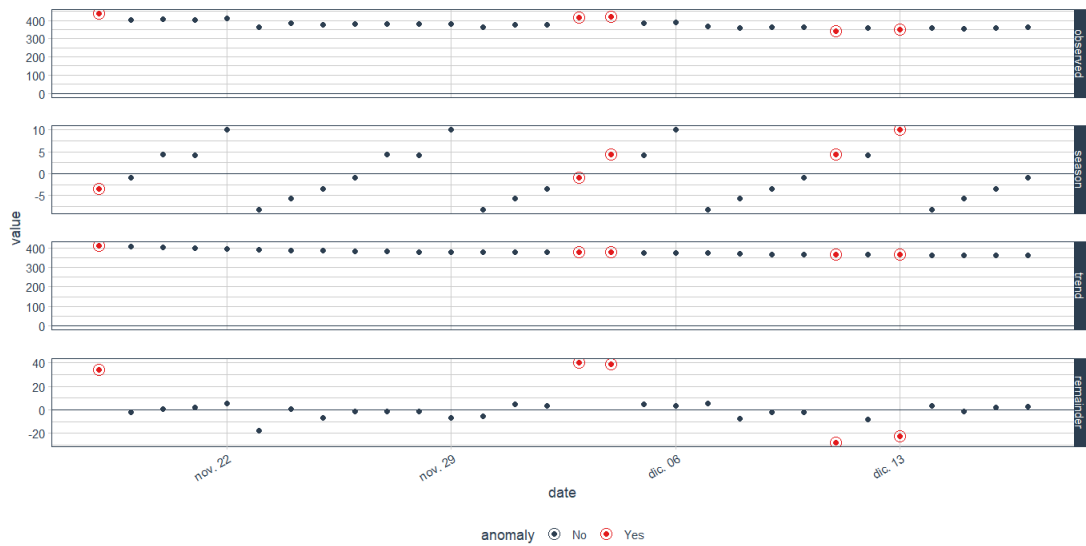


Figura 5-10.: s

	date	observed	season	trend	remain..'
	<date>	<dbl>	<dbl>	<dbl>	<dbl>
1	2021-11-18	440.	-3.48	409.	34.1
2	2021-12-03	417.	-0.978	377.	40.2
3	2021-12-04	420.	4.33	377.	38.8
"	-----				

Figura 5-11.: Tabla resumen atípicos

Puntos atípicos para el índice CRBQX:

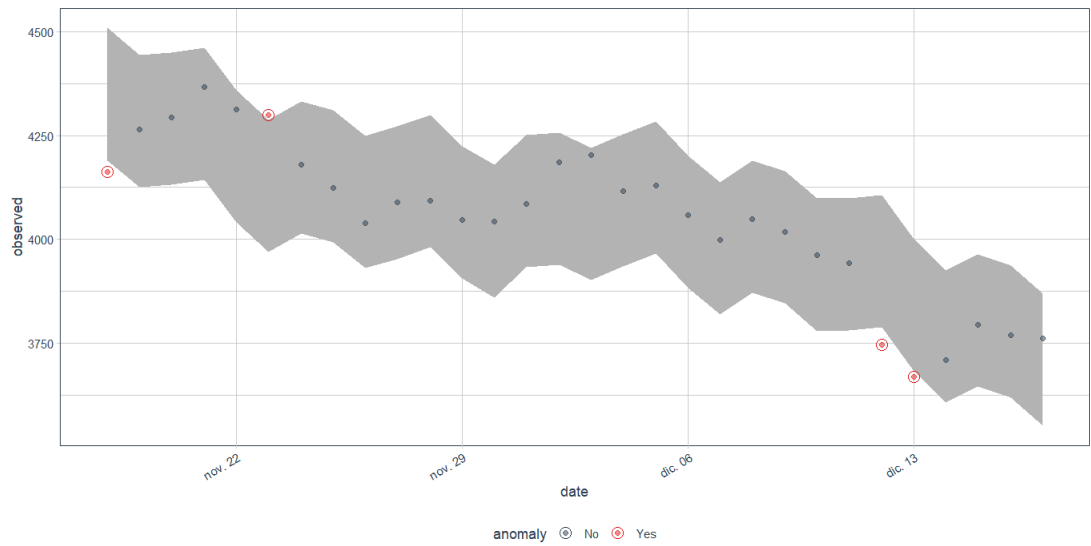


Figura 5-12.: s

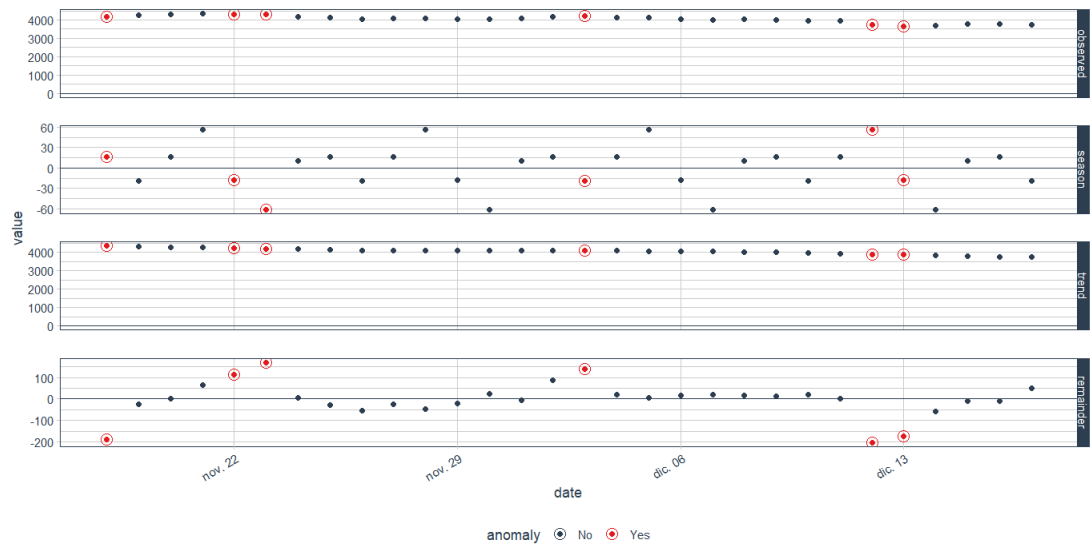
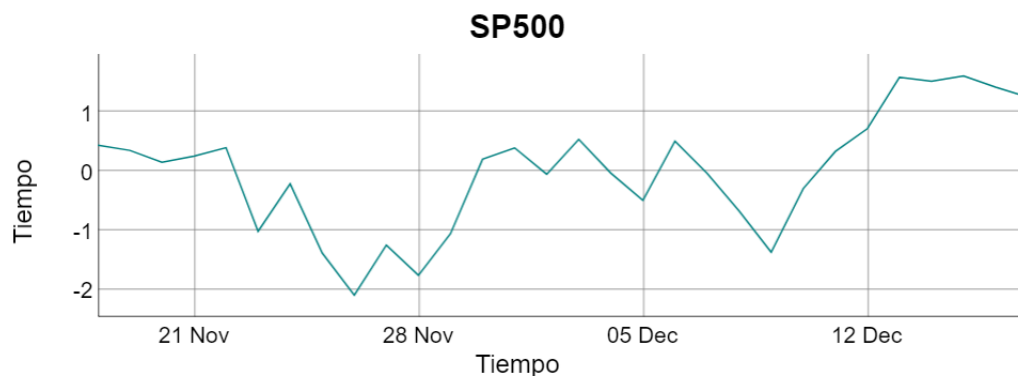


Figura 5-13.: s

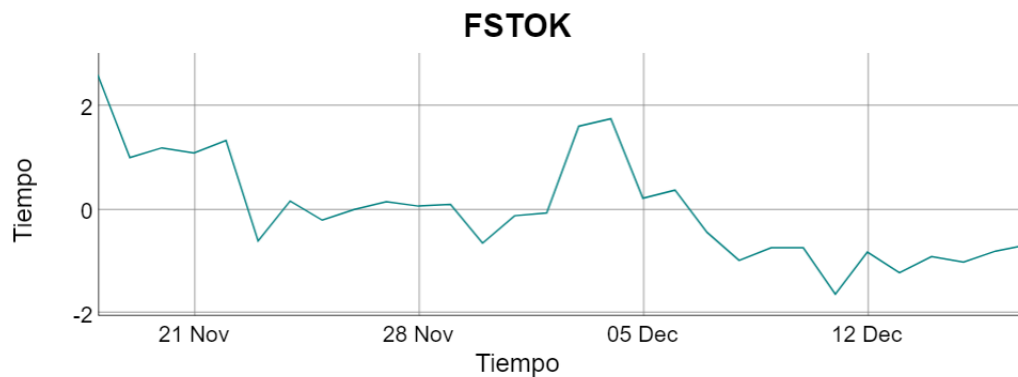
	date	observed	season	trend	remainder	remaide...!	remai...²
	<date>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2021-11-18	4162.	16.1	4334.	-188.	-161.	158.
2	2021-11-23	4300.	-61.4	4192.	170.	-161.	158.
3	2021-12-12	3747.	56.2	3893.	-202.	-161.	158.
4	2021-12-13	3669.	-18.0	3861.	-174.	-161.	158.

Figura 5-14.: Tabla resumen atípicos

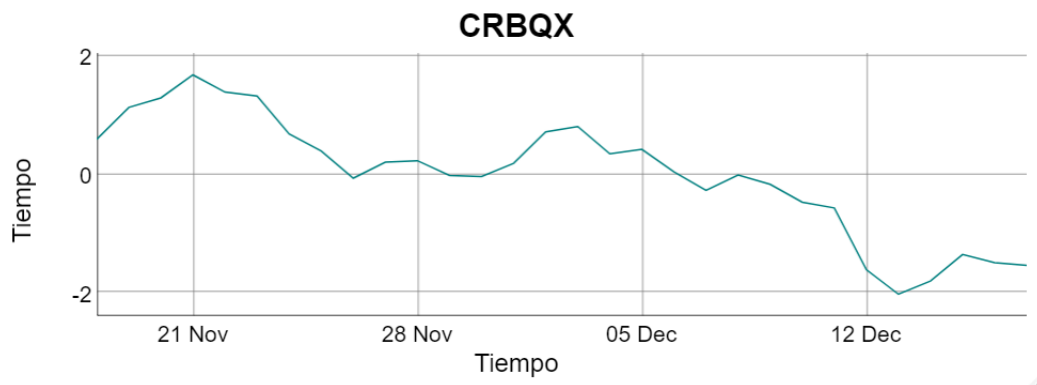
Aunque en las figuras 5.1, 5.2, y 5.3 se representan los valores originales de cada uno de los índices, esta no permite evidenciar las diferencias en términos de volatilidades o variaciones entre las mismas, ya que sus escalas (valores base) son diferentes. Para ello se generó las figuras 5.15, 5.16 y 5.17 en donde se detalla con sus valores normalizados, donde deja en evidencia que los cambios mas notorios con respecto a alzas o bajas de los precios los presenta el índice de las criptomonedas.



**Figura 5-15.:** Serie de los precios diarios normalizados SP&500



**Figura 5-16.:** Serie de los precios diarios normalizados FSTOK



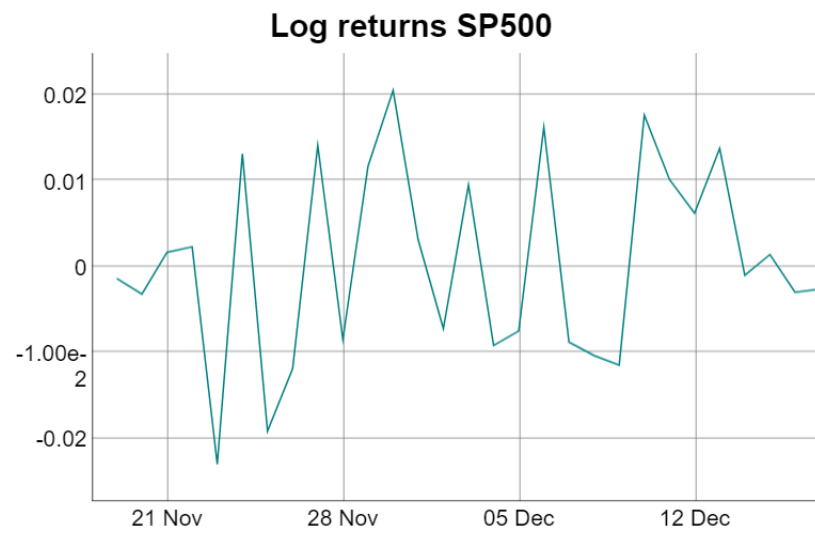
**Figura 5-17.:** Serie de los precios diarios normalizados CRBQX

Puesto que nuestro interés se centra en la volatilidad de los activos estudiados, dando énfasis en el índice de la criptomoneda que es el de principal interés para este estudio, dicha volatilidad puede cobrar más significancia medirla a través de los log retornos.

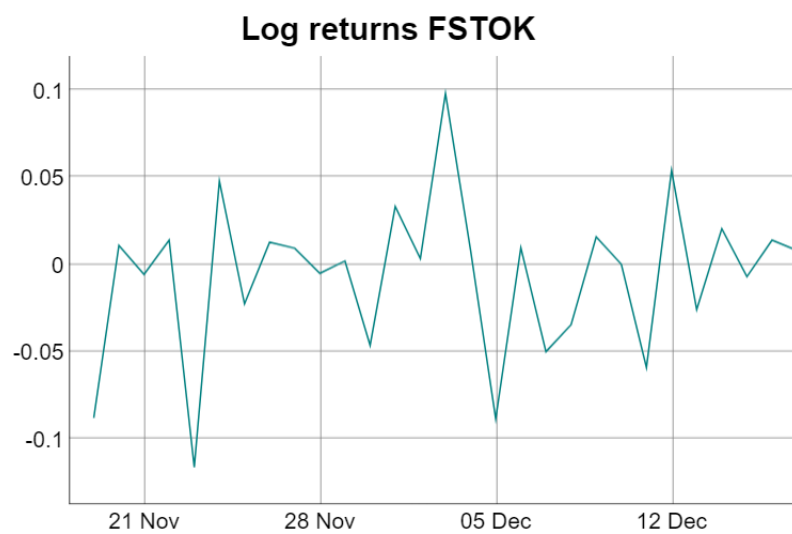
Se presenta algunas estadísticas descriptivas de los log retornos

**Tabla 5-2.:** Estadísticas descriptivas de los log retornos diarios del precio de cierre de los índices

Estadístico	SP500	FSTOK	CRBQX
n	29	29	29
Media	0.0004489	-0.006541	-0.003498
Desv. Estándar	0.01130612	0.04467452	0.01722563
Mínimo	-0.0229885	-0.116085	-0.051095
Q1 (25 %)	-0.0084850	-0.025863	-0.014272
Mediana (Q2)	-0.0010103	0.003234	-0.002921
Q3 (75 %)	0.0101298	0.013843	0.010321
Máximo	0.0204942	0.097824	0.024214
Asimetría	-0.03630881	-0.4381863	-0.4480838
Curtosis	-0.8034134	0.6409947	0.3300642

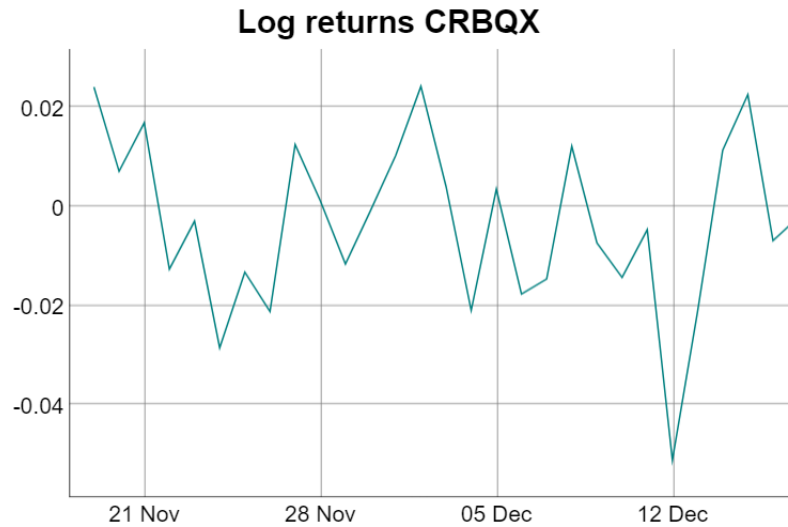


**Figura 5-18.:** Log retorno S&P500

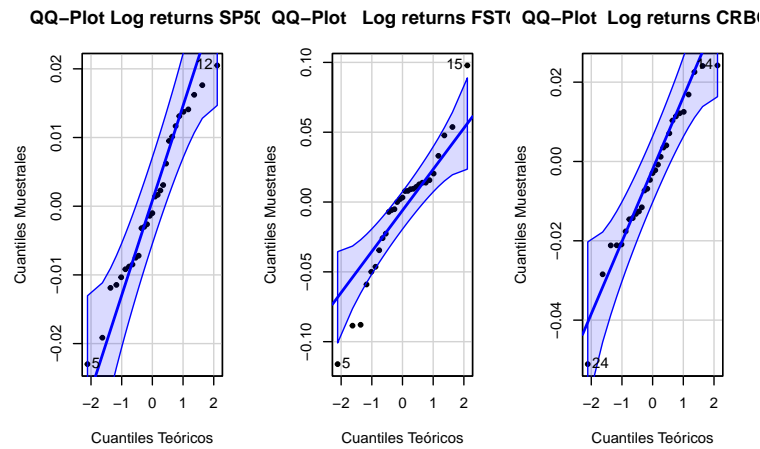


**Figura 5-19.:** Log retorno FSTOK10





**Figura 5-20.:** Log retorno CRBQX



**Figura 5-21.:** QQ-Plot de los Log retornos

Se realiza la prueba Shapiro-Wilks con el fin de evaluar la normalidad de los log retornos de cada índice, partiendo como  $H_0$ : los log retornos se distribuyen normal, versus  $H_1$ : los log retornos no se distribuyen normal.

En la tabla

**Tabla 5-3.:** Prueba Shapiro-Wilk

	SP500	FSTOK10	CRBQX
W	0.97181	0.93971	0.96694
Valor-p	0.6097	0.09855	0.48

## 5.2. Análisis Bayesiano

Através de lo expuesto en el marco teórico, se postula como previa la distribución no informativa de jeffreys, la cual brinda como posterior la distribución gamma inversa

Se presenta los resultados de las estimaciones de la distribución posterior para los índices en cuestión.

### 5.2.1. Índice S&P500

En la figura 5.22 muestra la densidad de la posterior, demarcando los limites de región de credibilidad al 95 %, evidenciando en donde se encuentra el verdadero valor del parámetro.

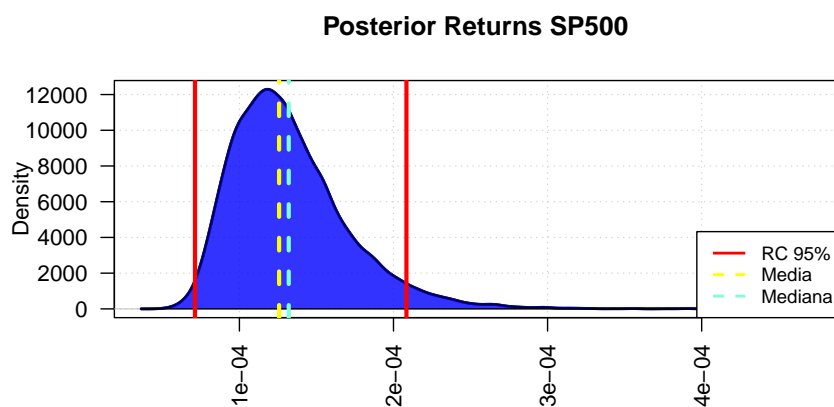


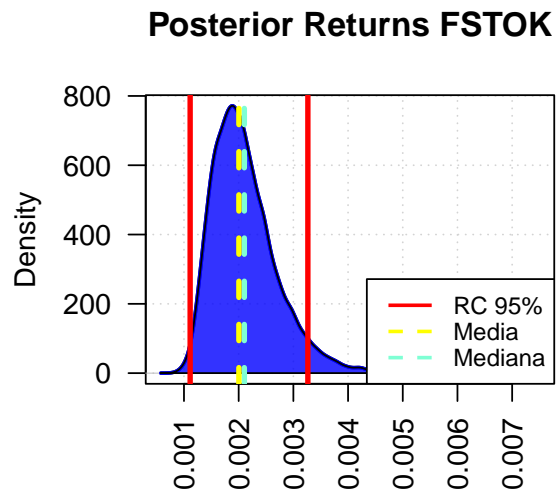
Figura 5-22.: Grafico posterior

En la tabla 5.4 se observa las estadísticas descriptivas para la estimación de la varianza a través de la distribución posterior, donde

**Tabla 5-4.:** Estadísticas Descriptivas

Estadístico	Estimación
Media	0.000132
Mediana	0.000126
Min	0.000055
Max	0.00 0413
Var	0.000000
Sd	0.000037
1st Qu.	0.000106
3rd Qu	0.000151
Coef.Var	0.280375
Error Estandar	0.0000003701

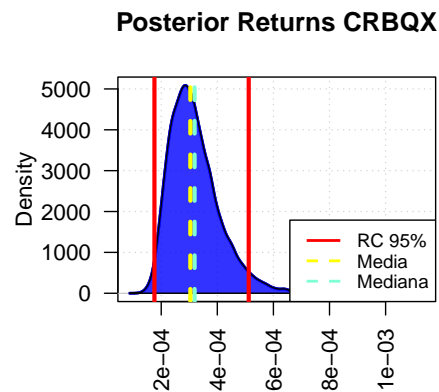
### 5.2.2. Índice FSTOK10

**Figura 5-23.:** Gráfico posterior FSTOK10

**Tabla 5-5.:** Estadísticas Descriptivas

Estadístico	Estimación
Media	0.002104
Mediana	0.002004
Min	0.000873
Max	0.006588
Var	0.000000
Sd	0.000590
1st Qu.	0.001687
3rd Qu	0.002406
Coef.Var	0.280375
Error Estandar	0.00000589779

### 5.2.3. Índice CRBQX

**Figura 5-24.:** Gráfico posterior FSTOK10

**Tabla 5-6.:** Estadísticas Descriptivas

Estadístico	Estimación
Media	0.000319
Mediana	0.000304
Min	0.000132
Max	0.000999
Var	0.000000
Sd	0.000089
1st Qu.	0.000256
3rd Qu	0.000365
Coef.Var	0.280375
Error Estandar	0.000000894427

## 6. Conclusiones y recomendaciones

### 6.1. Conclusiones

### 6.2. Recomendaciones

# Bibliografía

- Box, G. E. and Tiao, G. C. (2011). *Bayesian inference in statistical analysis*. John Wiley & Sons.
- Canavos, G. (1991). *Probabilidad y estadística: aplicaciones y métodos*. McGraw-Hill.
- Christensen R, C., Christensen, R., Johnson, W., Branscum, A., and Hanson, T. (2012). Bayesian ideas and data analysis.
- Gamerman, D. and Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition*.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- Hoff, P. D. U. o. W. (2009). *Springer Texts in Statistics: A First Course in Bayesian Statistical Methods*, volume 173.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Cryptography Mailing list at <https://metzdowd.com>*.
- Prieto, V. H., Quintero, C., and Rodríguez, I. (1995). Análisis de bayes empírico mediante un ejemplo. *Revista Colombiana de Estadística*, 16(31-32).
- Rojas, H. (2010). *Teoría estadística: aplicaciones y métodos*. Hugo Andrés Gutiérrez Rojas.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639.
- Tovar Cuevas, J. (2015). Inferencia Bayesiana e Investigación en salud: un caso de aplicación en diagnóstico clínico. *Revista Médica de Risaralda*, 21(1):9–16.
- Tsay, R. S. (2005). *Analysis of financial time series*. John wiley & sons.

A. Anexo: