



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Estudios de Postgrado  
Maestría en Estadística Aplicada

**ANÁLISIS DE LOS PATRONES ESTADÍSTICOS DE LAS CRIPTOMONEDAS DE MAYOR  
VALOR BURSÁTIL, PARA OPTIMIZAR EL RENDIMIENTO FINANCIERO RESPECTO AL  
INDICADOR DE RETORNO SOBRE LA INVERSIÓN (ROI)**

**Ing. Fabio Andrés Sandoval Girón**

Asesorado por el Mtro. William Eduardo Fagiani Cruz

Guatemala, octubre de 2021



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**ANÁLISIS DE LOS PATRONES ESTADÍSTICOS DE LAS CRIPTOMONEDAS DE MAYOR  
VALOR BURSÁTIL, PARA OPTIMIZAR EL RENDIMIENTO FINANCIERO RESPECTO AL  
INDICADOR DE RETORNO SOBRE LA INVERSIÓN**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA  
POR

**ING. FABIO ANDRÉS SANDOVAL GIRÓN**

ASESORADO POR EL MTRO. WILLIAM EDUARDO FAGIANI CRUZ

AL CONFERÍRSELE EL TÍTULO DE

**MAESTRO EN ESTADÍSTICA APLICADA**

GUATEMALA, OCTUBRE DE 2021



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Mtro. Edgar Dario Álvarez Cotí
EXAMINADOR	Mtro. Edwin Adalberto Bracamonte Orozco
EXAMINADOR	Mtro. Elmer Humberto Lemus Flores
SECRETARIO	Ing. Hugo Humberto Rivera Pérez



## HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**ANÁLISIS DE LOS PATRONES ESTADÍSTICOS DE LAS CRIPTOMONEDAS DE MAYOR VALOR BURSÁTIL, PARA OPTIMIZAR EL RENDIMIENTO FINANCIERO RESPECTO AL INDICADOR DE RETORNO SOBRE LA INVERSIÓN (ROI)**

Tema que me fuera aprobado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 23 de mayo de 2020.



Fabio Andrés Sandoval Girón



DTG. 571.2021

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Postgrado, al Trabajo de Graduación titulado: **ANÁLISIS DE LOS PATRONES ESTADÍSTICOS DE LAS CRIPTOMONEDAS DE MAYOR VALOR BURSÁTIL, PARA OPTIMIZAR EL RENDIMIENTO FINANCIERO RESPECTO AL INDICADOR DE RETORNO SOBRE LA INVERSIÓN (ROI)**, presentado por el **Ingeniero Fabio Andrés Sandoval Girón**, estudiante de la **Maestría en Estadística Aplicada**, y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

Inga. Anabela Cordova Estrada  
Decana



Guatemala, octubre de 2021.

AACE/cc



**Guatemala, octubre de 2021**

LNG.EEP.OI.067.2021

En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

**“ANÁLISIS DE LOS PATRONES ESTADÍSTICOS DE LAS  
CRIPTOMONEDAS DE MAYOR VALOR BURSÁTIL, PARA OPTIMIZAR  
EL RENDIMIENTO FINANCIERO RESPECTO AL INDICADOR DE  
RETORNO SOBRE LA INVERSIÓN (ROI)”**

presentado por **Fabio Andrés Sandoval Girón** quien se identifica con carné **201314415** correspondiente al programa de **Maestría en artes en Estadística aplicada** ; apruebo y autorizo el mismo.

Atentamente,

*“Id y Enseñad a Todas”*

  
**Mtro. Ing. Edgar Darío Álvarez Cotí**  
Director



**Escuela de Estudios de Postgrado  
Facultad de Ingeniería**



Guatemala 22 de marzo 2021.

**M.A. Edgar Darío Álvarez Cotí**  
**Director**  
**Escuela de Estudios de Postgrado**  
**Presente**

**M.A. Ingeniero Álvarez Cotí:**

Por este medio informo que he revisado y aprobado el Informe Final del trabajo de graduación titulado “ANÁLISIS DE LOS PATRONES ESTADÍSTICOS DE LAS CRIPTOMONEDAS DE MAYOR VALOR BURSÁTIL, PARA OPTIMIZAR EL RENDIMIENTO FINANCIERO RESPECTO AL INDICADOR DE RETORNO SOBRE LA INVERSIÓN (ROI)” del estudiante Fabio Andrés Sandoval Girón quien se identifica con número de carné 201314415 del programa de Maestría en Estadística Aplicada

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

Atentamente,

**MSc. Ing. Edwin Adalberto Bracamonte Orozco**  
**Coordinador**  
**Maestría en Estadística Aplicada**  
**Escuela de Estudios de Postgrado**

Guatemala, 06 de febrero de 2021.

Mtro. Ing. Edgar Darío Álvarez Cotí.  
Director de la Escuela de Estudios de Postgrado. FIUSAC.  
Presente.

Estimado Maestro Álvarez Cotí:

De la manera más atenta me dirijo a usted, esperando que sus actividades se desarrollen con éxito, y a la vez, hago de su conocimiento que el maestrando, **Fabio Andrés Sandoval Girón**, estudiante de la Maestría en Estadística Aplicada, quien se identifica con carné número **2013 14415**, me ha presentado el informe final de su trabajo de graduación titulado **“ANÁLISIS DE LOS PATRONES ESTADÍSTICOS DE LAS CRIPTOMONEDAS DE MAYOR VALOR BURSÁTIL, PARA OPTIMIZAR EL RENDIMIENTO FINANCIERO RESPECTO AL INDICADOR DE RETORNO SOBRE LA INVERSIÓN (ROI)”**.

Luego de revisar el informe que contiene los resultados de la investigación, manifiesto que le doy mi aprobación.

Sin otro particular, me suscribo a sus respetables órdenes.

Atentamente,



William Eduardo Fagiani Cruz  
Maestro en Artes en Estadística Aplicada

**INGENIERO QUÍMICO**  
**WILLIAM EDUARDO FAGIANI CRUZ**  
**COL. 1734**

## **ACTO QUE DEDICO A:**

### **Dios**

Por la bendición de poder concluir objetivos, plantear nuevos y darles continuidad a mis metas.

### **Mi madre**

Angélica Sandoval, por todo su apoyo incondicional, amor, motivación y estima. Por demostrarme con acciones que no existe brecha que no se pueda superar, su ejemplo siempre será mi inspiración.

### **Mis abuelos**

Salvador Sandoval (q. d. e. p) y Ofelia Girón, siempre tendrán un espacio especial en mi corazón.

### **Mi familia**

Por creer siempre en mí y ser de bendición para mi vida.

### **Mi novia**

Rosmin Marroquín, por todo su amor, amistad, apoyo y motivación, gracias por inspirarme a buscar la excelencia siempre.

### **Mis mentores**

Catedráticos, amigos, líderes y demás que han contribuido a mi formación académica, personal y profesional.



## **AGRADECIMIENTOS A:**

<b>Universidad de San Carlos de Guatemala</b>	Por ser mi casa de estudios y seguirme abriendo las puertas al desarrollo profesional y a la mejora continua.
<b>Escuela de Estudios de Posgrado</b>	Por la oportunidad de seguirme formando en el ámbito de mi interés.
<b>Compañero de clase</b>	Diego Cuté, por todos los momentos vividos, además de la invaluable amistad.
<b>Mis amigos</b>	Por su apoyo incondicional en cada decisión tomada, además de su estima.
<b>Docente de maestría</b>	Dr. Mayra Castillo, por su gran apoyo durante los estudios, además de compartir sus conocimientos.
<b>Coordinador de maestría</b>	Ing. Edwin Bracamonte, por su apoyo incondicional durante todo el programa, además de compartir su experiencia en el área.
<b>Mi asesor</b>	Ing. William Fagiani, por su invaluable apoyo en el desarrollo de esta investigación.



## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	VII
LISTA DE SÍMBOLOS .....	XIII
GLOSARIO .....	XV
RESUMEN.....	XVII
PLANTEAMIENTO DEL PROBLEMA.....	XIX
OBJETIVOS.....	XXIII
RESUMEN DEL MARCO METODOLÓGICO .....	XXV
INTRODUCCIÓN.....	XXXVII
1. MARCO REFERENCIAL.....	1
1.1. Marco contextual .....	9
2. MARCO TEÓRICO.....	11
2.1. Fundamentos estadísticos.....	11
2.1.1. Intervalos de confianza.....	11
2.1.2. Pruebas de hipótesis .....	14
2.1.3. Prueba de hipótesis para la media de la población. ....	17
2.1.4. Prueba de hipótesis para la diferencia de medidas de dos poblaciones.....	18
2.1.5. Análisis de varianza.....	20
2.1.5.1. Análisis de varianza de un factor ...	20
2.1.6. Regresión lineal.....	23
2.1.7. Regresión múltiple .....	23
2.1.8. Análisis de residuos.....	24
2.1.8.1. Independencia .....	25

	2.1.8.2.	Homocedasticidad.....	25
	2.1.8.3.	Normalidad.....	26
	2.1.8.4.	Prueba de Kolmogorov Smirnov....	27
	2.1.8.5.	Multicolinealidad.....	28
	2.1.8.6.	Coefficiente de determinación $R^2$ ...	29
2.1.9.		Series de tiempo .....	29
	2.1.9.1.	Componentes de una serie de tiempo.....	30
	2.1.9.2.	Prueba de Man Kendall para detectar tendencias.....	32
	2.1.9.3.	Prueba de estacionalidad de Dickey Fuller .....	33
	2.1.9.4.	Modelos de series de tiempo Autorregresivos integrados de medias móviles (ARIMA).....	34
	2.1.9.5.	Planteamiento del modelo ARIMA.	35
2.1.10.		Modelos de vector autorregresivo para series de tiempo multivariantes .....	39
	2.1.10.1.	Modelo de vector autorregresivo estacionario.....	40
	2.1.10.2.	Funciones de impulso respuesta...	42
2.1.11.		Modelos de aprendizaje de máquina.....	44
	2.1.11.1.	Tipos de problemas y tareas .....	44
	2.1.11.2.	Aprendizaje supervisado .....	45
	2.1.11.3.	Aprendizaje no supervisado .....	45
	2.1.11.4.	Algoritmos de clasificación .....	46
	2.1.11.5.	Árboles de clasificación .....	46
2.2.		Criptodivisas.....	47
	2.2.1.	Cadena de bloques .....	48

2.2.2.	Activos virtuales considerados en el estudio ....	49
2.2.2.1.	<i>Bitcoin</i> .....	49
2.2.2.2.	<i>Dash</i> .....	50
2.2.2.3.	<i>Litecoin</i> (LTC) .....	50
2.2.2.4.	<i>Monero</i> (XMR) .....	51
2.2.2.5.	<i>Ripple</i> .....	51
2.2.2.6.	<i>Ethereum</i> .....	52
2.2.3.	Métricas de red en la cadena de bloques .....	52
2.2.4.	Ratio de retorno sobre la inversión .....	54
2.2.5.	Análisis de valor en riesgo VaR .....	55
2.2.5.1.	Definición matemática del valor en riesgo VaR.....	56
3.	PRESENTACIÓN DE RESULTADOS .....	59
3.1.	Caracterización estadística de las criptodivisas estudiadas .....	59
3.1.1.	Estadísticas descriptivas de las divisas .....	59
3.1.1.1.	<i>Bitcoin</i> .....	59
3.1.1.2.	<i>Dash</i> .....	62
3.1.1.3.	<i>Ethereum</i> .....	65
3.1.1.4.	<i>Litecoin</i> .....	68
3.1.1.5.	<i>Monero</i> .....	71
3.1.1.6.	<i>Ripple</i> .....	74
3.1.2.	Caracterización de la tendencia.....	77
3.1.2.1.	Comparativa de la tendencia por criptodivisa.....	78
3.1.2.2.	Análisis de correlación en las criptodivisas .....	85
3.1.3.	Modelos de vector autorregresivo.....	87

	3.1.3.1.	Análisis de supuestos del modelo .92	
	3.1.3.2.	Análisis de impulso respuesta .....94	
	3.1.3.2.1.	Descomposición de la varianza del modelo ..... 103	
3.2.		Análisis de varianza ..... 105	
	3.2.1.	Análisis de varianza de varios factores (MANOVA) ..... 112	
	3.2.2.	Análisis de varianza para los distintos mercados de capital ..... 114	
3.3.		Análisis de series temporales..... 118	
	3.3.1.	Análisis de estacionalidad de la serie..... 129	
	3.3.2.	Diferencias estacionales de las series ..... 140	
	3.3.3.	Diseño de un modelo generalizado ..... 146	
	3.3.3.1.	Definición de los coeficientes del Modelo ARIMA ..... 149	
	3.3.3.2.	Aplicación del modelo sobre serie Generalizada ..... 151	
	3.3.3.3.	Resumen del modelo planteado.. 153	
	3.3.4.	Planteamiento de un modelo ARIMA por medio de aprendizaje de máquina ..... 158	
	3.3.4.1.	Comparativa entre ambos modelos..... 166	
	3.3.5.	Análisis predictivo de series temporales..... 168	
	3.3.5.1.	Análisis de valor en riesgo para <i>bitcoin</i> ..... 171	
	3.3.5.2.	Análisis de valor en riesgo para <i>ethereum</i> ..... 176	

3.3.5.3.	Análisis del valor en riesgo para <i>dash</i> .....	181
3.3.5.4.	Análisis del valor en riesgo para <i>litecoin</i> .....	185
3.3.5.5.	Análisis del valor en riesgo para <i>monero</i> .....	190
3.3.5.6.	Análisis del valor en riesgo para <i>ripple</i> .....	194
4.	DISCUSIÓN DE RESULTADOS .....	197
4.1.	Descripción de la tendencia de las criptodivisas .....	197
4.1.1.	Análisis de correlación de las divisas .....	198
4.1.1.1.	Análisis del modelo de vector autorregresivo para las seis divisas.....	200
4.2.	Comparativa del precio bursátil de las divisas.....	202
4.2.1.	ANOVA para factor de divisas y para el criterio de relación bursátil.....	202
4.2.2.	ANOVA para distintos mercados de capital....	203
4.3.	Modelo matemático que describe a las criptodivisas.....	204
4.3.1.	Modelo de series de tiempo generado.....	206
4.4.	Modelo de eficiencia financiera .....	208
4.4.1.	Rentabilidad esperada por predicción .....	209
4.4.1.1.	Rendimiento <i>bitcoin</i> .....	210
4.4.1.2.	Rendimiento <i>ethereum</i> .....	210
4.4.1.3.	Rendimiento <i>dash</i> .....	211
4.4.1.4.	Rendimiento <i>litecoin</i> .....	211
4.4.1.5.	Rendimiento <i>monero</i> .....	212
4.4.1.6.	Rendimiento <i>ripple</i> .....	212

4.4.2. Modelo de retorno sobre la inversión .....213

CONCLUSIONES.....217

RECOMENDACIONES .....219

REFERENCIAS .....221

## ÍNDICE DE ILUSTRACIONES

### FIGURAS

1.	Gráfico de normalidad con un 5 % de significancia.....	13
2.	Intervalos para el valor medio de apertura de <i>bitcoin</i> .....	60
3.	Intervalos para el valor medio de cierre de <i>bitcoin</i> .....	62
4.	Intervalos para el valor medio de apertura de <i>dash</i> .....	63
5.	Intervalos para el valor medio de cierre de <i>dash</i> .....	65
6.	Intervalos para el valor medio de apertura de <i>ethereum</i> .....	66
7.	Intervalos para el valor medio de cierre de <i>ethereum</i> .....	68
8.	Intervalo para el valor medio de apertura de <i>litecoin</i> .....	69
9.	Intervalo para el valor medio de cierre de <i>litecoin</i> .....	71
10.	Intervalo para el valor medio de apertura de <i>monero</i> .....	72
11.	Intervalo para el valor medio de cierre de <i>monero</i> .....	74
12.	Intervalo para el valor medio de apertura de <i>ripple</i> .....	75
13.	Intervalo para el valor medio de cierre de <i>ripple</i> .....	77
14.	Línea de tendencia para <i>bitcoin</i> .....	78
15.	Línea de tendencia para <i>dash</i> .....	79
16.	Línea de tendencia para <i>ethereum</i> .....	80
17.	Línea de tendencia para <i>litecoin</i> .....	81
18.	Línea de tendencia para <i>monero</i> .....	82
19.	Línea de tendencia para <i>ripple</i> .....	83
20.	Sobreposición de las dispersiones en cada criptodivisa.....	87
21.	Comportamiento de las seis criptomonedas.....	89
22.	Series de tiempo desestacionalizadas.....	90
23.	Residuos del modelo VaR para <i>bitcoin</i> .....	93

24.	Impulso-respuesta ortogonal de <i>bitcoin</i> .....	95
25.	Impulso-respuesta ortogonal de <i>ethereum</i> .....	96
26.	Impulso-respuesta ortogonal de <i>dash</i> .....	98
27.	Impulso-respuesta ortogonal de <i>litecoin</i> .....	99
28.	Impulso-respuesta ortogonal <i>monero</i> .....	101
29.	Impulso-respuesta ortogonal <i>ripple</i> .....	102
30.	Descomposición de la varianza explicada por el modelo.....	104
31.	Prueba de normalidad de residuos para el factor divisa .....	109
32.	Histograma de residuos para el factor divisa .....	109
33.	Intervalos de variación para el factor divisa .....	111
34.	Prueba de Dunnet para el factor divisa.....	112
35.	Serie de tiempo para <i>bitcoin</i> y <i>dash</i> .....	119
36.	Serie de tiempo para <i>ethereum</i> y <i>litecoin</i> .....	120
37.	Serie de tiempo para <i>monero</i> y <i>ripple</i> .....	121
38.	Componentes de la serie de tiempo de <i>bitcoin</i> .....	122
39.	Componentes de la serie de tiempo de <i>dash</i> .....	123
40.	Componentes de la serie de tiempo de <i>ethereum</i> .....	124
41.	Componentes de la serie de tiempo de <i>litecoin</i> .....	125
42.	Componentes de la serie de tiempo de <i>monero</i> .....	126
43.	Componentes de la serie de tiempo de <i>ripple</i> .....	127
44.	Transformación polar de <i>bitcoin</i> .....	130
45.	Estacionalidad interanual de <i>bitcoin</i> .....	131
46.	Transformación polar de <i>dash</i> .....	132
47.	Estacionalidad interanual de <i>dash</i> .....	133
48.	Transformación polar de <i>ethereum</i> .....	133
49.	Estacionalidad interanual de <i>ethereum</i> .....	134
50.	Transformación polar de <i>litecoin</i> .....	135
51.	Estacionalidad interanual de <i>litecoin</i> .....	135
52.	Transformación polar de <i>monero</i> .....	136

53.	Estacionalidad interanual de <i>monero</i> .....	137
54.	Transformación polar de <i>ripple</i> .....	138
55.	Estacionalidad interanual de <i>ripple</i> .....	139
56.	Diferencias estacionales ACF y PACF de <i>bitcoin</i> .....	141
57.	Diferencias estacionales ACF y PACF de <i>dash</i> .....	142
58.	Diferencias estacionales ACF y PACF de <i>ethereum</i> .....	143
59.	Diferencias estacionales ACF y PACF de <i>litecoin</i> .....	144
60.	Diferencias estacionales ACF y PACF de <i>monero</i> .....	145
61.	Diferencias estacionales ACF y PACF de <i>ripple</i> .....	146
62.	Árbol de Bayesiano para los atrasos del modelo teórico.....	150
63.	Serie generalizada de criptodivisas.....	152
64.	Residuos del modelo ARIMA teórico.....	155
65.	Prueba de normalidad para el modelo ARIMA planteado .....	156
66.	Pronóstico utilizando modelo ARIMA teórico .....	158
67.	Raíces polinomiales del modelo.....	160
68.	Visualización del algoritmo de clasificación.....	162
69.	Residuos del modelo ARIMA generado por ML .....	164
70.	Prueba de normalidad para modelo ARIMA de ML.....	165
71.	Pronóstico con modelo ARIMA (1,0,0) .....	166
72.	Pronóstico con modelo ARIMA para <i>bitcoin</i> .....	168
73.	Análisis de residuos para predicciones de <i>bitcoin</i> .....	169
74.	VaR de <i>bitcoin</i> .....	172
75.	Pronóstico con modelo ARIMA para <i>ethereum</i> .....	173
76.	Análisis de residuos para predicciones de <i>ethereum</i> .....	175
77.	VaR de <i>ethereum</i> .....	177
78.	Pronóstico con modelo ARIMA para <i>dash</i> .....	178
79.	Análisis de residuos para predicciones de <i>dash</i> .....	179
80.	VaR <i>dash</i> .....	181
81.	Pronóstico con modelo ARIMA para <i>litecoin</i> .....	182

82.	Análisis de residuos para predicciones de <i>litecoin</i> .....	184
83.	VaR <i>litecoin</i> .....	186
84.	Pronóstico del modelo ARIMA para <i>monero</i> .....	187
85.	Análisis de residuos para predicciones de <i>monero</i> .....	188
86.	VaR <i>monero</i> .....	190
87.	Pronóstico del modelo ARIMA para <i>ripple</i> .....	191
88.	Análisis de residuos de predicciones de <i>ripple</i> .....	193
89.	VaR <i>ripple</i> .....	195
90.	Distribución de pesos para la serie de modelación.....	205
91.	Serie de tiempo de ROI para cada criptomoneda.....	214
92.	Sobreposición de todas las criptomonedas en función del ROI.....	214
93.	Correlación policorica de los ROI de cada criptomoneda.....	215

## TABLAS

I.	Criptomonedas consideradas en el estudio.....	XXII
II.	Variables del estudio.....	XXVII
III.	Resumen de criterios para análisis de varianza.....	21
IV.	Estadísticas descriptivas de <i>bitcoin</i> .....	61
V.	Estadísticas descriptivas de <i>dash</i> .....	64
VI.	Estadísticas descriptivas de <i>ethereum</i> .....	67
VII.	Estadísticas descriptivas de <i>litecoin</i> .....	69
VIII.	Estadísticas descriptivas de <i>monero</i> .....	72
IX.	Estadísticas descriptivas de <i>ripple</i> .....	75
X.	Prueba de Kendall para <i>bitcoin</i> .....	79
XI.	Prueba de Kendall para <i>dash</i> .....	80
XII.	Prueba de Kendall para <i>ethereum</i> .....	81
XIII.	Prueba de Kendall para <i>litecoin</i> .....	82
XIV.	Prueba de Kendall para <i>monero</i> .....	83

XV.	Prueba de Kendall para <i>ripple</i> .....	84
XVI.	Análisis de correlación para todas las divisas con datos del mercado financiero de divisas.....	85
XVII.	Selección de atrasos para diseño del vector autorregresivo .....	91
XVIII.	Análisis de varianza para un factor .....	106
XIX.	Estadísticas descriptivas del volumen de capital total.....	107
XX.	Prueba de Tukey para las medias de las criptodivisas.....	108
XXI.	Prueba de Kolmogorov para residuales .....	108
XXII.	Análisis múltiple de varianza para el criterio de divisa .....	113
XXIII.	Análisis múltiple de varianza para el criterio de relación bursátil .	114
XXIV.	Mercados de capital para <i>bitcoin</i> .....	115
XXV.	ANOVA de <i>bitcoin</i> para mercados de capital .....	116
XXVI.	Mercados de capital para <i>dash</i> .....	116
XXVII.	ANOVA de <i>dash</i> para mercados de capital.....	116
XXVIII.	Mercados de capital para <i>ripple</i> .....	117
XXIX.	ANOVA de <i>ripple</i> para mercados de capital.....	117
XXX.	Prueba de Dickey Fuller para todas las divisas.....	140
XXXI.	Pesos para generar la base de datos generalizada .....	148
XXXII.	Resumen del modelo ARIMA planteado .....	154
XXXIII.	Prueba de Kolmogorov para modelo ARIMA teórico.....	157
XXXIV.	Resultados del modelo planteado por algoritmo .....	163
XXXV.	Métricas del aprendizaje de máquina.....	163
XXXVI.	Prueba de Kolmogorov para ARIMA generado por ML.....	166
XXXVII.	Comparativa de modelos ARIMA .....	167
XXXVIII.	Prueba de normalidad de residuos <i>bitcoin</i> .....	168
XXXIX.	Pronóstico a 24 meses de <i>bitcoin</i> con cifras en miles de millones de dólares.....	170
XL.	Resumen del modelo VaR <i>bitcoin</i> .....	172
XLI.	Prueba de normalidad de residuos de <i>ethereum</i> .....	174

XLII.	Pronóstico a 24 meses de <i>ethereum</i> con cifras en millones de dólares .....	175
XLIII.	Resumen modelo VaR <i>ethereum</i> .....	177
XLIV.	Prueba de normalidad de residuos de <i>dash</i> .....	178
XLV.	Pronóstico a 24 meses de <i>dash</i> con cifras en millones de dólares .....	180
XLVI.	Resumen del modelo VaR <i>dash</i> .....	182
XLVII.	Prueba de normalidad para <i>litecoin</i> .....	183
XLVIII.	Pronóstico a 24 meses de <i>litecoin</i> con cifras en millones de dólares.....	184
XLIX.	Resumen del Modelo VaR <i>litecoin</i> .....	186
L.	Prueba de normalidad de <i>monero</i> .....	187
LI.	Pronóstico a 24 meses de <i>monero</i> con cifras en millones de dólares .....	189
LII.	Resumen del modelo VaR <i>monero</i> .....	191
LIII.	Prueba de normalidad de <i>ripple</i> .....	192
LIV.	Pronóstico a 24 meses de <i>ripple</i> con cifras en millones de dólares .....	193
LV.	Resumen del modelo VaR <i>ripple</i> .....	195
LVI.	Asociaciones de correlación por divisa .....	199

## LISTA DE SÍMBOLOS

<b>Símbolo</b>	<b>Significado</b>
<i>L</i>	Atraso de una serie temporal
<i>C(t)</i>	Ciclo de una serie de tiempo
<i>AR</i>	Coefficiente de Autoregresión
<i>C<sub>i</sub></i>	Coefficiente de clúster
<i>SAR</i>	Coefficiente de estacionalidad
<i>β</i>	Coefficiente de regresión
<i>BIC</i>	Criterio de información Bayesiano
<i>AIC</i>	Criterio de información de Akaike
<i>ε<sub>t</sub></i>	Error de estimación
<i>S(t)</i>	Estacionalidad de una serie de tiempo
<i>F</i>	Estadístico de Fisher
<i>k<sup>-α</sup></i>	Grado de poder de la cadena de bloques
<i>GL</i>	Grados de libertad
<i>I(t)</i>	Independencia de una serie de tiempo
<i>α</i>	Nivel de significancia
<i>SCE</i>	Suma de cuadrados del error
<i>SCTR</i>	Suma de cuadrados del tratamiento
<i>T(t)</i>	Tendencia de una serie de tiempo



## GLOSARIO

<b>ACF</b>	Autocorrelación
<b>Bitc3in</b>	Criptomoneda de referencia en la comparativa de criptodivisas.
<b>Criptoactivo</b>	T3rmino financiero que acota como activos a las criptodivisas.
<b>Minitab</b>	Software comercial para el an3lisis de datos.
<b>NYSE</b>	Bolsa de valores de Nueva York
<b>PACF</b>	Autocorrelaci3n parcial
<b>RStudio</b>	Interfaz de usuario que utiliza el lenguaje de programaci3n R para el procesamiento de datos.
<b><i>Script</i></b>	Documento que contiene instrucciones escritas en c3digos de programaci3n, las cuales ejecutan diversas funciones en el interior de un lenguaje de programaci3n.
<b>SSE</b>	Bolsa de valores de Shangh3i



## RESUMEN

Las criptomonedas suponen una revolución a nivel financiero de la transaccionalidad de activos a través del mundo, además de presentar un nuevo reto para la industria financiera actual, en términos de estudio, auditoría y cuantificación de estos criptoactivos y la revolución de seguridad digital por medio del cual se comercializan, la cadena de bloques (*blockchain*).

El objetivo de la investigación es generar información con fundamentación estadística para caracterizar el comportamiento de las criptomonedas de mayor valor bursátil, y encontrar los patrones y comportamiento más significativos en su historia que permitan modelar dicho comportamiento histórico, esto mediante modelos matemáticos autorregresivos.

Se utilizó la metodología de modelado para el algoritmo de ARIMA, utilizando como referencia de optimización el criterio de información de Akaike. Se usó como variable de modelado histórico, al volumen total de capital bursátil de cada moneda.

Posteriormente por medio de un análisis de conglomerados en conjunto de pesos probabilísticos, se diseñó una serie generalizada de tiempo para poder predecir el efecto global de todas las monedas bajo un solo modelo generalizado, y con este posteriormente probar su eficiencia con cada moneda individualizada respecto a sus valores reales y los valores predichos por el modelo.

Se establece bajo el modelo planteado que los supuestos bajo los cuales se rigen económicamente las monedas se pueden probar estadísticamente, tales

como el valor homogéneo indiferente al factor país, y que el volumen de cada acción no está sujeta a caídas o regulaciones internacionales. Además, la variable que presenta el mayor efecto significativo es el volumen de capital bursátil, por ende, es la seleccionada para generar las predicciones.

El modelado es satisfactorio en la detección de la varianza característica de cada criptomoneda, además de generar predicciones eficientes respecto al valor real comparado en un conjunto de validación cruzada. Además, la tendencia en el comportamiento de la criptomoneda de mayor valor define en gran medida a la tendencia de las criptomonedas de menor valor de capital bursátil, por ende, una correlación positiva, aun proporcionalmente baja, refleja un potencial incremento en el volumen porcentual del retorno sobre la inversión, esto debido a la amplia diferencia de valor monetario, por lo que se demuestra lo valiosa que es la proyección de correlaciones entre divisas que aún están en fase de desarrollo y su relación con las divisas de mayor volumen de capital.

Finalmente, los residuos del modelo planteado presentan un comportamiento satisfactoriamente normal, además de pasar de forma correcta las pruebas estadísticas de independencia de los valores, lo que garantiza predicciones más eficientes o cercanas probabilísticamente al valor real.

## PLANTEAMIENTO DEL PROBLEMA

- Contexto general

Desde que surgió en 2008, el concepto abstracto de una moneda completamente virtual, no regulada y descentralizada, además de anónima; ha logrado que mercados de divisas, marcas internacionales de productos financieros y medios especializados en la materia, especulen sobre el origen, factores económicos y financieros involucrados en su comportamiento y diversas teorías enfocadas en la factibilidad de su uso, razón de ser y sobrevivencia en el actual mercado globalizado.

Estas divisas pasaron a categorizarse como criptodivisas, y desde el 2013 tras su abrupto crecimiento en participación de mercado y volumen de capital financiero, se convirtieron en activos intangibles de mucho atractivo para inversores de riesgo, con la expectativa de rendimientos atractivos, sin embargo, dada su naturaleza, no tienen un comportamiento económico sencillo de modelar, por lo que saber cuándo vender y cuándo comprar estas divisas se traduce en rendimientos financieros muy altos o pérdidas de capital completas.

- Descripción del problema

Las revistas financieras e instituciones que ya incluyen a las criptodivisas en sus resúmenes financieros y estados de resultados globales, describen a estas últimas bajo metodologías tradicionales determinísticas, y comparando sus alzas y bajas a otros activos, que puede y no sean la mejor referencia en cuanto a comportamiento bursátil, incluso englobando el desempeño de las criptodivisas

más apreciadas y con mayor participación de mercado, dentro de conceptos económicos como las burbujas económicas o mercados ficticios.

El riesgo extremadamente alto, y los desaciertos en los momentos de venta errónea de la divisa y compra en momentos de alzas no justificadas sugiere que una perspectiva estocástica es más adecuada para describir el comportamiento en el tiempo de las mismas, volviendo fundamental encontrar una caracterización adecuada en sus patrones estadísticos, por lo que esta investigación estudió el comportamiento de la divisa de mayor valor, así como de cinco divisas que aún conservan precios asequibles por medio de un análisis de series temporales, para el pronóstico de volumen de capital en el mercado, así como para la identificación de patrones que puedan interrelacionarlas.

- Formulación del problema

- Pregunta central

¿Cuáles son los patrones y tendencias de crecimiento y decremento del valor monetario real de las criptomonedas con mayor valor en el mercado bursátil, que afectan significativamente su precio y su retorno sobre la inversión en países que utilizan el dólar como moneda de referencia?

- Preguntas auxiliares

- ¿Qué tan significativo es el efecto del valor medio en el precio del *bitcoin*, sobre las principales criptodivisas en un horizonte temporal delimitado y en las mismas condiciones?

- ¿Cuál es la diferencia en el precio real, de las principales criptodivisas en los diferentes mercados financieros alrededor del mundo en sus denominaciones de entrada, salida y volumen de capital?
  - ¿Cuáles son los modelos estadísticos que posiblemente, generan una aproximación adecuada para las posibles decisiones de inversión en el mercado de criptodivisas?
  - ¿El riesgo que conlleva invertir activos en mercados altamente asimétricos puede ser respaldado por el retorno sobre la inversión asociado a la tendencia de crecimiento en el precio de mercado de la divisa?
- Delimitación del problema

La presente investigación se limitó a estudiar los principales mercados bursátiles, en donde se analizó el precio real de las principales criptodivisas en los últimos cuatro años, ya que es la ventana de tiempo utilizada para obtener una aproximación lo más homogénea posible en tiempo de existencia de las divisas consideradas. Únicamente se consideraron mercados financieros que utilizan el dólar como moneda de referencia, ya que los indicadores financieros brindan resultados más eficientes utilizando comparaciones en USD (acrónimo en inglés para dólar de los Estados Unidos).

Tabla I. **Criptodivisas consideradas en el estudio**

<b>Criptodivisa</b>	<b>Abreviación en mercados de capital</b>
<i>Bitcoin</i>	BTC
<i>Dash</i>	DASH
<i>Litecoin</i>	LTC
<i>Ethereum</i>	ETH
<i>Ripple</i>	XRP
<i>Monero</i>	XRM

Fuente: elaboración propia.

Mercados de capital evaluados:

- Bolsa de valores de Nueva York. Estados Unidos de Norteamérica.
- Bolsa de valores de Londres. Reino Unido.
- Bolsa de valores de Tokio. Japón.

## OBJETIVOS

- **General**

Describir el comportamiento estadístico del precio de mercado de las criptomonedas con mayor valor financiero, por medio de un análisis del comportamiento de esta variable en función de sus tendencias, ciclos y estacionalidades, utilizando modelos de series de tiempo.

- **Específicos**

- Describir la tendencia del valor monetario medio de bitcoin, así como el efecto e influencia de esta sobre el resto de criptodivisas estudiadas, por medio de un análisis de correlación en sus comportamientos a través del tiempo.
- Comparar el precio bursátil real de las criptodivisas estudiadas, utilizando los datos de precios de entrada, salida y volumen de capital, en los diferentes mercados financieros, mediante un análisis de varianza.
- Formular los modelos matemáticos que mejor describan los comportamientos históricos de cada divisa, por medio de algoritmos autorregresivos en conjunto con algoritmos iterativos de aprendizaje de máquina.

- Identificar cuáles criptomonedas podrían generar un mejor retorno sobre la inversión y están asociadas a un menor riesgo, por medio de un análisis predictivo de series temporales.

## RESUMEN DEL MARCO METODOLÓGICO

El enfoque del estudio realizado se basa en un enfoque cuantitativo, ya que las variables a medir y controlar corresponden a valores monetarios reales, asociados a las variaciones bursátiles en el precio de mercado de diferentes monedas electrónicas y sus diferentes factores económicos que desencadenan una fluctuación en el precio.

El alcance del estudio está acotado como descriptivo y correlacional, ya que se comprende de un análisis de la información histórica, que engloba datos asociados a diferentes precios del mercado bursátil de un periodo de tiempo específico, de las divisas del *blockchain* (sistema de interacción de criptodivisas) de mayor valor. Por lo que estas variaciones y sus asociaciones financieras, se transforman en una variable aleatoria continua, sujeta a un amplio alcance de tipo asociativo, el cual se utilizó como base para realizar predicciones.

El diseño utilizado en esta investigación es un diseño no experimental, del tipo analítico, ya que, a raíz de información bursátil histórica de la variable analizada; se planea hacer una serie de comparaciones en función de varias características económicas y financieras de las que derivan factores de interés que pueden ser medidos, esto permitió realizar una cuantificación de indicadores de salida del sistema, así también niveles de correlación y variabilidad, determinando las condiciones más probables de comportamiento en el mercado de divisas.

La observación de las variables de interés se realizó por medio una selección múltiple de distintas criptodivisas con ponderaciones significativas en

términos de valor de mercado, con las que se calcularon distintos indicadores de rendimiento monetario y con ello por medio de ponderaciones (*scores* en inglés) asociados a un mejor ajuste de correlación entre variables.

Se identificaron los horizontes temporales más significativos del estudio en el mercado bursátil, acorde a la homogeneidad de periodos entre las divisas, utilizando como referencia al menos tres mercados de capital en el mundo, con ello se minimizó el grado de incertidumbre de predicción.

- Unidades de análisis

Cada unidad de análisis es un estado del volumen de capital (en unidades monetarias) de las principales criptodivisas estudiadas, a su vez en el conjunto de situaciones económico-temporales involucradas.

Dado que las variables de medición y comparación están comprendidas por números reales, la población puede considerarse infinita, esto es así ya que considerando la cantidad de asociaciones posibles de todas las características de interés de una sola operación bursátil en un día, que comprenda o no la variable estudiada, y a su vez que esa misma iteración de compra o venta tenga influencia sobre alguna otra variable, genera un número infinito o no mesurable de combinaciones posibles, y un efecto en cadena sobre todas las variables, dados los distintos factores estudiados.

Para definir parámetros de comparación es necesario considerar una referencia; y una segmentación acorde a las características de la variable interés; por lo que diferentes asociaciones y combinaciones de precio de compra y rentabilidad generada, es el parámetro de la colección de datos generada en cada horizonte temporal.

- Variables

Tabla II. **Variables del estudio**

<b>Nombre de la variable</b>	<b>Definición teórica</b>	<b>Definición operativa</b>
Precio en el mercado bursátil. ( <i>P</i> )	Definición literal del precio o valoración de una acción o título económico con el objetivo de determinar su valor en el mercado. Es decir, el precio con el que se ha realizado la compraventa de un valor mobiliario, en una bolsa operativa o de valores (divisas), en un mercado local o en una moneda local o extranjera.	Número real, en unidades monetarias (papel moneda por territorio) que está sujeta a la variabilidad existente en un mercado económico.
Rentabilidad Económica. ( <i>ROI</i> )	Razón financiera que compara el beneficio o la utilidad obtenida en relación a la inversión realizada. Es decir, está asociado a una razón de análisis en el rendimiento que la empresa tiene desde el punto de vista financiero.	Número real. Adimensional. Se calcula a través de un operador de razón en donde el numerador puede admitir distintas definiciones de beneficios netos después de impuestos, o incluso un beneficio global.
Precio ajustado de cierre. ( <i>AdjC</i> )	Referencia con la que cierra un activo su precio, dentro del parámetro comprendido por un día en el mercado bursátil. Refleja el precio del activo al cierre del día para considerar la venta de acciones de lo que esté representando. Usualmente es considerado que es el valor de retorno real de un activo después de todas las variaciones en su precio, dadas las acciones del mercado de divisas del día.	Número real, asociado al valor asociado a un activo al cierre del mercado bursátil de ese día.

## Continuación tabla II.

---

Volumen del valor de mercado ( $V$ )	En cualquier día de negociación de un mercado bursátil, el precio del volumen total de acciones (redondeado al 1/10,000 más cercano por acción) de una acción total de las acciones ordinarias del mercado <i>Pinnacle</i> en la bolsa de valores de Nueva York.	Número real. Asociado al precio del volumen total de acciones de todos los atributos que componen a la variable estudiada.
--------------------------------------	--	--

---

Fuente: elaboración propia.

- Fases de estudio
  - Revisión de la literatura
    - Fundamentos de la informática e infraestructura asociada a la cadena de bloques

Se definió por medio de ejemplificaciones prácticas, asociaciones gráficas, ilustraciones y ejercicios resueltos el cómo funciona la conformación de las cadenas de bloques para la creación de criptodivisas. Para ello se estudiaron las relaciones de nodos por las que se conforma en todas las conexiones de servidores que son usados para tal fin, cómo operan las transacciones, y por ende los países en dónde es más frecuente analizar este tipo de relación.

- Fundamentos financieros

Se realizó una evaluación completa de todos los indicadores financieros que podrían intervenir como posibles ratios explicativos, respecto al comportamiento

en el mercado bursátil al que están sujetas las criptodivisas, como apoyo para el estudio, esto para la selección de los indicadores financieros que reflejen de forma más clara la posición económica y financiera de cada divisa, para lograr una mejor comparación respecto a su rentabilidad, y lograr mejores aproximaciones a nivel de inversión.

- Fundamentos estadísticos

Se realizó una evaluación de todos los conceptos asociados a las series de tiempo, modelos de asociación y comparación de variables, con el fin de encontrar un diseño con los recursos adecuados para cada etapa del análisis de la información.

- Obtención de la información

- Se consultó los distintos mercados financieros para identificar en qué rubros se encuentran las divisas a estudiar.

- ✓ Se identificó la disponibilidad en función del registro histórico de la información.

- ✓ Identificación de los distintos mercados financieros que guarden información histórica de libre consulta, en los que se pueda solicitar las bases de datos en relación al USD dólar. Para poder tener distintos parámetros de comparación de la información.

- ❖ Se solicitó la información del periodo de enero de 2014 a diciembre de 2019 para cada

criptodivisa, en función de las posibilidades económicas, y minimizando el efecto del tamaño de las bases de datos, se solicitarán tantas y como sean posibles.

- ❖ Solicitud, familiarización y adecuación del equipo de minería de datos. Dada la cantidad de datos a estudiar se alquiló un equipo de minería de datos, para generar una base de datos que pueda ser trabajable en un ordenador de Hardware regular, para ello debe configurarse los procesadores a las frecuencias adecuadas por el software a utilizar, así mismo el voltaje de las memorias RAM y la configuración de hilos del procesador del equipo, que trabaja bajo la plataforma de Windows Server 2012.

- Análisis y adecuación de las bases de datos obtenidas

Por medio del equipo de minería de datos, el cual es un ordenador portátil con las adecuaciones necesarias (MSI, GT80 Titan SLI), se analizaron las bases de datos, con las cuales se llevó a cabo el estudio global y descriptivo de la información, así como una limpieza en general de valores atípicos o faltantes. Dado que la información brindada por los sitios de mercados bursátiles estaba comprendida por cada variación a la que han estado sujetas las divisas estudiadas en un periodo diario, tenía características atípicas regulares de errores en motores de bases de datos, por lo que se realizó una ingeniería de

características de la información, además se realizaron las correcciones, transformaciones y adecuaciones necesarias para facilitar su estudio.

- ✓ Se aplicó un estudio a escala probabilística de la información, para inferir a qué distribución de probabilidad están sujetos los datos, si se puede conseguir una aproximación a la normal y bajo qué criterios de estudio puede facilitar su análisis posterior.
- ✓ Una vez analizada la fracción de datos se procedió a analizar el comportamiento de la distribución que describen por medio de un gráfico qq (cuantil-cuantil). Dadas muestras tan grandes de información y la ley de los grandes números, se concluyó normalidad en los datos, por lo que se probó la validez de pruebas paramétricas para su caracterización, fuera de los puntos extremos que se trataron por medio de las distribuciones de mejor ajuste, y en los casos no justificables fueron removidos de la base de datos.

- Exportar las nuevas bases de datos

Dado el volumen de información, se utilizó el equipo de minería de datos. Una vez analizadas las bases de datos, se procedió a realizar una limpieza de los datos, en donde existían datos extremos que conseguían que la variación de la serie sea demasiado significativa, así también en el caso el factor de inflación de varianza se encontraba inflado de forma artificial, para la corrección se utilizaron múltiples valores de referencia como el VIF (factor de inflación de la

varianza) y el criterio de información bayesiana. Posteriormente, ya segmentadas se trasladaron las bases de datos vía un medio de transferencia físico, por medio de un disco duro externo, la transferencia se realizó al equipo en dónde se desarrollaron las pruebas finales del estudio, que es un equipo de computación convencional.

- Análisis de la información
  - Análisis de las estadísticas descriptivas

Se caracterizaron descriptivamente a todas las criptodivisas. Para tener una referencia de cálculo, asociados a los test y los estadígrafos de prueba a los cuales se sometió a la información. Se calcularon todos los parámetros de centralización, posición y dispersión, además de los intervalos de confianza para el precio muestral por mes de cada divisa, a su vez se presentaron gráficamente la amplitud de estos intervalos para la media mensual del volumen de capital de cada divisa.

- Análisis de las series temporales que describen el fenómeno

Mediante un análisis gráfico de las series temporales que se forman a raíz de las bases de datos generadas para cada una de las criptodivisas, se obtuvieron conclusiones asociadas a su comportamiento y del mismo modo.

- ✓ Se segmentó cada serie de tiempo en función de su tendencia, su estacionalidad, ciclo y aleatoriedad para evaluar los patrones más significativos que describen el comportamiento de cada divisa.

- ✓ Se realizó una desestacionalización de la serie de tiempo, para analizar de mejor forma su comportamiento y encontrar patrones estadísticamente significativos en su tendencia.
- ✓ Se analizaron las distribuciones de los puntos extremos en las series en dónde existían, para valorar por medio de las pruebas de Fréchet y Gumbel si esos datos describían de alguna forma al fenómeno o era conveniente removerlos del conjunto.

- Análisis de correlación

Por medio de un análisis de dispersión, se compararon todas las asociaciones y combinaciones posibles de datos  $x \subset n$  con el fin de encontrar cuáles de las variables presentaban una mayor correlación, de esta forma se encontraron las mejores asociaciones posibles, posteriormente se sometieron las más consistentes a un análisis de varianza por medio de la distribución  $F$  de Fisher. Con ello se pudo probar el precepto de multicolinealidad en las variables y descartar aquellas que inflaban de forma artificial la varianza de las comparaciones, con ello se pudo encontrar el mejor modelo de ajuste de grado  $n$ .

- Análisis de varianza de un factor

Por medio de un análisis de comparación de medias utilizando de igual forma la distribución  $F$  de Fisher, se sometió a prueba la siguiente hipótesis a un 95 % de confianza:

$H_0$ : las medias de los precios asociados a las criptodivisas a través del tiempo provienen de la misma población, es decir son iguales.

$H_A$ : al menos un valor promedio proviene de una población distinta.

- Pruebas post-hoc para un factor

Para contrastar la hipótesis del inciso anterior se realizaron las siguientes pruebas, validando el supuesto de igualdad de varianzas, utilizando como referencia un análisis gráfico.

- Comparaciones en parejas de Tukey.
  - Método de la diferencia mínima de Fisher.
  - Prueba del rango mínimo (Duncan).
- Construcción de modelos de regresión (simple y múltiple)

Se construyó un modelo de regresión simple. Considerando el valor de cierre del día de cada divisa, además del tiempo transcurrido para conseguir al menos una variación en un momento cualquiera de la serie, incluyendo el error medio, dado el ajuste del mercado bursátil, ya que no es posible dadas las limitaciones actuales de información, el considerar todas las variaciones posibles de un día en la cadena de bloques.

Adicional, se buscó la transformación que presentó el mejor ajuste lineal, para describir matemáticamente el fenómeno y considerarlo como el modelo apto para realizar predicciones.

Además, se consideró la construcción de un modelo multivariado en dónde se tomó en consideración el factor de dos variables, el precio de entrada de la divisa, y la composición del volumen del precio total que conforma a la divisa, todo esto en sus iteraciones dentro del espectro temporal considerado. Sin embargo, dada la variabilidad de ambos valores, el modelo no presentó un ajuste aceptable en cuanto a su coeficiente de determinación y fue descartado.

- Interpretación de la información

Dados todos los resultados obtenidos, se procedió a realizar un análisis completo mostrar evidencia de logro de los objetivos propuestos para la investigación, Por medio de la asociación de los hallazgos estadísticos con las fluctuaciones en el valor-precio de cada divisa, así como la manera más eficiente de realizar predicciones utilizando el modelo matemático de series temporales generalizado encontrado.

- Elaboración del informe final



## INTRODUCCIÓN

Desde los comienzos de la web comercial, y la descentralización de dominios de Internet, se comenzaron a introducir métodos de pago electrónicos no centralizados por entidades financieras. Estos últimos manejan la idea de evitar los costos asociados a transacciones, tasas y tarifas bancarias además de intereses internacionales, aprovechando la baja de barreras que ofrecía Internet. Originalmente todo el concepto de pago electrónico sin tasas era muy teórico e ideal, ya que la mayor parte de proveedores de esta clase de servicio se valían de ser intermediarios para las compras y al final cobrar una tasa.

Esto ocurrió hasta que el modelo de las criptodivisas fue introducido; el cual propone un esquema digital en donde dos o más partes tienen autorización de realizar transacciones abiertas con otros, sin la necesidad de un tercero como ente intermediario; ya sea en transferencia bancaria, cambio de divisa, entre otros.

Por lo que, de manera piloto por medio de un artículo académico publicado en un sitio web de criptografía (*The Cryptography Mailing List*), surge *bitcoin* en el mercado, como la primera criptomoneda propuesta a principios del año 2008, para que sorprendentemente toda esta propuesta de cadena de bloques de transacción descentralizadas, fuese ampliamente reconocida como una nueva plataforma de distribución de rentas por parte de muchas instituciones internacionales, incluso de las instituciones financieras, admitiendo la impotencia de controlar o centralizar esta clase de divisas.

Para comenzar a dar contexto, las criptodivisas utilizan un modelo de encriptación especial de bloques de información para generar el denominado activo, estos a grandes rasgos son archivos computacionales a los que les fue imputado un valor.

Así que no es ninguna sorpresa que desde el año 2009 surgieran muchas criptodivisas nuevas con el auge de este mercado emergente tomando de referencia las premisas de *bitcoin*, estas surgieron con diferentes modelos de encriptación, modelos de guardado, transacción, entre otros. Pero con un fin en común, el pago descentralizado entre dos o más partes.

Posterior a todo ello, el pilar central de las transacciones de criptodivisas no es realmente la ausencia de tasas a las que están sujetas, sino radica en que toda transacción realizada dentro de la cadena de bloques supone un nivel de encriptación tal, que es anónima para cualquier parte que no esté involucrada.

Por lo que en la actualidad uno de los problemas más grandes a los que se enfrentan estas divisas, es la prohibición de su uso en las legislaciones de diferentes países, ya que el riesgo que suponen las transacciones 100 % anónimas no es algo que cualquier país está dispuesto a sobrellevar si no pueden tener injerencia en alguna parte del proceso.

De acuerdo con CoinMarketCap (2019) en el mercado existen alrededor de 914 criptomonedas, es importante mencionar que muchas de ellas aún se encuentran en un desarrollo de su estructura, por lo que no cuentan con un valor monetario real.

Pero si realizamos una combinación del valor de mercado de capitalización para todas las divisas, estamos en un mercado de aproximadamente US\$371

millones, en donde las 5 principales criptodivisas representan en torno al 83 % de su mercado. Por ende, uno de los principales atractivos de las criptodivisas es su precio, pero así mismo otro criterio que llama mucho la atención es la volatilidad que presentan siempre en términos del precio a través del tiempo. Muchos picos y relaciones asimétricas son las que suelen acompañar las tendencias de estas monedas.

Como menciona, Hayes (2015) el problema de investigación radica en la falta de análisis en relación a caracterización estadística para el fenómeno de las criptodivisas, fuera de los realizados en comparativa con otros activos de alta apreciación y los realizados por técnicas de algoritmos de clasificación de estadística multivariante para poder encapsularlos en ciertos factores económicos. Lo cual impide encontrar patrones de interés y la relación de los mismos que permitan encontrar asociaciones de interés entre estos activos, probar supuestos y realizar inferencia.

Entonces se planteó un estudio en el cual, por medio de validación de supuestos estadísticos y análisis de series temporales, se puede estimar el comportamiento probabilístico en el comportamiento bursátil de las criptodivisas de mayor valor actual, y respecto a sus variaciones, factores de autocorrelación y factores estacionales, predecir valores futuros de volumen de capital relacionado a las mismas por medio del diseño de un modelo matemático relacionado a la información histórica que tome en consideración a todas las criptodivisas consideradas en el estudio.

Se esperaba obtener cifras de volumen de capital con valores eficientes de predicción respecto a los valores reales, mediante los cuales, utilizando la métrica financiera de retorno sobre la inversión, se pueda cuantificar que tan interesante es en el corto plazo la inversión en determinado criptoactivo, con las cuales se

pueda estructurar un portafolio de sugerencias de inversión, de igual forma se esperaba el poder reflejar gráficamente los resultados antes mencionados, que apoye en la validación del análisis y facilitar su comprensión.

Los resultados obtenidos pueden cuantificar de forma técnica y probabilística las sugerencias de inversión relacionadas a estas criptodivisas y de igual forma, su asociación con las divisas de mayor valor bursátil, las cuales afirman ciertos supuestos de criterios económicos tradicionales. Sin embargo, el modelo propuesto sí puede ser aplicado como una medida de análisis de contención de capital de riesgo para potenciales inversionistas, en dónde se prueba que, a pesar de ser un mercado especulativo, el factor de autocorrelación se encuentra presente.

El informe final de esta investigación fue estructurado en 4 capítulos que se presentan a continuación: en el primer capítulo, se describen a nivel técnico, financiero y estadístico el funcionamiento y la economía de las criptodivisas, así también las variables que afectan su comportamiento bursátil.

En el segundo capítulo, se analizan los diseños de modelos de series de tiempo más utilizados, así también las pruebas estadísticas utilizadas para la validación de los mismos y la forma adecuada de interpretar sus resultados para generar modelos de predicción.

En el tercer capítulo, se exponen los resultados obtenidos de la investigación, donde gráficamente se pueden observar las series temporales que describen a los criptoactivos estudiados, sus tendencias, intervalos de confianza y descomposiciones, así también la validación de supuestos en su comportamiento financiero y económico.

El cuarto capítulo, es una discusión detallada, dónde por medio de la conceptualización estadística del problema, además de fundamentos financieros y económicos con los cuales se formulan las conclusiones descritas en los capítulos de cierre. Con ello se valida la solidez del conocimiento y sugerencias generadas a raíz del desarrollo de esta investigación.



## 1. MARCO REFERENCIAL

El análisis del comportamiento estadístico de las variables bursátiles relacionadas a cada criptomoneda, es un componente esencial de la caracterización de los factores económicos y financieros que impulsan la apreciación o depreciación de estos activos, esto a su vez es clave para poder encontrar patrones característicos de su evolución en el tiempo y garantizar la calidad en un proceso de inversión y con ello generar valor a través un rendimiento financiero óptimo.

Diversos autores han caracterizado este tipo de activos en los últimos años y a continuación se describe los más relevantes como puntos de referencia para esta investigación.

El mundo de las criptomonedas es realmente interesante, ya que su mayor atractivo yace en que brindan la disponibilidad acreedora de bienes o servicios, de una manera anónima y sin intermediarios, lo cual conlleva a que muchas empresas, las cuales tienen su giro de negocio completamente concentrado en prestar servicios de intermediario en transacciones web, comiencen a perder flujo de caja, por lo que es necesario considerar lo que es un activo como tal, y conceptualmente hace referencia a algo abstracto, tangible o intangible que siempre y cuando sea aceptado a cambio de una transacción de trueque (compra-venta), y que logre guardar valor para su propietario, puede ser llamado un activo.

Arango (2018) indica que la mayoría de criptomonedas no son un pasivo de ningún agente (público o privado); no tienen el estatus de ser de curso legal, es

decir, no tienen poder libratario ilimitado; no son reconocidos como divisas según definiciones del Fondo Monetario Internacional, es decir no son activos, ya que no son emitidos por un gobierno soberano. Se puede observar que es realmente interesante la contradicción que se presenta, ya que a pesar de que no tienen carácter de activos, funcionan como uno, de manera descentralizada y sin regulaciones, lo cual puede ser la brecha al inicio de una conceptualización nueva del dinero.

A raíz de esto surge una interrogante aún más interesante, la cual está asociada a cómo es producida una criptodivisa. En términos de macro estructura de tecnología de información, se hace referencia a un conjunto de datos, o cubos de información, cifrados de una manera en que cada hilo de datos se concatena de una manera distinta a la anterior, es decir, es un cifrado de complejidad denominada de alto nivel, para lograr tal grado de encriptación, es necesario realizar el proceso conocido como minería de datos.

Por medio de algoritmos de criptografía, y el uso de equipos de cómputo con disponibilidad de alta potencia gráfica, es posible realizar procesos complejos de minería de datos para crear activos. Hayes (2015) menciona que más del 84 % de la información significativa creada se puede explicar por tres variables involucradas, la potencia de la unidad computacional (la cual es el primer filtro en las dificultades de minado), el ratio de devolución de la moneda en producción, y la dificultad del algoritmo de minería programado para su realización.

Por lo que se puede observar que la creación de una criptodivisa tiene asociado un fuerte valor de producción, un tiempo de producción extremadamente alto.

Además de que se requieren conocimientos amplios en informática para la creación de esa clase de algoritmos y así mismo equipo computacional de alta potencia, si bien es subjetivo podría ayudar a justificar su valor en mercado, y comenzar a potenciar su uso.

A raíz de todo esto, la situación plantea la siguiente interrogante, ¿se está hablando de una única moneda? o, por el contrario, existen muchas criptodivisas. La segunda afirmación es la más acertada, ya que a raíz de que surgieron, las monedas tenían valores mínimos (en torno a centésimas de dólar estadounidense) por lo que muchos desarrolladores crearon sus propios algoritmos de producción, y así nació el mercado de criptodivisas, en 2019 existen centenas de estas divisas, aunque realmente para poder generar predicciones de inversión apropiadas se requiere evaluar a las que se han podido situar de manera correcta en el mercado.

Radovanov, Marcikić y Gvozdrenović (2018), mencionan que como objetivo para realizar un análisis comparativo de estas divisas, se debe analizar el precio diario en el mercado de las cuatro criptodivisas de mayor valor, las cuales son *bitcoin* (BTC), *ethereum* (ETH), *ripple* (XRP) y *litecoin* (LTC), por lo que se toma de referencia, los cierres diarios de distintos mercados financieros, el horizonte temporal contemplado por el autor está comprendido entre el 06 de agosto de 2015 a marzo 11 de 2018, en donde estas monedas son las que tienen un precio más significativo en el mercado.

Como menciona Radovanov *et al.* (2018), las criptodivisas que tienen un mayor precio de venta son las mencionadas anteriormente, por lo que las anteriores fueron utilizadas para referencias en esta investigación.

Al existir tantas monedas en el mercado, es fácil poder asumir que algunos algoritmos son mejores que otros, o plantear las interrogantes de por qué algunas monedas valen más que otras, estas últimas interrogantes pueden cuestionar el uso de referencia a monedas de alta concentración de flujo de efectivo frente monedas en surgimiento.

Gandal y Halaburda (2016), menciona que los efectos de red de seguimiento a favor de *bitcoin* continuaron hasta el fin del periodo del cual se dispone de información en el mercado financiero, por ejemplo, *bitcoin* mantuvo un precio en el mercado de alrededor de 400 dólares estadounidenses desde el final de abril de 2014 hasta el final de febrero de 2016, en donde existió una caída en el precio de manera masiva.

La mayoría de monedas de la muestra perdió al menos dos tercios de su precio de mercado, de manera similar a *bitcoin*. Por ejemplo, *Peercoin* que estaba cotizando a 2.21 USD durante el 30 de abril del 2014, costaba 46 céntimos de dólar a febrero de 2016, una variación porcentual muy similar a *bitcoin*.

Como comenta Gandal (2016), en su análisis se puede observar que las fluctuaciones en el precio de la moneda, aunque tengan variaciones, no son tan significativas como para impedir el análisis de su comportamiento a través del tiempo, independiente a si es una moneda en surgimiento o no, todo depende de las características de cada una.

En cuanto al incremento del volumen de transaccionalidad en algunas criptodivisas suele estar ligado a actividades fuera del marco legal (tal es el caso de *monero*) que dado con su alta capacidad de aleatorización de *Hash* (código de identificación único en informática) la vuelve una alternativa atractiva para

criminales informáticos, se estima una computadora de criptografía podría tardar hasta 100 años para encontrar una traza de su origen.

Sovbetov (2018) menciona que se debe tener especial consideración analítica en esta parte, dado a que a pesar de que la tendencia tienda a crecer, se trata de un crecimiento derivado a pagos de ilícitos como drogas, tráfico de armas, entre otros. Por lo que a medida que se vayan reduciendo estos ilícitos alrededor del mundo, la inflación artificial que presentan estas divisas reflejara una relación espuria con el resto.

Por otra parte, la orientación principal de esta investigación está enfocada a realizar mejores aproximaciones al momento para poder hacer decisiones inteligentes de inversión, esto por medio de un modelo que pueda caracterizar de forma estadística los patrones de incremento y decremento del valor monetario real de estas divisas, y sea posible visualizar en un horizonte temporal los rendimientos futuros que se podrían o no esperar.

Para ello se necesita una referencia con la cual correlacionar al resto de divisas de valor significativamente más pequeño de mercado, Cocco, Concas y Marchesi (2015) mencionan que el valor real en el precio de *bitcoin* puede ser bien reproducido en periodos cortos de tiempo, por lo que los ensayos en computadora definen claramente una negación a la hipótesis de que el mercado sigue un rumbo completamente aleatorio.

Por lo tanto, se estima que el rumbo del mercado de criptodivisas no es solo especulativo, está claramente relacionado al mercado financiero, con factores sociales, económicos y culturales incidiendo sobre su precio. Entonces está en claro que el objetivo es la inversión de divisas locales, en valores de cualquier criptodivisa para obtener un retorno a la inversión bastante favorable para el

inversionista. Así que entra en la investigación una variable bastante interesante, el riesgo.

¿Qué tan segura es la inversión de dinero en estos mercados?, no se sabe a ciencia cierta si podrían llegar a desaparecer, o si realmente en un futuro estas divisas tendrán aceptación en el mercado. Se puede definir que, al mismo tiempo, los retornos de las criptodivisas pueden ser predichos teóricamente por dos factores primordiales, estos son el momento de la inversión y la atención de los inversores.

Asimismo, Liu y Tsyvinski (2018) mencionan que se deben administrar las consultas populares de la formación de valor, tales como el costo de minería de datos, el precio dividido, la volatilidad de la moneda, y realmente todas estas variables siguen siendo útiles para caracterizar su comportamiento.

En resumen, la cadena de valor de las criptodivisas tiene una magnitud tal, que es imposible su desaparición inmediata del mercado financiero, así mismo tienen el potencial de reemplazar a las divisas tradicionales en negociaciones B2B (negocio a negocio por sus siglas en inglés) en transnacionales.

Dejando por un lado las consideraciones en el riesgo, también es importante mencionar, que una vez sea vea más normalizado el mercado de criptoactivos, será claro su crecimiento en cuanto a inversiones, por lo que a una proyección a futuro del circulante podría representar un problema, en lo referente al papel moneda, es decir, desde su surgimiento la volatilidad en el comportamiento de *bitcoin* y realmente de cualquier otra divisa, brindaron en un inicio un gran proporción de la moneda a algunas carteras cuando el algoritmo era de complejidad relativamente baja.

Por lo que un flotante de algunas divisas de alguna cartera grande podría poner en apuros a ciertas instituciones financieras, ya que la manera en que tan solo una pequeña porción de las mismas tiene un valor de alrededor de USD 10 MM, realmente no es claro cómo se validarían estas liquidaciones en caso el mercado que dispone de altas concentraciones de la moneda en su cartera comenzará a rotarlas en compra-venta de bienes.

Bakar, Rosbi y Uzaki (2017) comentan que al momento de comenzar las liquidaciones de estas divisas, las instituciones financieras tendrán que liquidar una divisa de la cual no disponen estandarizaciones, por lo que llegará un punto en el que a pesar de que el fin de la divisa es mantenerlos alejados, será necesario involucrar a las instituciones bancarias en algún punto además del poder realizar modelos de proyección estadísticos así mismo de clasificación para poder comenzar a generar sistemas de punteo y estandarización de sus burós de información internos.

Además, en función del estudio realizado y acorde al análisis de Conrad, Custovic y Ghysels (2018), como dato complementario, se sabe por estudios a largo plazo que realizar predicciones para activos de alta varianza se torna un problema con una incerteza un tanto difícil de manejar, por lo que es interesante el poder comparar el mercado de criptoactivos con un mercado que se comporte de manera similar, y de manera empírica poder comparar si existe correlación en los mismos, esto con el fin de poder definir si es posible explicar la variación mensual con alguna tendencia anual de otro mercado, o viceversa.

Los criterios de información de Akaike y Bayesiano el mejor modelo de ajuste, acorde al mercado de criptoactivos en función del modelo GARCH-MIDAS, está basado en el modelo Akaike de secado. Es decir, que alrededor del 65 % de la variación mensual de una serie asintótica puede ser ampliamente

explicada por movimientos en la volatilidad al largo plazo. A manera de comparación para el corto plazo, el modelo puede ser descrito con la volatilidad del modelo de largo plazo relacionado a bienes de lujo.

Así que a pesar de que predecir las variaciones en el corto plazo, por medio de un mercado correlacionado es posible, puede no ser la mejor aproximación a su valor monetario real, pero se tiene algo muy claro y es que los bienes de lujo están sujetos al valor de los criptoactivos, y por ende, los ajustes encontrados por autor brindan excelentes directrices para el modelado de estas series.

En función de lo anterior se puede concluir que la cartera está sujeta a los demás mercados económicos y financieros, por lo que es probable que sea muy predecible en función de muchas carteras de monedas alrededor del mundo.

Bianchi y Dickerson (2018) mencionan que, con una amplia variedad de pruebas basadas en regresión, ha sido posible coleccionar suficiente evidencia para probar que los resultados del mercado de criptoactivos no puede ser normalizado bajo los mismos estándares de riesgos y estrategias de portafolio de monedas de divisas tradicionales, además de que nunca sobrevivirían a los impuestos por transacción a las que están sujetas estas últimas.

Adicionalmente, gracias a muchas pruebas más robustas realizadas, se puede observar que los análisis empíricos no dependen ni siquiera de la muestra de criptoactivos seleccionados y del periodo calculado para medir su volumen, por lo que se evidencia que es extremadamente riesgoso asumir que el comportamiento de las divisas sigue un estándar actualmente impuesto, y que predecir su comportamiento está más ligado a conceptos abstractos e históricos, que a teorías económicas.

## **1.1. Marco contextual**

Los mercados de divisa considerados para la referencia comparativa de investigación son todos aquellos que tengan de forma nativa una transformación disponible al dólar estadounidense como comparativa, indiferente a la moneda de la localidad en dónde se encuentre la operación central de dicho mercado, con ello se homogeneizaron las divisas en relación a sus unidades monetarias para el estudio y con ello validar los supuestos de cada criptomoneda, dentro de estos mercados se encuentran la bolsa de valores de Estados Unidos de América, Londres y Japón.



## **2. MARCO TEÓRICO**

### **2.1. Fundamentos estadísticos**

A continuación, se describen los fundamentos de las técnicas estadísticas utilizadas en este estudio. Es importante acotar que análisis ulteriores más específicos son indispensables en algunos algoritmos robustos aplicados a las bases de datos. La investigación considera de forma intuitiva la aplicación de las siguientes técnicas estadísticas como los pilares fundamentales para la realización de este estudio. Sin embargo, la interpretación de resultados y validación de la eficiencia predictiva de los mismos será ahondada en los apartados correspondientes según el caso de uso.

#### **2.1.1. Intervalos de confianza**

La estimación de un parámetro de una población, con base a una muestra, puede realizarse por medio de distintos métodos y tratamientos estadísticos, los intervalos de confianza buscan encontrar un rango parametrizado dentro del cual se encuentre el estadístico a evaluar con una probabilidad de acierto denominada nivel de confianza el cual se denota como  $(1 - \alpha)$  en donde  $\alpha$  denota el nivel de significancia o nivel de confianza aceptado para el estudio.

Los niveles de confianza utilizados dentro de la mayoría de pruebas estadísticas son 99 %, 95 % o 90 %, entre otros. Estos representan la probabilidad de que el estadístico no se encuentre dentro del rango especificado.

Mendenhall, Scheaffer y Wackerly (2010), sugieren que para modificar la holgura de un intervalo se puede realizar por dos métodos distintos: cambiando la magnitud del error que se está dispuesto a cubrir, es decir el nivel de confianza de la prueba realizada. Cambiando el tamaño de la muestra, esto por una relación de variabilidad, a menor tamaño de muestra, se crea un intervalo de confianza mucho más amplio, dado a que no se disponen de las características suficientes para realizar inferencias más precisas.

- Intervalos de confianza para la media con desviación estándar poblacional conocida.

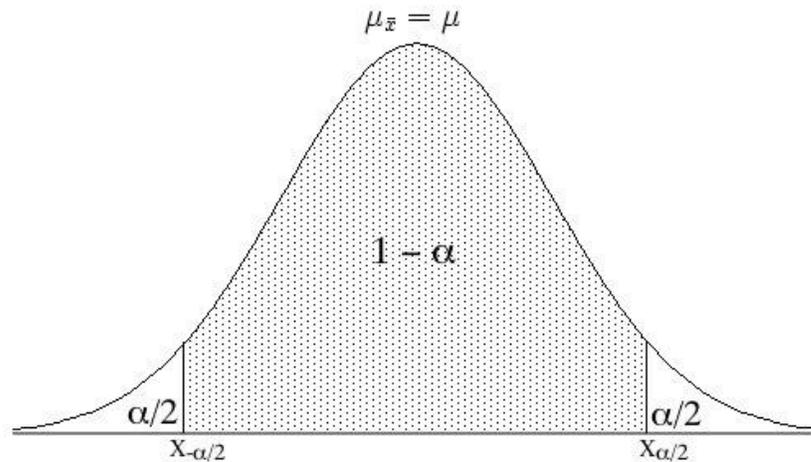
Cuando se conoce la desviación estándar poblacional y la muestra es mayor o igual a 30 registros, se puede hacer uso de la aproximación a la normal por medio del teorema del límite central, en dónde se puede hacer uso de la distribución normal Z.

Para realizar el cálculo de la distribución Z se utiliza la siguiente expresión:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ (Ec. 1)}$$

En donde  $Z$  hace referencia a la distribución de probabilidad normal,  $\bar{x}$  a la media aritmética de la muestra estudiada,  $\mu$  al parámetro de prueba en el estudio,  $\sigma$  a la desviación estándar poblacional y  $n$  al tamaño de la muestra.

Figura 1. **Gráfico de normalidad con un 5 % de significancia**



Fuente: elaboración propia.

Lo que se busca es que el valor  $Z$  se encuentre dentro del intervalo planteado. Con una confianza conocida como  $1 - \alpha$ .

Por su parte, Klenke (2014) sugiere que para el cálculo de un intervalo de confianza se utilice la siguiente expresión:

$$IC \text{ para la media: } P\left(\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = (1 - \alpha) \quad (Ec. 2)$$

En dónde:

- $\bar{X}$  hace referencia a la media aritmética de la muestra.
- $Z_{\alpha/2}$  es el estadístico de prueba para la prueba  $Z$  del valor de  $\alpha/2$ .
- $\bar{X} - \mu$  es la media poblacional,  $\sigma$  es la desviación estándar y  $n$  es el tamaño de la muestra.

Para efectos de cálculo se observa que se utiliza únicamente el valor de  $\alpha/2$ , esto se fundamenta en la simetría de una curva de datos normales. Por otro lado, se debe considerar que se puede trabajar bajo la premisa de que existen poblaciones que pueden ser mesurables y poblaciones que no pueden ser mesurables (o infinitas teóricamente). Para ellas el cálculo del error del intervalo será el siguiente:

Error de estimación para poblaciones infinitas:

$$\pm Z \alpha/2 \frac{\sigma}{\sqrt{n}} \quad (Ec. 3)$$

En donde  $\alpha$  hace referencia al nivel de significancia,  $\sigma$  a la desviación estándar poblacional y  $n$  al tamaño de la muestra.

Error de estimación para poblaciones finitas:

$$\pm Z \alpha/2 \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad (Ec. 4)$$

En donde  $\alpha$  hace referencia al nivel de significancia,  $\sigma$  a la desviación estándar poblacional,  $n$  al tamaño de la muestra y  $N$  al tamaño de la población.

### **2.1.2. Pruebas de hipótesis**

En las pruebas de hipótesis, se utilizan para validar o invalidar por medio de supuestos o afirmaciones de lo que se desea probar sobre una población o un conjunto de datos, puede seguirse una metodología para probar la veracidad de

las mismas. La metodología propuesta por la mayoría de autores para poder efectuar de forma correcta una prueba de hipótesis es la siguiente:

- Definir una hipótesis nula y los parámetros que suponen la veracidad de la misma.
- Definir una hipótesis alterna y los parámetros que suponen la veracidad de la misma, es decir, el rango que abarca que la hipótesis nula sea rechazada.
- Definir el nivel de significancia con el que se está dispuesto a trabajar.
- Establecer el área de rechazo o aceptación de la hipótesis nula, de acuerdo con la distribución de probabilidad con la que se está trabajando el ensayo.
- Aplicar el criterio o regla de decisión, en dónde en relación a si el estadígrafo de prueba cae en la zona de rechazo, se rechaza la hipótesis alterna y se falla en rechazar la hipótesis alterna y viceversa.

La presentación de las pruebas de hipótesis a nivel de la notación de este trabajo de graduación es la siguiente:

*H<sub>0</sub>: 'referencia a hipótesis nula'*

*H<sub>A</sub>: 'referencia a hipótesis alterna'*

Siempre que se elabore un análisis por medio de una prueba de hipótesis se debe considerar que existe un margen de error al considerar estas afirmaciones, Hayes (2015) indica que para elaborar la correcta validación por

medio de hipótesis el analista debe tener en consideración dos posibles supuestos, los cuales son llamados error del tipo I y error tipo II.

El error tipo I, denotado por la letra griega  $\alpha$ , consiste en rechazar una hipótesis cuando en realidad nada invalidaba su veracidad o su nivel de significancia; por contraparte el error del tipo II consiste en no rechazar una hipótesis cuando en realidad nada valida su veracidad, es decir, es falsa en la población, es denotada por la letra  $\beta$ .

Las pruebas de hipótesis pueden ser realizadas de distintas formas, asociado al nivel de significancia que se está probando, es decir, pueden ser bilaterales o unilaterales. En función de cómo esté definida la hipótesis alterna se plantea la región de rechazo de la hipótesis nula. Entonces, se realiza un ensayo bilateral cuando se esté comparando una hipótesis alterna que suponga todo lo contrario al parámetro de referencia utilizado. Por otro lado, Suárez y Bautista (2016) mencionan que un ensayo unilateral se realiza cuando se compare con hipótesis alternas que indiquen qué parámetro será mayor o menor que lo que se está suponiendo.

Para un correcto criterio de decisión, se debe considerar de qué forma se está realizando la prueba. En caso de que las pruebas sean bilaterales la regla de decisión es: si el estadístico que se está comparando es mayor o menor según los valores críticos del error, se rechazará la hipótesis nula, es decir lo que se desea probar.

En cuanto a las pruebas unilaterales, se debe evaluar en qué área de la curva se está fijando el límite del error tipo I, ya sea del lado derecho o sea del lado izquierdo. Dependerá si es a la cola derecha, y si el estadístico de prueba es mayor estará incluido en el área de error, y por el lado de la cola izquierda si

el estadístico de prueba es menor que el valor crítico de igual forma, en ambos casos no se tendrá evidencia estadística suficiente para rechazar la hipótesis que se está validando.

### 2.1.3. Prueba de hipótesis para la media de la población

Las pruebas de hipótesis para la media de una población buscan probar supuestos o afirmaciones a nivel de características fijas de una población, en este caso la prueba de medias. Klenke (2014) supone dos casos para poder realizar este tipo de prueba de hipótesis: cuando se conozca  $\sigma$  y se tenga una muestra de tamaño mayor o igual a 30 datos y en caso se desconozca  $\sigma$  y se tenga una muestra menor a 30 datos.

Dado que durante la realización de esta investigación se analizaron grandes series de datos, se puede utilizar la aproximación a una normalización de datos dado el teorema del límite central, para verificar que los datos se comportan de forma normal, por lo que utilizando un valor Z teórico, calculado por medio del nivel de significancia y un estadístico de prueba Z. Para ello se hace mención de la siguiente expresión:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (Ec. 5)$$

Se sugiere utilizar el siguiente procedimiento:

- $H_0: \mu = \mu_0$ .
- $H_1: \mu \neq \mu_0; H_1 \mu < \mu_0; H_1 \mu > \mu_0$ .
- $\alpha = 1 \%, 5 \%, 10 \%, etc$  si así se desea.

- Establecer la región crítica utilizando la tabla de la distribución normal. Para un ensayo bilateral  $\pm Z_{\frac{\alpha}{2}}$  y para ensayos unilaterales  $-Z_{\alpha}$  si  $H_1 \mu < \mu_0$  o  $+Z_{\alpha}$  si  $H_1 \mu > \mu_0$ .
- Estadístico de prueba (ecuación 5).
- Aplicar el parámetro de decisión, para un ensayo bilateral si  $z < -Z_{\frac{\alpha}{2}}$  o  $z > Z_{\frac{\alpha}{2}}$  se rechaza la hipótesis nula.

En el caso de ensayos unilaterales si  $z < -Z_{\alpha}$  o  $z > Z_{\alpha}$  se rechaza la hipótesis nula de igual forma.

#### **2.1.4. Prueba de hipótesis para la diferencia de medidas de dos poblaciones**

Esta prueba siempre se basa en un análisis que utiliza la media como parámetro para realizar inferencias estadísticas, se busca realizar la comparación de dos poblaciones independientes, lo cual contribuye a comprender el comportamiento individual de cada media, así como el comportamiento de la brecha que las diferencias. (Suarez y Bautista, 2016).

Al igual que los ensayos de una población, la hipótesis nula siempre considera parámetros comparativos cerrados, es decir una igualdad, sea con una cota fija en el valor, o una cota fija en el valor sea mayor o menor, estas expresiones pueden diferenciarse como  $\leq$ ;  $\geq$  o  $=$ .

La hipótesis alterna por su parte busca lo opuesto a la hipótesis nula, por lo que no está buscando establecer verificaciones puntuales sobre los valores, por lo que las comparaciones suelen estar definidas con la simbología:  $<$ ;  $>$  o  $\neq$ .

Usualmente el ensayo que se desea probar está asociado a la igualdad estadística de las muestras estudiadas, por ende, en normas generales (más no específicas) se definen a las pruebas de hipótesis de la siguiente forma:

$$H_0: \mu_A = \mu_B$$

$$H_A: \mu_A \neq \mu_B$$

- Pruebas de hipótesis para la diferencia entre dos medias con  $\sigma_1$  y  $\sigma_2$  conocidas.

Como se mencionó anteriormente, se consideran muestras grandes como aquellas en las que se disponga una distribución de más de 30 observaciones, por lo que cuando la muestra es superior a 30 observaciones y por el teorema del límite central se puede aproximar el conjunto de datos a una distribución normal. Por lo tanto, se utiliza la distribución normal para la validación de las hipótesis. Para este tipo de prueba debe definirse un parámetro  $d_0$ , el mismo representa la diferencia supuesta entre las medias de las poblaciones.

Por otra parte, Mendenhall *et al.* (2010), indican que para calcular el estadístico de referencia se puede utilizar la siguiente expresión.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sigma(\bar{x}_1 - \bar{x}_2)} \quad (Ec. 6)$$

En dónde:

- $(\bar{x}_1 - \bar{x}_2)_1$  hace referencia a la diferencia entre las medias muestrales.
- $d_0$  a la diferencia supuesta entre las medias poblacionales.

- $\sigma(\bar{x}_1 - \bar{x}_2)_2$  es la desviación de la distribución muestral de las diferencias.

Se sugiere utilizar el procedimiento indicado en la prueba de hipótesis de una población.

### **2.1.5. Análisis de varianza**

A las pruebas de hipótesis para tres o más poblaciones simultáneas, se les conoce como el método de análisis de varianza (ANOVA), este método es utilizado específicamente para determinar si las medias de poblaciones de datos son iguales o distintas.

#### **2.1.5.1. Análisis de varianza de un factor**

En el caso de este ANOVA, se utiliza un solo factor o parámetro para varias poblaciones de datos, con ello se puede estudiar la influencia de este factor como característica de interés en dónde se desea definir los puntos de estudio para los cuales estas poblaciones cumplen el supuesto dado. Este análisis debe considerar la variación entre las muestras, así como la variación a lo interno de las mismas. Este análisis consiste en la comparación entre estos dos tipos de variaciones por medio de la distribución F de Fisher, la que indica si las medias son iguales o son distintas. (Klenke, 2014). Los criterios principales para realizar un análisis de varianza se enlistan en la siguiente tabla:

Tabla III. **Resumen de criterios para análisis de varianza**

<b>Fuente de variación</b>	<b>Suma de cuadrados</b>	<b>Grados de libertad</b>	<b>Cuadrado medio</b>	<b>Valor F</b>
Tratamiento	SCTR	c-1	$\frac{SCTR}{c - 1}$	$\frac{CMTR}{CME}$
Error	SCE	n-c	$\frac{SCE}{n - c}$	
Variación total	SCT	n-1		

Fuente: elaboración propia.

De la misma forma, Sovbetov (2018) indica que un correcto análisis de varianza requiere determinar la suma de cuadrados asociado a tal variación entre los tratamientos o muestras, y su suma de cuadrados formada por la variación entre los tratamientos o muestras, para lo cual sugiere el uso se la siguiente expresión:

$$SCT = SCTR + SCE \text{ (Ec. 7)}$$

En donde la suma de los cuadrados del total, es considerada la variación de todas las observaciones en un todo respecto a la media global, es decir, la media de todas las medias.

$$SCTR = \sum r_j (\bar{x}_j - \bar{\bar{X}})^2 \text{ (Ec. 8)}$$

En dónde  $r_j$  hace referencia al número de observaciones de cada tratamiento,  $\bar{x}_j$  a la media de cada tratamiento y  $\bar{\bar{X}}$  a la gran media.

La suma de los cuadrados de los tratamientos es el estimador más puntual de la variación de cada tratamiento alrededor su promedio total:

$$SCE = \sum \sum (X_{ij} - \bar{X}_j)^2 \quad (Ec. 9)$$

En dónde  $X_{ij}$  hace referencia a las observaciones de cada tratamiento y  $\bar{X}_j$  a la media de cada tratamiento.

En relación al parámetro de cálculo, Klenke (2014) menciona que se debe definir los grados de libertad asociados. Define qué número de grados de libertad es el número de valores en el cálculo final de un estadístico de prueba que pueden variar libremente.

El número de elementos independientes de información que se incluyen en la estimación de un parámetro se denominan grados de libertad.

Los grados de libertad de una estimación son el número de datos independientes que se utilizaron para calcular la estimación. No es lo mismo que el número de elementos en la muestra. Para obtener los grados de libertad para la estimación, se debe restar 1 del número de elementos por cada estimador a calcular. Klenke (2014) menciona que un ejemplo que se suele utilizar es del cálculo de la media de pérdida de peso para una dieta baja en carbohidratos de un grupo de personas.

$$gl \text{ para } SCTR = c - 1 \quad (Ec. 10)$$

$$gl \text{ para } SCE = n - c \quad (Ec. 11)$$

A partir de ello se pueden definir reglas de decisión para establecer el nivel de significancia que se usará y determinar, si se podrá rechazar una hipótesis nula, en dónde se comparará la igualdad de todas las medias involucradas, o se

rechazará esa hipótesis dado que el valor estudiado es mayor que un valor crítico de Fisher.

### **2.1.6. Regresión lineal**

Como mencionan Kuhn y Johnson (2013), es una técnica estadística que tiene como finalidad establecer un modelo que explique la relación en función de su comportamiento, de dos variables,  $X; Y$ . Con dicho modelo se puede hacer predicciones sobre los valores de una variable independiente o exógena, sobre una variable dependiente o endógena. Cuando la asociación de las variables es alta supone la generación de un modelo de predicción que puede llegar a estimar con cierto grado de error, valores fuera del rango del compendio de datos.

$$y_i = \alpha + \beta(X_i) + \varepsilon_t \text{ (Ec. 12)}$$

En dónde  $\alpha$  describe el intercepto en  $y$  de la recta modelada,  $\beta$  hace referencia a la pendiente de la recta, y  $\varepsilon$  se refiere al error de estimación, es decir a los parámetros para los cuales el modelo no es un buen estimador.

### **2.1.7. Regresión múltiple**

Algunos casos de estudio necesitan más de una variable exógena para conseguir una mayor robustez en el modelo planteado. Por ende, necesita el aporte de datos que puedan brindarle muchas más variables independientes que estén asociadas al estudio (Kuhn y Johnson, 2013).

Los modelos de regresión lineal múltiple, creados a partir de un conjunto de datos, pueden ser definidos por parte de la siguiente ecuación general:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \text{ (Ec. 13)}$$

En la ecuación anterior, se tiene una cantidad de  $n$  variables independientes, que a su vez están acompañadas de un coeficiente  $\beta$ , el cual indica cual es el grado de relación o el peso que tiene cada una de las variables independientes sobre el valor dependiente  $Y$  o endógena. Además de poder encontrar un valor  $\beta_0$  y  $\varepsilon$  en el cual el primero representa a la coordenada de origen, el valor de la variable dependiente  $Y$ , cuando todos los predictores son cero y el segundo representa la dispersión de los errores en el modelo, ambas variables son utilizadas para representar lo que los coeficientes de las variables independientes no tienen el alcance de explicar dentro del modelo.

### **2.1.8. Análisis de residuos**

Posterior al desarrollo de un modelo lineal, los coeficientes y parámetros estimados deben enfrentar un proceso de validación a raíz de hipótesis, estas hipótesis parten de los supuestos que como mínimo debe cumplir los ajustes de dicho modelo lineal para considerarse válido, los supuestos principales a verificar son la independencia, la homocedasticidad, la normalidad y la eficiencia de la explicación de los ajustes del modelo acorde a la varianza de los datos con los que está construido.

La validez de estos supuestos es necesaria para que los resultados sean significativos. Si se incumplen estos supuestos, el resultado puede ser incorrecto y tener graves consecuencias, Kuhn y Johnson (2013) mencionan entre ellas el sobreajuste a la muestra del diseño o inferencias poco correlacionadas a la realidad de la población dado un conjunto de datos sesgado.

### **2.1.8.1. Independencia**

El supuesto de independencia se asocia al comportamiento que deben presentar los residuos de un modelo de regresión lineal, sea simple o múltiple. En dónde lo que se busca probar es total independencia entre ellos, es decir, que no sean producto de una tendencia o concentración dentro del mismo modelo, como mencionan los autores Kuhn y Johnson (2013).

La prueba de independencia sirve para establecer si existe alguna colinealidad entre las variables del modelo estudiado, es decir, si existe correlación fuerte entre determinadas variables independientes en el modelo.

Entonces, la independencia significa que es necesario que todas las variables que se pretenden incluir en el modelo como variables predictoras, no dependan de ninguna otra variable incluida en el modelo, si no existiese independencia, dicha variable no podría ser utilizada.

Cuando una variable independiente, tiene alguna relación con cualquier otra del modelo, o depende del comportamiento de otra, estas variables muy probablemente presentan colinealidad o multicolinealidad, por ende, se debe evitar a toda costa dichas variables en la generación de modelos, ya que solo presentarían ruido.

### **2.1.8.2. Homocedasticidad**

El supuesto de homocedasticidad (que significa misma varianza) es fundamental para los modelos de regresión lineal. La homocedasticidad describe una situación en la que el término de error (es decir, el ruido o error aleatorio en la relación entre las variables) es el mismo en todos los valores de las variables

independientes. La heterocedasticidad (la violación de la homocedasticidad) está presente cuando el tamaño del término de error difiere entre los valores de una variable independiente. Kuhn y Johnson (2013) mencionan que el impacto de violar el supuesto de homocedasticidad es una cuestión de grado, que aumenta a medida que aumenta la heterocedasticidad.

$$V_{\varepsilon} = \sigma^2 \quad (\text{Ec. 14})$$

La homocedasticidad en un modelo multivariado, plantea la hipótesis que la variación de los residuos será independiente de toda variable del modelo.

### **2.1.8.3. Normalidad**

En la regresión lineal múltiple se puede establecer una hipótesis que plantea que todos los residuos en comparación con el valor esperado, así mismo con todos los demás valores observados de una serie de datos, es decir que tendrán un comportamiento significativamente homogéneo al de una distribución de probabilidad normal, con una media igual a cero (Klenke, 2014).

$$\varepsilon \approx N(0, \sigma) \quad (\text{Ec. 15})$$

Para determinar la normalidad de una serie de residuos se puede utilizar una prueba de Kolgomorov-Smirnov. Esta prueba estadística, plantea una hipótesis, para el supuesto en el que los datos de la distribución estudiada siguen un comportamiento probabilístico normal. Esta hipótesis de falta de normalidad, se valida utilizando un estimador del valor  $p$ , menor que la significancia  $\alpha$  establecida del modelo.

#### 2.1.8.4. Prueba de Kolmogorov Smirnov

La prueba de Kolmogorov-Smirnov es una prueba de bondad de ajuste no paramétrica y se utiliza para determinar si dos distribuciones difieren, o si una distribución de probabilidad subyacente difiere de una distribución hipotética. Se utiliza cuando tenemos dos muestras procedentes de dos poblaciones que pueden ser diferentes. La prueba tiene la ventaja de considerar las funciones de distribución colectivamente. También se puede usar como prueba de bondad de ajuste.

En este caso, solo se debe de disponer de una muestra aleatoria obtenida de una población en donde la función de distribución es específica y conocida. En el caso del análisis de residuos de modelos lineales, es una prueba significativamente robusta para validar que la distribución de varianza no explicada por el modelo se encuentra insesgada y distribuida de forma normal (Dodge, 2010).

El criterio de decisión se relaciona a la validación de dos subconjuntos de una muestra estudiada tienen la misma distribución de probabilidad. Para ello se debe realizar un contraste de hipótesis, la cual se plantea de la siguiente forma:

*$H_0$ : Los subconjuntos tienen una misma distribución.*

*$H_A$ : Los subconjuntos tienen una distribución distinta.*

En otras palabras, se tiene que:

$$H_0: F_1(x) = F_2(x)$$

$$H_A: F_1(x) \neq F_2(x)$$

El estadístico a probar es el siguiente:

$$KS = \max|F_1(x) - F_2(x)| \quad (Ec. 16)$$

Siendo el valor crítico:

$$VC_a = vc_a \sqrt{\frac{n_1 + n_2}{n_1 * n_2}} \quad (Ec. 17)$$

En dónde  $n$  hace referencia al tamaño de cada muestra, y  $vc_a$  al valor crítico de la prueba estadística acorde al valor de significancia (Dodge, 2010). El criterio de decisión para la prueba es; si  $KS > VC_a$ , se rechaza la hipótesis nula, y se puede concluir que la distribución de datos no sigue un comportamiento normal en torno a su media.

#### **2.1.8.5. Multicolinealidad**

Esta es una metodología conocida como el supuesto de la no colinealidad. Se refiere a que, dentro de las variables endógenas de un modelo de regresión lineal, no debe existir ningún tipo +- cuando una de las variables independientes tiene una regresión lineal con una o más variables independientes.

- Multicolinealidad parcial: este tipo de colinealidad resulta cuando existe una alta concentración de variables independientes.

Uno de los preceptos para evaluar la multicolinealidad, se observa cuando las variables independientes del modelo creado se expresen como una combinación lineal.

$$0 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \text{ (Ec. 18)}$$

La forma más fácil en cuanto a definir la colinealidad entre las variables independientes, es aplicarles una prueba de la diferencia entre el coeficiente de determinación  $R^2$ , para cada variable con el resto.

#### **2.1.8.6. Coeficiente de determinación $R^2$**

El coeficiente de determinación, denotado  $R^2$ , es el cociente de la variación explicada (suma de cuadrados debido a la regresión) a la variación total (suma total de cuadrados total SS (TSS) en un modelo de regresión lineal simple o múltiple, está descrito de forma algebraica por la siguiente razón (Dodge, 2010).

$$R^2 = \frac{VE}{TV} \text{ (Ec. 19)}$$

Dónde  $VE$  hace referencia a la variación explicada por el modelo lineal, y  $TV$  al total de la variación del modelo.

#### **2.1.9. Series de tiempo**

Una serie de tiempo es un conjunto de datos secuencial, medido usualmente por un periodo acotado de tiempo, por medio de una frecuencia secuencial. Está matemáticamente definido como el conjunto de vectores  $X(t)$ ,  $t = 0,1,2,3 \dots$  en dónde  $t$  representa el tiempo transcurrido. La variable  $x(t)$  es tratada como una variable aleatoria. Como expone Radovanov, *et al.* (2018), las medidas capturadas durante un evento en una serie de tiempo, siempre están ordenadas en función del orden cronológico en el que sucedieron las eventualidades.

Una serie de tiempo contiene el registro en términos de una sola variable, así como en términos multivariados. Una serie de tiempo además puede ser continua o discreta. En las series de tiempo continuas las observaciones son medidas a cada momento del tiempo, por otro lado, las series de tiempo discretas contienen mediciones de tiempo en momentos discretos de la serie.

Usualmente en series de tiempo discretas, las observaciones siempre son secuenciales, guardados por medio de intervalos de tiempo, como horas, días, semanas, meses o incluso años.

#### **2.1.9.1. Componentes de una serie de tiempo**

Las series de tiempo en general, están sujetas a ser afectas por cuatro componentes principales. Los cuales pueden ser separados por los datos observados (Cryer y Chan, 2008). Estos componentes son la tendencia, la frecuencia de ciclos, la estacionalidad y los componentes irregulares. A continuación, se brinda una pequeña descripción de estos componentes.

La tendencia general de una serie de tiempo a crecer, decrecer o estancarse en un periodo de tiempo específico es conocida usualmente como la tendencia secular de una serie de tiempo.

Las variaciones estacionales en una serie de tiempo son fluctuaciones durante un periodo seleccionado para el estudio, el cual puede estar comprendido desde horas, días, años, décadas, entre otros. Los factores importantes que pueden causar fluctuaciones recurrentes en una serie de tiempo son: el clima y las condiciones meteorológicas, las aduanas, los hábitos tradicionales, entre otros.

Las variaciones cíclicas de una serie de tiempo describen a mediano plazo, los cambios en una serie de tiempo causado por circunstancias ajenas a la serie, esto se puede ver representado de forma cíclica. Muchas de las series de tiempo que describen situaciones financieras y económicas suelen mostrar cierto punto de variación cíclica. En cuanto a las variaciones irregulares o aleatorias en una serie de tiempo, estas suelen causar influencias impredecibles a lo largo de todo el modelo.

Las cuales, en muchos casos, se suele caer en el error de estimación y clasificar dichas observaciones como regulares y no como una variación muy asimétrica de la serie, Cryer y Chan (2008) mencionan que es complicado analizar la recurrencia de valores atípicos en un modelo que contenga valores atípicos ya que, no se tiene definida una técnica estadística específica para medir las variaciones fluctuantes y este tipo de correcciones suelen estar frecuentemente desarrolladas por medio de una amplia experiencia en el tema estudiado, además de la realización de las transformaciones estadísticas correspondientes.

Por ende, considerando los efectos de estos cuatro componentes en general, dos tipos diferentes de modelos son usados para series de tiempo. Los modelos de series de tiempo multiplicativos y los modelos de series de tiempo aditivos.

$$\text{Modelo multiplicativo: } Y(t) = T(t) * S(t) * C(t) * I(t) \quad (\text{Ec. 20})$$

$$\text{Modelo aditivo: } Y(t) = T(t) + S(t) + C(t) + I(t) \quad (\text{Ec. 21})$$

En donde  $Y(t)$  es la observación y  $T(t), S(t), C(t)$  y  $I(t)$  son respectivamente la tendencia, la estacionalidad, el ciclo y las variaciones irregulares en un tiempo determinado  $t$ .

Los modelos multiplicativos están basados en asumir que los componentes de una serie de tiempo no son necesariamente independientes, es decir, que afectan directamente al otro, por otro lado, el modelo de sumas asume que los cuatro componentes que componen a la serie son independientes entre sí.

### **2.1.9.2. Prueba de Man Kendall para detectar tendencias**

El propósito de la prueba de Mann-Kendall abreviada usualmente como (MK), es evaluar estadísticamente si existe una tendencia monótonica hacia arriba o hacia abajo de la variable de interés a lo largo del tiempo. Una tendencia monótonica hacia arriba significa que la variable aumenta constantemente (visualmente disminuye) a lo largo del tiempo, pero a su vez ejemplifica si tendencia puede o no ser lineal.

La prueba Mann Kendall se puede usar en lugar de un análisis de regresión lineal paramétrico, que se puede usar para probar si la pendiente de la línea de regresión lineal estimada es diferente de cero. Esto porque el análisis de regresión requiere que los residuos de la línea de regresión ajustada se distribuyan normalmente; una suposición no requerida por la prueba, es decir, es no paramétrica es decir sin distribución (Cryer y Chan, 2008).

Las hipótesis a probar por la prueba de Mann Kendall son las siguientes:

*H<sub>0</sub>: No existe una tendencia asociada a la variable estudiada.*

*H<sub>A</sub>: Existe una tendencia de cualquier tipo, asociada a la variable estudiada.*

Adicional la prueba de Mann Kendall se basa en un coeficiente para la estimación de la dirección futura de la tendencia estudiada, esta es  $\tau_k$  en dónde si:

$0 \geq \tau_k \leq 1$  la tendencia proyecta hacía el crecimiento futuro, es decir una tendencia positiva.

$-1 \geq \tau_k \leq 0$  la tendencia proyecta a la caída futura, es decir a una tendencia negativa.

### **2.1.9.3. Prueba de estacionalidad de Dickey Fuller**

La prueba de Dickey Fuller es una prueba estadística, que busca validar la hipótesis nula de que una raíz unitaria está presente en una muestra de serie temporal. La hipótesis alternativa es diferente según la versión de la prueba utilizada, pero generalmente es estacionariedad o tendencia-estacionaria, es decir que existe al menos una parte estacional en la serie estudiada (Cryer y Chan, 2008).

Es una versión aumentada de la prueba Dickey para un conjunto de modelos de series temporales más grande y complicado.

La estadística aumentada de Dickey-Fuller (ADF), utilizada en la prueba, es un número negativo. Cuanto más negativo es, más fuerte es el rechazo de la hipótesis de que hay una raíz unitaria en algún nivel de confianza.

Las pruebas de hipótesis a validar por medio de la prueba de Dickey Fuller son las siguientes:

$H_0$ : La serie de tiempo no tiene un componente estacional.

$H_A$ : La serie de tiempo tiene al menos un componente estacional.

#### **2.1.9.4. Modelos de series de tiempo Autorregresivos integrados de medias móviles (ARIMA)**

El modelo de autorregresivo integrado de media móvil (ARIMA) Se expresa como:

$$y'_t = I + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} \dots + \alpha_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_q \varepsilon_{t-q} \text{ (Ec. 22)}$$

Donde  $\alpha$  hace referencia al estimador del modelo para cada momento en el que se considera a la variable de salida,  $y$  al momento en cuestión para el cual se está estimando a la variable, además  $\varepsilon$  al error o ruido blanco derivado de cada atraso que el modelo genere (Cryer y Chan, 2008).

La parte AR de ARIMA muestra que la serie temporal se regresa en sus propios datos pasados. La parte MA de ARIMA indica que el error de pronóstico es una combinación lineal de errores respectivos pasados. La parte I de ARIMA muestra que los valores de los datos se han reemplazado con valores diferenciados de orden  $d$  para obtener datos estacionarios, que es el requisito del enfoque del modelo ARIMA.

El modelo ARIMA es efectivo para ajustar datos valuados con este enfoque combinado y ayuda a pronosticar puntos futuros en una serie de tiempo.

La ecuación 22 muestra que los predictores son los puntos de datos  $p$  rezagados para la parte autorregresiva y los errores  $q$  rezagados son para la

parte de promedio móvil, que están todos diferenciados. La predicción es el  $y_t$  diferenciado en el orden  $d$ . Esto se llama el modelo ARIMA ( $p, d, q$ ). Estimar los coeficientes  $\alpha$  y  $\theta$  para un determinado  $p, d, q$  es lo que hace ARIMA cuando aprende de los datos de un conjunto de entrenamiento en una serie de tiempo. Especificar  $p, d, q$  puede ser complicado (y una limitación clave) pero dentro de la metodología de trabajo de estos modelos se pueden probar diferentes combinaciones y evaluar el rendimiento en relación al error.

Para los modelos ARIMA se necesita especificar el valor de  $p, d, q$ , estos coeficientes necesitan ser estimados de forma iterativa. La forma más común de estimar es a través de la estimación de máxima verosimilitud. Es similar a la estimación de mínimo cuadrado para la ecuación de regresión, excepto que un estimador por máxima verosimilitud encuentra los coeficientes del modelo de tal manera que maximiza las posibilidades de encontrar los datos reales, este ejercicio suele realizarse de forma óptima por medio de ejercicios de aprendizaje de máquina (Cryer y Chan, 2008).

#### **2.1.9.5. Planteamiento del modelo ARIMA**

Se describe un modelo ARIMA planteado de forma teórica, es decir, en función de las características intrínsecas de los datos estudiados, y no con el fin de optimizar algún parámetro de criterios de información estadística, los motores de análisis de datos que actualmente se encuentran en el mercado suelen crear una función de optimización para el criterio de información de Akaike y encontrar un modelo ARIMA de forma automática (Cryer y Chan, 2008).

Para plantear esta clase de modelos se debe realizar una combinación de un modelo mixto, con una parte autorregresiva (AR), además de una parte de medias móviles (MA).

Se debe considerar que los modelos autorregresivos siempre son la representación de un proceso aleatorio, que tiene la capacidad de caracterizar una cantidad considerable de datos a través del tiempo a pesar de que tenga una varianza muy alta.

Se sabe de forma teórica que la forma general de un modelo estocástico AR está definida cómo:

$$X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \text{ (Ec. 23)}$$

En donde  $\phi_{i,p}$  son parámetros del modelo,  $c$  es una constante y  $\varepsilon$  hace referencia a un ruido blanco.

Utilizando la estructura del modelo de series temporales anteriormente planteado, en donde todos los valores  $t$  son enteros, asumiendo que el error es independiente y distribuido uniformemente por toda la serie de datos, y que la distribución de la muestra se comporta (gracias al teorema del límite central) con media cero.

$$\varepsilon_t = N(0, \sigma^2) \text{ (Ec. 24)}$$

Se pueden integrar ambas partes del modelo, tanto su parte autorregresiva como su parte de medias móviles para generar un modelo ARIMA.

Dadas las consideraciones anteriores del estudio, el comportamiento de la serie de tiempo en sus retrasos o rezagos tiene un comportamiento más o menos uniforme, esto quiere decir que a pesar de que los atrasos se distribuyan de forma abrupta en su totalidad, tienden a normalizarse de forma parcial (Francés, 2018).

$$\text{Rezago} = L \text{ (Ec. 25)}$$

Entonces:

$$X = \{X_1, X_2, \dots\} \text{ (Ec. 26)}$$

En dónde:

$$LX_t = X_{t-1} \text{ siempre que } t > 1 \text{ (Ec. 27)}$$

Utilizando la teoría de modelos autorregresivos, y que todos los rezagos tienen un comportamiento polinomial, además de que tienen una raíz unitaria, es decir un factor del retraso anterior  $(1 - L)$  en dónde  $L$  es el máximo atraso a considerar de la serie, con multiplicidad  $d$  (Cryer y Chan, 2008).

Considerando los factores de alta variabilidad intermensual que presentan las medias entre un atraso inicial y el máximo atraso bajo la multiplicidad de la cantidad de series estudiadas se tiene que:

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) = \left(1 - \sum_{i=1}^{p'-d} \alpha_i L^i\right) (1 - L)^d \dots \text{ (Ec. 28)}$$

Para representar la expresión de forma polinomial y que además tenga en consideración al menos dos atrasos significativos, basado en la propiedad de  $p = p' - d$ , en dónde se supuso que la serie tiene al menos una varianza máxima en cada reinicio de los atrasos, lo cual es una realidad en las fluctuaciones en el precio de cada criptodivisas de forma anual; con ello se mermo la falla en las

aproximaciones del valor real versus el predicho y hacer el modelo más robusto. Generalizando el modelo se puede expresar de la siguiente forma:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (Ec. 29)$$

Utilizando como notación a las siglas  $(p, q, d)$  por literatura general como base para diferenciar el orden del modelo que se está generando, es decir, la parte autorregresiva, las diferenciaciones involucradas y la parte de medias móviles. Además, de considerar un proceso de rezagos intermitente que siempre utilizará de estimador a los rezagos de los atrios de más alta variabilidad para considerar en el modelo.

Por otra parte, Radovanov, *et al.* (2018), plantean un modelo para los rezagos asociados a un modelo GARCH, que también aplican a una metodología de modelos autorregresivos, la cual se define la siguiente expresión para todos los retrasos de la serie de datos:

$$Rezagos = \frac{\delta}{(1 - \sum \varphi_i)} \quad (Ec. 30)$$

En dónde  $\delta$  hace referencia al total de tendencias aparentes de gráfico de correlación total, y  $\varphi_i$  hace referencia los elementos dentro del límite.

De forma que el modelo generalizado para realizar las pruebas de integración de datos y generar predicciones es el siguiente dadas las características de las 6 divisas.

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1-L)^d X_t = \delta + \left(1 + \sum_{i=1}^q \theta_i L^i\right) \epsilon_t \quad (Ec. 31)$$

Donde se puede observar que se pretende mermar el efecto de las alzas y bajas en la variable de respuesta añadiendo un cuantificador de los rezagos contemplando que se mantendrán constantes los saltos de varianza alta a lo largo de toda la serie (Francés, 2018).

### **2.1.10. Modelos de vector autorregresivo para series de tiempo multivariantes**

El modelo de vector autorregresivo (VAR) es un modelo de series de tiempo multivariadas que ha demostrado flexibilidad y practicidad en su uso para brindar conclusiones generalizadas a partir de series temporales multivariantes. Es una extensión natural del modelo autorregresivo univariado dinámico de series de tiempo (Cryer y Chan, 2008).

El modelo VAR ha demostrado ser especialmente útil para describir el comportamiento dinámico de las series de tiempo económicas y financieras y para la previsión de activos bursátiles. A menudo proporciona pronósticos con métricas del error superiores a los modelos univariantes de series temporales.

Los pronósticos de los modelos VAR son bastante flexibles debido a que pueden condicionar las posibles trayectorias futuras de variables especificadas en el modelo. Además de la descripción de datos y la previsión, el modelo VAR también se utilizó para inferencia estructural y análisis de políticas económicas. En el análisis estructural, ciertos supuestos sobre la estructura causal de los datos investigados se imponen, y los impactos causales resultantes de choques inesperados o se resumen las innovaciones de las variables especificadas en las

variables del modelo. Estos impactos causales generalmente se resumen con respuesta de impulso funciones y descomposiciones de varianza de error de pronóstico (Cryer y Chan, 2008).

### 2.1.10.1. Modelo de vector autorregresivo estacionario

Estableciendo  $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$  denote a un vector  $(n * 1)$  de series temporales. De esa forma el vector básico de atrasos para el vector autorregresivo  $VAR(p)$  tiene la siguiente forma:

$$y_t = c + \prod_1 Y_{t-1} + \prod_2 Y_{t-2} + \dots + \prod_p Y_{t-p} + \varepsilon_t, t = 1, \dots, T \text{ (Ec. 32)}$$

En dónde  $\prod_i$  es la matriz de coeficientes  $(n * n)$  y  $\varepsilon_t$  es el vector de un proceso de ruido blanco con una media inobservable de valor cero  $(n * 1)$  sin autocorrelación serial o independencia con la matriz de covarianzas de la serie temporal.

Entonces un proceso bivariado tiene la forma:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & \pi_{12}^1 \\ \pi_{21}^1 & \pi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \pi_{11}^2 & \pi_{12}^2 \\ \pi_{21}^2 & \pi_{22}^2 \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \text{ (Ec. 33)}$$

En dónde la covarianza entre  $(\varepsilon_{1t}, \varepsilon_{2t}) = \sigma_{12}$ . Se debe prestar especial importancia a que cada ecuación tiene los mismos regresores bajo la salvedad del atraso comprendido entre  $y_{1t}$  y  $y_{2t}$ . Por lo tanto, el modelo VAR es solo un modelo de regresión aparentemente no relacionado (SUR) con variables y términos deterministas como regresores comunes.

En la notación del operador de atraso, el VAR (p) se escribe como:

$$\prod(L) Y_t = c + \varepsilon_t \quad (Ec. 34)$$

En dónde  $\prod(L) = I_n - \prod_1(L) - \dots - \prod_p(L)^p$  es decir, el modelo  $VAR(P)$  es estable si las raíces del determinante  $(I_n - \prod_1(Z) - \dots - \prod_p(Z)^p) = 0$  se encuentran fuera del círculo unitario complejo (tienen un módulo mayor que uno) o, de manera equivalente, si los valores propios de la matriz complementaria (Cryer y Chan, 2008).

$$F = \begin{pmatrix} \Pi_1 & \Pi_2 & \dots & \Pi_n \\ I_n & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & I_n & 0 \end{pmatrix} \quad (Ec. 35)$$

La última matriz tiene un módulo menor que uno. Suponiendo que el proceso se ha inicializado en el pasado infinito, entonces un proceso VAR (p) estable es estacionario y ergódico con medias, varianzas y autocovarianzas invariantes en el tiempo.

Si el proceso es de covarianza estacionario, entonces la media incondicional es dada por:

$$Y_t - \mu = \Pi_1(Y_{t-1} - \mu) + \Pi_2(Y_{t-2} - \mu) + \dots + \Pi_p(Y_{t-p} - \mu) + \varepsilon_t \quad (Ec. 36)$$

El modelo básico VAR (p) puede ser demasiado restrictivo para representar de forma eficiente las principales características de los datos. En particular, otros términos deterministas tales como una tendencia de tiempo lineal o variables ficticias estacionales pueden ser necesarias para representar los datos correctamente. Además, las variables exógenas estocásticas también pueden

ser necesarias. La forma general del modelo VAR (p) con términos deterministas y variables exógenas viene dada por:

$$Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \dots + \Pi_p Y_{t-p} + \Phi D_t + G X_t + \varepsilon_t \quad (Ec. 37)$$

Donde  $D_t$  representa una matriz  $(I * 1)$  de componentes deterministas,  $X_t$  representa una matriz  $(m * 1)$  de variables exógenas, y  $\Phi$  y  $G$  son matrices de parámetros (Cryer y Chan, 2008).

### 2.1.10.2. Funciones de impulso respuesta

Una función impulso-respuesta describe la evolución de la variable de interés a lo largo de un horizonte de tiempo especificado después de un choque en un momento dado. Para hacer que el pensamiento sea más intuitivo, se debe centrar primero en el caso de un proceso AR (1) univariado (Cryer y Chan, 2008).

$$x_t = \phi x_{t-1} + u_t \quad (Ec. 38)$$

En dónde  $x_t$  es un escalar,  $\phi < 1$  (lo que lo vuelve estacionario) y  $u_t$  es un escalar aleatorio sin varianza y de media 0.

Para representar el efecto de las medias móviles dado que el proceso anterior es estacionario, podemos encontrar la representación de media móvil infinita de  $L$  como el operador de retardo tal que  $LX_t = X_{t-1}$  y  $L^2 X_t = X_{t-2}$  bajo esa premisa se tiene la siguiente expansión

$$\begin{aligned}
x_t &= \phi x_{t-1} + u_t \\
(1 - \phi L)x_t &= u_t \\
x_t &= (1 - \phi L)^{-1}u_t \\
x_t &= \frac{1}{1 - \phi L}u_t && \text{(Ec. 39)} \\
x_t &= (\phi L)^0 u_t + (\phi L)^1 u_t + (\phi L)^2 u_t + (\phi L)^3 u_t + \dots \\
x_t &= u_t + \phi u_{t-1} + \phi^2 u_{t-2} + \phi^3 u_{t-3} + \dots
\end{aligned}$$

Supongamos ahora que  $x_t$ , en lugar de ser un escalar, es un vector columna con dimensiones  $(n * 1)$ , es decir, ahora tenemos un VAR (1) en lugar de un AR (1). Sin embargo, estamos interesados en la evolución de  $x_t$  después de un choque estructural, más que después de una innovación en la proyección de uno de los integradores. Si se piensa en  $x_t$  como innovaciones de forma reducida tal que una combinación mixta de algunos choques estructurales  $t$ , se puede asumir la siguiente relación

$$u_t = B\varepsilon_t \quad \text{(Ec. 40)}$$

Donde  $B$  es una matriz de  $(n*n)$  y  $t$  es un vector de columna  $(n*1)$  que contiene el  $n$  estructural del choque (la relación entre esta matriz  $B$  y la matriz  $A_0$  que ahora pueden evidenciarse como un conjunto concatenado de conjuntos es decir  $B = A_0^{-1}$ ). Entonces podemos escribir la representación de la media móvil anterior como

$$\begin{aligned}
x_t &= B\varepsilon_t + \phi B\varepsilon_{t-1} + \phi^2 B\varepsilon_{t-2} + \phi^3 B\varepsilon_{t-3} + \dots \\
x_t &= C_0\varepsilon_t + C_1\varepsilon_{t-1} + C_2\varepsilon_{t-2} + C_3\varepsilon_{t-3} + \dots && \text{(Ec. 41)}
\end{aligned}$$

### **2.1.11. Modelos de aprendizaje de máquina**

Es un subcampo de las ciencias computacionales, el cual converge en el estudio y reconocimiento de patrones, así como en el estudio del aprendizaje computacional en la inteligencia artificial.

El aprendizaje de máquina (o *machine learning*) por sus siglas en inglés ML, explora la construcción de algoritmos que pueden aprender de una serie específica de información y generar inferencias o predicciones relacionadas a las mismas (Kuhn y Johnson, 2013).

Estos algoritmos están contruidos en función de una serie de datos definidos, es decir información histórica de un suceso, y operan creando un modelo a raíz de ejemplo de variables y datos de entrada en función de crear mejores aproximaciones a predicciones o decisiones más eficientes.

El aprendizaje de máquina suele confundirse o en muchos casos sobreponerse a las estadísticas computacionales, es una disciplina que se enfoca en la modelización de predicciones, además de operaciones de optimización matemática por computación.

#### **2.1.11.1. Tipos de problemas y tareas**

El aprendizaje de máquina suele ser categorizado en tres categorías principales, esto depende mucho de la categoría o naturaleza del problema a resolver. Estas categorizaciones están encapsuladas por el aprendizaje supervisado y el aprendizaje no supervisado (Kuhn y Johnson, 2013).

### **2.1.11.2. Aprendizaje supervisado**

Al algoritmo se le presentan las variables de entrada deseadas y ejemplos del tipo de variable de salida que se desean, por regla general se conoce a este método como que sigue un algoritmo de aprendizaje en función de una muestra o una referencia, en dónde la prioridad será que después de aprender del conjunto de entrenamiento pueda igualar dichos resultados con otro tipo de información. Kuhn y Johnson (2013) mencionan que el aprendizaje supervisado es ideal para problemas aplicativos de estadística computacional, en dónde todas las combinaciones posibles de parámetros de un modelo llevarían a una cantidad significativamente grande de posibles soluciones, por ende, su aplicación optimiza los tiempos de procesado.

### **2.1.11.3. Aprendizaje no supervisado**

No se le brindan ningún tipo de etiqueta ni tampoco algún tipo de guía al algoritmo, dejándolo únicamente con un conjunto definido de variables de entrenamiento.

El aprendizaje no supervisado suele tener asociado un coste computacional muchísimo más alto, Kuhn y Johnson (2013) mencionan que esto se debe a que se deja procesar infinitas o al menos una cantidad no parametrizada, de iteraciones, las cuales podrían o no llegar a converger en un resultado que resulte conveniente en términos de optimizar (maximización o minimización, usualmente del error) de una determinada variable respecto a un ajuste. El aprendizaje no supervisado es muy utilizado para encontrar patrones en los datos que no pueden ser apreciados a simple vista.

#### **2.1.11.4. Algoritmos de clasificación**

En el aprendizaje automático, la clasificación es un enfoque de aprendizaje supervisado en el que el programa informático aprende de los datos de entrada y luego utiliza este aprendizaje para clasificar nuevas observaciones. Este conjunto de datos puede ser simplemente binario o puede ser multiclase (Kuhn y Johnson, 2013).

Dentro de los algoritmos que se utilizarán están comprendidos los algoritmos de árboles de clasificación, esto con el fin de estimar los parámetros para los modelos autorregresivos planteados, de forma que presenten el menor valor de estimación posible con los datos de entrada.

#### **2.1.11.5. Árboles de clasificación**

El árbol de decisión crea modelos de clasificación o regresión en forma de estructura de árbol. Desglosa un conjunto de datos en subconjuntos cada vez más pequeños, mientras que al mismo tiempo se desarrolla un árbol de decisión asociado. El resultado final es un árbol con nodos de decisión y nodos hoja.

Un nodo de decisión tiene dos o más ramas y un nodo hoja representa una clasificación o decisión. El nodo de decisión superior en un árbol que corresponde al mejor predictor llamado nodo raíz. Los árboles de decisión pueden manejar datos categóricos y numéricos (Kuhn y Johnson, 2013).

Existen dos tipos fundamentales de árboles de decisión estadísticos aplicables a problemas con información de índole financiera, estas son:

- Árboles de clasificación para variables de índole categórica, en dónde la función tiene como objetivo la clasificación de una variable del tipo factor.
- Árboles de clasificación para una variable del tipo continua, en dónde la función se optimiza para poder clasificar una variable continua.

Kuhn y Johnson (2013) mencionan que para entornos tan variables como lo son los escenarios de variación financiera, el criterio fundamental para la toma de decisión y clasificación de estos árboles es la entropía. La entropía es una medida de la aleatoriedad de la información que se procesa. Cuanto mayor sea la entropía, más difícil será sacar conclusiones.

Matemáticamente la entropía esta descrita por la siguiente expresión:

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (Ec.42)$$

En dónde  $p_i$  hace referencia a la frecuencia probabilística dentro de un modelo estadístico de un elemento de clase  $i$  en la muestra analizada.

## 2.2. Criptodivisas

Introducido y documentado por primera vez en 2009, *bitcoin* es una forma de criptodivisa o activo virtual, un sistema de pago electrónico basado en pruebas criptográficas, en lugar del sistema financiero de banca central.

Arango (2018) señala que la compra y venta en línea se ha vuelto dependiente casi exclusivamente de instituciones financieras que actúan como

terceros confiables para procesar pagos electrónicos. En otras palabras, los pagos de transacciones en línea deben pasar por una empresa, como un banco o un emisor de tarjetas de crédito, para que se verifiquen factores como el fraude y el pago exitoso. Este tipo de sistema se basa en la confianza; sin embargo, estos activos están centralizados por múltiples entes y sistemas, de forma en que toda transacción realizada puede regularse.

En cuanto a los activos virtuales se encuentran encriptados, descentralizados y tienen valor adquisitivo, lo cual denota que pueden realizarse transacciones del todo anónimas sin una traza específica y sin costos asociados a la intermediación de instituciones financieras. “La mayoría de criptodivisas no son un pasivo de ningún agente (público o privado); no tienen estatus de ser de curso legal, no son reconocidos como divisas según definiciones del FMI, es decir no son emitidos por un gobierno soberano” (Arango, 2018, p. 3).

Desde una perspectiva más amplia, los activos virtuales no están actualmente controlado por un organismo de gobierno central, lo que reduce las preocupaciones de privacidad. Además, no están vinculado con ningún tipo de mercancía, por ejemplo, oro o plata como en fenómenos económicos tradicionales (Arango, 2018).

### **2.2.1. Cadena de bloques**

Para saber cuánto dinero tiene cada usuario en el sistema financiero tradicional se utilizan los registros históricos en un núcleo informático bancario; donde se generan estados de cuenta. En cuanto a los activos virtuales se necesita un registro de todas las transacciones de cada criptodivisas, y además que este registro esté disponible de forma pública y con toda la trazabilidad que ha tenido cada activo virtual a lo largo de la historia.

Este registro público que contiene toda la información histórica, bajo la salvedad de que la información sensible que se encuentra completamente encriptada por medio de algoritmos de criptografía, se encuentra a cargo en su totalidad de los usuarios del sistema y, se le conoce como cadena de bloques.

Se debe recordar que cada una de las cuentas son anónimas, si bien existe la traza, no se sabe a quién pertenecen realmente. Muchas personas trabajan desarrollando las transacciones que genera cada activo virtual, a estas personas que añaden registros a la cadena de bloques se les conoce como Mineros de datos y ellos obtienen a cambio por su trabajo criptodivisas (Cocco *et al.*, 2015).

## **2.2.2. Activos virtuales considerados en el estudio**

Se seleccionaron las criptodivisas que cubren los volúmenes de capital más significativos, así como los de las proyecciones más favorables en apreciación para este estudio, debido a que como expone CoinMarketCap (2019) ya que entre ellas conforman el 90 % del mercado de capitalizaciones a octubre de 2019. En el siguiente apartado se dará una pequeña descripción de las seis divisas seleccionadas.

### **2.2.2.1. Bitcoin**

*Bitcoin* fue la primera realización de la idea de un nuevo tipo de moneda, en otras palabras, el primer activo virtual completamente descentralizado.

Esta descentralización significa que la red de *bitcoin* es controlada y adueñada por sus propios usuarios, y todo el set de usuarios deben adherirse a

las mismas reglas, es decir un sistema cooperativo que beneficia a todos, sin la necesidad de la intermediación de un ente financiero (Arango, 2018).

*Bitcoin* utiliza la tecnología de la cadena de bloques, lo cual deja un registro único y permanente de cada transacción involucrada con el activo virtual, así mismo del proceso de autenticación transaccional, entonces se puede inferir que no existirán nunca dos divisas exactamente iguales y a su vez no pueden ser clonadas o duplicadas dada la complejidad del algoritmo de la cadena de bloques.

#### **2.2.2.2. *Dash***

En su concepción conocida como *darkcoin* o *xcoin* es una divisa de transacciones instantáneas con privacidad centralizada en la misma, a pesar de que está basada en *bitcoin*, y comparte muchas de sus propiedades esta es ínfimamente distinta (CoinMarketCap, 2019).

#### **2.2.2.3. *Litecoin (LTC)***

Es una divisa que se creó durante el año 2011 con el soporte de toda la comunidad de *bitcoin*, en un principio sería una homóloga, aunque se tornó en un camino completamente distinto. Está basada en el mismo protocolo *de peer-to-peer* (P2P) que *bitcoin*, presenta mejoras significativas respecto a las deficiencias principales de este último las cuales eran la poca disponibilidad de su registro histórico, un validador de datos y su dificultad de formación (CoinMarketCap, 2019).

#### **2.2.2.4. Monero (XMR)**

Es la divisa segura, privada e irrastreable, como denota en su marca promocional, tiene su fuerte en la descentralización y el índice de crecimiento de la misma. Fue creada en abril del 2014 (CoinMarketCap, 2019).

La divisa está completamente basada en donaciones, es decir la comunidad de minería de datos han generado completamente el trasfondo de la operación de la moneda, con aportes de mejora continua; asociada al funcionamiento de criptodivisas obtenidas a lo largo de la historia (CoinMarketCap, 2019).

#### **2.2.2.5. Ripple**

Originalmente un concepto de 2012, fue la primera divisa en tener un estatuto estándar de funcionamiento a nivel global, con un tipo de red en particular (RTGS), lo cual significa que permite los bancos u otras entidades financieras enviar, en tiempo real, pagos internacionales por medio de la red especializada para tal fin.

En resumen, comparada con *bitcoin* es una divisa que ha sido mucho más centralizada y para un uso más particular, pero no pierde ninguna de las características que por naturaleza tiene cada criptodivisa. Al estar centralizada por tantas entidades financieras, se tiene la ventaja que se tiene mayor control sobre su uso por lo que no está sujeta a una volatilidad tan impredecible. Una de sus desventajas es que, por su uso tan específico, no es una moneda tan atractiva para otro tipo de usuarios, además que tiene un proceso de autorización de información durante su minería tan complicado que llega a ser hasta redundante (CoinMarketCap, 2019).

#### **2.2.2.6. *Ethereum***

*Ethereum* representa una cadena de bloques que tiene un desarrollo completo en cuanto a su propio lenguaje de programación, provee un diagrama abstracto en dónde cualquiera puede crear sus propias reglas de pertenencia, por lo que ninguna divisa de *ethereum* tendrá una relación directa con los métodos de transferencia de otra. Esto se volvió muy interesante en los contratos inteligentes en negocios en los que no se deseaba tener ningún tipo de trazabilidad (CoinMarketCap, 2019).

#### **2.2.3. Métricas de red en la cadena de bloques**

La estructura de la cadena de bloques está basada en métricas de red, es decir el cómo están cuantificados todos los parámetros de cada divisa dentro de la cadena de bloques para realizar transacciones.

Estas métricas funcionan como criterios de agrupación para las divisas utilizadas en este estudio, en otras palabras, es la forma en la que pueden diseñarse grupos homogéneos para la consolidación de los volúmenes de capital de los mercados financieros.

Klenke (2014) menciona que dentro de las métricas de red consideradas como significativas para valorar la homogeneidad entre divisas, se pueden mencionar los grados de distribución y leyes de poder, los conglomerados de criptodivisas a raíz de la homogeneidad de distribuciones, la ruta más corta hacia la cual se puede concentrar el volumen de capital y la centralidad individual que posee cada divisa.

Generalizando el funcionamiento de la cadena de bloques se debe de considerar un enfoque micro, es decir el de un eslabón de la cadena para comprender el cómo funciona la estructura completa, Klenke (2014) menciona que el grado de distribución no es más que el cálculo de pesos matemáticos que capturan la estructura de la red en un momento definido cualquiera.

Esta fotografía numérica de la red genera toda la evidencia necesaria para certificar que cada nodo de la cadena se encuentra acorde al histórico de la misma, y las nuevas entradas o salidas son auténticas y no existe fraude, cabe mencionar que esta validación la realizan en paralelo distintos mineros a lo largo del día. Bakar y Rosbi (2017) mencionan que pueden llegar a ser miles de millones de validaciones lo cual brinda robustez al método, como si se tratase de un sistema de cumplimiento bancario generado por los mismos usuarios.

Gracias a la validación del cumplimiento de los supuestos de la red, una metodología estadística de aprendizaje automático es aplicada, esto con el fin de diferenciar la transaccionalidad de cada divisa por separado y asociar el valor que está generando cada una específicamente al nodo al que corresponde, esta metodología denominada como conglomerados funciona para poder inferir el valor real asociado a cada nodo que origina cada divisa, de esa forma es posible apreciar la inflación o deflación que genera cada una, Klenke (2014) expone que como es lógico el conglomerado de nodos de *bitcoin* es significativamente más grande al resto y denota el alza de valor adquisitivo que han tenido estos activos virtuales.

Por otra parte, Bakar y Rosbi (2017) mencionan que para que una transacción con activos virtuales sea eficiente debe ser igual de rápida que realizar una transacción con cualquier institución financiera, sin embargo, dada la complejidad de la cadena de bloques se optó por algoritmos estadísticos de

aprendizaje automático, en el caso concreto de la cadena de bloques se utiliza la metodología de la ruta más corta; definida matemáticamente como la distancia euclidiana más corta entre los últimos nodos generados por una divisa en concreto, de esta forma a pesar de mantener un total anonimato si es posible transferir un activo virtual entre carteras, como si se tratase de una tarjeta de crédito en un tiempo relativamente homogéneo.

Finalmente, Klenke (2014) menciona la centralidad de las criptodivisas como la medida de la influencia sobre el mercado bursátil que tiene el conjunto de cada activo virtual individualizado, de esta forma es como puede adquirir o perder valor en el mercado de divisas, medido exclusivamente por la cantidad de nodos proporcionales comparados de un momento a otro, es decir, entre más transacciona, tiende a apreciarse más (Bakar y Rosbi, 2017).

#### **2.2.4. Ratio de retorno sobre la inversión**

El retorno sobre la inversión o ROI (por sus siglas en inglés) es una medida del desempeño de una inversión, en otras palabras, calcula la eficiencia monetaria de una inversión para poder realizar una comparativa teórica con otras inversiones, para poder realizar un cálculo eficiente del retorno sobre la inversión.

Rachlin (2019) expone que se necesitan dos parámetros los cuales en términos generales son, el retorno o beneficio neto que está otorgando la inversión además del coste asociado a haber realizado dicha inversión, el resultado es presentado como un ratio o un porcentaje.

En la literatura el retorno sobre la inversión puede ser definido como una métrica (financiera o no financiera) del retorno, efecto o impacto en cuanto al

resultado monetario de algún negocio o unidad de negocio (Castro y Forero, 2018).

Para el cálculo del retorno de la inversión se pueden considerar muchas variantes, tal es el caso de la unidad de negocio que se esté evaluando y el tipo de unidades de ingreso que está manejando. Por ende, de forma general se puede establecer que el modelo general del retorno sobre la inversión está definido por:

$$ROI |v| = \omega\{ROI[E], B\} \text{ (Ec. 43)}$$

Generalizando la ecuación para cualquier tipo de unidad de negocio que genere beneficio del tipo monetario tenemos que:

$$ROI = \frac{\sum FinRet(i) - \sum Costo(i)}{\sum_i^j Costo(i)} \text{ (Ec. 44)}$$

En dónde  $FinRet(i)$  y  $costo(i)$  hacen referencia al beneficio neto y a los costos de operación respectivamente.

### **2.2.5. Análisis de valor en riesgo VaR**

El valor en riesgo, o VaR, intenta medir el riesgo de cambios inesperados en los precios o la tasa de retorno de registros dentro de un período determinado.

Es una forma muy simple y popular de medir el riesgo de mercado. Dado que proporcionará una cantidad que resume la mayoría del riesgo de mercado al que se enfrenta una inversión, VaR se ha convertido en una implementación

necesaria en cualquier gestión profesional de riesgos corporativos (Tsyvinski, 2018).

El VaR no solo es aplicable para explorar el riesgo de mercado sino también en la gestión de todos los demás tipos de riesgo. En síntesis, el sistema está diseñado principalmente para riesgos financieros de gestión y cumplimiento.

Es ampliamente utilizado por la mayoría de las instituciones financieras, bancos comerciales y bancos de inversión para estimar la pérdida potencialmente máxima de sus 2 carteras durante un período de tiempo dado para una condición de mercado determinada. Por otro lado, desde el punto de vista de cumplimiento, el VaR se puede definir como un margen establecido, que es el mínimo pérdida bajo el intervalo de confianza de un tiempo horizontal.

Según la definición básica del VaR, es la pérdida potencial máxima esperada en la cartera del horizonte de tiempo dado para un intervalo de confianza en condiciones normales de mercado (Tsyvinski, 2008). En otras palabras, hay tres elementos clave para describir el valor en riesgo (VaR):

- El periodo de tiempo.
- El monto a evaluar (Portafolio, activos, entre otros.).
- Una referencia de intervalos normales en el mercado (o en su defecto un intervalo de confianza).

#### **2.2.5.1. Definición matemática del valor en riesgo VaR**

Dado un nivel de confianza de  $p \in 0,1$ , y asumiendo el índice de tiempo de  $t + \alpha$ , desea encontrar el activo de cambio de  $\Delta V (\alpha)$  en la posición financiera

durante el período de tiempo  $\alpha$ . Dejar  $F! x$  sea la función de distribución acumulada (CDF) de  $\Delta V \alpha$ . Dado que la situación financiera es  $\Delta V \alpha \leq 0$ , entonces podemos definir el VaR de una posición larga en el horizonte temporal  $\alpha$  para un  $p$  dado como:

$$p = \mathbb{P}[\Delta V(\alpha) \leq VaR] = F_{\alpha}(VaR) \quad (Ec. 45)$$

Considerando al *momentum* de una posición corta, en un tiempo dado  $\alpha$  con probabilidad  $p$ , y la posición  $\Delta V \alpha \geq 0$ , el VaR se define como:

$$p = \mathbb{P}[\Delta V(\alpha) \geq VaR] = 1 - \mathbb{P}[\Delta V(\alpha) \leq VaR] = 1 - F_{\alpha}(VaR) \quad (Ec. 46)$$

A continuación, se define el  $p$ -cuantil de  $F! x$ , eso para cualquier CDF de  $F! x$  y la confianza dada el nivel de  $p \in 0,1$  es

$$VaR_p = x_p = \inf\{x \mid F_{\alpha}(x) \geq p\} \quad (Ec. 47)$$

En dónde *inf* es el número real más pequeño y  $x_p$  es el equivalente a  $VAR_p$  cuando esta sea conocida.



### **3. PRESENTACIÓN DE RESULTADOS**

#### **3.1. Caracterización estadística de las criptomonedas estudiadas**

Para describir de forma integral el comportamiento histórico de los activos virtuales estudiados, se presenta a continuación un análisis preliminar basado primordialmente en estadísticas descriptivas y análisis gráfico del comportamiento de los valores medios de distintas variables de cada activo, con el fin de brindar al lector un panorama más claro del comportamiento financiero individualizado y su evolución según un momento puntual basado en la referencia temporal del año calendario.

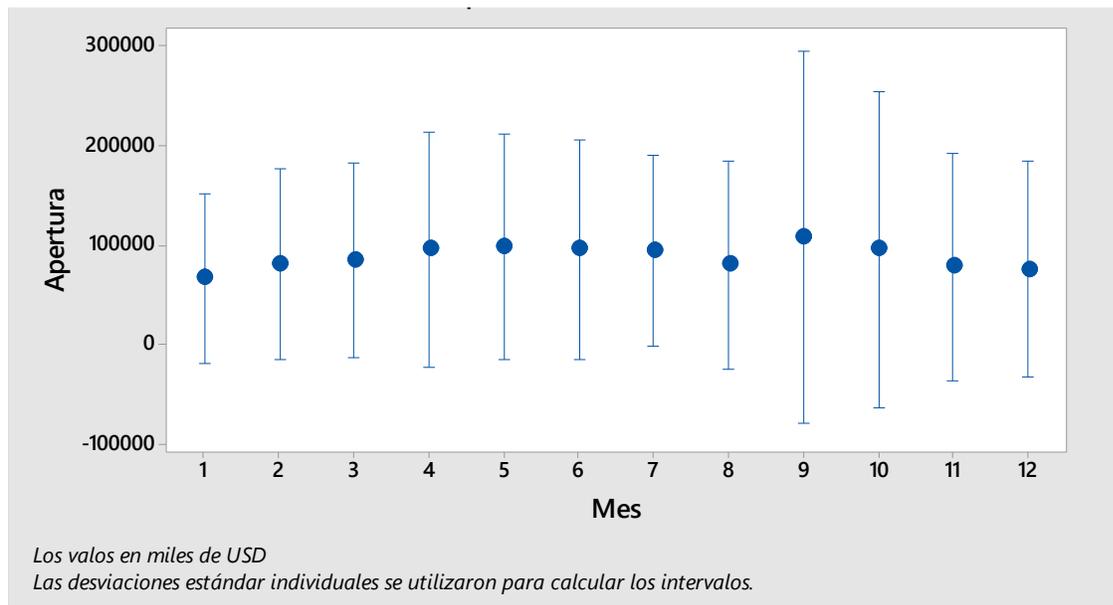
##### **3.1.1. Estadísticas descriptivas de las divisas**

A continuación, se presentan los resultados de las estadísticas descriptivas asociadas a los volúmenes de capital, apertura y cierre en el precio del mercado bursátil de cada divisa, para calcular el mismo se realizó por medio de un conglomerado ponderado mensual utilizando la ecuación 25, con ello se puede reflejar el comportamiento mensual, dentro del horizonte temporal considerado.

##### **3.1.1.1. *Bitcoin***

Se presenta un análisis técnico de las estadísticas descriptivas de *bitcoin*, además de un análisis gráfico de su comportamiento consolidado anual, representado por la amplitud de los intervalos de confianza para las variables bursátiles de precio medio de apertura y cierre.

Figura 2. Intervalos para el valor medio de apertura de *bitcoin*



Fuente: elaboración propia.

Se aprecia una proyección de intervalos de confianza bastante estables en relación a la media, es decir, la moneda ha podido conservar de forma eficiente su valor para el horizonte temporal considerado, adicionalmente tiene una amplitud del intervalo de confianza mucho más grande en los meses de septiembre y octubre.

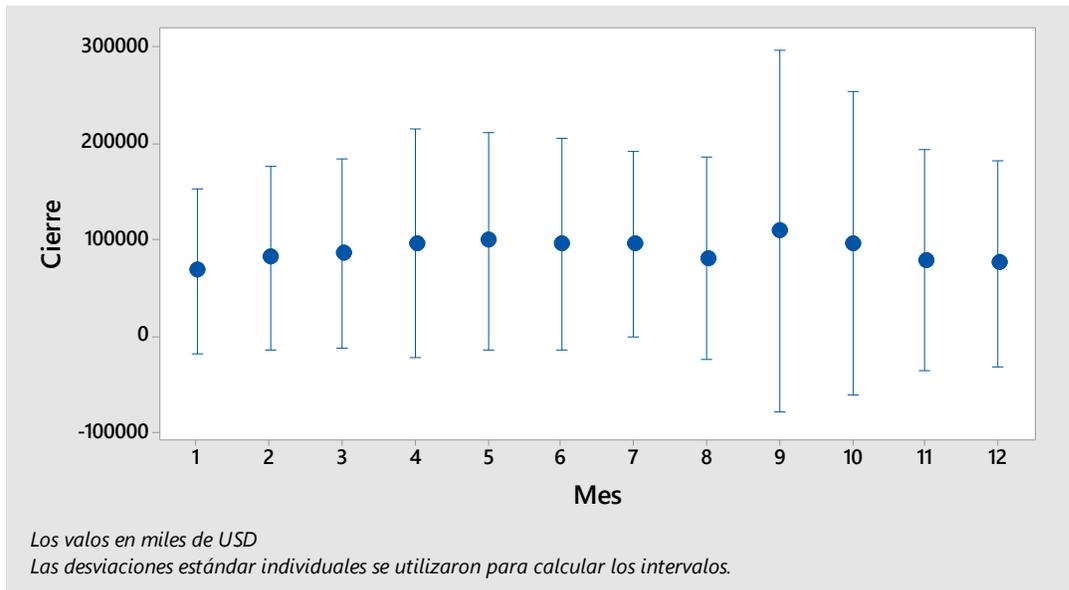
Como menciona Rauchs (2017), esto se debe a que efectivamente las divisas están ganando territorio en la transaccionalidad de compra-venta de activos por internet, en donde las compras de artículos de lujo se concentran según las marcas de tarjetas de crédito en los meses de septiembre y octubre esto a su vez se ve claramente reflejado en los cierres asociados en la figura 2.

Tabla IV. **Estadísticas descriptivas de *bitcoin***

	<b>Valor bursátil de apertura (dólares americanos USD)</b>	<b>Valor bursátil de cierre (dólares americanos USD)</b>
Valores válidos	2371	2371
Valores perdidos	0	0
Media	29027.2	29030.2
Mediana	672.4	673.3
Moda	418.4	104
Desviación estándar	3732.8	3733.5
Varianza	13933470.7	13938684
Asimetría	1.4	1.4
Error estándar de asimetría	0.1	0.1
Curtosis	1.4	1.4
Error estándar de curtosis	0.1	0.1
Rango	19407.3	19429
Mínimo	68.5	68.4
Máximo	19475.8	19497.4
Suma	6940476	6947606.9
Percentiles	Apertura	Cierre
25	370.5	371.3
50	672.4	673.3
75	5251.5	5268.3

Fuente: elaboración propia.

Figura 3. Intervalos para el valor medio de cierre de *bitcoin*

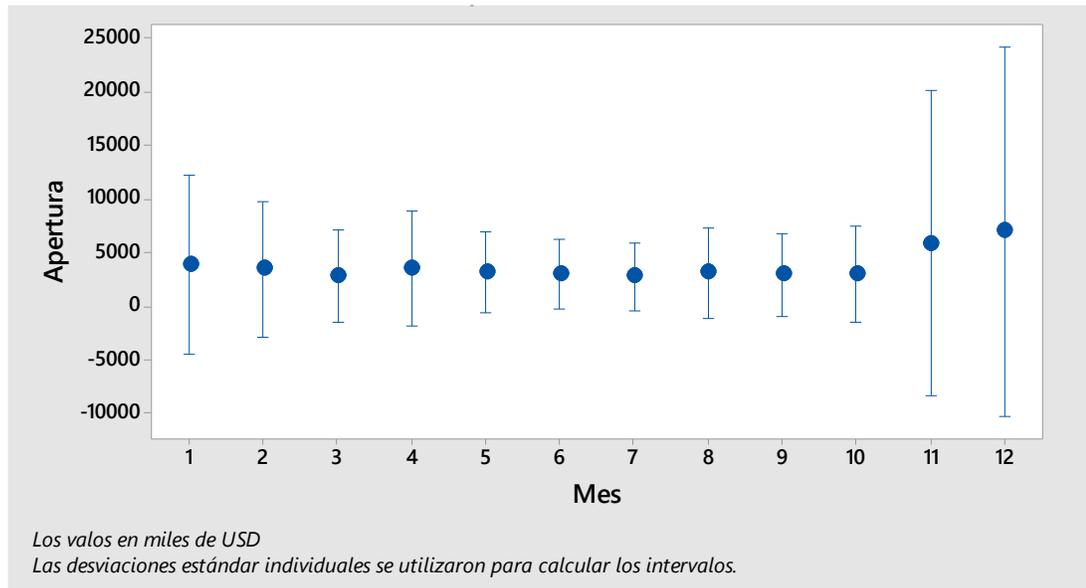


Fuente: elaboración propia.

### 3.1.1.2. *Dash*

A continuación, se presenta un análisis técnico de las estadísticas descriptivas de *dash*, además de un análisis gráfico de su comportamiento consolidado anual, representado por la amplitud de los intervalos de confianza para las variables bursátiles de precio medio de apertura y cierre.

Figura 4. Intervalos para el valor medio de apertura de *dash*



Fuente: elaboración propia.

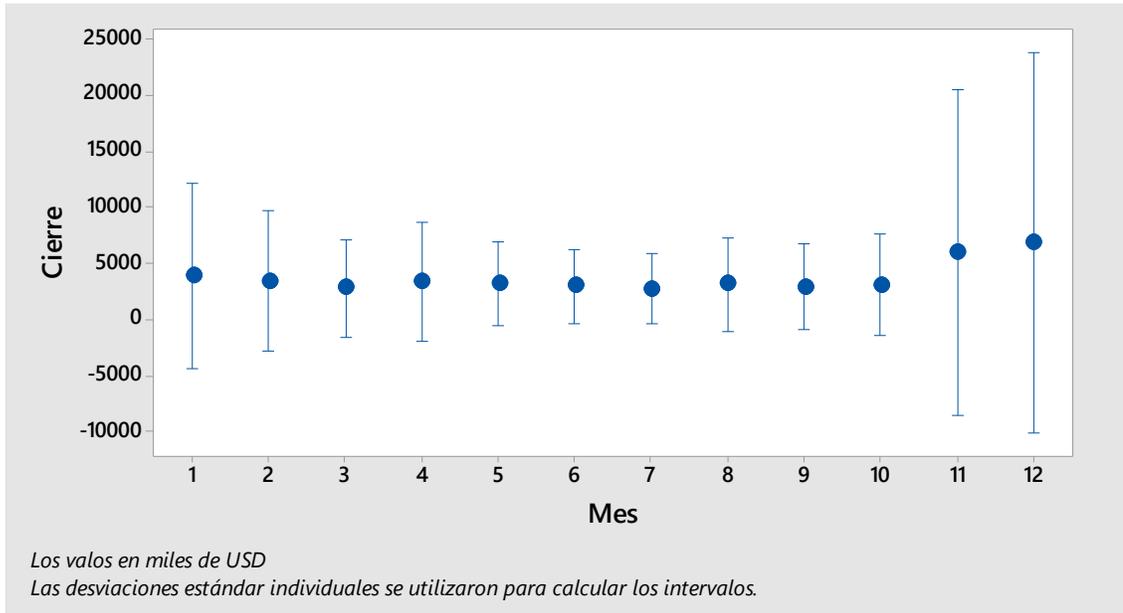
En contraparte con *bitcoin* en la figura 4, se puede apreciar como *dash* abre con valores si bien constantes alrededor de su valor medio, lo hace de una forma más cercana al 0 y con amplitudes más marcadas hacia el eje negativo, es decir, se denotan pérdidas significativas en donde la pérdida es neta en cuanto sus posibles rendimientos, se contrasta de forma homogénea son los valores de cierre que maneja los cuales se aprecian en la figura 4. El diferencial de amplitudes en el intervalo de confianza mensual está de igual forma atado a su volumen transaccional, Rauchs (2017) menciona que *dash* tiene características de moneda de compra digital, por ende, en volúmenes transaccionales más fuertes por festividades mundiales en los meses de noviembre y diciembre, su valor de volumen de capital se tiende a inflar.

Tabla V. **Estadísticas descriptivas de *dash***

	<b>Valor bursátil de apertura (dólares americanos USD)</b>	<b>Valor bursátil de apertura (dólares americanos USD)</b>
Valores válidos	2118	2118
Valores perdidos	0	0
Media	31733.9	31567.4
Mediana	4687	4694
Moda	237	234
Desviación estándar	118578.5	118122.7
Varianza	14060862137	13952969861
Asimetría	5.8	5.8
Error estándar de asimetría	0.1	0.1
Curtosis	33.8	34
Error estándar de curtosis	0.1	0.1
Rango	987089	987454
Mínimo	103	103
Máximo	987192	987557
Suma	67212412	66859688
Percentiles	Apertura	Cierre
25	375.8	375.8
50	4687	4694
75	16846.8	16799.3

Fuente: elaboración propia.

Figura 5. Intervalos para el valor medio de cierre de *dash*

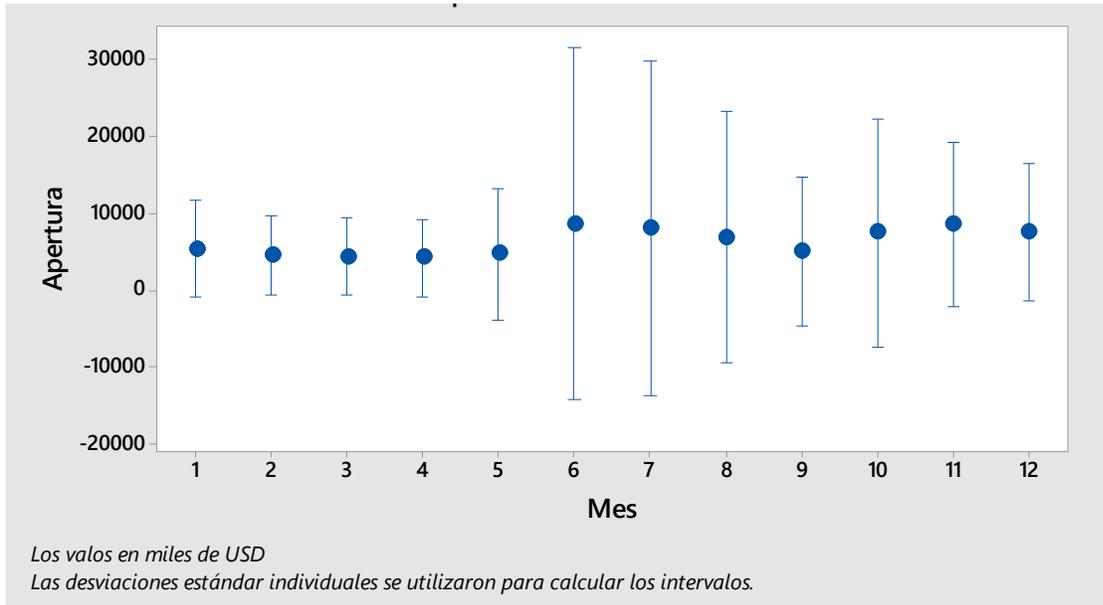


Fuente: elaboración propia.

### 3.1.1.3. *Ethereum*

Se presenta un análisis técnico de las estadísticas descriptivas de *ethereum*, además de un análisis gráfico de su comportamiento consolidado anual, representado por la amplitud de los intervalos de confianza para las variables bursátiles de precio medio de apertura y cierre.

Figura 6. Intervalos para el valor medio de apertura de *ethereum*



Fuente: elaboración propia.

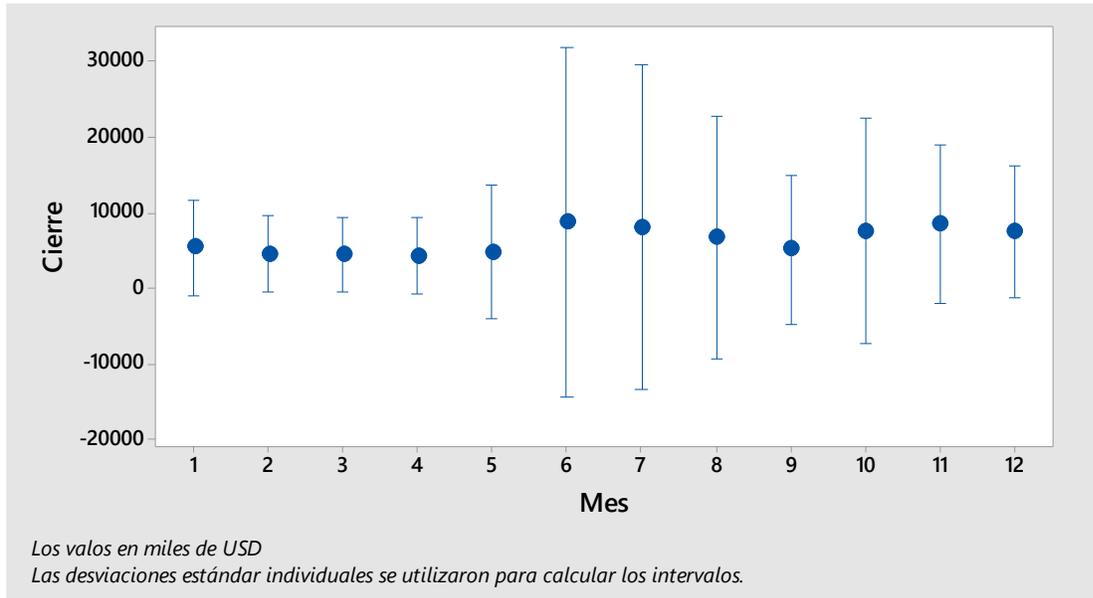
En relación a *ethereum* se puede apreciar una no continuidad; un tanto más marcada en cuanto a la proyección horizontal de la media, lo cual sugiere que sus variaciones tienden a ser más definidas, en la tabla VI se puede apreciar que el rango y por ende la varianza son mucho más inestables que las divisas analizadas anteriormente, por otro lado el comportamiento de los intervalos de confianza sufren del mismo fenómeno de ampliación en meses en concreto, la explicación a ello tiene un supuesto en relación a la transaccionalidad en tecnología, concentrada en los meses centrales del año, Rauchs (2017) menciona que si bien no puede validarse el supuesto dada la naturaleza de las divisas, la mayoría de sitios de tecnología al mayoreo admiten ETH como medio de pago digital.

Tabla VI. **Estadísticas descriptivas de *ethereum***

	<b>Valor bursátil de apertura (dólares americanos USD)</b>	<b>Valor bursátil de cierre (dólares americanos USD)</b>
Valores válidos	1579	1579
Valores perdidos	0	0
Media	80862.5	80873.5
Mediana	18335	18333
Moda	122.00 <sup>a</sup>	1100
Desviación estándar	213569.9	213574.1
Varianza	45612116642	45613917217
Asimetría	3.4	3.4
Error estándar de asimetría	0.1	0.1
Curtosis	9.9	9.9
Error estándar de curtosis	0.1	0.1
Rango	999116	999178
Mínimo	100	100
Máximo	999216	999278
Suma	127681942	127699302
Percentiles	Apertura	Cierre
25	1290	1290
50	18335	18333
75	35004	34789

Fuente: elaboración propia.

Figura 7. Intervalos para el valor medio de cierre de *ethereum*

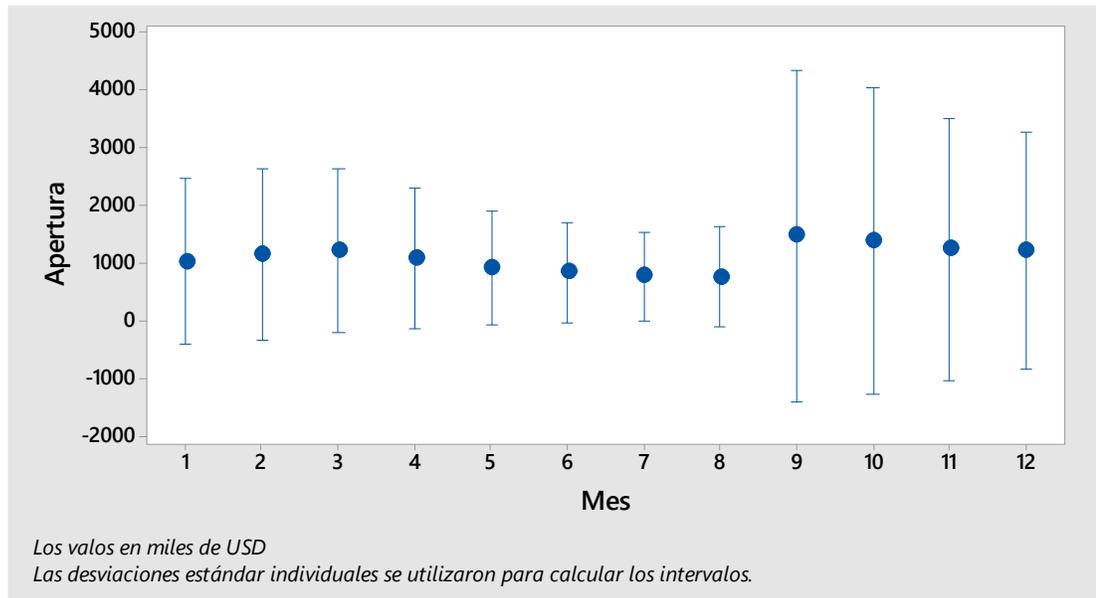


Fuente: elaboración propia.

#### 3.1.1.4. *Litecoin*

A continuación, se presenta un análisis técnico de las estadísticas descriptivas de *litecoin*, además de un análisis gráfico de su comportamiento consolidado anual, representado por la amplitud de los intervalos de confianza para las variables bursátiles de precio medio de apertura y cierre.

Figura 8. Intervalo para el valor medio de apertura de *litecoin*



Fuente: elaboración propia.

Tabla VII. Estadísticas descriptivas de *litecoin*

	Valor bursátil de apertura (dólares americanos USD)	Valor bursátil de cierre (dólares americanos USD)
Valores válidos	2410	2410
Valores perdidos	0	0
Media	3678.9	3679.6
Mediana	837.5	846
Moda	384	384
Desviación estándar	5290.8	5287.5
Varianza	27992287.2	27958091.9
Asimetría	2.3	2.3
Error estándar de asimetría	0	0
Curtosis	6.3	6.3
Error estándar de curtosis	0.1	0.1

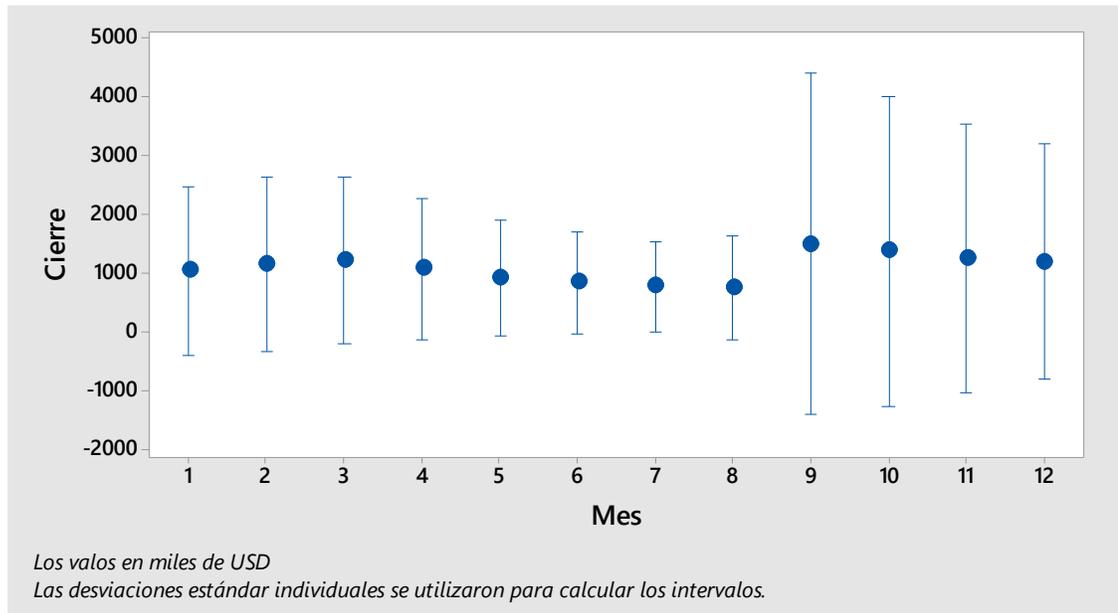
Continuación tabla VII.

Rango	35798	35718
Mínimo	115	116
Máximo	35913	35834
Suma	8866098	8867812
Percentiles	Apertura	Cierre
25	352	352
50	837.5	846
75	5583.3	5579

Fuente: elaboración propia.

Los valores de *litecoin* se encuentran bastante homogéneos a la media, en donde su valor medio de 3678.9 dólares, tabla (VII) se encuentra presente tanto en las amplitudes del intervalo de cierre y apertura, además se aprecia una clara inflación en los últimos meses del año, los cuales desplazan su valor medio, Rauchs (2017) atribuye la fuente de variación al volumen transaccional que presenta la venta de esta criptodivisa en estos periodos del año.

Figura 9. Intervalo para el valor medio de cierre de *litecoin*

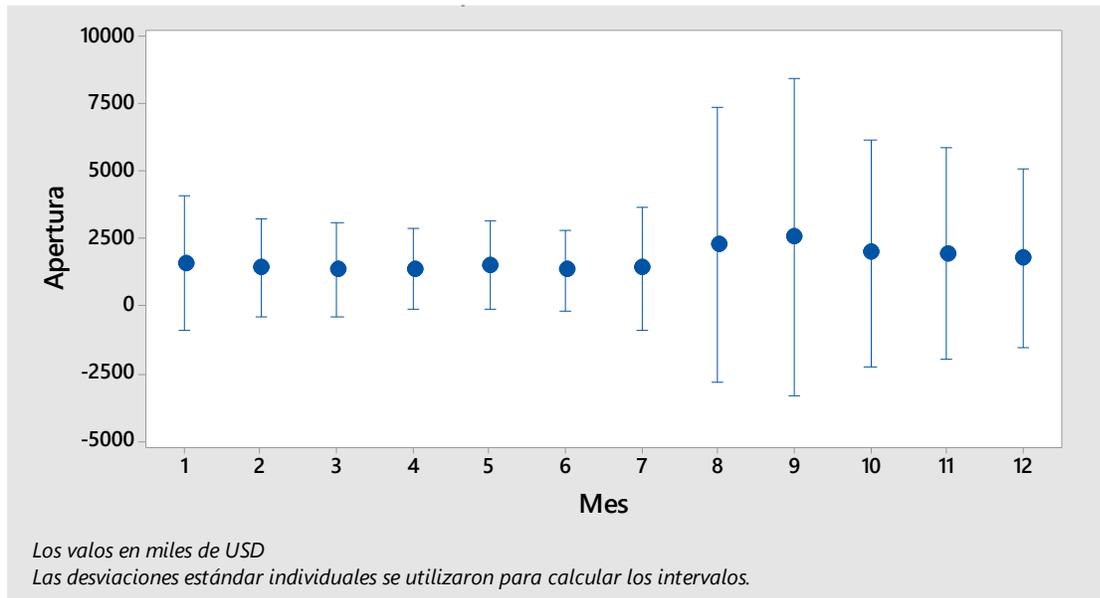


Fuente: elaboración propia.

### 3.1.1.5. *Monero*

A continuación, se presenta un análisis técnico de las estadísticas descriptivas de *monero*, además de un análisis gráfico de su comportamiento consolidado anual, representado por la amplitud de los intervalos de confianza para las variables bursátiles de precio medio de apertura y cierre.

Figura 10. Intervalo para el valor medio de apertura de *monero*



Fuente: elaboración propia.

Tabla VIII. Estadísticas descriptivas de *monero*

	Valor bursátil de apertura (dólares americanos USD)	Valor bursátil de cierre (dólares americanos USD)
Valores válidos	1982	1982
Valores perdidos	0	0
Media	160118.8	160093
Mediana	9251.5	9239.5
Moda	198	201
Desviación estándar	262985.2	262974.6
Varianza	69161233410	69155618632
Asimetría	1.5	1.5
Error estándar de asimetría	0.1	0.1
Curtosis	0.9	0.9
Error estándar de curtosis	0.1	0.1

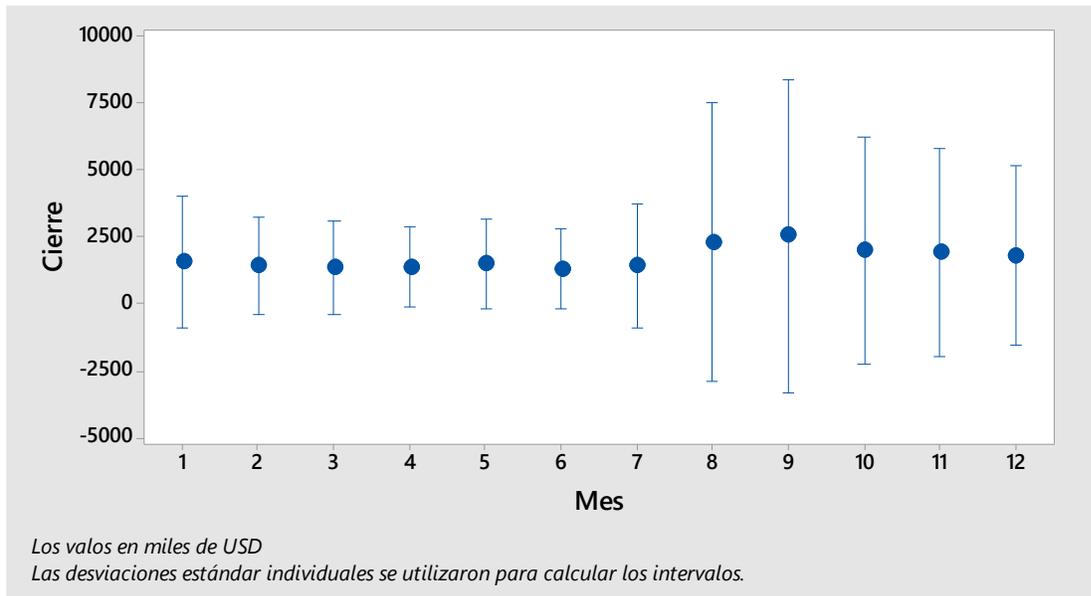
Continuación tabla VIII.

Rango	998618	999852
Mínimo	100	100
Máximo	998718	999952
Suma	317355486	317304335
Percentiles	Apertura	Cierre
25	2045.3	2055.5
50	9251.5	9239.5
75	379242.3	376708

Fuente: elaboración propia.

Se aprecia en el caso de *monero* que la amplitud del intervalo es mucho más fuerte, derivando en que la relación entre el rango de valores (tabla VIII) si difiere mucho entre los máximos y mínimos, Rauchs (2017) menciona que *monero* es un activo de los denominados capital de riesgo, porque si bien presenta el fenómeno de amplitudes en su intervalo más amplias (meses de agosto y septiembre) no existe una causa atribuible (al menos no del mercado financiero) a esta inflación, por lo que a pesar de que los rendimientos pudiesen ser atractivos, conllevan un fuerte riesgo.

Figura 11. Intervalo para el valor medio de cierre de *monero*

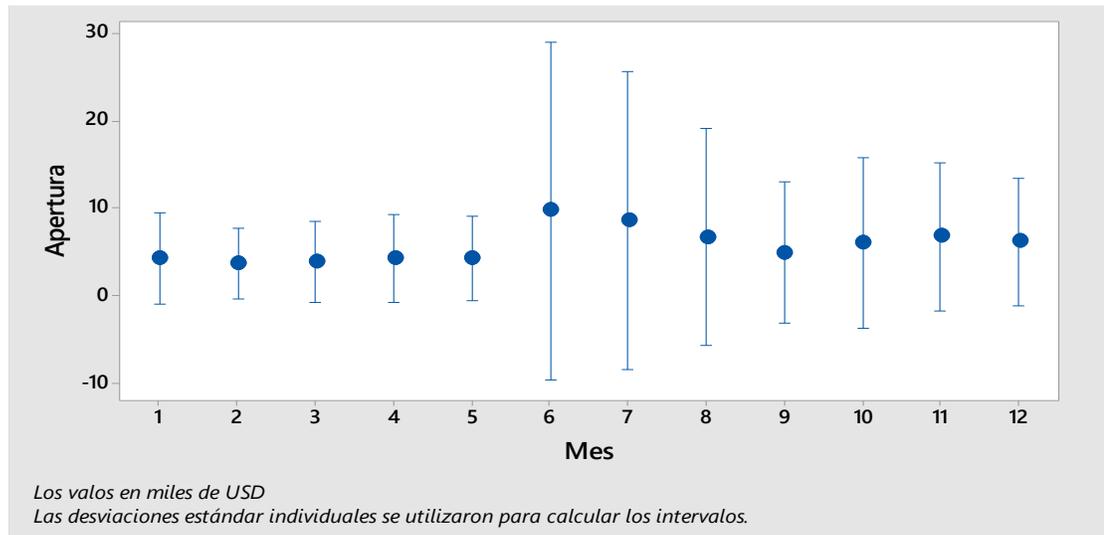


Fuente: elaboración propia.

### 3.1.1.6. *Ripple*

Se presenta un análisis técnico de las estadísticas descriptivas de *ripple*, además de un análisis gráfico de su comportamiento consolidado anual, representado por la amplitud de los intervalos de confianza para las variables bursátiles de precio medio de apertura y cierre.

Figura 12. Intervalo para el valor medio de apertura de *ripple*



Fuente: elaboración propia.

Tabla IX. Estadísticas descriptivas de *ripple*

	Valor bursátil de apertura (dólares americanos USD)	Valor bursátil de cierre (dólares americanos USD)
Valores válidos	2312	2312
Valores perdidos	0	0
Media	150071.1	150165.3
Mediana	11043	11068
Moda	114.00 <sup>a</sup>	114
Desviación estándar	209252.5	209270.5
Varianza	43786592500	43794127973
Asimetría	1.5	1.5
Error estándar de asimetría	0.1	0.1
Curtosis	1.9	1.9
Error estándar de curtosis	0.1	0.1

Continuación tabla IX.

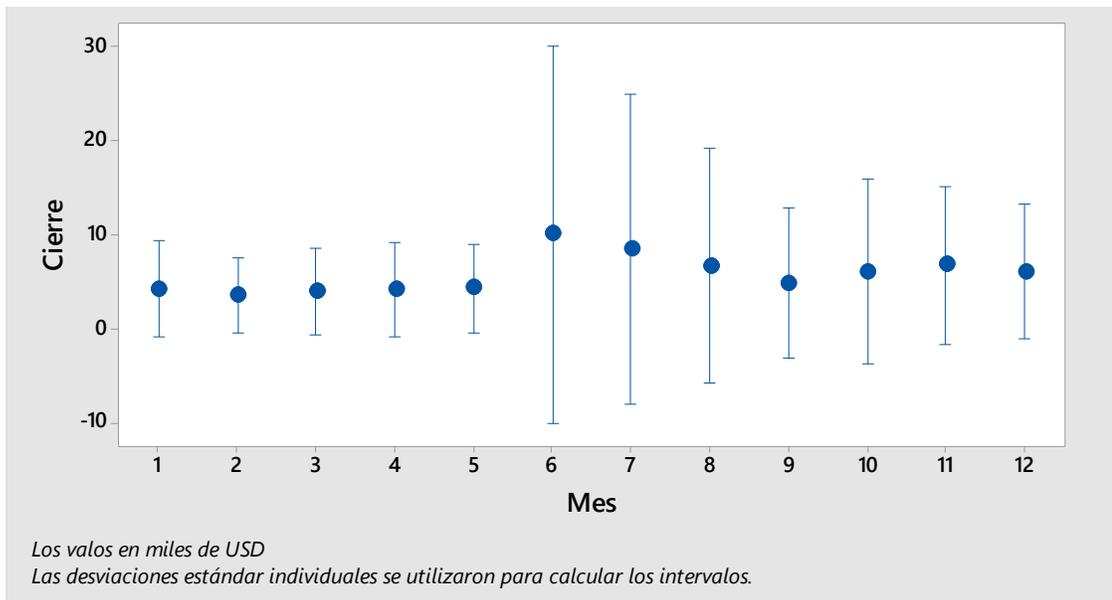
Rango	997558	998755
Mínimo	101	100
Máximo	997659	998855
Suma	346964355	347182169
Percentiles	Apertura	Cierre
50	11043	11068
75	291640	291615.3

Fuente: elaboración propia-

*Ripple* según sus estadísticas se sitúa como una de las monedas más inestables además de la de menor valor adquisitivo en comparación al dólar americano (tabla IX), el fenómeno se repite, en los meses en los que su transaccionalidad es alta, su volumen de apertura tiende a subir y se mantiene hasta el cierre bursátil (figura 12), Rauchs (2017) menciona que su bajo valor de compra la vuelve atractiva para quienes desean interactuar con una criptomoneda por primera vez.

En términos generales se aprecia que una de las premisas de las criptomonedas, en específico en el que su finalidad es la utilización como moneda digital si se refleja en sus estadísticas transaccionales históricas, al menos de las divisas consideradas en este estudio.

Figura 13. Intervalo para el valor medio de cierre de *ripple*



Fuente: elaboración propia.

### 3.1.2. Caracterización de la tendencia

Debido a ciertos rasgos en el comportamiento histórico de algunos activos virtuales, se tiende a creer que más allá de los factores económicos y financieros bajo los que se rigen por lo general los activos bursátiles, el comportamiento de estos últimos tiende a ser extremadamente aleatorio o especulativo (Rauchs, 2017).

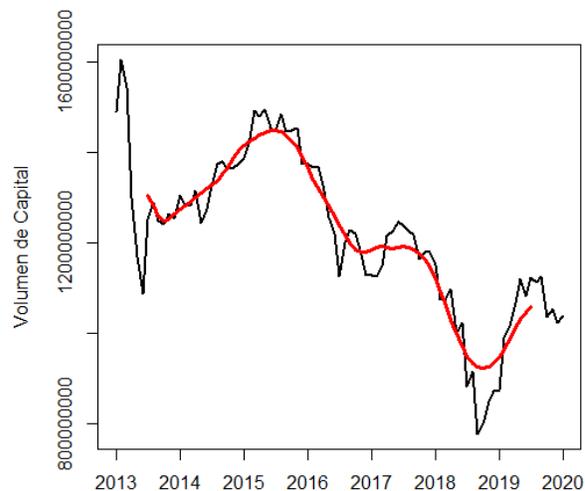
En seguimiento a esa premisa se analizó por medio de pruebas estadísticas y un análisis gráfico la tendencia histórica de los activos virtuales estudiados, con el fin de poder identificar patrones, quiebres estructurales fuertes de la tendencia o cambios bruscos en la estructura de su varianza, si bien por sí mismo el análisis no es lo suficientemente robusto para generar conclusiones de proyección futura,

podrán apreciarse comparaciones con potenciales impactos e inferencias realizadas para cada divisa.

### 3.1.2.1. Comparativa de la tendencia por criptodivisa

Se presentarán una serie de figuras a continuación, en las cuales se ilustran las series de tiempo para la variable de volumen consolidado de capital de todas las divisas estudiadas, en dónde adicional está sobrepuesta una línea de color rojo, la cual es generada a partir de la prueba de Mann-Kendall para detectar tendencias y su proyección más probable, a su vez se incluyen los resultados de dicha prueba, tanto para el coeficiente  $\tau_k$  como para el valor  $p$ .

Figura 14. Línea de tendencia para *bitcoin*



Fuente: elaboración propia.

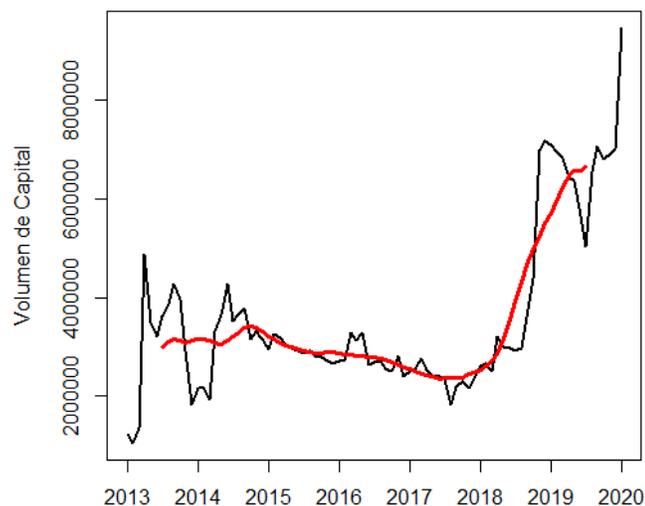
Tabla X. Prueba de Kendall para *bitcoin*

	<b>Bitcoin</b>
<b>Tau de Kendall</b>	0.5320
<b>P-valor</b>	0.0001
<b>Valor de alfa</b>	0.0500

Fuente: elaboración propia.

Como se aprecia en la figura 14 y en la tabla X, la línea de tendencia para los periodos 2015 hasta a mediados de 2018 fue decreciente, sin embargo, presenta una recuperación significativa, el cual se evidencia por un valor positivo de tau de Kendall, además el valor  $p$  es significativo, lo que denota que efectivamente existe una tendencia.

Figura 15. Línea de tendencia para *dash*



Fuente: elaboración propia.

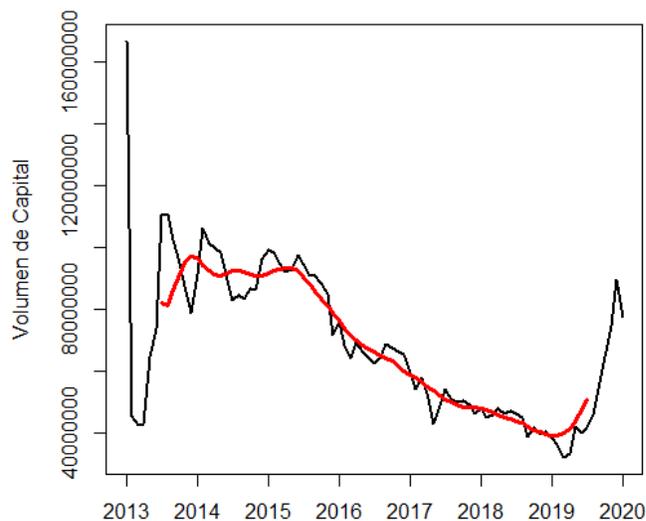
Tabla XI. Prueba de Kendall para *dash*

	<i>Dash</i>
Tau de Kendall	0.8100
P-valor	0.0001
Valor de alfa	0.0500

Fuente: elaboración propia.

En el caso de la figura 15 y la tabla XI, se aprecia un comportamiento bastante simétrico durante el nacimiento de la divisa, es decir del periodo 2013 a 2017, sin embargo, gráficamente se aprecia un alza significativa, y esta se refleja en un valor positivo de tau de Kendall, además que el  $p$  valor es significativo, lo que denota que existe una tendencia, y claramente es positiva.

Figura 16. Línea de tendencia para *ethereum*



Fuente: elaboración propia.

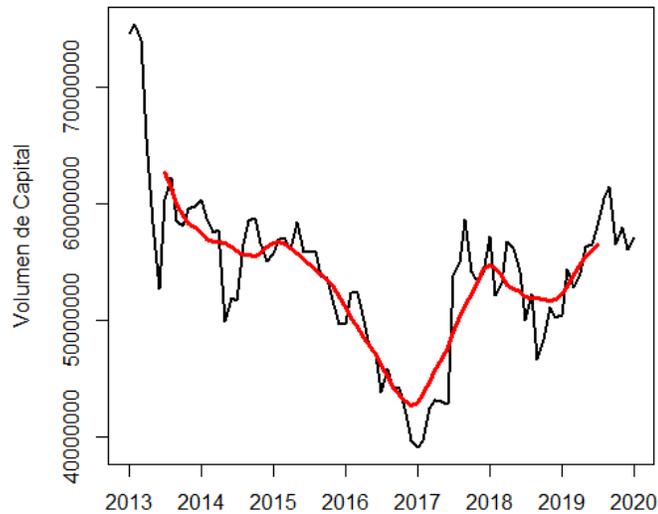
Tabla XII. Prueba de Kendall para *ethereum*

	<i>Ethereum</i>
Tau de Kendall	0.7140
P-valor	0.0001
Valor de alfa	0.0500

Fuente: elaboración propia.

En el caso de *ethereum* como se aprecia en la figura 16 y la tabla XII se aprecia una baja significativa en su capitalización durante el 2013, esta tendencia continua hasta el 2019, cuando tiene un alza significativa, que se refleja en un tau de Kendall positivo, además de que el  $p$  valor de la prueba es significativo denota que sí existe una tendencia en la serie, sin embargo, esta históricamente ha sido más negativa que positiva, al menos en un primer análisis gráfico.

Figura 17. Línea de tendencia para *litecoin*



Fuente: elaboración propia.

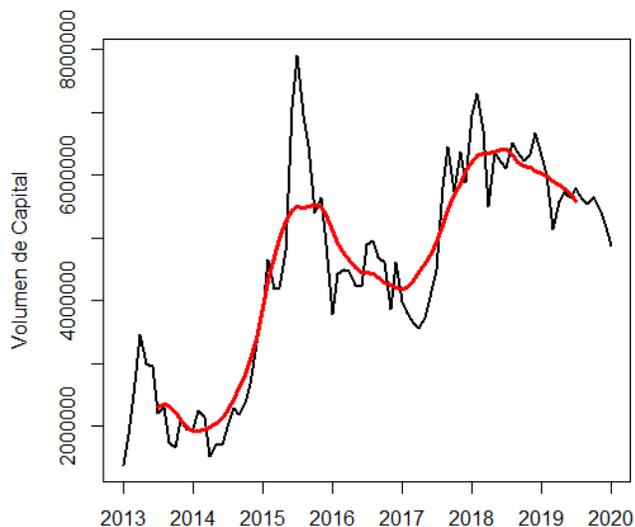
Tabla XIII. Prueba de Kendall para *litecoin*

	<b>Litecoin</b>
<b>Tau de Kendall</b>	0,6120
<b>P-valor</b>	0,0001
<b>Valor de alfa</b>	0,0500

Fuente: elaboración propia.

Se aprecia para *litecoin* en la figura 17 y la tabla XII, que es una divisa que ha tenido una tendencia fraccionada, es decir durante su nacimiento en el periodo 2013 a 2017 su comportamiento fue a la baja, sin embargo, a partir de 2017 se ha comportado al alza, reflejado por una tendencia positiva dado el coeficiente de tau de Kendall de 0.612 y un  $p$  valor significativo, que corrobora el supuesto de que existe una tendencia.

Figura 18. Línea de tendencia para *monero*



Fuente: elaboración propia.

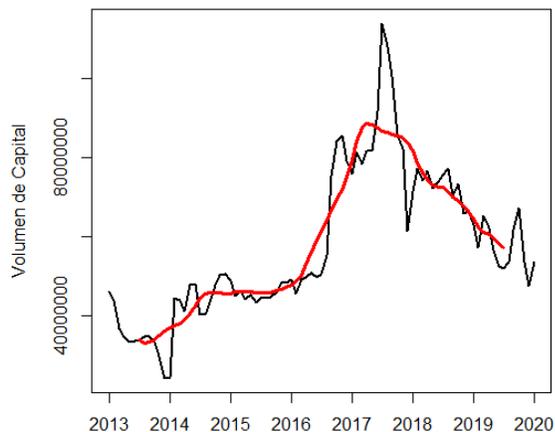
Tabla XIV. Prueba de Kendall para *monero*

	<i>Monero</i>
<b>Tau de Kendall</b>	0.432
<b>P-valor</b>	0.001
<b>Valor de alfa</b>	0.050

Fuente: elaboración propia.

En el caso de *monero*, en la figura 18 y la tabla XIV se aprecia que ha sido una divisa con evolución oscilatoria al alza al menos de forma gráfica, sin embargo durante el último periodo evaluado 2019-2020 presenta un comportamiento más propenso a la baja o a una tendencia negativa, esto se refleja con una tau de Kendall positivo pero de menor magnitud (que las divisas estudiadas anteriormente), sin embargo, la prueba está en la brecha de significancia en relación al  $p$  valor, en dónde aún puede concluirse que existe una tendencia para la serie.

Figura 19. Línea de tendencia para *ripple*



Fuente: elaboración propia.

Tabla XV. **Prueba de Kendall para *ripple***

	<b><i>Ripple</i></b>
<b>Tau de Kendall</b>	0.351
<b>P-valor</b>	0.001
<b>Valor de alfa</b>	0.050

Fuente: elaboración propia.

En cuanto a *ripple* se aprecia en la figura 19 y la tabla XV, que su comportamiento final es hacia la baja, lo que se denota con un tau de Kendall significativamente más pequeño, pero positivo, por lo que se puede concluir que la tendencia no es negativa, es decir no destruye valor, además la tendencia sigue presente dado el nivel de significancia del  $p$  valor.

Asumir una relación directa con factores clásicos de comportamiento económico a estas variables carece de mucho sentido; dada la relación de crecimiento que se refleja en las figuras 14 a 19, en dónde se aprecia la línea de tendencia en color rojo asociada a la serie temporal del volumen de capital por divisa, la prueba de Kendall refleja valores significativos para todas las divisas, además de un coeficiente  $\tau$  positivo para todas las divisas, lo que sugiere que a nivel de apreciación visual, las series sugieren caídas, la tendencia es hacia el crecimiento, por ende se puede considerar como un buen preámbulo para considerar estas divisas como objetos de inversión financiera.

El análisis anterior sugiere que una aproximación adecuada para una caracterización integral de las divisas es por medio de un análisis de series de tiempo, con ello es viable describir patrones de interés, así como las fluctuaciones en su varianza, para evitar caer en redundancia u variables que no aporten a la

variabilidad general en el comportamiento de las divisas, dado que este análisis pretende utilizar esta caracterización para crear un modelo predictivo.

### 3.1.2.2. Análisis de correlación en las criptodivisas

En relación al apartado anterior, es mucho más eficiente trabajar el análisis de todas las divisas por medio del volumen de capital de mercado dado a que reflejan de mejor forma los criterios de apertura y cierre bursátil asociado a las mismas, por ende, para el análisis de correlación de las divisas se utilizó esta cifra, para analizar si realmente existe una relación directa en el incremento o decremento del valor de volumen de mercado entre todas las divisas.

Castro y Forero (2018) comentan en su análisis de potenciales inversiones en criptoactivos que un punto porcentual de relación con la divisa de mayor volumen de capital bursátil vuelve a la divisa objeto atractivo de potencial inversión, además que según el autor se debe reflejar en su rendimiento financiero (ROI), esta relación es dada a la amplitud en el volumen de capital entre divisas, como puede apreciarse en la sección de estadísticas descriptivas, para analizar la correlación entre divisas se procedió a realizar la siguiente matriz de correlación utilizando el método de *Pearson* para correlacionar la variable de volumen de capital bursátil.

Tabla XVI. **Análisis de correlación para todas las divisas con datos del mercado financiero de divisas**

	<i>Bitcoin</i>	<i>Dash</i>	<i>Ethereum</i>	<i>Litecoin</i>	<i>Monero</i>	<i>Ripple</i>
<i>Bitcoin</i>	1.00	0.12	0.15	-0.31	-0.06	0.01
<i>Dash</i>	0.12	1.00	-0.10	-0.03	0.08	-0.04
<i>Ethereum</i>	0.15	-0.10	1.00	0.02	-0.05	0.00

Continuación tabla XVI.

<b><i>Litecoin</i></b>	-0.31	-0.03	0.02	1.00	-0.11	0.08
<b><i>Monero</i></b>	-0.06	0.08	-0.05	-0.11	1.00	0.09
<b><i>Ripple</i></b>	0.01	-0.04	0.00	0.08	0.09	1.00

Fuente: elaboración propia.

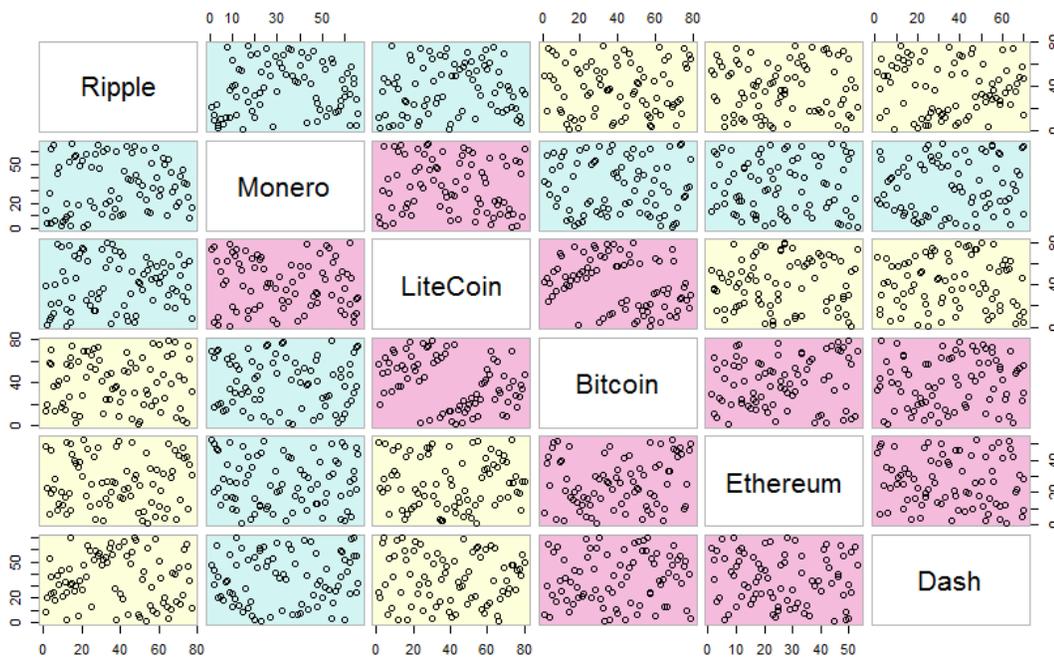
En términos generales la correlación de todas las criptodivisas es bastante baja contemplando que las cifras de referencia parten de la premisa de una relación no significativa asociada al valor 0 y una relación cuasi perfecta asociada al 1, sin embargo, utilizando dichos valores como criterio de asociación por valor proporcional, se puede observar que algunas se asocian de forma más significativa con la divisa *bitcoin*, siendo *litecoin* y *ethereum* tales candidatos, lo cual se refleja en su volumen alto de mercado.

Según Castro y Forero (2018) sus rendimientos podrían estar sujetos a un rendimiento atractivo, sin embargo, basarse en definir rendimientos únicamente por su factor de correlación con la divisa más atractiva; supondría un sesgo en el estudio dado que algunas divisas tienen características intrínsecas que aún se encuentran en etapa de maduración, por lo que si bien un análisis de correlación puede dar una muy buena referencia inicial en su comportamiento, se debe abordar por medio de sus componentes temporales para un análisis más integral

Para representar de mejor forma las correlaciones de las variables se realizó un diagrama de dispersión de datos en forma de matriz, para todas las monedas, en dónde las variables se reordenaron y se representan en ese orden particular debido a que aquellas con mayor correlación están cerca de la diagonal

principal, además de una identificación por código de colores por variable para facilitar su identificación.

Figura 20. **Sobreposición de las dispersiones en cada criptodivisa**



Fuente: elaboración propia.

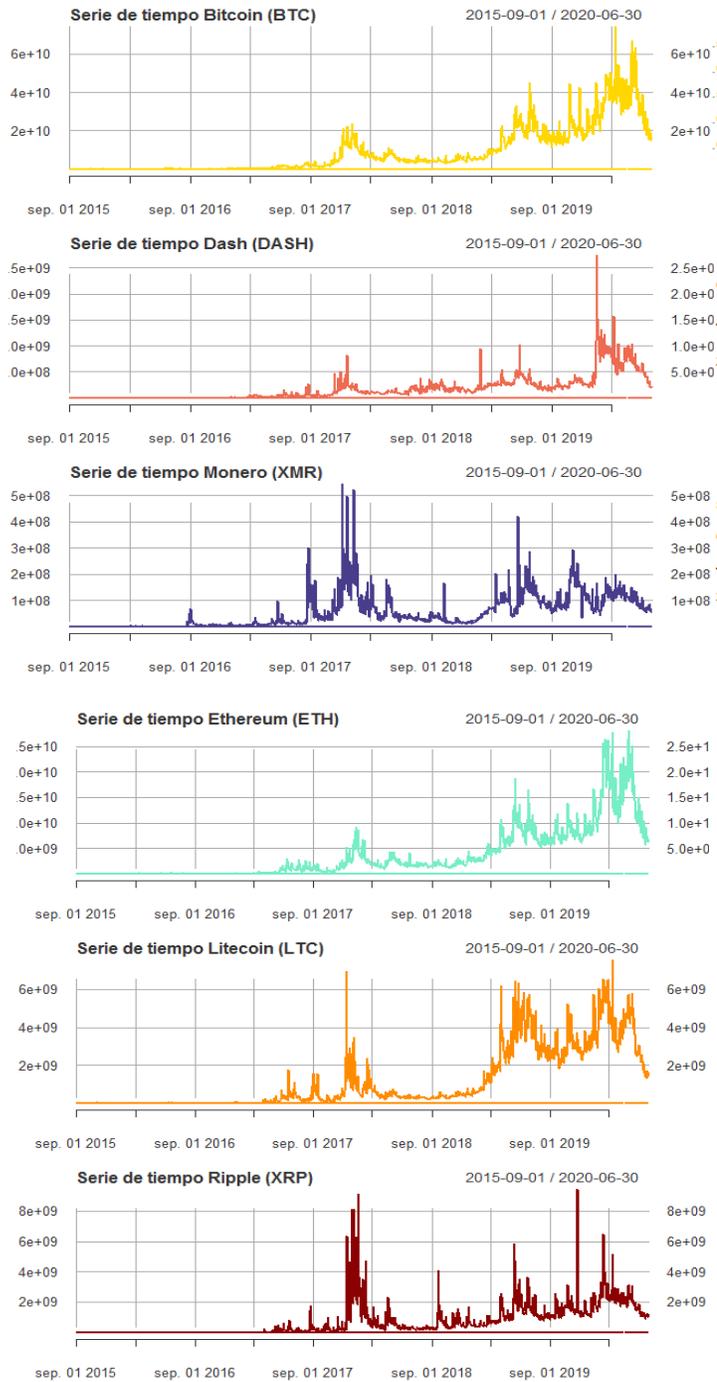
### 3.1.3. Modelos de vector autorregresivo

Se desea complementar el análisis de la influencia que ejerce *bitcoin* sobre el resto de activos virtuales estudiados, dado a que si bien es cierto la correlación lineal es un buen estimador de las caídas y apreciaciones de una divisa respecto a otra en el horizonte temporal estudiado, es evidente que no es eficiente si se desea analizar de forma integral las relaciones del mercado, en este caso en particular la influencia directa que tiene la variación de un activo respecto al comportamiento de sus pares.

Para ello un modelo de vector autorregresivo supone una excelente alternativa para estudiar dichas interacciones, dado a que ofrece la bondad del análisis de series temporales multivariadas, en dónde la consideración de los atrasos para la estimación del modelo está directamente correlacionados al supuesto de que al menos una proporción de la varianza de una variable está influenciada por otra de un clúster de espacios muestrales homogéneo, dado que a que puntualmente en seres temporales, los valores más influyentes de las series estacionarias que componen un modelo de vector autorregresivo tienden a ser influyentes con pares del mismo espacio muestral (Cryer y Chan, 2008).

Dado a que el modelo tiene la capacidad de tomar como insumo para las estimaciones distintas variables, se toman de referencia los precios ajustados de cierre para los seis activos virtuales para un periodo de estudio homogéneo, con el fin de llegar a un análisis gráfico de impulso respuesta para explicar como un impacto en las distintas divisas se homologa con el resto, y de igual forma describir gráficamente como es la composición proporcional de la varianza del modelo, detectando así el supuesto de cambios estructurales significativos en la varianza a raíz de una variación porcentual de *bitcoin*.

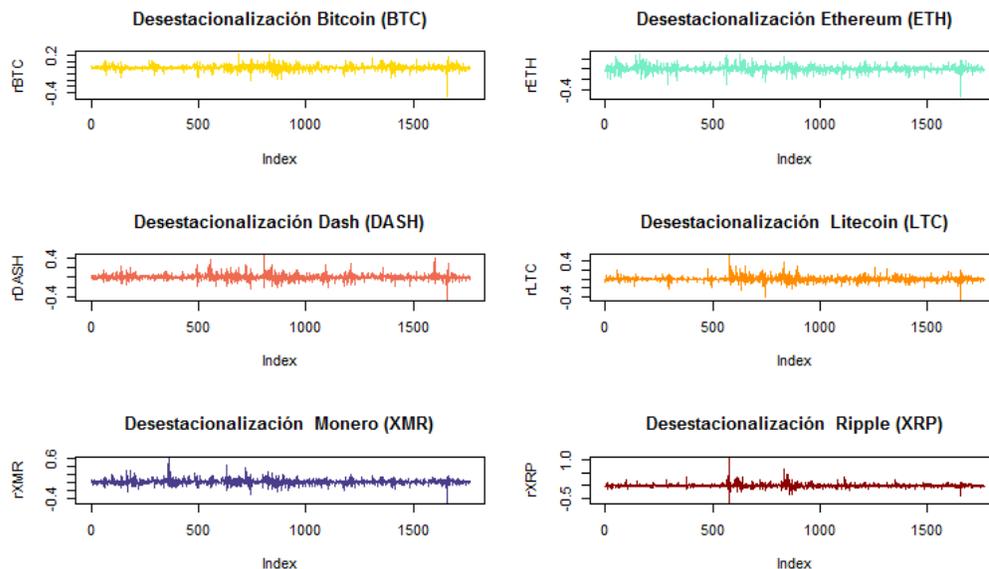
Figura 21. Comportamiento de las seis criptomonedas



Fuente: elaboración propia.

Dada la naturaleza del análisis es necesario que las series de tiempo a involucrar en el modelo sean estacionarias, indiferente a si se trata de estacionariedad en media o en varianza, el modelo es capaz de inferir los cambios estructurales influyentes únicamente bajo este supuesto, además, se garantiza que los atrasos seleccionados no se encuentren sesgados hacia una tendencia o estacionalidad significativamente superior (Cryer y Chan, 2008).

Figura 22. **Series de tiempo desestacionalizadas**



Fuente: elaboración propia.

En la figura 22 se presentan las series desestacionalizadas, todas las divisas se convirtieron en estacionarias aplicándoles una segunda diferenciación estacional, la cual no es más que la aplicación de una segunda transformación lineal sobre el primer logaritmo de la serie, ante esto ninguna presenta evidencia de raíz unitaria, este supuesto se validará de forma práctica bajo la prueba aumentada de Dickey Fuller en un apartado posterior del estudio.

Antes de la construcción del modelo, se utilizó la información de las series de tiempo desestacionalizadas como insumo para la estimación adecuada de atrasos y/o coeficientes de rezago en un motor de datos, para la identificación del modelo de vector autorregresivo se utilizan como máximo 12 rezagos para punto de referencia (dado a que las series de tiempo estudiadas son anuales) de acuerdo a distintos criterios de información (AIC, HQ, SC y FPE) en dónde la tabla XVII sugiere en función del error de estimación que 6 atrasos son eficientes para la estimación del modelo.

Tabla XVII. **Selección de atrasos para diseño del vector autorregresivo**

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
AIC												
(n)	-36.578	-36.583	-36.647	-36.633	-36.650	-36.660	-36.643	-36.623	-36.630	-36.63	-36.628	-36.605
HQ												
(n)	-36.529	-36.493	-36.516	-36.459	-36.44	-36.403	-36.350	-36.284	-36.249	-36.211	-36.164	-36.099
SC												
(n)	-36.446	-36.339	-36.291	-36.164	-36.069	-35.966	-35.837	-35.705	-35.599	-35.491	-35.372	-35.24
FPE	1.30	1.29	1.21	1.21	1.21	1.20	1.22	1.24	1.24	1.24	1.24	1.27
(n)	* 10 <sup>-16</sup>											

Fuente: elaboración propia

En función de lo anterior se ajusta el modelo VAR para 6 rezagos, en dónde se pretende analizar las raíces del polinomio característico, así como los estadísticos necesarios para llevar a cabo la inferencia estadística.

$$\text{Raíces polinomiales} = 0.7395, , 0.7326, 0.7326, 7.7228, 0.7199, 0.7199, 0.6927, 0.6786, 0.6786 \dots, \\ 0.5658, 0.5433, 0.5433, 0.3062, 0.3062, 0.1857$$

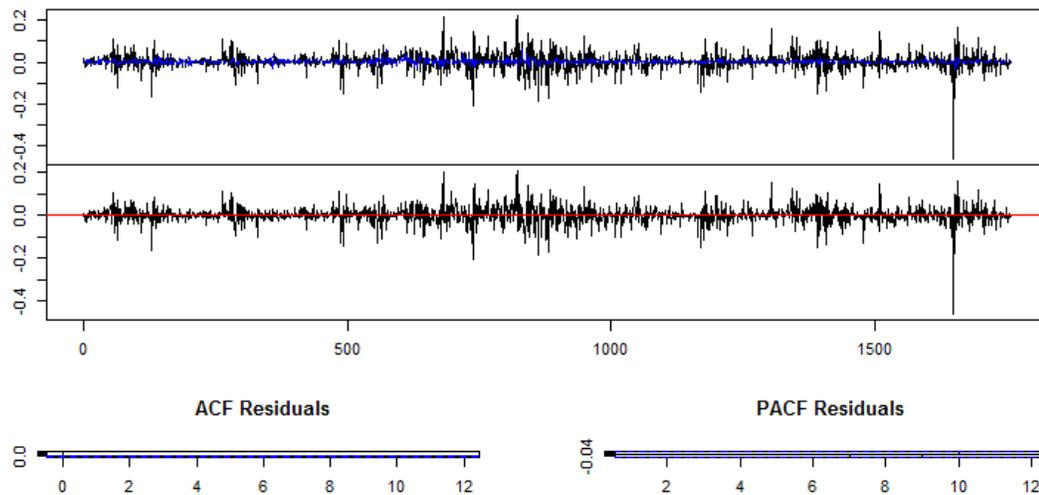
Se tiene que las raíces del polinomio característico son menores a uno, por lo que el VAR estimado con 6 rezagos satisface la condición de estabilidad o estacionariedad.

### **3.1.3.1. Análisis de supuestos del modelo**

Para garantizar la validez del modelo se deben analizar los residuos del mismo, esto para validar que no existe autocorrelación serial entre los residuales, que los mismos se distribuyan de forma normal, que existe homocedasticidad en su distribución, entre otros. Dado a que si alguno de estos supuestos no se cumple el modelo no es apto para generar predicciones ni para aproximar una proyección de influencia de uno respecto al resto o los supuestos de cambios estructurales de varianza se verían sesgados.

Dado que el modelo está construido con las seis criptodivisas estudiadas se dispone de seis distribuciones de residuos, los cuales tienen se pueden visualizar por medio del diagrama de puntos ajustados vs predichos y los diagramas de autocorrelación parcial y total, a fines ilustrativos se muestra solo la distribución de residuos de *bitcoin*, sin embargo, para los análisis se emplea la totalidad de los mismos.

Figura 23. Residuos del modelo VaR para *bitcoin*



Fuente: elaboración propia.

Para probar la autocorrelación serial se utiliza una prueba de Portmanteau, la cual resulta no significativa para el modelo de las seis divisas planteadas, lo cual concluye que efectivamente no hay presencia de correlación serial de los residuales (Klenke, 2014).

$$p - value = 0.2055$$

Dada la naturaleza del modelo para poder estimar la normalidad en sus residuos se calcula una prueba de Jarque-Bera, esta es una prueba de bondad de ajuste que se utiliza para probar si una muestra de datos tiene la asimetría y curtosis de una distribución normal. El estadístico de Jarque-Bera se distribuye asintóticamente como una chi-cuadrado con dos grados de libertad y puede

utilizarse para probar la hipótesis nula de que los datos pertenecen a una distribución normal o cuasi-normal (Klenke, 2014).

$$P - value_{Asimetría} = 0.3120$$

$$P - value_{curtosis} = 0.1541$$

Al presentar resultados no significativos para ninguna de las dos pruebas, la distribución de residuos del ajuste para la serie multivariante puede considerarse con una correcta bondad de ajuste a la normal.

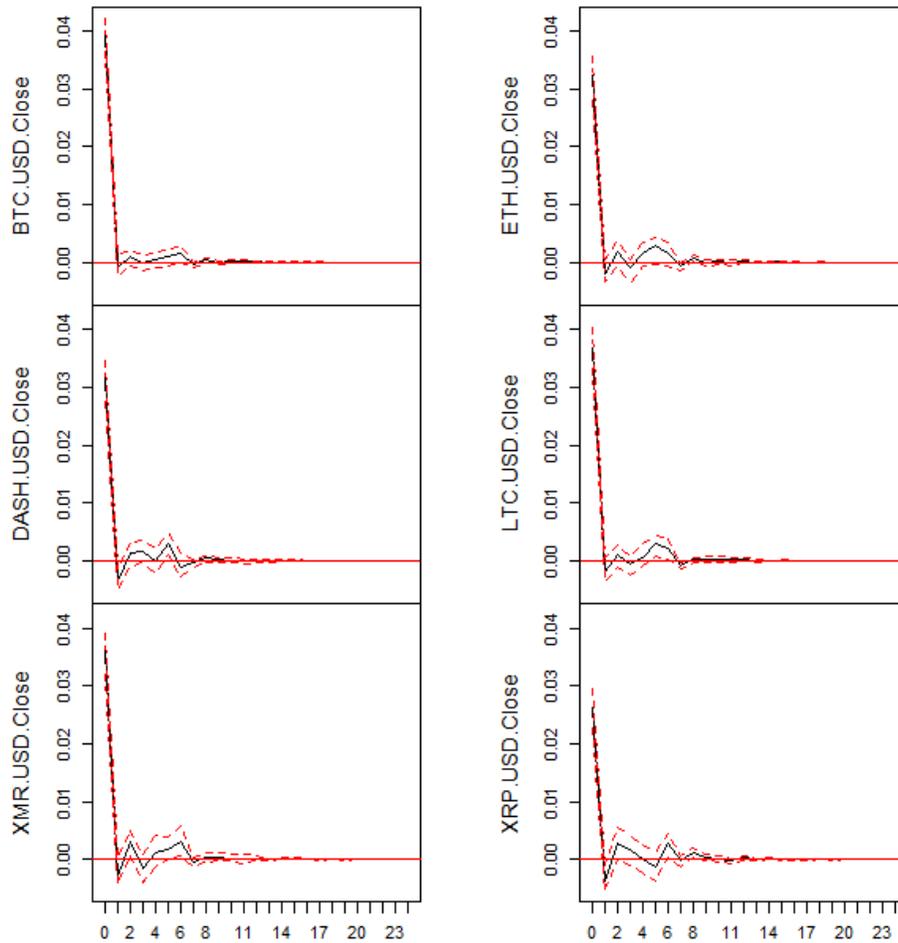
Como punto final de los supuestos se utiliza un modelo ARCH de vector autorregresivo para estimar si puntos aleatorios de inferencias influyentes tienen una varianza homocedástica, presentando resultados no significativos por lo que se satisface el supuesto de la varianza constante dada la falta de evidencia de rechazo hacia la hipótesis nula

$$p - value_{ARCH} = 0.1140$$

### **3.1.3.2. Análisis de impulso respuesta**

Una vez se han llevado a cabo las pruebas de validación de supuestos del modelo y verificar que el modelo VAR las satisface, se procede a realizar el análisis de impulso respuesta y con ello observar la trayectoria de la variable de estudio. Para la proyección de este análisis se obtienen los multiplicadores para 24 meses de impulso respuesta para los seis activos virtuales estudiados.

Figura 24. Impulso-respuesta ortogonal de *bitcoin*

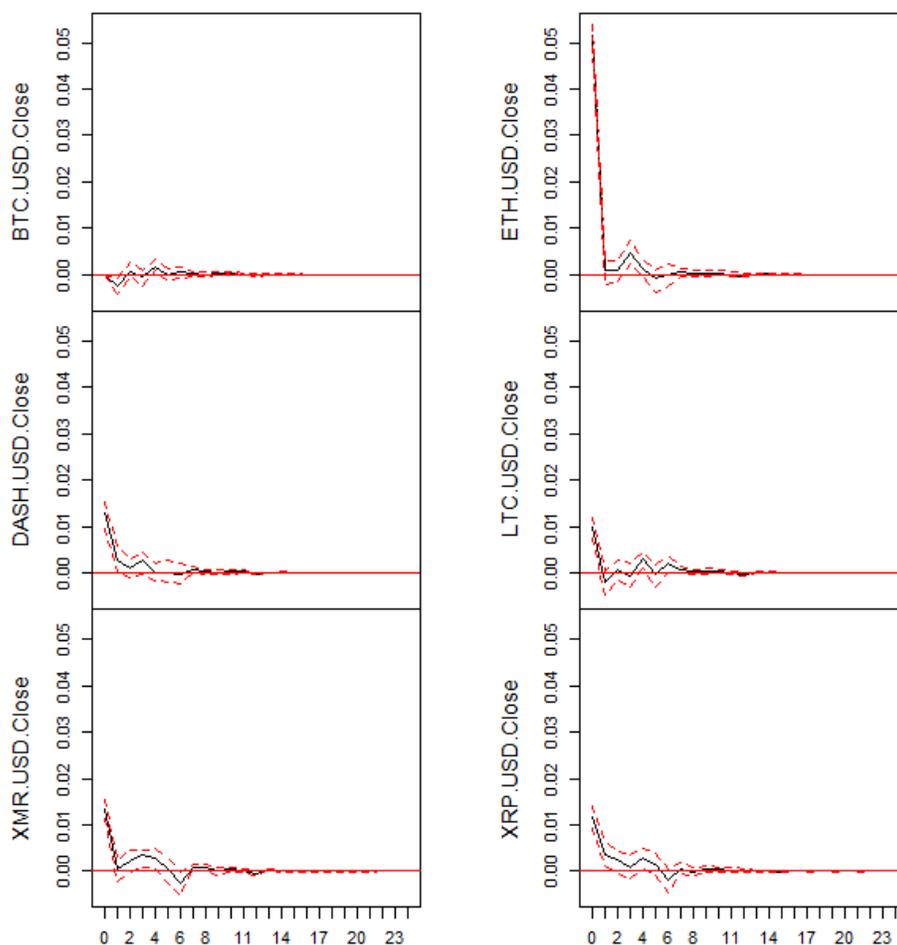


Fuente: elaboración propia.

En el primer análisis se representa el análisis del impulso-respuesta de *bitcoin* a 24 meses, se observa que ante un *shock* del mismo *bitcoin* causa un impacto positivo, al tercer mes del impacto tiende a retomar su equilibrio estacionario teniendo ciertos quiebres estructurales durante el mes 8 y 10, por otro lado ante un *shock* en *bitcoin* el resto de divisas tienen un impacto positivo quebrado, es decir el *shock* de la divisa de mayor apreciación influye

positivamente en el primer trimestre de todas las divisas, evidenciando un quiebre estructural en el mes 4 y de igual forma a excepción de XRP *ripple*, todas tienen una influencia de nuevo positiva al mes 5 retomando su equilibrio estacionario en el mes 8. En el caso de *ripple* se tiene el mismo efecto, pero a partir del mes 8, lo cual brinda evidencias de la homogeneidad en el comportamiento con *bitcoin*.

Figura 25. **Impulso-respuesta ortogonal de *ethereum***



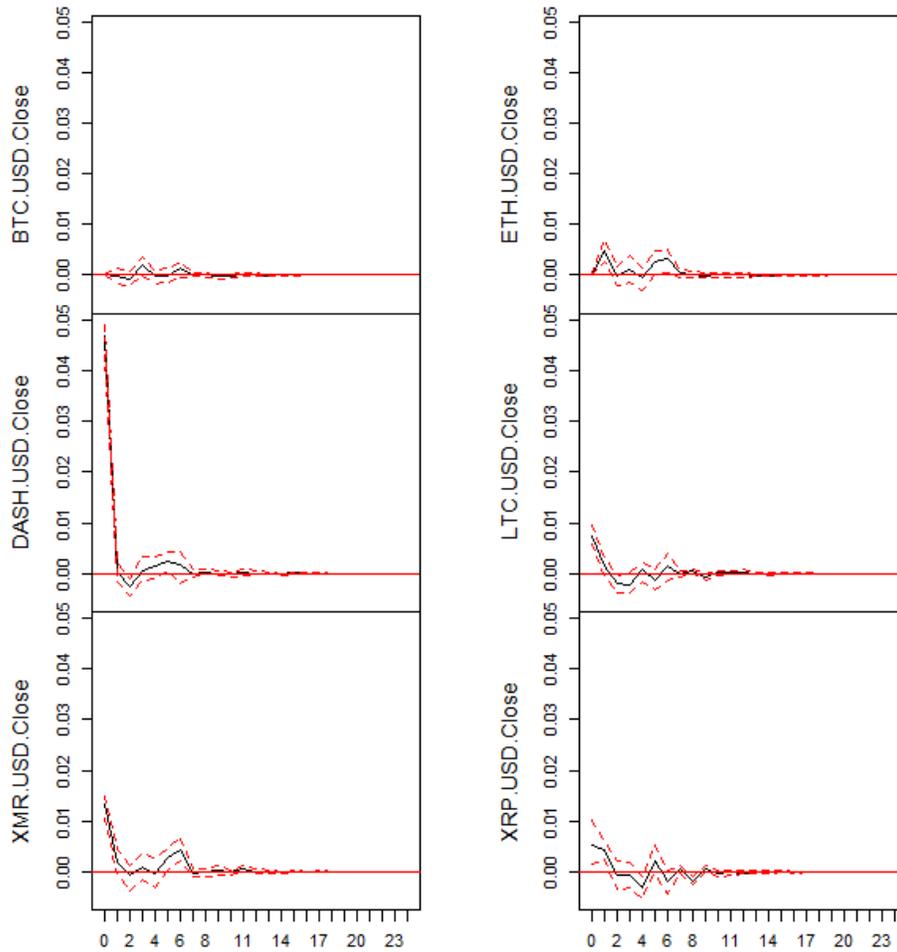
Fuente: elaboración propia.

En el caso de *ethereum* se tiene un escenario bastante distinto al de bitcoin, dado que un *shock* en esta variable no representa ningún quiebre estructural significativo para *bitcoin*, es probable que este sea un efecto común en el resto de divisas, si bien es cierto se presenta una ligera baja durante el primer mes del *shock*, la estabilidad estacionaria casi se reestablece inmediatamente.

En el caso del *shock* comparada con ella misma tiene un efecto positivo, *ethereum* se aprecia durante los primeros dos meses tendiendo a estabilizarse a partir del mes 3 logrando una estabilidad estacionaria a partir del mes 5 un caso de apreciación similar supone *dash* dónde el impacto en *ethereum* sugiere una ligera apreciación que se estabiliza casi inmediatamente.

Por otro lado, *litecoin*, *monero* y *ripple* suponen un cambio estructural negativo hacia el *shock* de *ethereum* teniendo un efecto parabólico negativo entre el primer mes y el mes 6, en donde a partir del mes 8 el equilibrio estacionario se vuelve homogéneo, entonces un *shock* en *ethereum* supone un periodo de depreciación para los activos de menor valor.

Figura 26. Impulso-respuesta ortogonal de *dash*

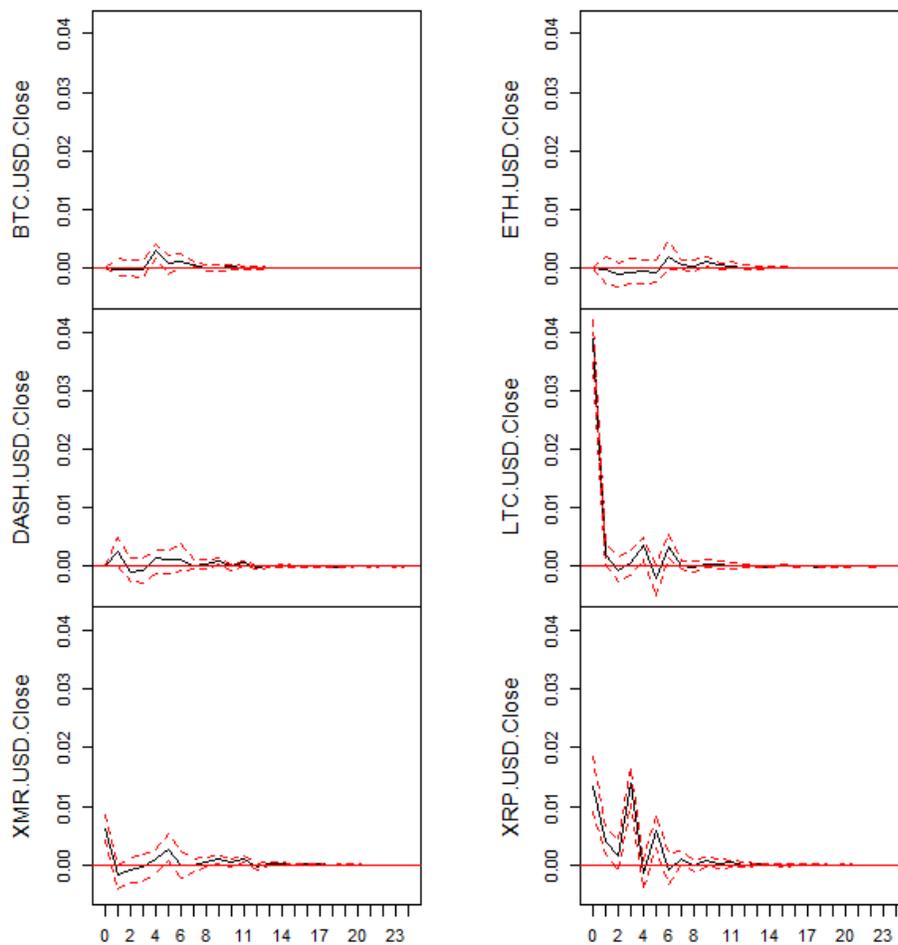


Fuente: elaboración propia.

En el caso de *dash* se aprecia que un *shock* en la misma apenas causa quiebres estructurales en la composición estacional de *bitcoin*, por otro lado, *ethereum* y *monero* suponen comportamientos interesantes dónde el equilibrio estacional es constante al inicio del *shock* pero comparten ligeras apreciaciones a partir del mes 5, las cuales logran volver al equilibrio estacional después del mes 9.

Por otro lado, en cuanto a la misma divisa, *litecoin* y *ripple*, un *shock* en *dash* supone una influencia negativa, es decir tiene un impacto negativo, dónde *dash* tiene el escenario de equilibrio más favorecedor durante el primer trimestre, siendo *ripple* la divisa que más se desestabiliza ante el *shock* logrando un equilibrio estacionario después del mes 10, al ser una divisa de baja apreciación y utilidad equivalente supone ese riesgo.

Figura 27. **Impulso-respuesta ortogonal de *litecoin***



Fuente: elaboración propia.

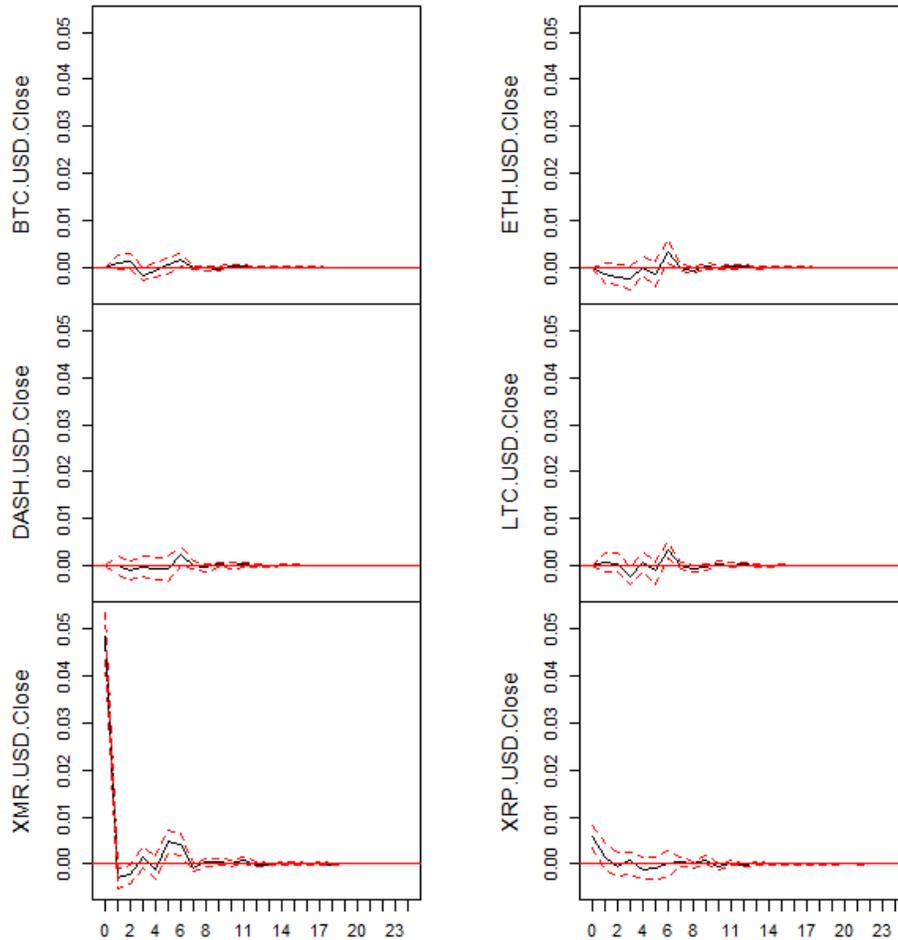
En el caso de un *shock* en *litecoin* se aprecia una influencia poco significativa en *bitcoin*, *ethereum* y *dash*, si bien *ethereum* presenta un comportamiento por debajo del equilibrio estacionario el quiebre en el mes 7 logra un comportamiento estacional por lo que la influencia no es demasiado influyente.

Sin embargo, un shock en la misma divisa tiene un impacto negativo en la misma teniendo el quiebre más pronunciado en el mes 6, logrando posteriormente el equilibrio estacionario, caso contrario a *XRP ripple* que evidencia un impacto positivo ante el *shock*, llegando a un equilibrio estacionario después del quiebre de *litecoin*, evidenciando potenciales apreciaciones ante la caída de la divisa, es decir son divisas con comportamientos relativamente contrarios.

Por otro lado, *XMR monero* supone de igual forma apreciaciones o un impacto positivo ante el *shock*, pero es bajo, siendo el mes 6 de igual forma el quiebre estructural más alto y posteriormente se obtiene el equilibrio estacional.

*Litecoin* supone el inicio del análisis de las divisas de menor valor medio global del análisis, esto se evidencia como la respuesta ante el *shock* de la misma en relación a las divisas más apreciadas como *bitcoin* y *ethereum* es virtualmente nulo, sin embargo es interesante que dadas las condiciones del mercado el mismo *shock* ante las divisas de menor apreciación tiene un impacto positivo en periodos homólogos (semestres) donde en mayor o menor medida tiende a influenciar apreciaciones en estas divisas.

Figura 28. Impulso-respuesta ortogonal *monero*



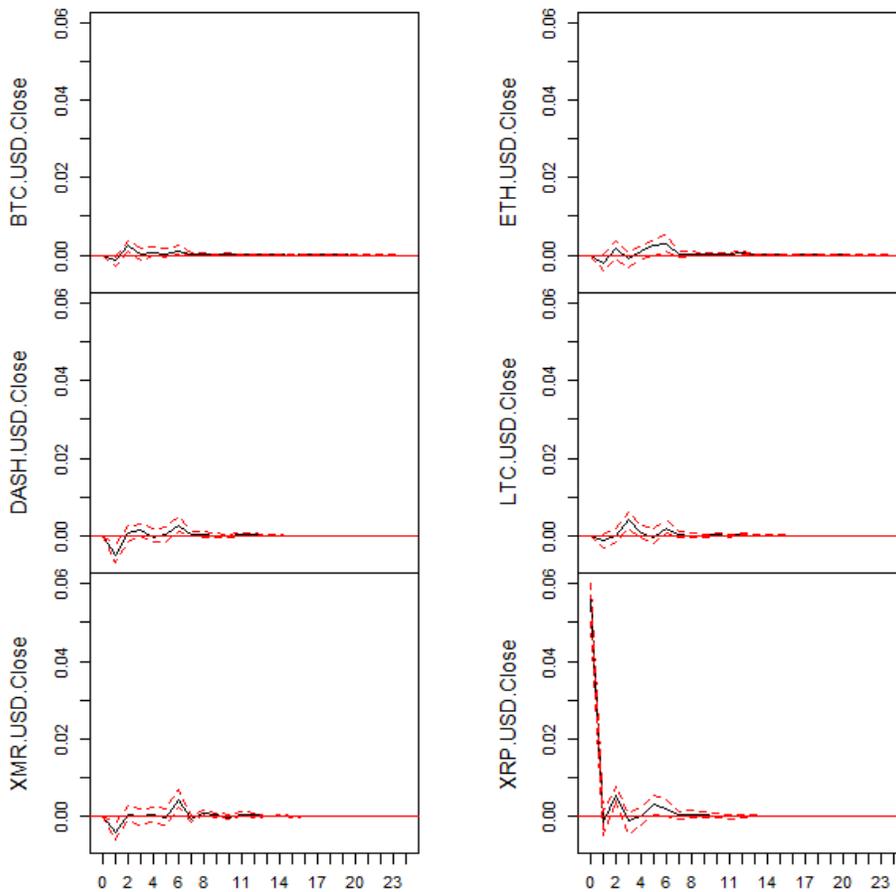
Fuente: elaboración propia.

Como se evidencio anteriormente, los *shocks* en las divisas de menor apreciación no tienden a tener influencias significativas en la estructura estacional de las de mayor valor, por ende, es un excelente indicio que la estructura de la varianza del modelo tiene una estructura basada en *bitcoin* y *ethereum* para todas las divisas, ese supuesto se validará posteriormente en el estudio, sin embargo, tanto para *bitcoin*, *ethereum*, *dash* y *ripple*, un *shock* en la

divisa no supone un impacto significativo, sin embargo es interesante visualizar como un quiebre en el mismo activo supone un impacto positivo en su valor.

Por otro lado, en el caso de *litecoin* se evidencia lo anterior, solo que en este caso el impacto es positivo para ambas divisas, demostrando una ligera estabilidad superior en XMR *monero* que el LTC *litecoin* ante los cambios estacionarios.

Figura 29. Impulso-respuesta ortogonal *ripple*



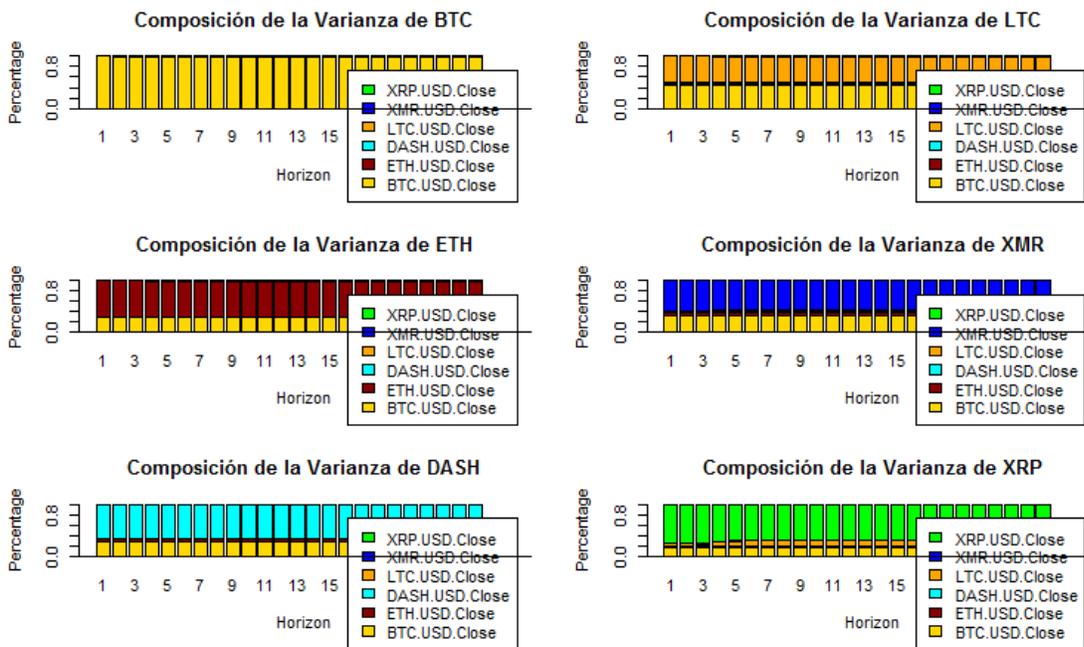
Fuente: elaboración propia.

Por otro lado, se evidencia como la estructura de la variación del modelo no se ve influenciada por *ripple*, esto está relacionado a que al ser la divisa de menor valor un *shock* en la estructura de la misma no causa un cambio significativo en el resto, evidenciando que el mercado sigue un comportamiento estocástico fuertemente influenciado por la divisa de mayor valor, esto se puede evidenciar de mejor forma si se descompone visualmente la varianza que conforma al modelo (Radovanov *et al.*, 2018).

#### **3.1.3.2.1. Descomposición de la varianza del modelo**

Descomponiendo la variabilidad explicada por el modelo de forma visual se puede evidenciar de forma visual para un periodo de 24 meses, que variables influyen la varianza del modelo

Figura 30. Descomposición de la varianza explicada por el modelo



Fuente: elaboración propia.

En la figura 30 se evidencia de forma gráfica y por código de colores el cómo se conforma la variabilidad de la serie, es claro evidenciar la participación de *bitcoin* en la composición estructural de la varianza del resto (color amarillo).

Es importante evidenciar como en divisas de un espacio muestral más cercano (valores medios más próximos dada la matriz de distancias) la estructuración proporcional casi es homogénea entre la divisa misma y *bitcoin*, tal es el caso de *litecoin*, *monero* y *ethereum*, sin embargo, también es importante evidenciar como en divisas más distantes como en el caso de *XRP ripple* la composición de la varianza conformada por *bitcoin* es significativamente inferior.

Por ende, es claro evidenciar que a pesar de que un cambio en *bitcoin* no supone un cambio homologo o equivalente en el resto de divisas estudiadas, se evidencia que ante un *shock* en las mismas es la que supone una influencia, indiferente a si es un impacto positivo o negativo en el comportamiento más significativo sobre el resto.

### **3.2. Análisis de varianza**

Se realizó un análisis de varianza de un factor para el análisis de las divisas, utilizando como factores cada divisa *per se* y utilizando como factor de comparación el volumen total del mercado de capital, con el fin de facilitar el estudio, dado a que no se consideran factores de apertura y cierre bursátil, sino más bien el volumen de capital monetario, que en síntesis refleja la media armónica entre los volúmenes ponderados del capital a través del tiempo como mencionan Salwach y Perz (2020).

Con este análisis se analizó la variabilidad real que presenta cada una de las divisas en relación a la cuota de mercado que le corresponde, además de identificar qué tan distante se encuentra su volumen de financiación en relación una de otra, lo cual brinda una buena aproximación por medio de estadísticos de prueba para determinar si realmente este tipo de variaciones pueden ser predecibles con una confianza significativamente importante.

Tabla XVIII. **Análisis de varianza para un factor**

Fuente de variación	Grados de libertad	Suma de cuadrados	de Contribución	Cuadrado medio	<i>F</i>	<i>V p</i>
Criptodivisa	5	$4.2225 * 10^{24}$	27.24 %	$8.45019 * 10^{23}$	56.91	0.00
Error	12766	$1.12732 * 10^{25}$	72.76 %	$8.83068 * 10^{20}$		
Total	12771	$1.549863 * 10^{25}$	100.00 %			

Fuente: elaboración propia.

La prueba *F* de Fisher, muestra que el factor de criptodivisas supone un efecto significativo sobre el valor de mercado de capital, por ende, se procedió a evaluar las diferencias de todos los niveles con la prueba de Tukey (tabla XVIII), con lo que se pudo evaluar a un 95 % de confianza, y se encontró que tres grupos son significativamente iguales, y de igual forma son los grupos que suponen un nacimiento más tardío en la cadena de bloques, es decir que comparten características de crecimiento de valor muy similares como se aprecia en la sección anterior dada la tendencia en sus primeros años, por lo que es, una conclusión viable es la reafirmación de que se trata de un mercado financiero con un crecimiento de proyección favorable, esto puede respaldarse con las tablas X a la XV, en dónde la prueba de Kendall sugiere una tendencia creciente positiva.

Asimismo, se denota que para *bitcoin* se tiene una referencia de mejor correlación en relación al resto de divisas, dado que es la divisa que presenta una variabilidad mucho más marcada, como puede apreciarse en las desviaciones estándar para la media en la tabla XIX, en dónde el intervalo de confianza en unidades monetarias presentado para *bitcoin* es:

(48465900829, 50858393200) USD

Siendo estas íntegramente positivas, además teniendo hasta un 99 % más de variabilidad en relación a la desviación estándar (es decir mayor probabilidad de dispersión) en relación a la segunda de mayor peso que es *ethereum* como se aprecia en la tabla.

El intervalo de confianza (tabla XIX) para *bitcoin* sugiere el valor más favorable en relación dividendos positivos, lo cual la supone como candidata para ser considerada como una variable de control adecuada.

En la siguiente tabla puede apreciarse el resumen de las métricas descriptivas de los volúmenes de capital asociadas a las principales criptodivisas, las cuales son:

Tabla XIX. **Estadísticas descriptivas del volumen de capital total**

<b>Criptodivisa</b>	<b>N</b>	<b>Media</b>	<b>Desviación estándar</b>	<b>Intervalos de confianza</b>
<b><i>Bitcoin</i></b>	2371	$4.9662 * 10^{10}$	$6.4845 * 10^{10}$	(48465900829, 50858393200)
<b><i>Dash</i></b>	2118	948496077	1594106065	(-317182313, 2214174467)
<b><i>Ethereum</i></b>	1579	$2.0388 * 10^{10}$	$2.4026 * 10^{10}$	(18922398139, 21854136931)
<b><i>Litecoin</i></b>	2410	2032513917	3025550516	(845986381, 3219041454)
<b><i>Monero</i></b>	1982	893251426	1308427875	(-415130445, 2201633297)
<b><i>Ripple</i></b>	2312	7439229148	$1.2586 * 10^{10}$	(6227815596, 8650642701)

Fuente: elaboración propia

Tabla XX. **Prueba de Tukey para las medias de las criptodivisas**

<b>Criptodivisa</b>	<b>N</b>	<b>Media</b>	<b>Agrupación</b>
<i>Bitcoin</i>	2371	49662147015	<b>A</b>
<i>Ethereum</i>	1579	20388267535	<b>B</b>
<i>Ripple</i>	2312	7439229148	<b>C</b>
<i>Litecoin</i>	2410	2032513917	<b>D</b>
<i>Dash</i>	2118	948496077	<b>D</b>
<i>Monero</i>	1982	893251426	<b>D</b>

Fuente: elaboración propia.

Los supuestos del modelo general planteado fueron probados en primer lugar, por la normalidad de los residuos que generó el ajuste, a través de una prueba de Kolmogorov Smirnov, los resultados se presentan a continuación:

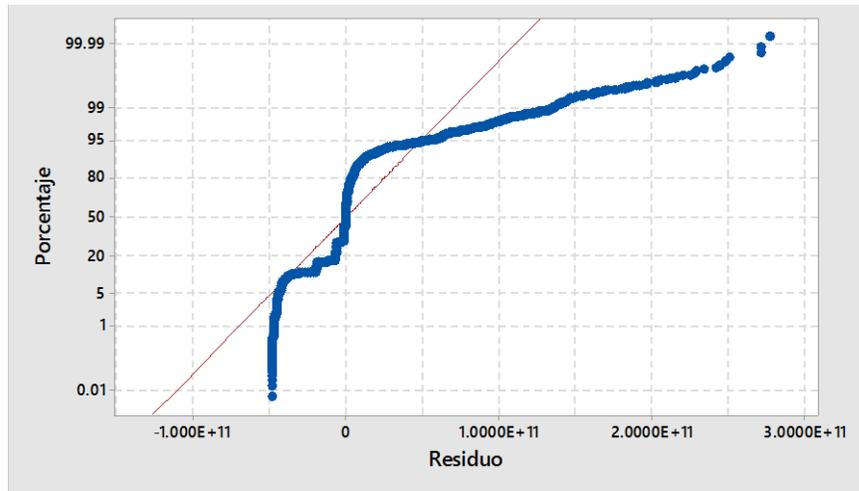
Tabla XXI. **Prueba de Kolmogorov para residuales**

<b>Modelo</b>	<b>ANOVA criptodivisas</b>
<b>D</b>	0.46
<b>Valor-P</b>	$1.29 * 10^{-9}$

Fuente: elaboración propia

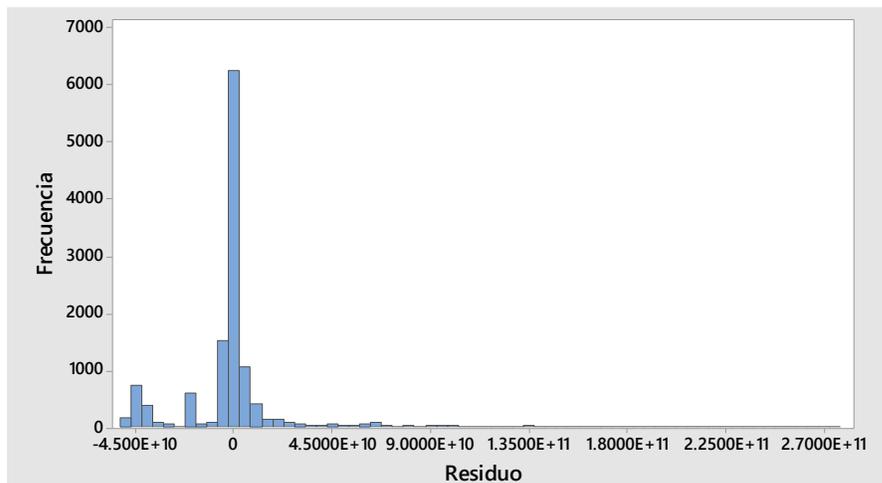
Estos no siguen una distribución normal como se puede apreciar gráficamente en las figuras 31 y 32 presentadas a continuación:

Figura 31. Prueba de normalidad de residuos para el factor divisa



Fuente: elaboración propia.

Figura 32. Histograma de residuos para el factor divisa



Fuente: elaboración propia.

Esto último quiere decir que el valor más probable podría ser modificando añadiendo factores que mejoren su ajuste, es decir, un análisis de variables adicionales, tales como el precio de apertura de mercado, así como el precio de cierre de mercado.

Sin embargo, no merece la pena, dado que estas cifras presentan una variabilidad aun mayor como se pudo observar en el resumen de estadísticas descriptivas de la tabla XIX, en dónde se presenta la desviación estándar y varianza para la media de cada divisa, en dónde a nivel de variación porcentual se tienen amplitudes probabilísticas (en relación a la desviación estándar de la curva) de hasta un 98 % tal es el caso de *ripple* en la tabla IX y *bitcoin* en la tabla IV, lo cual es un factor concluyente en primera instancia que inversiones en esta clase de activos pueden ser muy atractivas en términos de rendimiento financiero, sin embargo, reflejando una amplitud de apertura de los intervalos por debajo del valor 0 es decir pérdida neta.

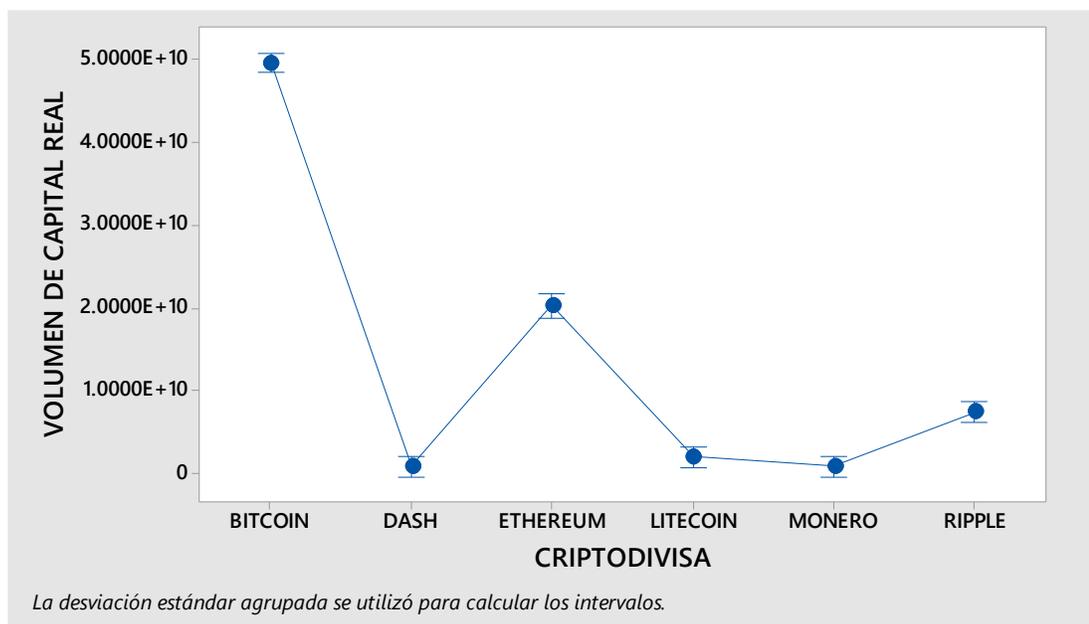
Lo anterior se puede apreciar en el caso de la divisa *monero* como referencia en la tabla XVII, el riesgo es alto debido a que existen momentos históricos en dónde el valor medio de la divisa en lugar de crear valor lo destruye, es decir está por debajo del 0, lo cual se traduce en que vender en un momento inadecuado la divisa supondrá una pérdida o una utilidad poco eficiente y, por ende, maximiza el riesgo de inversión.

Se confirma que los residuos del modelo lineal para el factor divisa de los datos no sigue una distribución normal, lo cual era un supuesto claro dada la variabilidad de cada criptodivisa que presenta un comportamiento más asimétrico conforme pasa cada año, lo cual es compensado por el factor de crecimiento que presenta cada divisa en relación a su tendencia, la cual se refleja en factores de

correlación bajos pero existentes a la divisa de *bitcoin*, esto es una señal de alerta, dado que el fin es determinar un modelo de predicción.

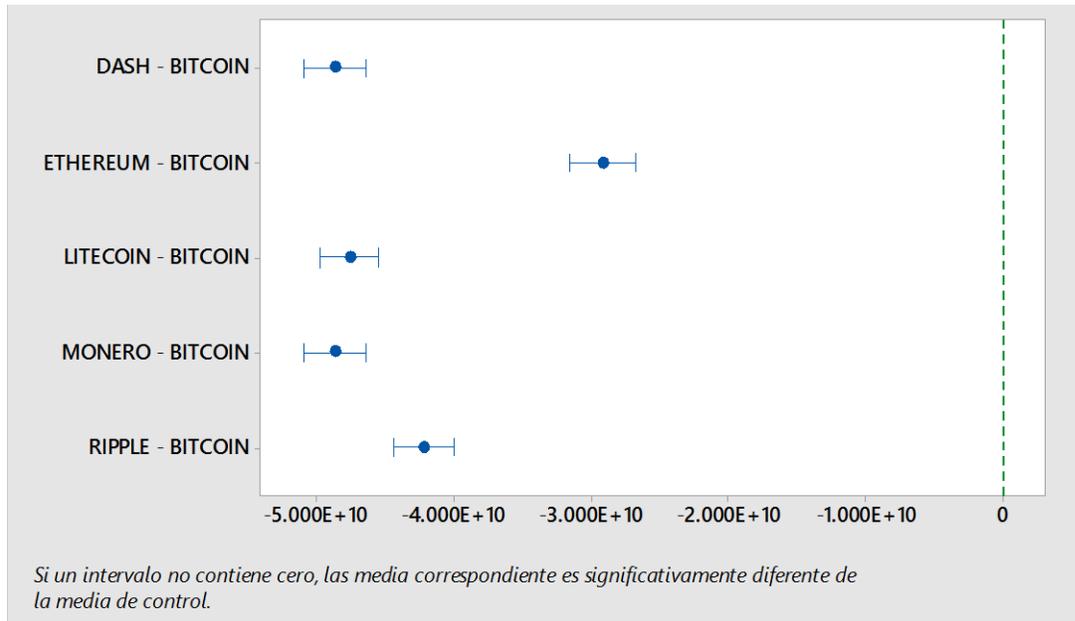
Radovanov *et al.* (2018) comentan una alternativa dada su alta varianza presente en general en el mercado de criptoactivos, el cual es realizar un modelo estadístico, cualquiera que fuese, bajo el supuesto de una proyección ponderada, en dónde el peso mayor lo supondrían las poblaciones de datos significativamente más homogéneas.

Figura 33. Intervalos de variación para el factor divisa



Fuente: elaboración propia.

Figura 34. Prueba de Dunnet para el factor divisa



Fuente: elaboración propia.

Para finalizar esta sección de cálculo, se realizaron dos pruebas post-hoc para el modelo asociado a la serie de datos considerando el factor divisa, esto con la finalidad de estimar como se encuentran los contrastes en relación a la divisa de mayor valor bursátil, como se aprecia en la figura 34, todas las variaciones se encuentran por debajo del 0, lo que demuestra que efectivamente la divisa de mayor valor es *bitcoin* y además lo es con creces, por ende sí puede ser un buen estimador en el comportamiento del resto.

### 3.2.1. Análisis de varianza de varios factores (MANOVA)

Analizando por medio del precio medio bursátil real de las principales criptodivisas para cuantificar la variación, se pretende encontrar la asociación o diferencia que existe en los patrones de alza y baja del valor medio real, en donde

se consideran como factores de diferenciación para un análisis de varianza multivariado las diferentes criptomonedas y los patrones de inflación y deflación en dónde la relación es una variable del tipo *dummy*, en relación a cuando el valor bursátil es mayor, si el de apertura o el de cierre bajo la siguiente relación:

$$RB = \text{Valor ajustado de cierre} > \text{Valor ajustado de Apertura}$$

MANOVA para el criterio de criptodivisa.

Tabla XXII. **Análisis múltiple de varianza para el criterio de divisa**

Criterio	Estadístico prueba	de Aproximación Fisher	GL		P-valor
			Número	Denominador	
De Wilks	0.62267	439.753	15	35233	0
Lawley-Hotelling	0.55302	470.498	15	38285	0
De Pillai	0.41032	404.509	15	38295	0
De Roy	0.42975				

Fuente: elaboración propia.

MANOVA para el criterio de relación bursátil.

Tabla XXIII. **Análisis múltiple de varianza para el criterio de relación bursátil**

Criterio	Estadístico		GL		P-Valor
	de prueba	Aproximación Fisher	Número	Denominador	
<b>De Wilks</b>	0.98763	53.284	3	12763	0
<b>Lawley-Hotelling</b>	0.01252	53.284	3	12763	0
<b>De Pillai</b>	0.01237	53.284	3	12763	0
<b>De Roy</b>	0.01252				

Fuente: elaboración propia.

Se observa que no existe una relación directa, ni bajo criterios de alza o baja del valor medio bursátil, tampoco por tipo de criptomoneda, por lo que el comportamiento debe diferenciarse en función de otro tipo para generar las mejores aproximaciones a su valor medio real, esta diferenciación podría trabajarse por medio de medias de valor real ponderado, en donde se tomen en consideración factores de las divisas de bajo valor con una ponderación más alta, dado que como se identifica en las tablas XXII y XXIII de análisis de varianza múltiple ninguno de los factores considerados es representativo en cuanto al valor medio bursátil real.

### **3.2.2. Análisis de varianza para los distintos mercados de capital**

Dado que una de las premisas de esta investigación es probar que, no existe diferencia significativa entre los mercados de capital para cada una de las divisas, se planteó analizar a distintas divisas en distintos mercados de capital por medio de un análisis de varianza, para tal fin se seleccionaron de forma aleatoria tres divisas a estudiar, y de igual forma se seleccionaron 15 fechas por

medio de muestreo por conglomerados de capital homogéneo es decir se seleccionaron a razón anual, para cada divisa en cada mercado de capital, el análisis no se hizo para todas las divisas dada la dificultad y el coste asociado a poder coleccionar los datos, sin embargo, las conclusiones se consideran estadísticamente significativas dados los resultados obtenidos; los cuales afirman que las divisas tienen el mismo valor de mercado así como precio; esto indiferente a cuál es el mercado de valores en dónde se haga la consulta, esa es una de sus principales virtudes, estos resultados se presentan en las tablas a continuación.

Las divisas evaluadas fueron *bitcoin*, *dash* y *litecoin*, además los mercados de capital o bolsas de valores consideradas son: la bolsa de valores de Nueva York NYSE, la bolsa de valores de Shanghái SSE, y la bolsa de valores de Londres LSE. Los resultados obtenidos son los siguientes.

Tabla XXIV. **Mercados de capital para *bitcoin***

<b>Información del factor</b>		
Factor	Niveles	Valores
Mercado de Capital	3	<i>bitcoin</i> LSE, <i>bitcoin</i> NYSE, <i>bitcoin</i> SSE

Fuente: elaboración propia.

Tabla XXV. **ANOVA de *bitcoin* para mercados de capital**

Análisis de Varianza					
Fuente	GL	SC ajustado	MC ajustado	Valor F	Valor p
Mercado de Capital	2	10457	5228	0.13	0.879
Error	12	481607	40134		
Total	14	492064			

Fuente: elaboración propia

Tabla XXVI. **Mercados de capital para *dash***

Información del factor		
Factor	Niveles	Valores
Mercado de capital	3	<i>DASH</i> LSE, <i>DASH</i> NYSE, <i>DASH</i> SSE

Fuente: elaboración propia.

Tabla XXVII. **ANOVA de *dash* para mercados de capital**

Análisis de Varianza					
Fuente	GL	SC Ajustado	MC Ajustado	Valor F	Valor p
Mercado de capital	2	21.16	10.58	0.36	0.702
Error	12	348.45	29.04		
Total	14	369.61			

Fuente: elaboración propia.

Tabla XXVIII. **Mercados de capital para ripple**

Información del factor		
Factor	Niveles	Valores
Mercado de Capital	3	RIPPLE LSE, RIPPLE NYSE, RIPPLE SSE

Fuente: elaboración propia.

Tabla XXIX. **ANOVA de ripple para mercados de capital**

Análisis de Varianza					
Fuente	GL	SC Ajustado	MC Ajustado	Valor F	Valor p
Mercado de Capital	2	0.0001	0.00005	0.16	0.85
Error	12	0.003631	0.000303		
Total	14	0.003731			

Fuente: elaboración propia.

Se deseaba probar el supuesto que el valor bursátil de las divisas era el mismo, sin importar la bolsa de valores en la que se tomara la referencia o la moneda principal del país en dónde se realiza el análisis, los resultados son bastante contundentes, dado que los supuestos para el análisis de varianza de un factor son los siguientes:

*Ho: No existe diferencia significativa en las medias consideradas en el análisis.*

*HA: Existe diferencia significativa en las medias consideradas en el análisis*

Como muestran las tablas XXVI, XXVII, XXVIII se falla en rechazar la hipótesis nula planteada, y el valor p denota una significancia considerable, por lo que se puede concluir que en efecto, no existe diferencia significativa en el valor bursátil real de todas las criptodivisas, lo cual prueba que uno de sus

supuestos principales es real, el cual es no estar sujeto a ninguna entidad financiera regulatoria, sea privada o estatal, por ende, no supondría la existencia de una variación en su precio en cualquier país en dónde se realice el análisis.

### **3.3. Análisis de series temporales**

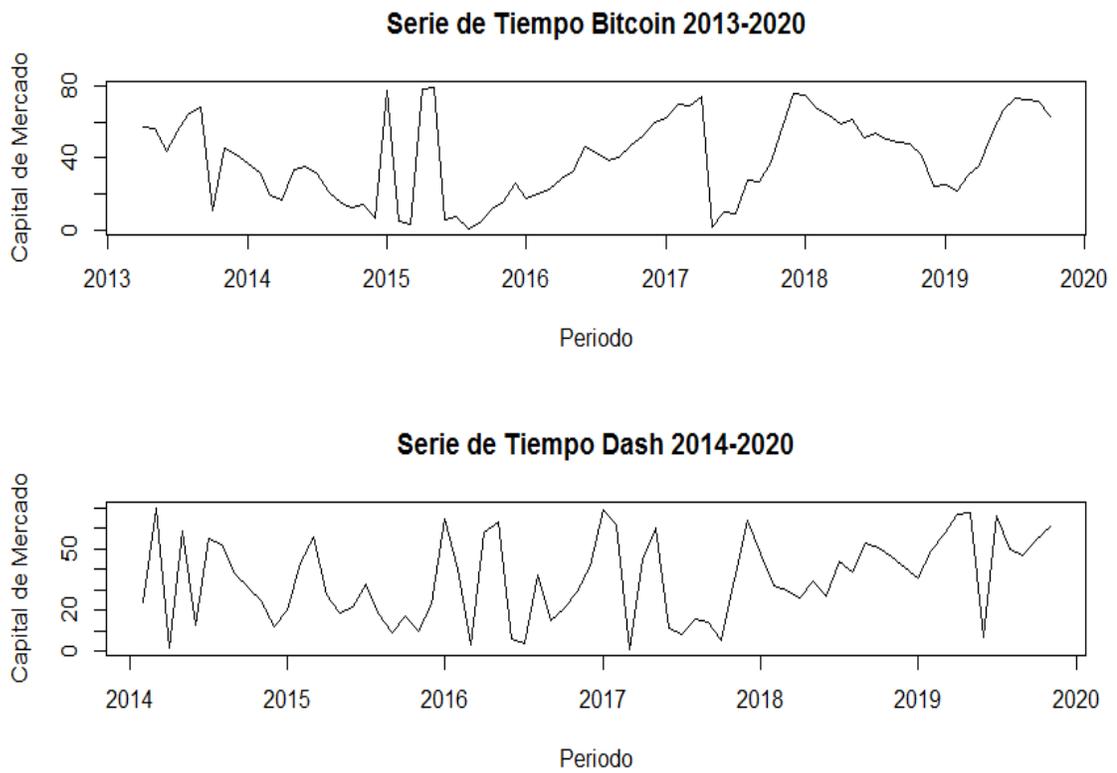
Utilizando los valores de volumen de mercado se realizó el análisis de series temporales para un análisis integral a través del tiempo de todas las divisas, bajo los supuestos que los datos son completamente independientes, distribuidos de forma aleatoria, sin correlación serial y sin heterocedasticidad.

Las series de tiempo se presentan en las figuras siguientes, en donde se puede ver un comportamiento particular para cada divisa, la diferenciación en cuanto al tiempo del análisis es debido a que las divisas al ser relativamente nuevas, no han nacido durante el mismo periodo de tiempo, es decir algunas son más antiguas que otras, sin embargo a simple vista se puede observar que los supuestos relacionados a aleatoriedad y valores extremos se cumplen, dado que a que visualmente no se perciben puntos con una caída significativamente brusca para ninguna divisa, fuera de las pérdidas de valor naturales en el mercado de divisas.

Dado ese comportamiento tan aleatorio, es difícil identificar si existen ciertas condiciones similares, homogéneas o simétricas en su comportamiento que favorezcan el planteamiento de un modelo predictivo por medio de series de tiempo, para ello se planteó una descomposición de cada serie en sus componentes de tendencia, estacionalidad, ciclo y aleatoriedad, los resultados de ambos análisis se presentan en las gráficas siguientes en dónde posteriormente se analizará a detalle una descomposición exhaustiva de cada

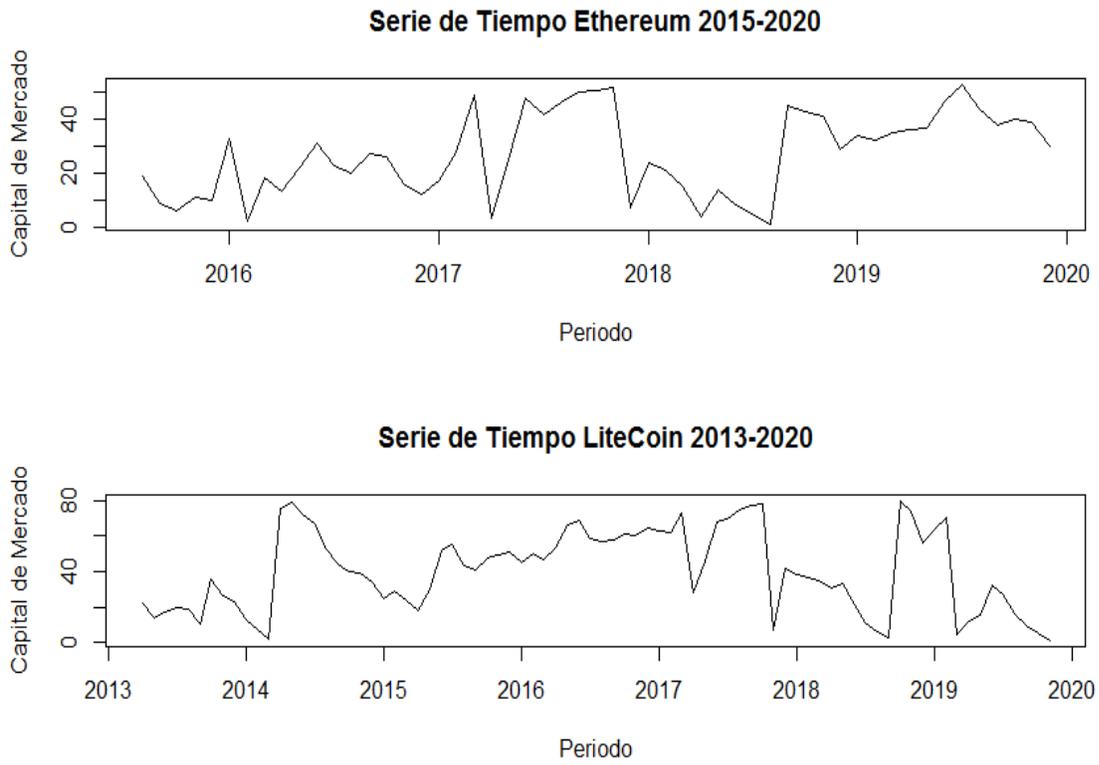
serie, sin embargo, a fines visuales es un apoyo significativo para comenzar a visualizar cada fenómeno.

Figura 35. **Serie de tiempo para *bitcoin* y *dash***



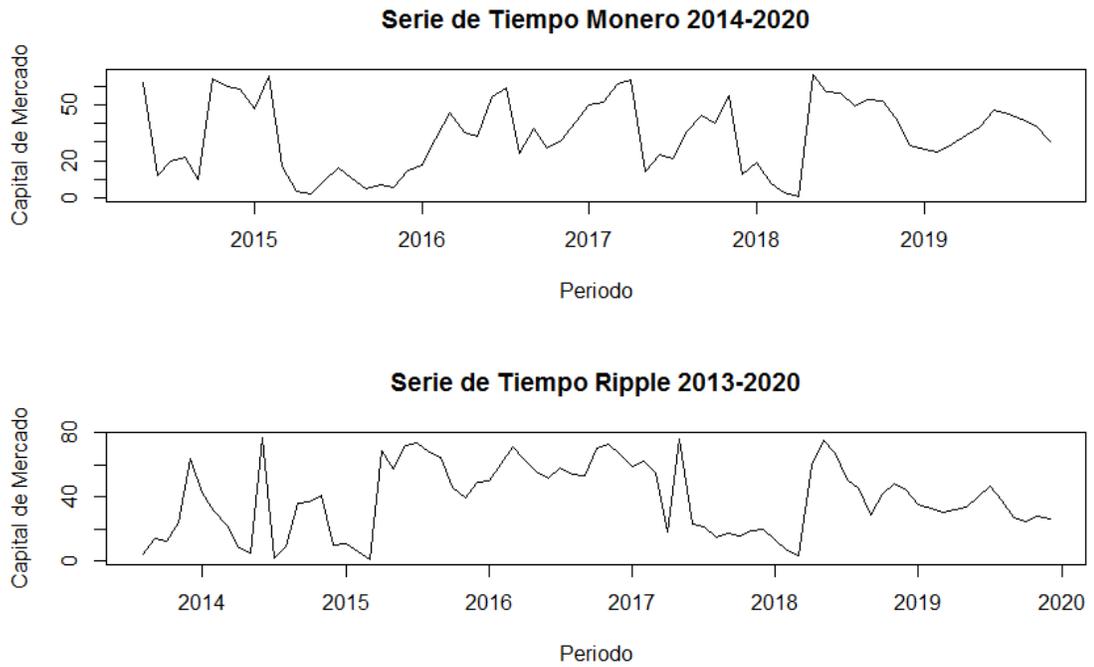
Fuente: elaboración propia.

Figura 36. **Serie de tiempo para *ethereum* y *litecoin***



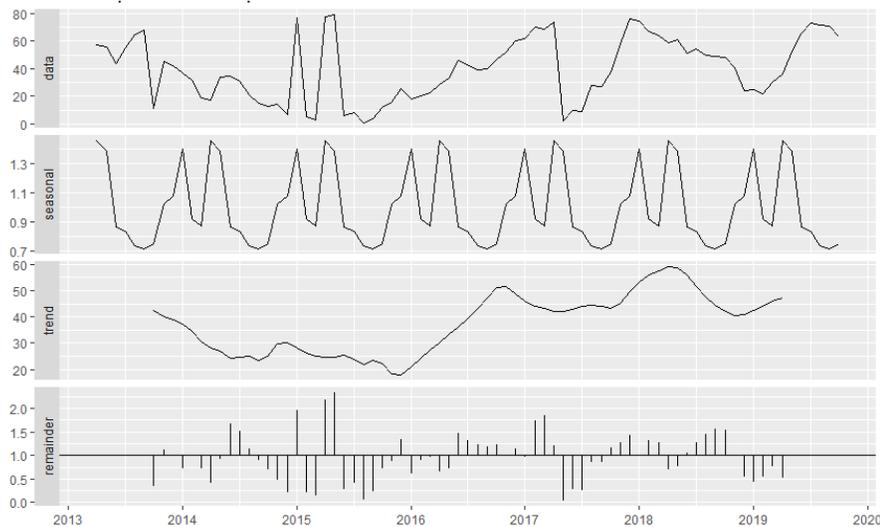
Fuente: elaboración propia.

Figura 37. **Serie de tiempo para *monero* y *ripple***



Fuente: elaboración propia.

Figura 38. Componentes de la serie de tiempo de *bitcoin*

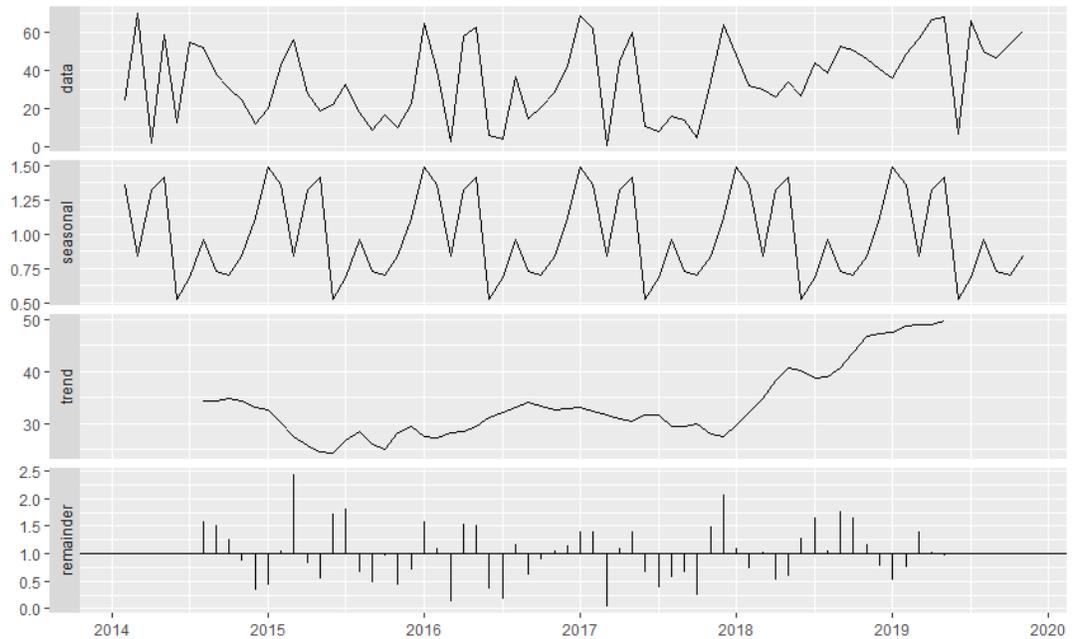


Fuente: elaboración propia.

En la figura 38, puede apreciarse una figura dividida en cuatro bloques, que de arriba hacia abajo son: la serie de datos, la componente estacional gráfica, la tendencia gráfica y los atrasos o rezagos de la serie, este formato de gráfico se presentará en adelante hasta la figura 43, con la finalidad de presentar un análisis integral sobre cada serie de tiempo.

En el caso de *bitcoin*, se puede apreciar que sí existe una tendencia visible, una tendencia hacia la baja en sus primeros años en el mercado de divisa, que ahora gráficamente presenta un comportamiento hacia el alza, y rezagos poco simétricos con varias alzas, más adelante se analizarán a detalle por medio de diferencias estacionales.

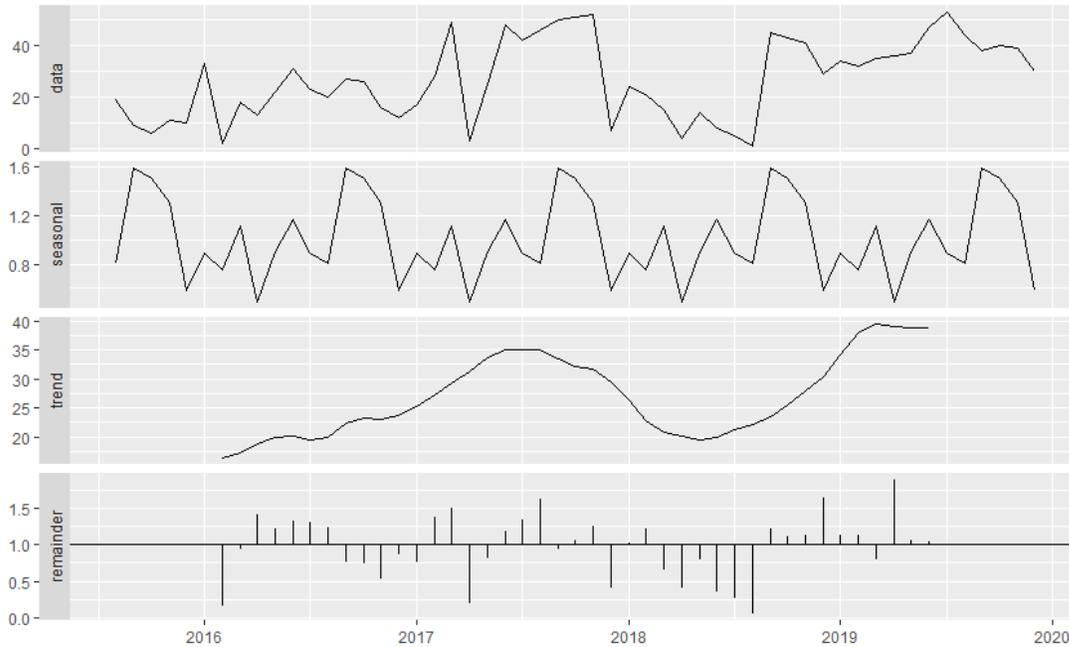
Figura 39. Componentes de la serie de tiempo de *dash*



Fuente: elaboración propia.

En el caso de *dash* se aprecia una clara componente estacional gráfica, por lo que en primera instancia se puede asumir estacionalidad para la serie, esto se validará posteriormente con una prueba estadística, además como un análisis preliminar se tiene una tendencia bastante simétrica durante gran parte de su comportamiento bursátil histórico, teniendo una componente hacia el crecimiento durante los últimos dos años, los atrasos por su parte se ven de forma homogénea, fuera de algunas diferencias en los años 2015 y 2017.

Figura 40. Componentes de la serie de tiempo de *ethereum*

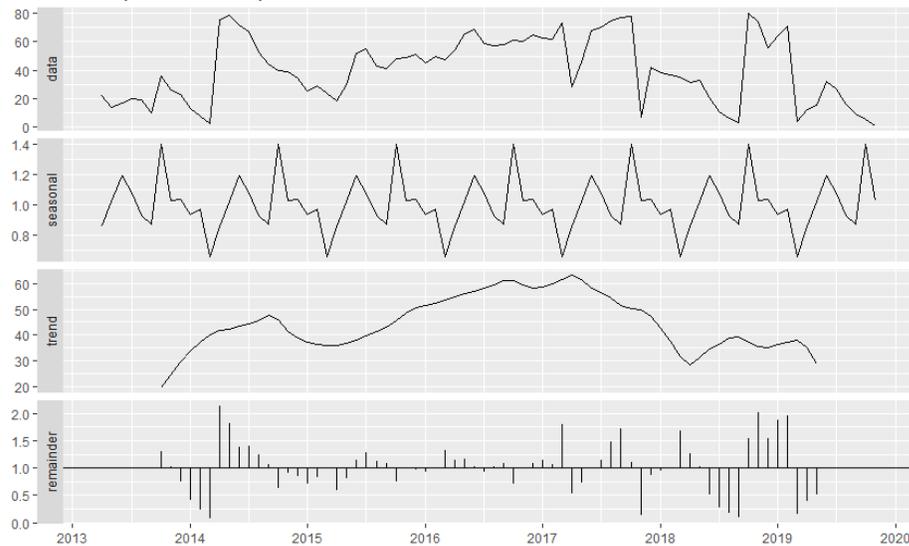


Fuente: elaboración propia.

*Ethereum* presenta una componente estacional gráfica clara, al menos como un análisis preliminar de estacionalidad, es un hallazgo favorecedor para poder modelar el comportamiento, en cuanto a la tendencia, se tiene un efecto parabólico, lo cual demuestra que ha presentado caídas significativas de valor, sin embargo, durante los últimos dos periodos presenta una tendencia hacia el alza.

En cuanto a los atrasos y rezagos se aprecia homogeneidad en los mismos, lo cual favorece a que un posterior análisis de estacionalidad, es probable que pueda modelarse a la serie sin necesidad de realizar muchas diferencias estacionales.

Figura 41. Componentes de la serie de tiempo de Litecoin

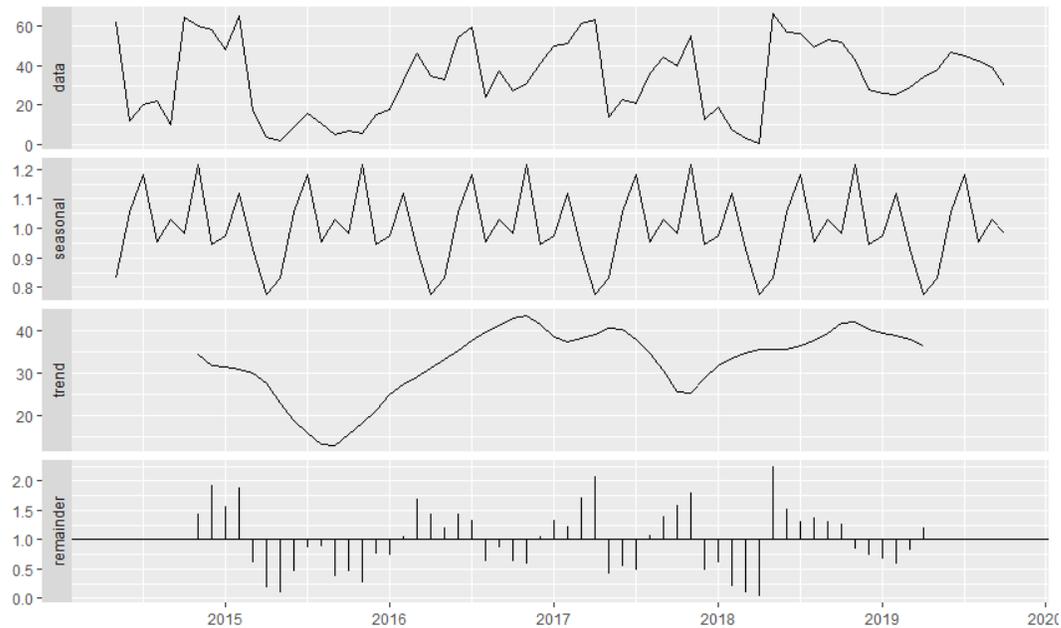


Fuente: elaboración propia.

En el caso de *litecoin* al igual que en las series anteriores se presenta una componente estacional al menos de forma gráfica, por lo que de manera preliminar se puede concluir que efectivamente existe una componente estacional, además, en cuanto a la tendencia se aprecia que ha tenido un comportamiento mayoritario al alza desde su nacimiento, sin embargo, presenta un comportamiento mayoritario a la baja durante el último año.

En cuanto a los atrasos tiene un comportamiento muy uniforme durante el surgimiento de la divisa, además de un periodo en dónde no existen alteraciones bruscas (2015-2017), sin embargo, durante los últimos dos años si presenta alteraciones significativas en los atrasos, lo cual demuestra que es probable que la divisa se haya apreciado, y para poder modelar de mejor forma el comportamiento, sea necesario realizar algún ajuste por diferencias estacionales.

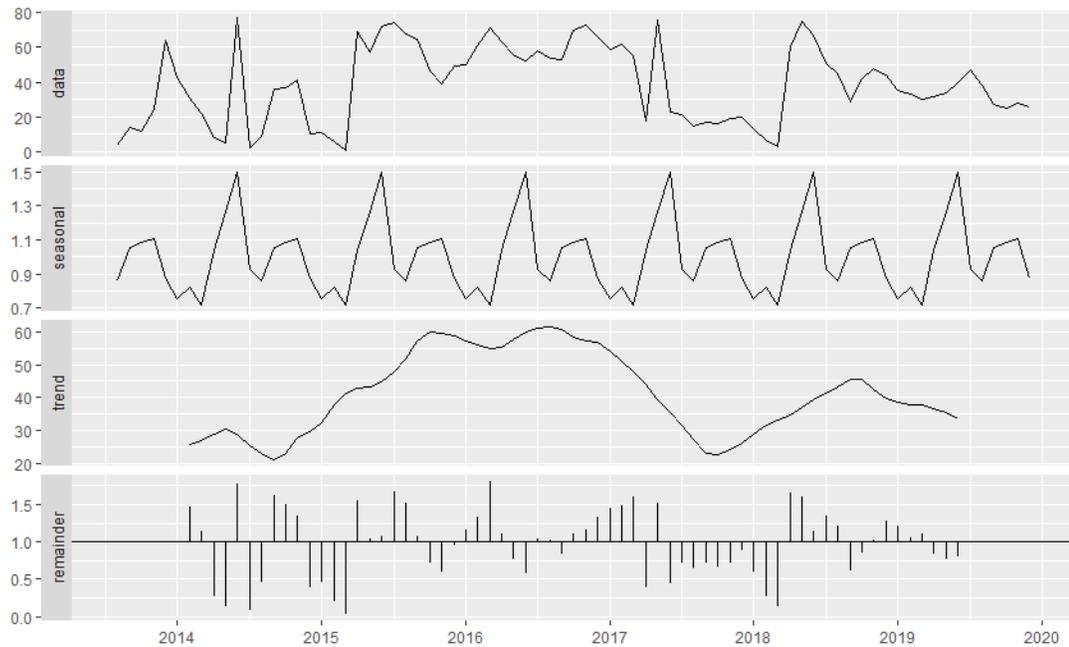
Figura 42. Componentes de la serie de tiempo de *monero*



Fuente: elaboración propia.

En el caso de *monero* la componente estacional es bastante evidente, con cuatro picos estacionales inter anuales, además de una caída significativa al final del periodo, esto brinda una excelente perspectiva de inversión, al menos de forma gráfica, por lo que de forma preliminar también se concluye que efectivamente la serie cuenta con una componente estacional, además, en relación a la tendencia, al final del periodo 2020 se presenta una componente visual positiva, lo cual favorece los supuestos de apreciación de la divisa.

Figura 43. Componentes de la serie de tiempo de Ripple



Fuente: elaboración propia.

En el caso de *ripple*, es clara una componente gráfica estacional para todos los años, con un pico estacional bastante bien definido, además, en cuanto a su tendencia, a pesar de que se presenta una caída de capitalización considerable durante el periodo 2018, la divisa pudo nivelarse y apreciarse con una tendencia positiva posteriormente. Los atrasos finales favorecen al estudio, dado que, al no presentar variaciones significativamente bruscas, es probable que se pueda modelar con la serie de datos integra.

Para condensar los hallazgos de gráficos anteriores se puede sintetizar que el comportamiento de la serie indistinto a la moneda analizada, tiene ciertos patrones particulares, a pesar de evaluar periodos relativamente no homogéneos dado a que algunas divisas nacieron antes que otras, sin embargo, pueden

apreciarse ciertos comportamientos específicos, tales como cambios abruptos en la varianza.

Asimismo, como se aprecia en los rezagos, la variabilidad es alta y volátil, pero con atrasos en periodos específicos lo cual es una bondad, debido a que es un indicio claro de autocorrelación al menos de forma gráfica, el factor clave a estudiar en todas las series es que como se aprecia en las figuras 38 a la 43 en su tercer apartado, existe un factor estacional en todas las divisas, es decir no se aprecia continuidad en el horizonte temporal.

Por lo que evaluar la significancia de los periodos que denotan estacionalidad en determinados periodos es crucial para generar un modelo de autocorrelación adecuado dado a que validara si efectivamente es necesario diferenciar o no a la serie de tiempo en cuestión, además, de validar si matemáticamente estos patrones estacionales podrían optimizar la recomendación de cuando comprar o vender o ayudar al modelo a predecir de una mejor forma, así mismo como es el caso de XRP *ripple*, a pesar de que su tendencia en un momento llegó a ser decreciente, la serie tiende a normalizarse en un punto y comenzar a crecer de nuevo, como lo refleja la prueba de Mann Kendall en el apartado anterior.

Dados los patrones visuales de estacionalidad presentados en las series; es necesario profundizar un poco en este aspecto para encontrar las mejores asociaciones que mejor describan su comportamiento histórico, definir si es necesario diferenciar a la serie en cuestión y además comprender de mejor forma el comportamiento de cada divisa.

### 3.3.1. Análisis de estacionalidad de la serie

Se realizó un gráfico de la estacionalidad y estacionariedad, para ello se recurrió a la técnica de transformación de datos por medio de curvas paramétricas planteada por los autores Cryer y Chan (2008), en dónde plantean la practicidad en el estudio de estacionalidad en series temporales transformando los gráficos por medio de coordenadas polares, utilizando como ejes radiales los meses o periodos que componen la serie, en el caso ya que las series se presentan por medio de datos anuales, presentarlo por medio de estas cifras es ideal.

Para este análisis se necesita realizar ciertas transformaciones en cuanto a los ejes gráficos, los cuales están planteados por las siguientes expresiones trigonométricas.

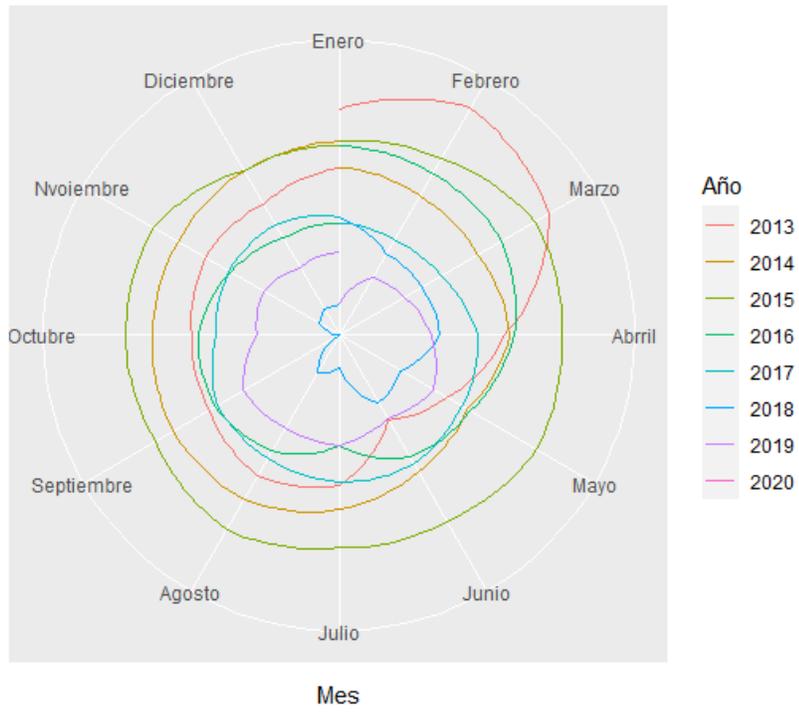
$$x = 2 * \text{Sen}(nt) * \text{Cos}(nt) \text{ (Ec. 48)}$$

$$y = 2 * \text{Sen}(nt) * \text{Sen}(nt) \text{ (Ec. 49)}$$

$$\text{Función } (rx) = 2 * \text{Sen}(nx) \text{ (Ec. 50)}$$

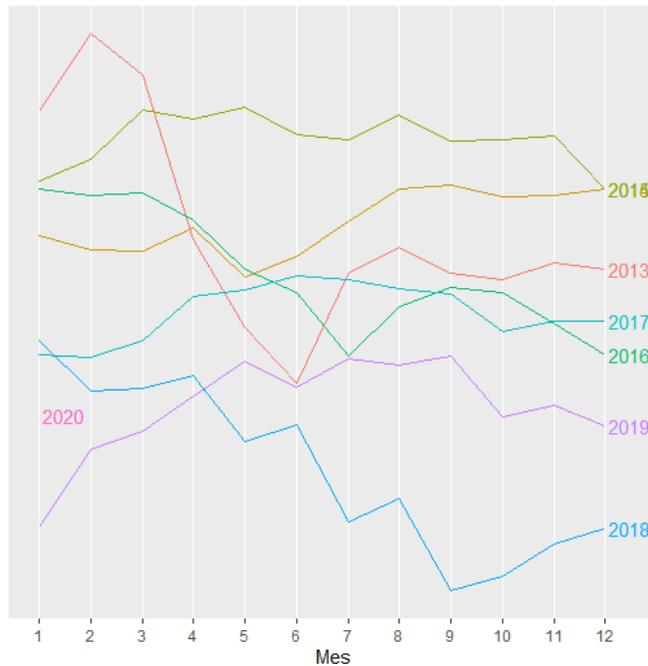
Con esto definido, se creó un *Script* computacional que facilite el ingreso del gran volumen de datos y con ello analizar las estacionalidades de la serie en conjunto para todos los años por medio de este método.

Figura 44. Transformación polar de *bitcoin*



Fuente: elaboración propia.

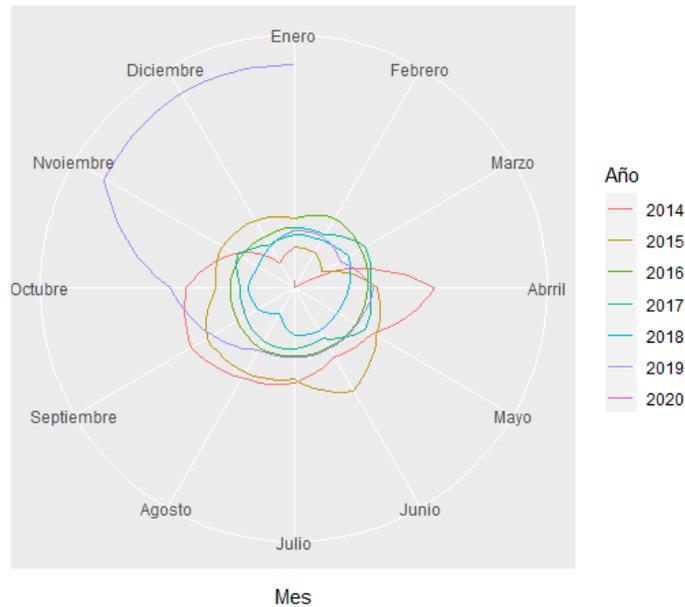
Figura 45. **Estacionalidad interanual de *bitcoin***



Fuente: elaboración propia.

Se aprecia en la figura 44 que gráficamente la serie aparenta tener una componente estacional, esto posteriormente se validará con la prueba de Dickey-Fuller, sin embargo, es muy probable que, si exista un factor estacional, dado a que como se aprecia en la figura 45, el año 2018 es el único que refleja una pérdida significativa de valor, el resto de año reflejan cruces entre curvas.

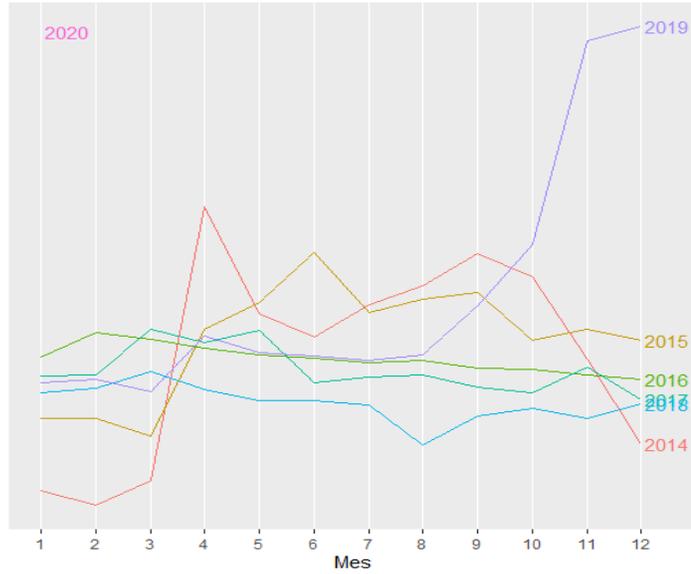
Figura 46. Transformación polar de *dash*



Fuente: elaboración propia.

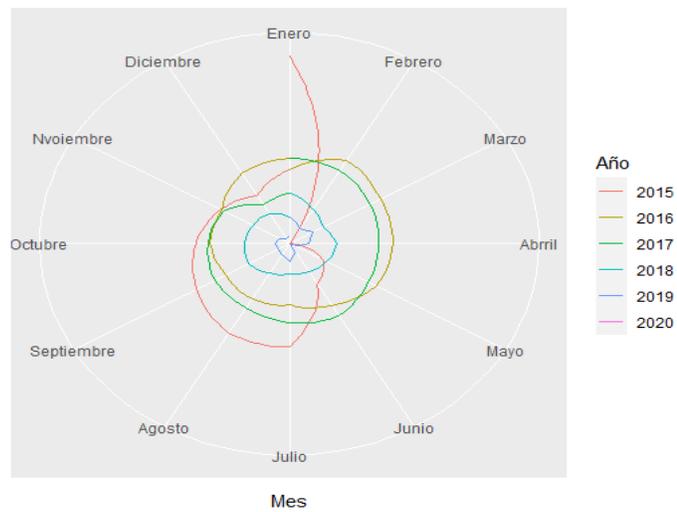
La estacionalidad en *dash* es más evidente que en el caso de *bitcoin*, dado que la mayoría de años, la figura circunscrita es bastante estable al centro como se aprecia en la figura 46, la cola de valor extremo corresponde al año 2019, que como se aprecia en la figura 47 se debe a la inflación en el valor bursátil asociado a la divisa, de forma gráfica si se aprecia una componente estacional.

Figura 47. **Estacionalidad interanual de dash**



Fuente: elaboración propia.

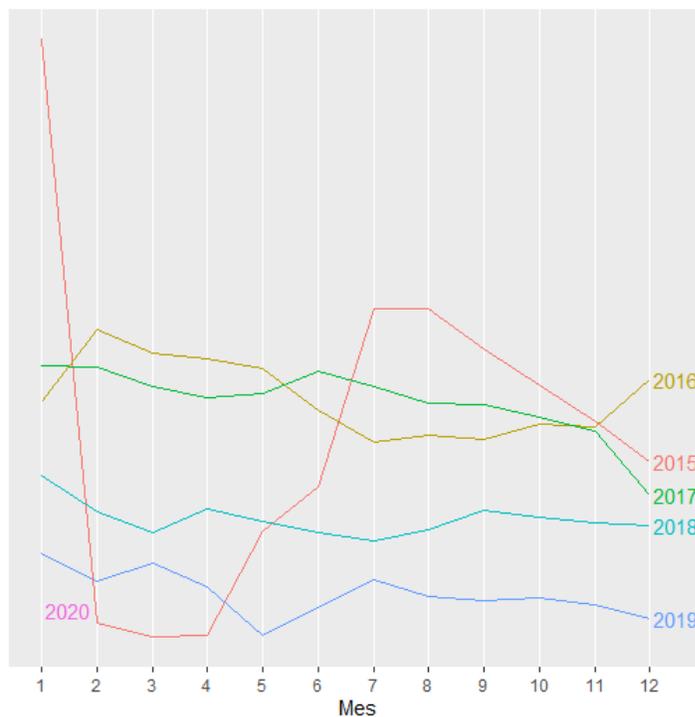
Figura 48. **Transformación polar de *ethereum***



Fuente: elaboración propia.

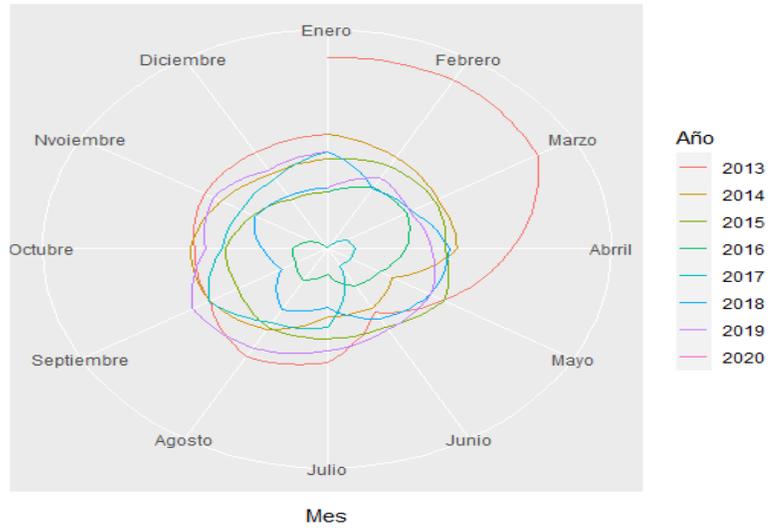
En cuanto a la componente estacional de *ethereum* no se tiene una referencia gráfica contundente para afirmar si existe o no, como se aprecia en la figura 48 y 49 no existe un corte entre curvas demasiado evidente, además que las variaciones en su valor hacia la baja reflejan una distancia entre curvas significativa, por lo que una prueba de estacionalidad es necesaria para validar este supuesto.

Figura 49. **Estacionalidad interanual de *ethereum***



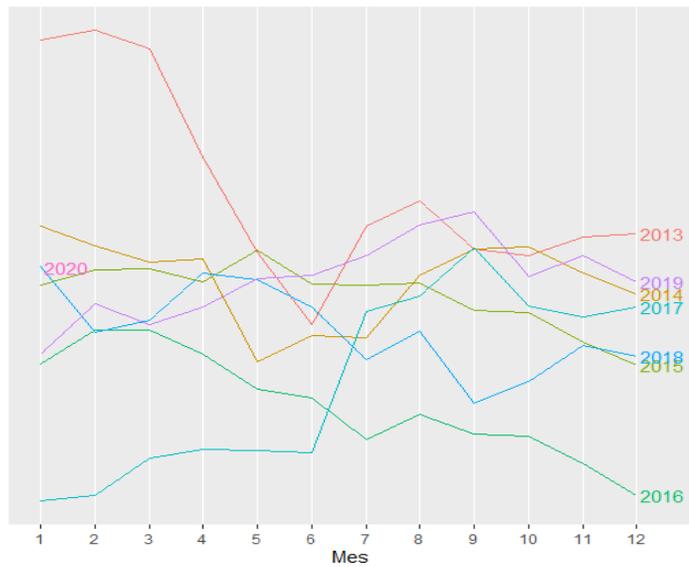
Fuente: elaboración propia.

Figura 50. **Transformación polar de *litecoin***



Fuente: elaboración propia.

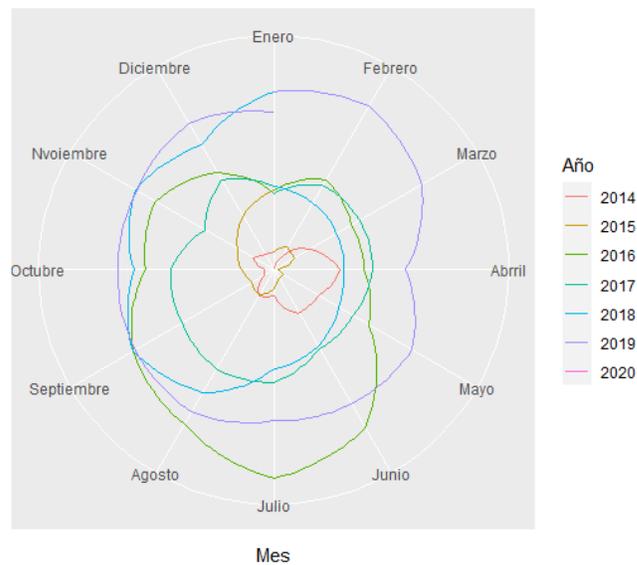
Figura 51. **Estacionalidad interanual de *litecoin***



Fuente: elaboración propia.

*Litecoin* muestra de forma gráfica un claro componente estacional, dado a que como se aprecia en la figura 51, existen sobreposiciones continuas en la figura que circunscribe la gráfica polar, esto tiende a incrementar en los meses de julio, agosto y septiembre, además que la cola de valor extremo que se aprecia en el gráfico polar tiene su explicación dado el valor extremadamente alto de introducción bursátil de la divisa.

Figura 52. **Transformación polar de *monero***

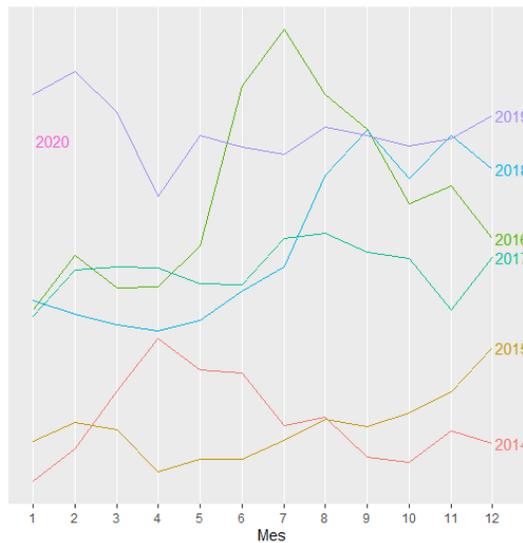


Fuente: elaboración propia.

En la figura 52 para la divisa *monero* se aprecia que es una divisa en donde dada la distancia entre curvas ha sufrido de cambios significativos en el valor de volumen de capital bursátil en los últimos años, sin embargo, a pesar de ello, aun circunscribe una figura, por lo que es muy probable que exista una componente estacional, la prueba de ello para esta y todas las divisas estudiadas se presenta a continuación del análisis gráfico de estacionalidad.

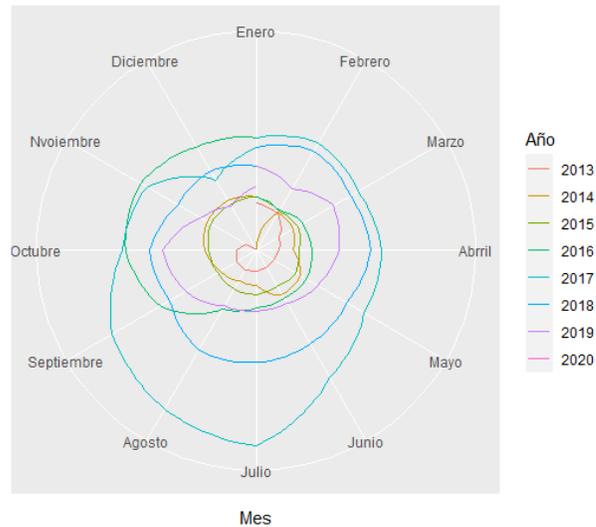
La figura 53 muestra que *monero* en efecto ha sufrido variaciones significativas de valor desde su nacimiento, lo cual la vuelven un potencial activo de interés para inversión.

Figura 53. **Estacionalidad interanual de *monero***



Fuente: elaboración propia.

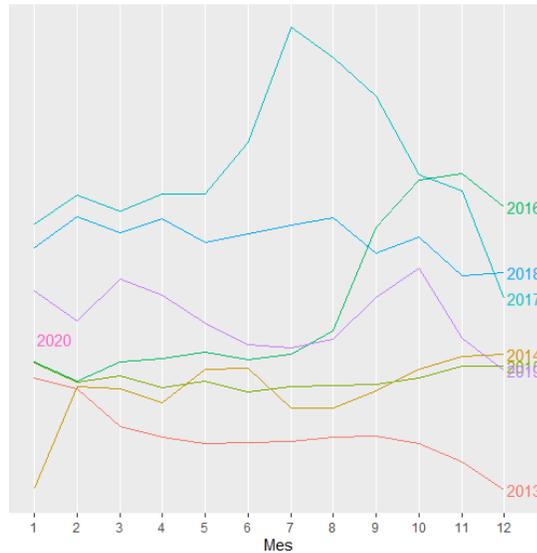
Figura 54. **Transformación polar de *ripple***



Fuente: elaboración propia.

La figura 54 muestra un evidente factor estacional ya que las circunscripciones mensuales son bastante homogéneas, además que la amplitud entre curvas bajo la misma apariencia refleja que la serie tiene una componente estacional y que se ha apreciado de forma anual, lo cual puede corroborarse en la figura 55.

Figura 55. **Estacionalidad interanual de *ripple***



Fuente: elaboración propia.

Para sintetizar los hallazgos comentados de las componentes estacionales para las seis divisas, se concluye que si bien los factores de alza y baja son muy particulares y específicos para cada serie de datos, no se puede evitar observar que existen factores secuenciales en el comportamiento de determinadas series, tal es el caso de *dash* en el gráfico en dónde existen caídas secuenciales en junio, o en *ripple*, en dónde se observa una secuencialidad con desfase a partir de marzo en el gráfico, por otro lado, existen ciertos momentos en el que los patrones de altas y bajas se vuelven extremadamente variables, tal es el caso de *bitcoin* para años más recientes.

Para poder validar los supuestos de estacionalidad que se presentan de forma gráfica en el apartado anterior, se aplicó una prueba de estacionalidad de Dickey Fuller la cual además de probar la hipótesis de que las series efectivamente tienen una componente estacional, servirá como guía para saber

si es necesario diferenciar las series antes de comenzar el proceso de creación de un modelo. A continuación, se presentan los resultados de la prueba:

Tabla XXX. **Prueba de Dickey Fuller para todas las divisas**

<b>Serie</b>	<b>Bitcoin</b>	<b>Serie</b>	<b>Dash</b>
Dickey-Fuller	-12.5969	Dickey-Fuller	-6.9857
Orden del atraso	6	Orden del atraso	9
Valor-P	0.0035	Valor-P	0.001
<b>Serie</b>	<b>Litecoin</b>	<b>Serie</b>	<b>Ethereum</b>
Dickey-Fuller	-8.547	Dickey-Fuller	-11.201
Orden del atraso	6	Orden del atraso	6
Valor-P	0.001	Valor-P	0.001
<b>Serie</b>	<b>Monero</b>	<b>Serie</b>	<b>Ripple</b>
Dickey-Fuller	-6.999	Dickey-Fuller	-10.113
Orden del atraso	6	Orden del atraso	6
Valor-P	0.0038	Valor-P	0.001

Fuente: elaboración propia.

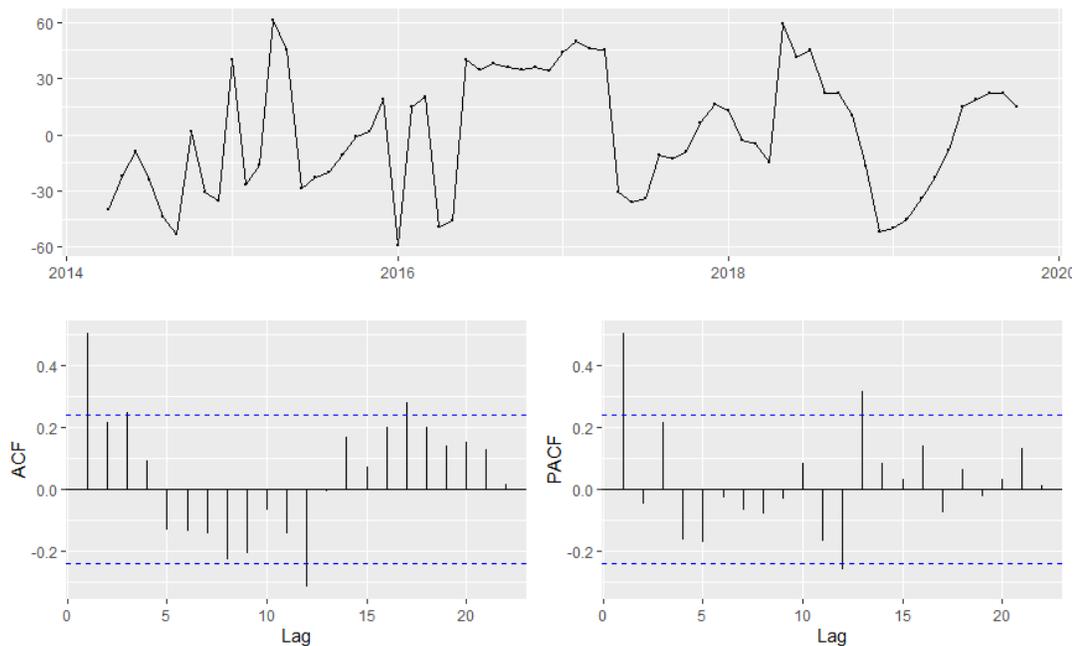
Dado a que para todas las pruebas se falla en rechazar la hipótesis nula dada la significancia del valor  $p$ , se concluye que todas las divisas tienen una componente estacional asociada, por lo que es necesario diferenciar las series para eliminar el componente estacional de la serie además de su tendencia, y poder comenzar a crear un modelo autorregresivo de forma adecuada.

### 3.3.2. Diferencias estacionales de las series

Se plantea el uso de modelos autorregresivos no asintóticos para generar predicciones de las series de datos. Para validar el uso de este tipo de modelos y realizar iteraciones teóricas sobre los mismos, es necesario comprobar que se ha eliminado de forma adecuada el componente estacional, así como a la

tendencia de la serie, de esa forma se conserva de forma íntegra la variabilidad de los datos. Para ello se presentan a continuación las diferenciaciones más eficientes para las series de las criptodivisas.

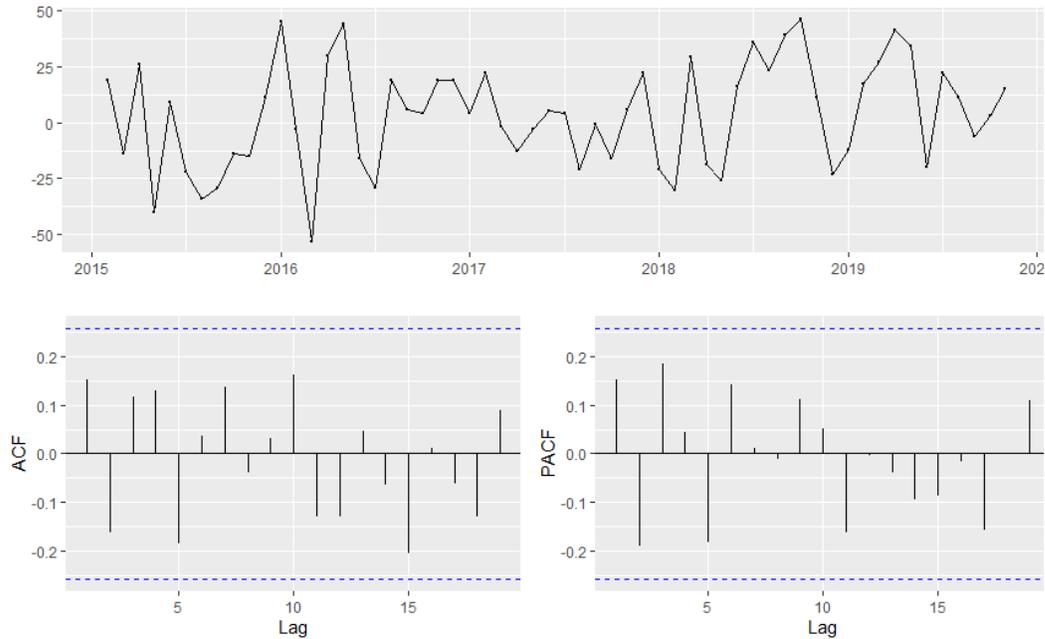
Figura 56. **Diferencias estacionales ACF y PACF de *bitcoin***



Fuente: elaboración propia.

Se aprecia el efecto de la diferencia estacional en *bitcoin* con una curva mucho más alisada, y con atrasos homogéneos tanto para los gráficos de autocorrelación total como parcial, como se aprecia en la figura 56, estos saltos en los atrasos corresponden a las altas variaciones mensuales que tiene *bitcoin* como se aprecia en el inciso 3.1.1.1 en la figura 2.

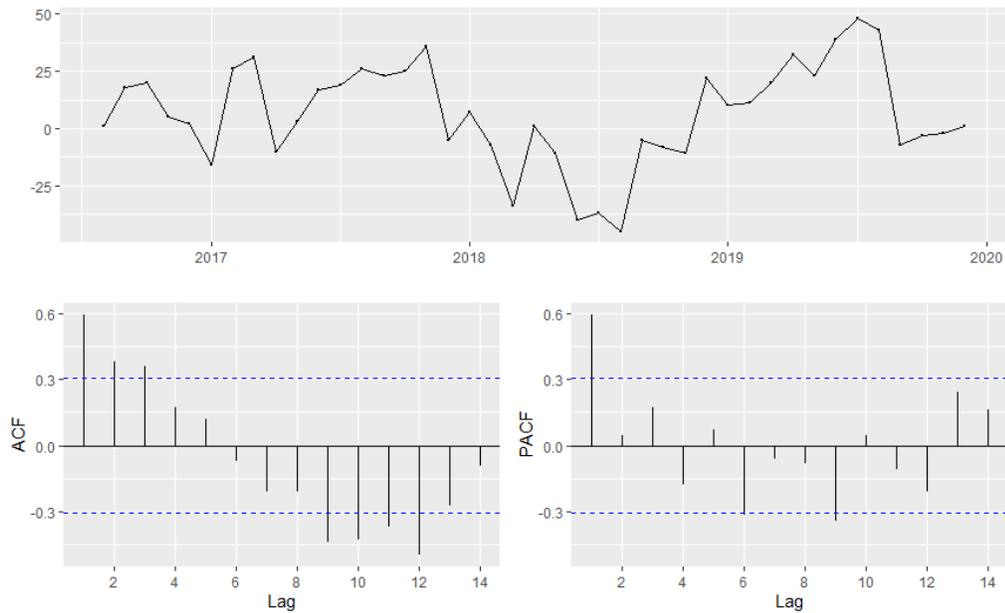
Figura 57. Diferencias estacionales ACF y PACF de *dash*



Fuente: elaboración propia.

En cuanto a *dash* se puede apreciar una curva mucho más alisada después de las diferencias estacionales óptimas, además de que, tanto para el gráfico de autocorrelación total como parcial, ya ningún atraso se encuentra fuera de las bandas, por ende, no debería generar ningún inconveniente la serie al momento de plantear un modelo autorregresivo.

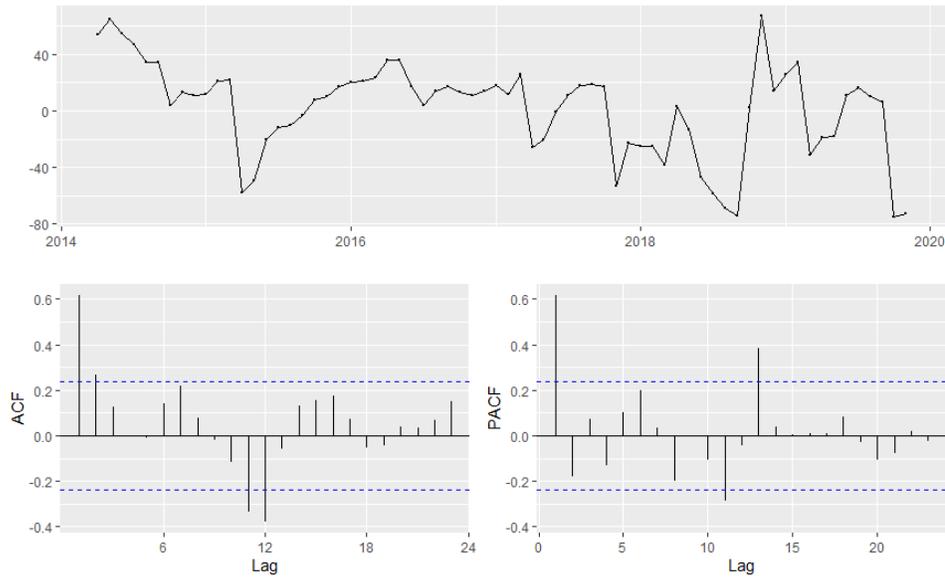
Figura 58. **Diferencias estacionales ACF y PACF de *ethereum***



Fuente: elaboración propia.

En cuanto a *ethereum* se aprecia el efecto de caídas homogéneo a la sección anterior de una mejor forma, la cual se refleja también en el gráfico de autocorrelación total cada tres atrasos, sin embargo, el gráfico de autocorrelación parcial se encuentra dentro de los límites normales, lo cual supone que la serie es adecuada para generar un modelo relacionado a la misma.

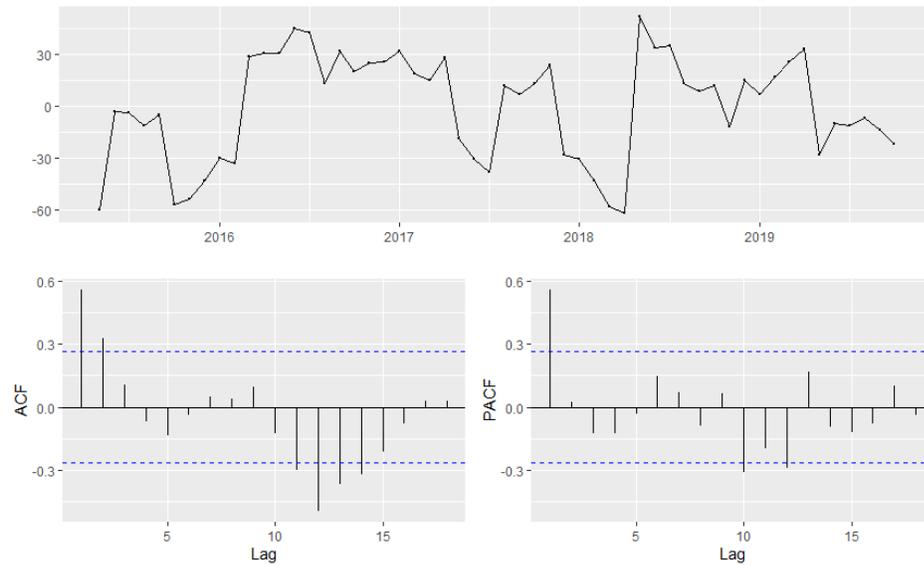
Figura 59. **Diferencias estacionales ACF y PACF de *litecoin***



Fuente: elaboración propia.

Las diferencias estacionales óptimas para *litecoin* reflejan una curva completamente alisada, además de que los atrasos se encuentran en su mayoría dentro de los límites, a excepción de los meses en los que presenta un intervalo de valor más amplio como se aprecia en la línea de tendencia para *litecoin* que se presenta en la figura 17, sin embargo, este efecto se ve mermado en el gráfico de autocorrelación parcial, lo cual es un buen indicio de independencia entre los datos y que el factor estacional fue eliminado de forma satisfactoria.

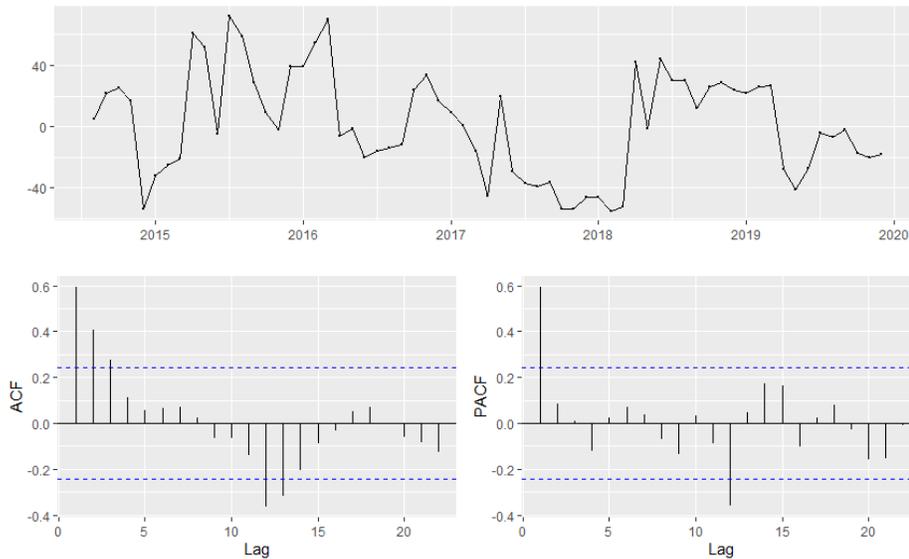
Figura 60. **Diferencias estacionales ACF y PACF de *monero***



Fuente: elaboración propia.

En el caso de *monero* se aprecian atrasos constantes en el gráfico de autocorrelación total, sin embargo, el efecto se ve normalizado en el gráfico de autocorrelación parcial, por lo que es posible atribuirle este efecto a la amplitud aumentada en el intervalo del valor bursátil asociado en ciertos meses, el cual se puede apreciar de mejor forma en la línea de tendencia de *monero* en la figura 18.

Figura 61. **Diferencias estacionales ACF y PACF de *ripple***



Fuente: elaboración propia.

En el caso de *ripple* se aprecia un fenómeno parecido, en dónde el gráfico de autocorrelación parcial normaliza el efecto de los atrasos.

En general en todas las series de datos se puede comprobar que existe estacionalidad y que esta fue tratada de forma adecuada por medio de las diferenciaciones, además, se pudo comprobar de qué forma se comparten los rezagos, es decir, que los mismos salen de la media cada cierta cantidad de periodos tanto para ACF y PACF. Sin embargo, ahora se cuentan para todas las criptodivisas con rezagos en los atrasos estabilizados respecto a la media.

### 3.3.3. **Diseño de un modelo generalizado**

Para generalizar el comportamiento de las divisas en un único modelo integrando todas las características de interés, y por ende, no formular un modelo

en específico para cada divisa, sino uno aplicable a todas las consideradas en este estudio, se necesita hacer un análisis de minería de datos con el fin de crear una base de datos generalizada, que a pesar de que por la naturaleza del estudio no es posible no perder información, los datos siempre mantendrán las características intrínsecas de interés asociadas a las criptodivisas, esto para comenzar a diseñar un modelo de aproximación predictivo a su valor real.

Para tal fin se aisló el valor neto de volumen de capital total de las criptodivisas estudiadas de forma mensual, para de esa forma tener una comparativa intermensual. Esto con el fin de calcular un promedio móvil ponderado del valor total por mes, y con ello generar una nueva serie de tiempo.

Sin embargo, generar una serie de tiempo a partir de datos totales y de forma aritmética crearía un sesgo en el estudio, dado que el volumen de capital de *bitcoin* (es decir la divisa de referencia) puede llegar a ser hasta un 300 % más grande que el del resto de criptodivisas como se apreció en el análisis descriptivo de las divisas.

El planteamiento que se realizó para evitar este inconveniente es el de realizar un conglomerado o categorización por agrupación de los datos, por medio de una relación de pesos en cuanto al comportamiento del capital deseado dadas las leyes de orden que rigen la distribución de estos datos.

Para la referencia de esta metodología se utilizaron las conclusiones del análisis de varianza realizado en la sección anterior en relación al factor de volumen de capital, en el cual sabemos que para efectos de la media de volumen de capital las monedas *litecoin*, *dash* y *monero* comparten un comportamiento simétrico, esto se puede validar en el estudio de estacionalidad y de series de tiempo, en el cual se puede observar los saltos estacionales que dan las

criptodivisas, también es apreciable en el resto de las mismas en los momentos cuando aún su volumen de capital no había alcanzado valores de inflación tan grandes.

Por lo que posterior a un análisis de representatividad en el volumen financiero de capital, como menciona Bakar, Rosbi y Uzaki (2017), que siempre se tendrá un efecto más favorable en relación al modelado cuando se asigna un peso probabilístico mayor a las divisas con mayor potencial de crecimiento, es decir a la de menor valor bursátil, y viceversa en relación a la de mayor valor.

*Relación: 3: 3: 3: 2: 2: 1*

La cual se puede representar de mejor forma de la siguiente manera:

Tabla XXXI. **Pesos para generar la base de datos generalizada**

<b>Criptodivisa</b>	<b>Media</b>	<b>Peso probabilístico en el estudio (3:2:1)</b>
<i>Bitcoin</i>	$4.9662 * 10^{10}$	5 %
<i>Ethereum</i>	$2.0388 * 10^{10}$	10 %
<i>Ripple</i>	7439229148	10 %
<i>Litecoin</i>	2032513917	25 %
<i>Dash</i>	948496077	25 %
<i>Monero</i>	893251426	25 %

Fuente: elaboración propia.

### **3.3.3.1. Definición de los coeficientes del Modelo ARIMA**

Dada la aleatoriedad en la distribución de los valores monetarios de las criptodivisas, además de la independencia que muestran, se diseñó un modelo de solución para encontrar el modelo de integración más eficiente de forma teórica que pudiese realizar las mejores aproximaciones al valor real de las divisas, haciendo uso de la base de datos ponderada, para ello fue necesario hacer uso de las funciones de autocorrelación y autocorrelación parcial.

Para determinar el orden del modelo se realizaron distintos análisis iterativos por diferencias estacionales, basado en la teoría de intervalos de autocorrelación y autocorrelación parcial (ACF, PACF, respectivamente por sus siglas en inglés), se pudo observar un comportamiento normalizado en sus diferenciaciones por lo general en el orden dos de la serie (figuras 56 a 61), esto indica que la parte de media móviles está representada un orden dos.

Adicional se construyó un diagrama de decisión o un árbol de decisión Bayesiano para los retrasos de la serie planteada, en donde se establezca por medio de su significancia estadística cuál o cuáles son los valores más significativos para presentar el orden de autocorrelación del modelo planteado.

De esta forma se expresó una versión más generalizada de forma gráfica del modelo ARIMA en dónde el factor de mejor ajuste está descrito por el criterio de información de Akaike.



$$(1 - \phi_1 - L^1 - \phi_2 L^2)(1 - L)^1 X_t = c + (1 + \theta_1 L^1 + \theta_2 L^2) \epsilon_t \quad (Ec. 51)$$

$$(1 - \phi_1 - L^1 - \phi_2 L^2)(x_t - x_{t-1}) = c + (1 + \theta_1 L^1 + \theta_2 L^2) \epsilon_t \quad (Ec. 52)$$

$$(1 - \phi_1 - L^1 - \phi_2 L^2) \Delta x_t = c + (1 + \theta_1 L^1 + \theta_2 L^2) \epsilon_t \quad (Ec. 53)$$

$$\Delta x_t - \phi_1 L^1 \Delta x_t - \phi_2 L^2 \Delta x_t = c + \epsilon_t + \theta_1 L^1 \epsilon_t + \theta_2 L^2 \epsilon_t \quad (Ec. 54)$$

$$\Delta x_t - \phi_1 \Delta x_{t-1} - \phi_2 \Delta x_{t-2} = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \quad (Ec. 55)$$

En donde  $X_t$  hace referencia al estimador del modelo para cada momento en el que se considera a la variable de salida,  $\phi_2$  al momento en cuestión para el cual se está estimando a la variable es decir la parte autorregresiva, además  $\epsilon$  al error o ruido blanco derivado de cada atraso que el modelo genere y  $L$  a los atrasos contemplados para el modelo (Cryer y Chan, 2008).

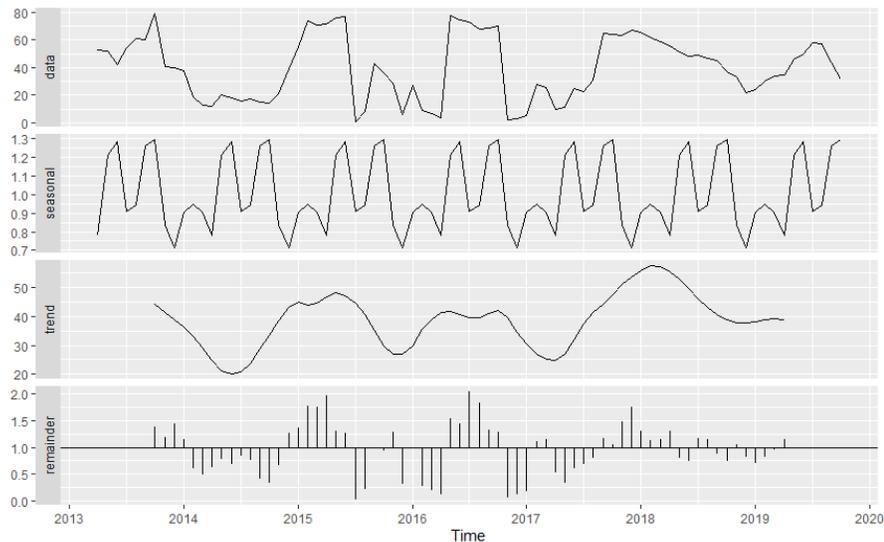
El modelo *ARIMA* (2,1,0)(2,1,0)[12] bajo el modelo planteado para las criptodivisas se puede expresar matemáticamente de la siguiente forma, la cual se diferencia de la estructura estándar del modelo *ARIMA* estándar, en que se buscaran al menos 2 atrasos en cada uno de las variaciones de los estimadores es decir los deltas ( $\Delta x_t$ ).

$$\Delta x_t = c + \phi_1 \Delta x_{t-1} + \phi_2 \Delta x_{t-2} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t \quad (Ec. 56)$$

### 3.3.3.2. Aplicación del modelo sobre serie Generalizada

A continuación, se presentan las características de la serie generalizada por ponderación a la cual se realizaron predicciones con el modelo *ARIMA* planteado.

Figura 63. **Serie generalizada de criptodivisas**



Fuente: elaboración propia.

La serie presentada en la figura 63 nace a partir de la estructura que se hizo en la tabla XXXI, en dónde se deseaba asignar características de interés ponderando cada divisa en periodos homogéneos, con ello se observa que la curva presenta un comportamiento estacional similar a *bitcoin*, además de que la curva presenta una tendencia bastante similar gráficamente, lo cual demuestra de que a pesar de que tenía únicamente un 5 % de participación en la generalización de la serie de tiempo, confirma otro de los supuestos mencionados por los autores Castro y Forero (2018) que los factores de comportamiento son bastante similares.

el cual es que las decisiones de inversión sobre divisas que siguen un comportamiento semejante a *bitcoin* o divisas de participación grande en el mercado podrían suponer rendimientos interesantes, lo cual muestra que las divisas involucradas en el estudio, a pesar de tener un valor significativamente

menor al de *bitcoin* comparten semejanzas en su comportamiento, si se desea comparar esta semejanza de forma más clara se puede comparar la figura 63 con la figura 38 si se desea una comparativa gráfica.

### **3.3.3.3. Resumen del modelo planteado**

A continuación, se presentan los criterios de salida del modelo ARIMA planteado en la sección anterior de forma tabular en donde se encuentran los coeficientes del modelo y los estimadores de su eficiencia, además un análisis de residuos, en donde se desea observar qué tal es el ajuste del modelo bajo el supuesto de normalidad de los residuos del modelo, en donde se desea validar el supuesto de que los residuos del mismo se distribuyen de forma normal, es decir tiene la capacidad predictiva de simular el fenómeno.

Esto a su vez también funciona de referencia para realizar un análisis comparativo respecto a un modelo entrenado por aprendizaje de máquina supervisado.

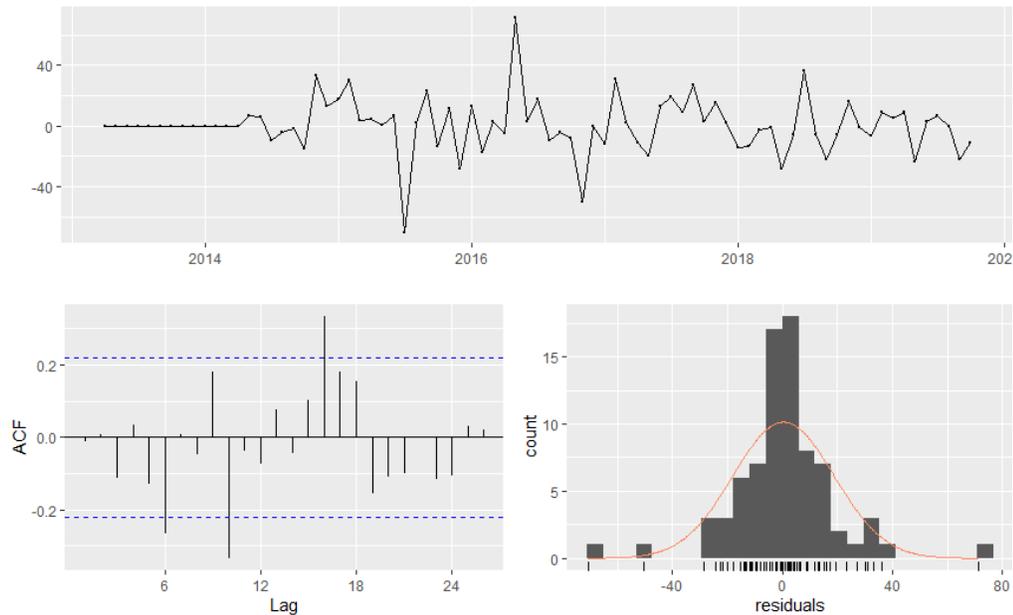
Tabla XXXII. **Resumen del modelo ARIMA planteado**

<b>MODELO</b>				
<b>ARIMA (2,1,0)(2,1,0)[12]</b>				
Coeficientes				
Referencia	AR1	AR2	SAR1	SAR2
Valor	-0.017	-0.1167	-0.7855	-0.4964
Cuadrado medio del error	0.1269	0.1476	0.1295	0.1226
<b>Criterios de información</b>				
$\sigma^2$				424.3
AIC				603.19
AICC				604.19
BIC				614.14

Fuente: elaboración propia.

El modelo ARIMA planteado de forma teórica presenta los siguientes resultados al aplicarlos, presenta un coeficiente de Akaike eficiente en relación al modelo generado de forma automática (presentado más adelante) además como era esperado, el modelo presenta dos componentes autorregresivos, que eran los estimados dados los primeros cuatro atrasos del gráfico 52, el modelo también presenta un componente integrado, es decir el orden de diferenciación que presenta, lo cual también es eficiente para la serie de datos a trabajar, las cual parte de las series diferenciadas para cada divisa, por lo que es coherente la necesidad de esta diferenciación, no posee una componente de medias móviles, y adicional presenta un componente autorregresivo estacional, que de igual forma se aprecia en la figura 2, en donde existen componentes estacionales hasta en el último atraso es decir el 12, dado a que la series son anuales.

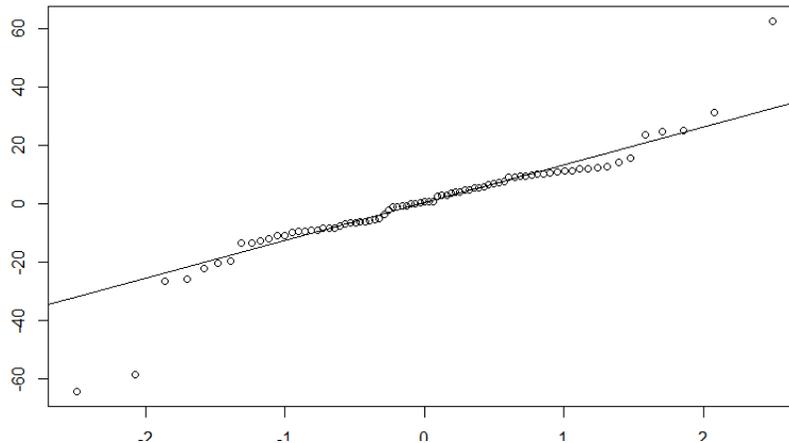
Figura 64. Residuos del modelo ARIMA teórico



Fuente: elaboración propia.

Utilizando el modelo presentado en la tabla XXXI sobre la serie generalizada de datos, se obtienen los siguientes residuos, en forma de resumen, se presentan en la figura 64, en donde se aprecia en el primer renglón de la gráfica la distribución de residuos en el tiempo y en el segundo los atrasos de los residuos y un consolidado frecuentista en formato de histograma. Se aprecia una continuidad no estacional en los residuos que tiende a ser virtualmente nula al final de la serie, lo cual es un buen indicio de normalidad, además de que la distribución acorde a la curva de densidad es gráficamente satisfactoria. Para completar el análisis se realiza un gráfico en papel normal y una prueba de Kolmogorov Smirnov para probar normalidad sobre estos residuos.

Figura 65. **Prueba de normalidad para el modelo ARIMA planteado**



Fuente: elaboración propia.

En la figura 65 se puede observar un gráfico de probabilidad normal, en donde entre menos sea la concentración de las observaciones muestrales sobre la línea continua, existe menor probabilidad de que se distribuyan normalmente, este análisis gráfico presenta resultados satisfactorios.

En donde los valores extremos pueden ser atribuidos a la presencia de valores tan altos como los de *bitcoin* en el estudio, este supuesto se validará a continuación con una prueba de Kolmogorov Smirnov la cual se encuentra en la tabla XXXIII.

Considerando que se trata de la serie generalizada de las divisas, el patrón histórico sigue un patrón homogéneo al de la figura 64, por ende, visualmente no existe una continuidad no esperada de la curva, existe un comportamiento singular en el gráfico de autocorrelación en el que en los atrasos 6, 10 y 16 existen salidas del umbral, esto podría ser atribuible a los valores extremos que puede

presentar la serie, dado a que tiene componentes de *bitcoin* que supera con creces el valor de las divisas y podría generar esa distribución de puntos extremos.

Para poder validar de forma definitiva la eficiencia del modelo se sometió la serie de residuos a una prueba de normalidad de Kolmogorov Smirnov la cual se presenta a continuación:

Tabla XXXIII. **Prueba de Kolmogorov para modelo ARIMA teórico**

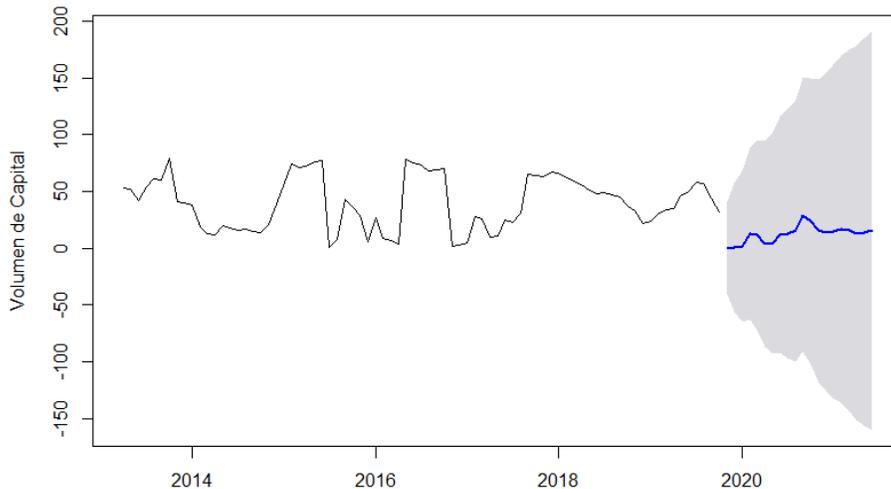
<b>Modelo</b>	ARIMA teórico
<b>D</b>	0.1733
<b>Valor-P</b>	0.5528

Fuente: elaboración propia.

Como se aprecia en la tabla XXXIII y en el gráfico de papel normal en la figura 65, efectivamente existe normalidad en la distribución de los residuos, lo cual garantiza que puede generar aproximaciones eficientes al fenómeno estudiado, los valores extremos que se pueden observar en la figura seguramente están sujetos a que *bitcoin* dispone de periodos de volumen de capital de valor elevado.

Bajo esta premisa, se desea validar el modelo teórico planteado utilizando una metodología de modelado distinta, por lo que se planteó montar los datos de la misma serie y generar un modelo ARIMA con distintas premisas, con ello se realizó una comparativa entre la eficiencia de ambos modelos y se seleccionó bajo los criterios de información disponibles al de mejor bondad de ajuste.

Figura 66. **Pronóstico utilizando modelo ARIMA teórico**



Fuente: elaboración propia.

Se aprecia de forma gráfica que el modelo tiene la capacidad de encontrar los atrasos en el modelo y generar aproximaciones en el pronóstico. Por otra parte, Cryer y Chan (2008) mencionan qué modelos formulados inadecuadamente, generan aproximaciones lineales sobre la media, lo cual es un indicio de un modelo no eficiente para la serie de datos en cuestión.

#### **3.3.4. Planteamiento de un modelo ARIMA por medio de aprendizaje de máquina**

Se planteó realizar una comparación o *benchmark* (término en inglés) del modelo teórico, para ello se definió utilizar un modelo derivado de un algoritmo de aprendizaje de máquina que tome como criterio de optimización bajo la metodología estadística de iteraciones hacia adelante, el valor del AIC (criterio estadístico de Akaike), por lo que se planteó un algoritmo ML (*Machine Learning*

por sus siglas en inglés) autorecursivo que tiene como finalidad la minimización del error en cuanto a criterios del valor AIC.

Dada la naturaleza de este planteamiento no se incluyó el desarrollo de este algoritmo ni tampoco las pruebas realizadas, dado que la finalidad del mismo es solo servir como un comparativo para evaluar el desempeño y la eficiencia para el modelo planteado de forma teórica, sin embargo, sí se incluyeron generalidades del mismo. El modelo fue planteado como un *Script* simple en el lenguaje de programación R.

Para encontrar un modelo con el cual realizar comparaciones en relación al modelo teórico, se planteó utilizar un algoritmo supervisado, para homogeneizar métricas se utilizó únicamente la variable de volumen de capital que es la utilizada en el diseño del modelo teórico, sin embargo, es probable que exista un mejor modelo bajo las interacciones múltiples entre todas las variables bursátiles disponibles, sin embargo para hacer una comparación eficiente se platearon ambos modelos en igualdad de condiciones.

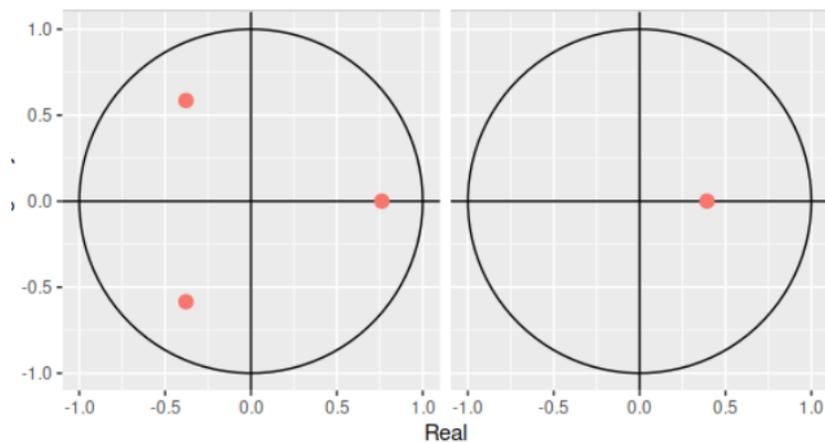
El algoritmo de aprendizaje utilizado fue el de árboles de clasificación, este fue planteado con la delimitación de variables de entrenamiento las cuales son las mismas contempladas para el modelo original, para la creación de los conjuntos de datos se estableció la generalización del 70/30 es decir, se utilizaron el 70 % de los datos para generar un conjunto de entrenamiento y validación y un 30 % para probar la eficiencia del modelo.

Para la clasificación de datos, es decir, clasificar los datos en relación al criterio de Akaike se utilizó un árbol de clasificación probabilístico, el cual puede apreciarse en la figura 68, en el mismo se puede observar la amplia cantidad de nodos que genera el árbol, reflejando los saltos entre atrasos que presentan las

series de datos originales, la mejor selección de hiperparámetros para el árbol seleccionado presenta que sí existe viabilidad en llegar a conclusiones respecto a los nodos iniciales como se aprecia en la figura 69.

La variable de salida del modelo es el valor de la raíz polinomial que optimiza el error según el coeficiente de integración del modelo ARIMA, en dónde como criterio de decisión se pretende que en la circunferencia de raíces las mismas se encuentren lo más alejadas posibles de la circunferencia y con ello se pueda generar un modelo estable que tenga capacidad de inferencia eficiente.

Figura 67. **Raíces polinomiales del modelo**



Fuente: elaboración propia.

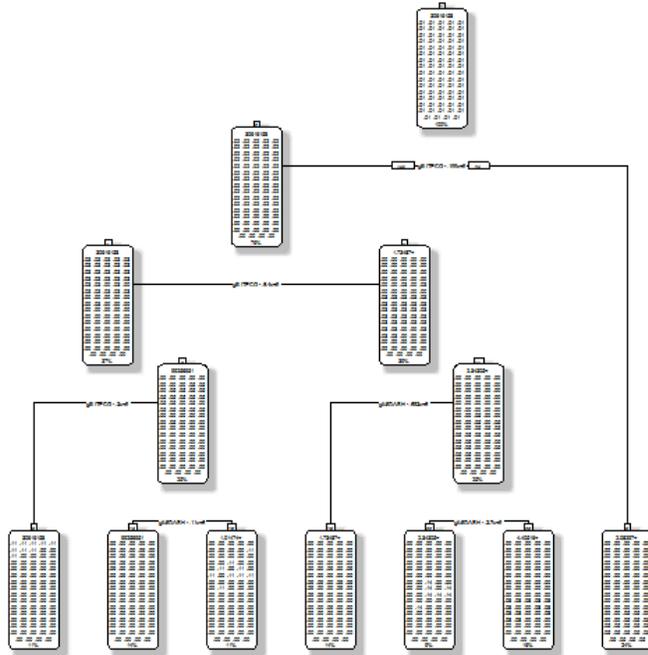
Para las iteraciones el algoritmo se utilizó una función activadora del tipo no lineal por cada clasificación propuesta con el fin de encontrar la raíz que minimizara el error en función del coeficiente seleccionado, las iteraciones se realizaban variando los hiperparámetros del modelo de un total pre seleccionado.

Sin embargo, al solo considerar un modelo no lineal donde el discriminante estaba asociado a una única variable, es probable que no se esté explotando todo el potencial que ofrecen estos modelos, dada la capacidad de integración de múltiples variables y conjuntos de parámetros más robustos, sin embargo, la idea es generar una aproximación automática lo más cercana posible al planteamiento teórico realizado para una única variable.

Dado a que se desea consolidar el modelo con las condiciones más homogéneas al planteado originalmente se utiliza únicamente la variable de volumen de capital para el análisis.

Además, muchos algoritmos automatizados de búsqueda de modelos ARIMA en los *softwares* actuales utilizan una metodología similar, en dónde la búsqueda es la minimización del error de estimación, sin importar si el modelo planteado presenta estimaciones extremadamente conservadoras o inestabilidad en sus raíces en algunos casos (Cryer y Chan, 2008).

Figura 68. Visualización del algoritmo de clasificación



Fuente: elaboración propia.

El modelo presentado por el algoritmo tiene la forma ARIMA (1,0,0) es decir con una parte autorregresiva, pero sin la parte integrada ni de medias móviles, tampoco una parte estacional. A grandes rasgos puede atribuírsele a su objetivo de minimizar el error del parámetro ingresado, por consiguiente, puede que el modelo suponga un buen ajuste probabilístico.

Sin embargo, es extraño que no se incluya una diferenciación de las series, dado el componente estacional observado en la sección anterior, esto tiene un efecto directo en la minimización del error dado un mejor ajuste sobre el conjunto de entrenamiento, pero limitando a la creación de predicciones futuras.

A continuación, se presenta el resumen del modelo planteado por computadora:

Tabla XXXIV. **Resultados del modelo planteado por algoritmo**

<b>MODELO</b>			
<b>ARIMA (1,0,0)</b>			
Coeficientes			
Referencia	AR1	Media	
Valor	0.6849	40.1304	
Cuadrado medio del error	0.0801	5.7266	
<b>Criterios de información</b>			
$\sigma^2$			278.5
AIC			673.51
AICC			673.83
BIC			680.62

Fuente: elaboración propia.

Tabla XXXV. **Métricas del aprendizaje de máquina**

<b>Métricas de entrenamiento</b>	<b>ME</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>MASE</b>	<b>ACF1</b>
<b>Conjunto de entrenamiento</b>	0.156	16.471	11.174	147.038	166.166	0.350	0.095

Fuente: elaboración propia

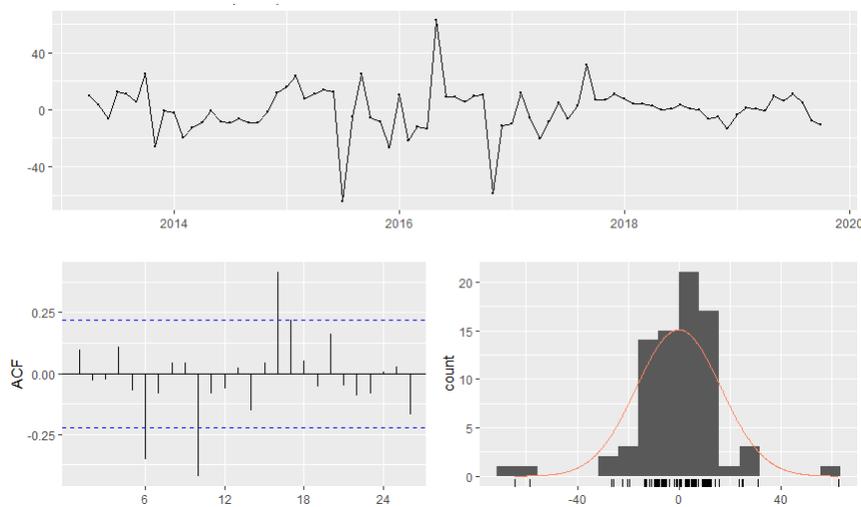
Se observa en la tabla XXXV que efectivamente existe una componente autorregresiva del modelo planteado de forma teórica, sin embargo, no existen componentes de medias móviles ni de integración.

Se observa un valor para el criterio de información de Akaike bastante tolerable, aunque menos eficiente incluso que el modelo teórico del modelo

presentado en la tabla XXXII, por otra parte las métricas de entrenamiento del modelo son poco eficientes, teniendo un error medio absoluto de estimación de más del 100 % (MAPE) como se aprecia en la tabla XXXV, dado a que estas métricas son sobre el conjunto de validación, sugiere que existe un sobre ajuste a los datos de entrenamiento, y que este tipo de optimizaciones están sujetas a este tipo de error.

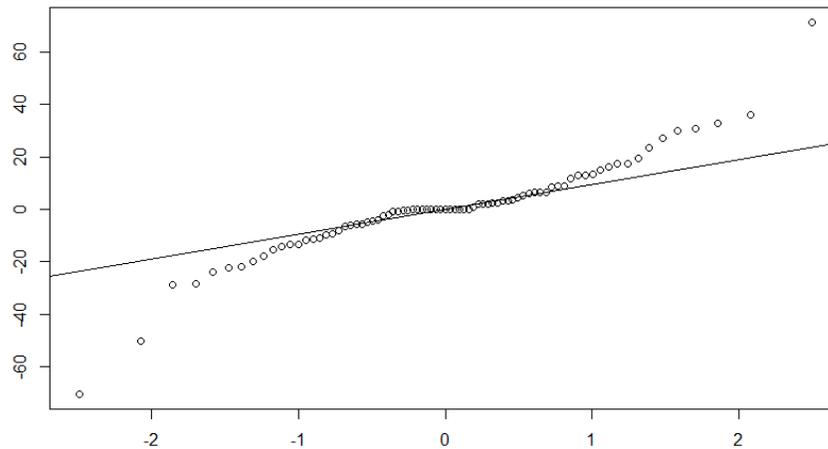
Gráficamente el modelo generado tiene la siguiente forma.

Figura 69. **Residuos del modelo ARIMA generado por ML**



Fuente: elaboración propia.

Figura 70. Prueba de normalidad para modelo ARIMA de ML



Fuente: elaboración propia.

Se aprecia que la distribución de residuos en la figura 70 tiene un comportamiento bastante homogéneo al presentado por el modelo teórico en la figura 64, sin embargo, en cuanto a la distribución de los residuos se aprecia visualmente en la figura inferior derecha que existe una concentración anormal sobre la media, es decir que las estimaciones podrían no comportarse de forma normal, para validar ese supuesto se aplicó la prueba de normalidad de Kolmogorov-Smirnov, la cual presenta un p-valor significativo de  $1.292 \times e^{-09}$ , los resultados completos se presentan en la tabla a continuación.

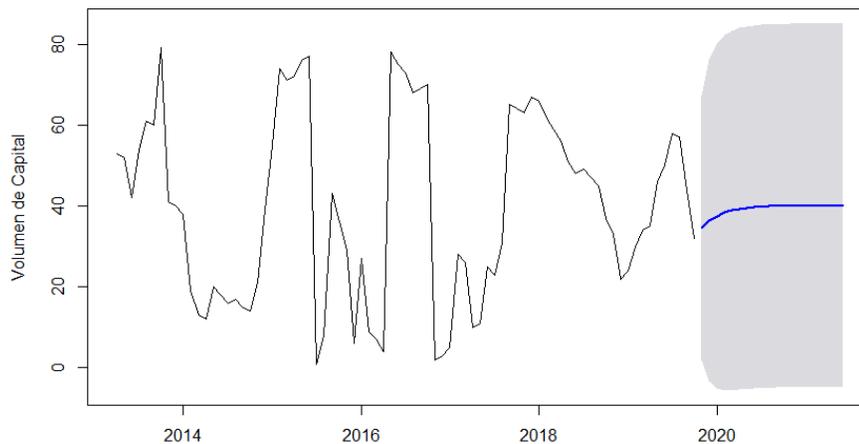
Desde este análisis el modelo no es funcional, dado que, al no presentar normalidad en la distribución de los residuos, viola uno de los supuestos para considerar al modelo adecuado para generar predicciones, esto además se puede observar de forma gráfica en la figura 71 en donde se evidencia que el modelo no es capaz de encontrar las variaciones de la serie de forma adecuada y generar predicciones eficientes fuera de una curva lineal en el eje central.

Tabla XXXVI. **Prueba de Kolmogorov para ARIMA generado por ML**

Modelo	ARIMA Aprendizaje automático
D	0.46
Valor-P	1.292e-09

Fuente: elaboración propia.

Figura 71. **Pronóstico con modelo ARIMA (1,0,0)**



Fuente: elaboración propia.

### 3.3.4.1. Comparativa entre ambos modelos

Desde un inicio se puede descartar al modelo generado por aprendizaje supervisado dado a que no se pueden validar los supuestos relacionados a la distribución de residuos al modelo como se aprecia en la tabla XXXVI, lo cual lo vuelve un modelo poco eficiente para generar predicciones, sin embargo, dada

la premisa se compararan ambos métodos con la finalidad de validar si efectivamente existe cambio en alguna métrica de estimación.

Tabla XXXVII. **Comparativa de modelos ARIMA**

<b>MODELO (2,1,0)(2,1,0)[12]</b>		<b>MODELO (1,0,0)</b>		<b>Eficiencia Comparativa</b>
<b><math>\sigma^2</math></b>	424.3	<b><math>\sigma^2</math></b>	278.50	<b>34.36 %</b>
<b>AIC</b>	603.19	<b>AIC</b>	673.51	<b>-11.66 %</b>
<b>AICC</b>	604.19	<b>AICC</b>	673.83	<b>-11.53 %</b>
<b>BIC</b>	614.14	<b>BIC</b>	680.62	<b>-10.82 %</b>

Fuente: elaboración propia.

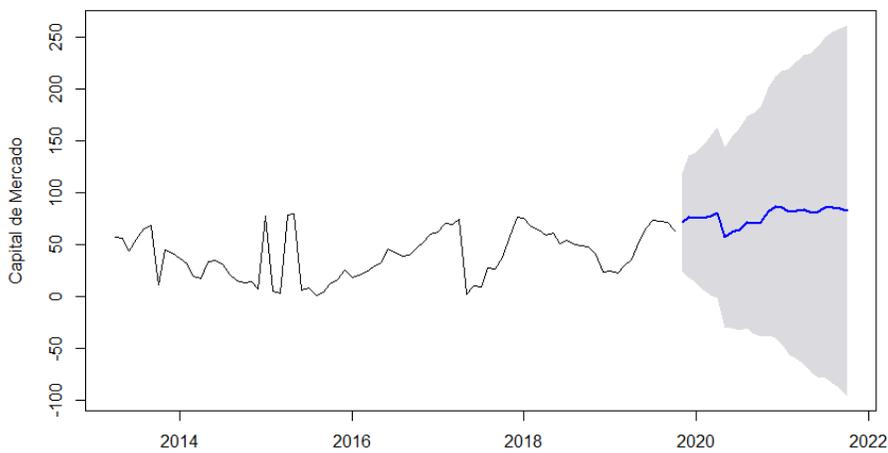
En la tabla XXXVII se aprecia las métricas de estimación de ambos modelos además de la eficiencia del primer modelo en relación al segundo, se aprecia que el objetivo del modelo de aprendizaje automático es decir el ARIMA (1,0,0), es realizar estimaciones disminuyendo la varianza, aun sin tomar la consideración de diferenciación para los atrasos significativos que presenta la generalidad de las series, por ende, el factor de autocorrelación por sí solo no es un estimador tan robusto para crear estimaciones óptimas, a pesar de que las generadas sobre los datos de entrenamiento son eficientes como se aprecia en la tabla XXXV sobre métricas del aprendizaje de máquina.

Por ende, el modelo generado por aprendizaje de máquina no es eficiente para generar aproximaciones a esta serie de datos y demuestra algunas limitantes en los estimadores automáticos por software como mencionan los autores Cryer y Chan (2008) en el apartado de predicciones por software.

### 3.3.5. Análisis predictivo de series temporales

Se realiza un pronóstico con modelo ARIMA para cada una de las criptomonedas analizadas.

Figura 72. Pronóstico con modelo ARIMA para *bitcoin*



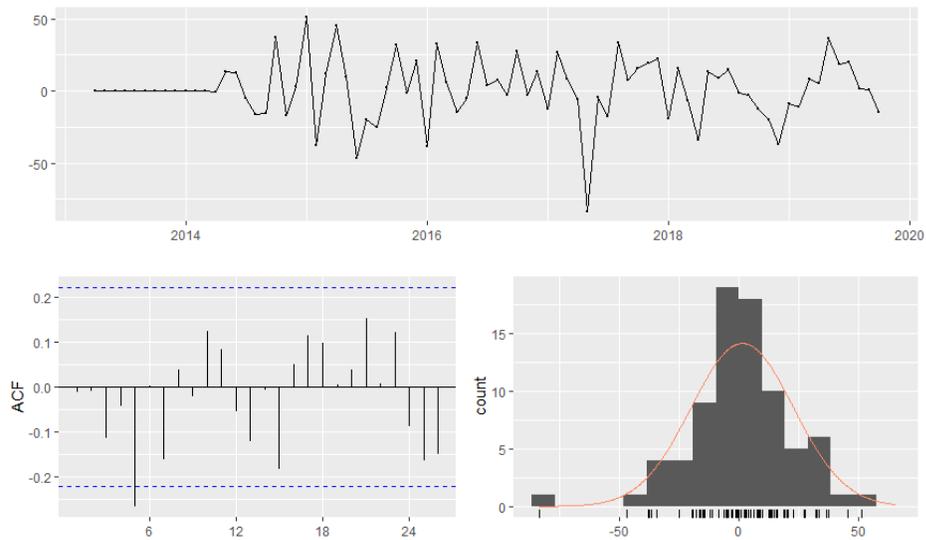
Fuente: elaboración propia.

Tabla XXXVIII. Prueba de normalidad de residuos bitcoin

Modelo	Residuos <i>bitcoin</i>
D	0.0600
Valor-P	0.2236

Fuente: elaboración propia.

Figura 73. **Análisis de residuos para predicciones de *bitcoin***



Fuente: elaboración propia.

En la figura 72 puede apreciarse el pronóstico de *bitcoin* utilizando el modelo generalizado sobre la divisa individual, los resultados aparentan ser satisfactorios, dado a que como se aprecia en el la figura 73 para el análisis de los residuos, en la esquina inferior izquierda, se aprecia que solo un atraso sale de la serie, lo cual presenta una eficiencia considerable en relación a la diferenciación por si sola aplicada a la serie, la cual se puede apreciar en la figura 45, además, el histograma de residuos aparenta normalidad en el extremo inferior derecho, lo cual se corrobora en la tabla XXXVII con una prueba de normalidad no significativa, por lo que el modelo es válido para realizar predicciones, el objetivo principal de este planteamiento era realizar un pronóstico de al menos 24 meses al futuro y estimar posibles decisiones de inversión eficientes, estos se presentan a continuación para *bitcoin* y para todas las divisas.

Tabla XXXIX. **Pronóstico a 24 meses de *bitcoin* con cifras en miles de millones de dólares**

Periodo predicho	Valor más probable	IC. Límite inferior	IC. Límite superior
1	USD 71.63	USD 24.50	USD 118.76
2	USD 76.94	USD 18.53	USD 135.35
3	USD 75.85	USD 12.92	USD 138.77
4	USD 75.75	USD 5.94	USD 145.57
5	USD 77.96	USD 1.23	USD 154.69
6	USD 80.70	-USD 1.43	USD 162.84
7	USD 56.68	-USD 30.59	USD 143.94
8	USD 61.88	-USD 30.48	USD 154.23
9	USD 64.58	-USD 32.51	USD 161.67
10	USD 71.11	-USD 30.45	USD 172.68
11	USD 70.11	-USD 35.77	USD 175.99
12	USD 71.79	-USD 38.24	USD 181.83
13	USD 81.22	-USD 37.87	USD 200.30
14	USD 86.18	-USD 39.88	USD 212.25
15	USD 85.66	-USD 45.92	USD 217.23
16	USD 81.54	-USD 55.97	USD 219.05
17	USD 83.04	-USD 60.30	USD 226.38
18	USD 83.38	-USD 65.34	USD 232.10
19	USD 80.53	-USD 73.40	USD 234.45
20	USD 81.36	-USD 77.66	USD 240.39
21	USD 85.27	-USD 78.67	USD 249.22
22	USD 85.53	-USD 83.18	USD 254.23
23	USD 84.52	-USD 88.83	USD 257.87
24	USD 82.65	-USD 95.22	USD 260.52

Fuente: elaboración propia.

### 3.3.5.1. Análisis de valor en riesgo para *bitcoin*

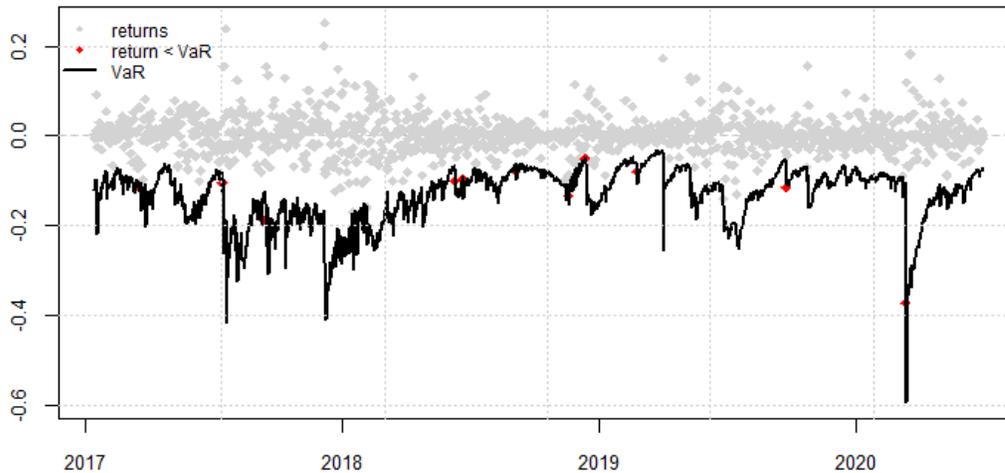
Para validar el rendimiento del modelo ARIMA estacional, es útil realizar una prueba de respaldo histórica para comparar el valor en riesgo estimado (VaR) para las pérdidas y ganancias durante el horizonte temporal estudiado. Si el retorno es menor que el VaR en la gran mayoría de casos se tiene un exceso de valor en riesgo. Dada la volatilidad de las series estudiadas se desea analizar el valor en riesgo siendo restrictivo en las pérdidas y ganancias de las mismas, por lo tanto, se realizará el análisis a un nivel de significancia del 1 % para todas las divisas en adelante. El periodo de la prueba retrospectiva estimado es anual, de 360 días después del comienzo de la serie (año 2015 para todas las divisas).

Las estimaciones de VaR producidas por el modelo ARIMA son evaluadas por la prueba de cobertura incondicional de Kupiec y la prueba de independencia de Christoffersen (Klenke, 2014).

El modelo de predicción se puede interpretar bajo estos criterios en función de:

- Si la cobertura media  $> \alpha$ : se tienen demasiadas superaciones, el cuantil pronosticado debería ser más negativo. Se ha subestimado el riesgo de perder dinero.
- Si la cobertura media  $< \alpha$  muy pocas superaciones, el cuanto pronosticado también fue negativo. Se ha sobreestimado el riesgo de perder dinero.

Figura 74. VaR de *bitcoin*



Fuente: elaboración propia.

Los retornos de los datos alcanzan el 1 % es decir el VaR (puntos rojos) 12 veces en comparación con las 12.6 veces esperadas, además, como se evidencia en la tabla XL las superaciones son correctas e independientes. Esto implica que se espera que una pérdida del 0.94 % se supere solo el 1 % del tiempo en función de la inversión inicial, dado este criterio las inversiones al corto plazo siguen siendo interesantes.

Tabla XL. Resumen del modelo VaR *bitcoin*

Modelo estudiado	ARIMA Estacional
Metadata	1264 registros
$\alpha$	1 %
VaR esperado	12.6
VaR excedido realmente	12

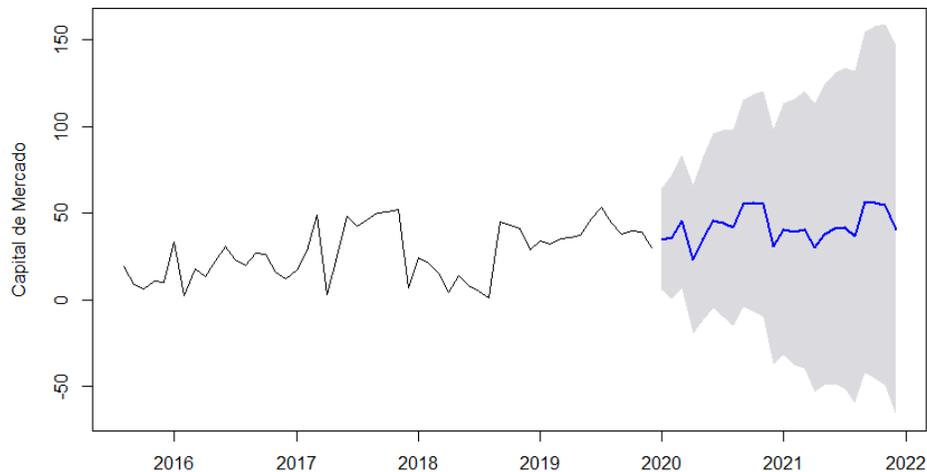
Continuación tabla XL.

% de VaR	0.9 %
Cobertura incondicional (Kupiec)	$p - value = 0.855$
Independencia de las coberturas (Christoffersen)	$p - value = 0.877$

Fuente: elaboración propia.

Como se estimaba desde un inicio las predicciones para el valor bursátil de las criptodivisas tenderían a tener intervalos de confianza para el valor medio amplio, en dónde en algunos meses podría generar rendimientos de incluso el 300 %, tal es el caso del pronóstico para el mes 20 de *bitcoin* en la tabla XXXVIII, y de igual forma en ese mismo mes tener una pérdida en bruto de casi toda la proporción invertida acorde al límite inferior, las sugerencias por periodo para la inversión por divisa se aborda de manera más amplia en el siguiente capítulo.

Figura 75. **Pronóstico con modelo ARIMA para *ethereum***



Fuente: elaboración propia.

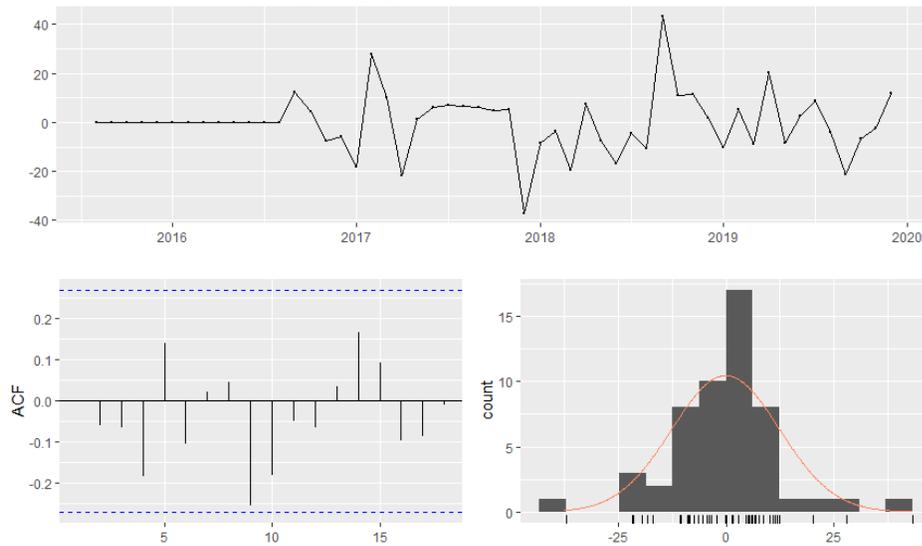
Tabla XLI. **Prueba de normalidad de residuos de *ethereum***

<b>Modelo</b>	<b>Residuos <i>ethereum</i></b>
<b>D</b>	0.110
<b>Valor-P</b>	0.523

Fuente: elaboración propia.

En la figura 75 se aprecia de forma gráfica el pronóstico para 24 meses orientados a *ethereum*, se aprecia de forma gráfica que el modelo tiene la capacidad de simular la varianza de la serie, además que como se aprecia en la figura 76 en el gráfico de autocorrelación en el extremo inferior izquierdo, ningún atraso se sobrepone a las bandas por lo que se considera el modelo como eficiente para controlar a los mismos, además el histograma de residuos en la esquina inferior derecha presenta un comportamiento no del todo tradicional, la prueba de normalidad en la tabla XLI muestra que efectivamente los residuos se distribuyen de forma normal, por ende, el modelo es eficiente para generar predicciones.

Figura 76. **Análisis de residuos para predicciones de *ethereum***



Fuente: elaboración propia.

Tabla XLII. **Pronóstico a 24 meses de *ethereum* con cifras en millones de dólares**

Periodo pronosticado	Valor más probable	IC. Límite inferior	IC. Límite superior
1	USD 34.69	USD 5.28	USD 64.11
2	USD 36.07	USD 0.32	USD 71.82
3	USD 45.27	USD 6.61	USD 83.93
4	USD 23.27	-USD 19.83	USD 66.37
5	USD 35.20	-USD 12.03	USD 82.43
6	USD 45.67	-USD 4.87	USD 96.20
7	USD 44.03	-USD 9.73	USD 97.79
8	USD 41.84	-USD 15.05	USD 98.72
9	USD 55.55	-USD 4.24	USD 115.34
10	USD 55.91	-USD 6.64	USD 118.47
11	USD 55.41	-USD 9.81	USD 120.64

Continuación tabla XLII.

12	USD	30.76	-USD	37.02	USD	98.54
13	USD	40.58	-USD	32.24	USD	113.41
14	USD	39.12	-USD	37.59	USD	115.82
15	USD	40.17	-USD	39.78	USD	120.12
16	USD	29.85	-USD	53.55	USD	113.26
17	USD	37.58	-USD	49.16	USD	124.32
18	USD	41.17	-USD	48.68	USD	131.02
19	USD	41.37	-USD	51.52	USD	134.25
20	USD	36.31	-USD	59.54	USD	132.15
21	USD	56.23	-USD	42.47	USD	154.93
22	USD	56.14	-USD	45.34	USD	157.61
23	USD	54.87	-USD	49.31	USD	159.05
24	USD	40.28	-USD	66.53	USD	147.09

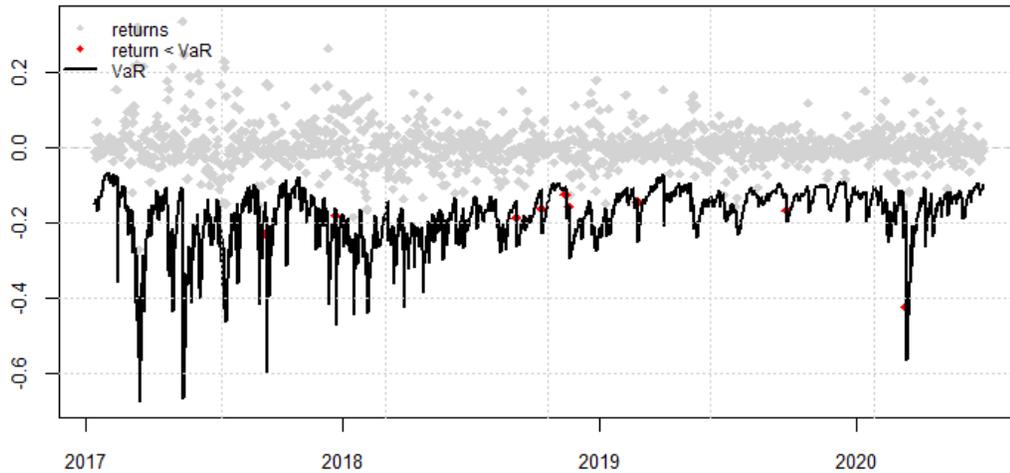
Fuente: elaboración propia.

Para *ethereum* las cifras se presentan con intervalos de confianza menos amplios como se aprecia en la tabla XLII, sin embargo, sí existe el factor de rendimiento atractivo implícito, de hasta un 100 % en rendimientos, como se puede apreciar en el pronóstico para el mes 9.

### 3.3.5.2. Análisis de valor en riesgo para *ethereum*

Para el análisis VaR de *ethereum* se tiene la siguiente visualización

Figura 77. VaR de *ethereum*



Fuente: elaboración propia.

Los retornos de los datos alcanzan el 1 % es decir el VaR (puntos rojos) 9 veces en comparación con las 12.6 veces esperadas, además, como se evidencia en la tabla XLII las superaciones son correctas e independientes. Esto implica que se espera que una pérdida del 0.71 % se supere solo el 1 % del tiempo en función de la inversión inicial, dado este criterio las inversiones al corto plazo siguen siendo interesantes, en mayor escala, aunque de *bitcoin*.

Tabla XLIII. Resumen modelo VaR *ethereum*

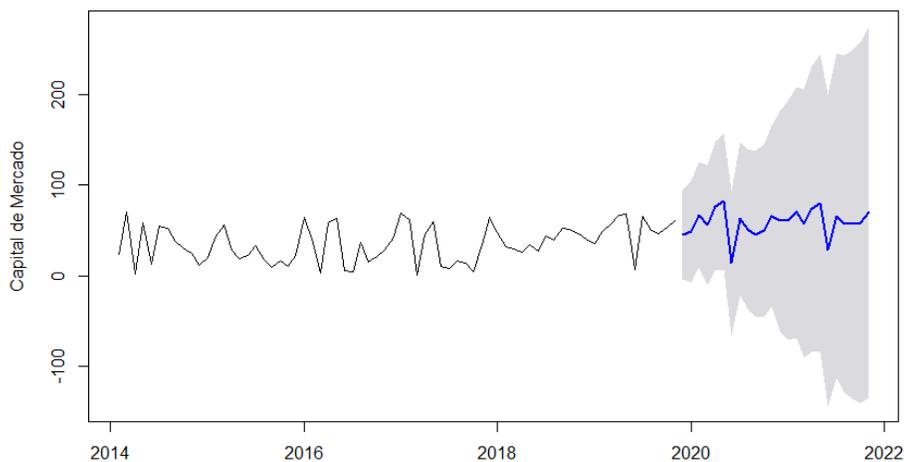
Modelo estudiado	ARIMA estacional
Metadata	1264 registros
$\alpha$	1 %
VaR esperado	12.6
VaR excedido realmente	9

Continuación tabla XLIII.

% de VaR	0.7 %
Cobertura incondicional (Kupiec)	$p - value = 0.278$
Independencia de las coberturas (Christoffersen)	$p - value = 0.52$

Fuente: elaboración propia.

Figura 78. **Pronóstico con modelo ARIMA para *dash***



Fuente: elaboración propia.

Tabla XLIV. **Prueba de normalidad de residuos de *dash***

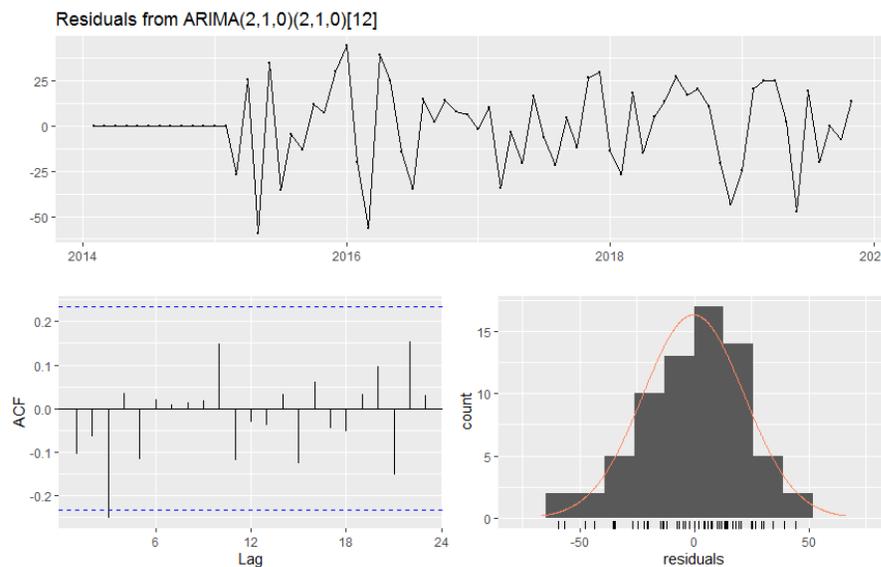
Modelo	Residuos <i>dash</i>
D	0.078
Valor-P	0.612

Fuente: elaboración propia.

En el caso de *dash* se aprecia como el modelo es capaz de predecir los atrasos de forma autorregresiva, como se puede apreciar gráficamente en la figura 78, dado a que circunscribe patrones bastante uniformes, además en la figura 79 se puede apreciar en la esquina inferior izquierda como los atrasos se comportan de forma uniforme en el gráfico de autocorrelación, el histograma de residuos por su parte en el extremo inferior izquierdo aparenta ser normal, el cual es validado por la prueba de normalidad en la tabla XLIV, la cual brinda un resultado no significativo.

Por lo que se puede concluir que el modelo es adecuado para simular de forma individual a la divisa y es eficiente para realizar predicciones.

Figura 79. **Análisis de residuos para predicciones de *dash***



Fuente: elaboración propia.

Tabla XLV. **Pronóstico a 24 meses de *dash* con cifras en millones de dólares**

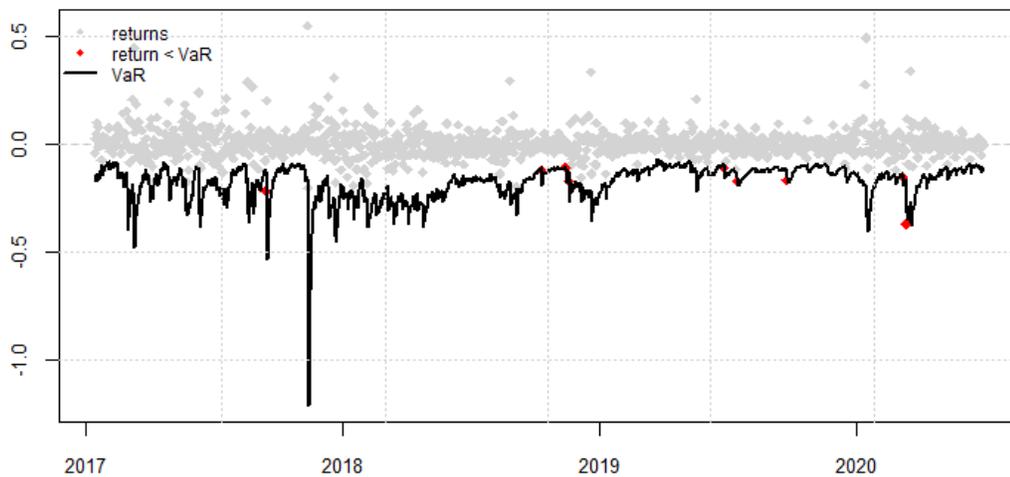
Periodo pronosticado	Valor más probable	IC. Límite inferior	IC. Límite superior
1	USD 45.04	-USD 4.37	USD 94.44
2	USD 49.02	-USD 7.19	USD 105.22
3	USD 67.22	USD 8.85	USD 125.59
4	USD 56.19	-USD 9.71	USD 122.09
5	USD 77.10	USD 5.26	USD 148.93
6	USD 82.00	USD 6.55	USD 157.45
7	USD 13.72	-USD 66.24	USD 93.68
8	USD 62.80	-USD 21.69	USD 147.29
9	USD 51.69	-USD 36.55	USD 139.93
10	USD 45.73	-USD 46.24	USD 137.70
11	USD 49.79	-USD 45.93	USD 145.51
12	USD 65.77	-USD 33.42	USD 164.97
13	USD 60.54	-USD 61.56	USD 182.63
14	USD 60.57	-USD 71.65	USD 192.78
15	USD 70.02	-USD 68.27	USD 208.32
16	USD 57.79	-USD 90.70	USD 206.27
17	USD 73.75	-USD 83.67	USD 231.18
18	USD 80.32	-USD 84.04	USD 244.68
19	USD 28.10	-USD 143.80	USD 199.99
20	USD 66.02	-USD 113.30	USD 245.35
21	USD 57.64	-USD 128.37	USD 243.65
22	USD 56.79	-USD 135.78	USD 249.36
23	USD 58.52	-USD 140.53	USD 257.58
24	USD 70.31	-USD 134.91	USD 275.54

Fuente: elaboración propia.

### 3.3.5.3. Análisis del valor en riesgo para *dash*

En el caso de *dash* los rendimientos tienen una prospección a ser bastante eficientes, como se muestra en la tabla XLV, para el mes 9 existe más de un 500 % de rendimiento futuro probable, sin embargo, para ese mismo mes, la banda de confianza inferior denota una pérdida bruta completa, por ende, el análisis gráfico del valor en riesgo es el siguiente.

Figura 80. VaR *dash*



Fuente: elaboración propia.

Los retornos de los datos alcanzan el 1 % es decir el VaR (puntos rojos) 9 veces en comparación con las 12.6 veces esperadas, además, como se evidencia en la tabla XLVI las superaciones son correctas e independientes. Esto implica que se espera que una pérdida del 0.71 % se supere solo el 1 % del tiempo en función de la inversión inicial, dado este criterio las inversiones al corto

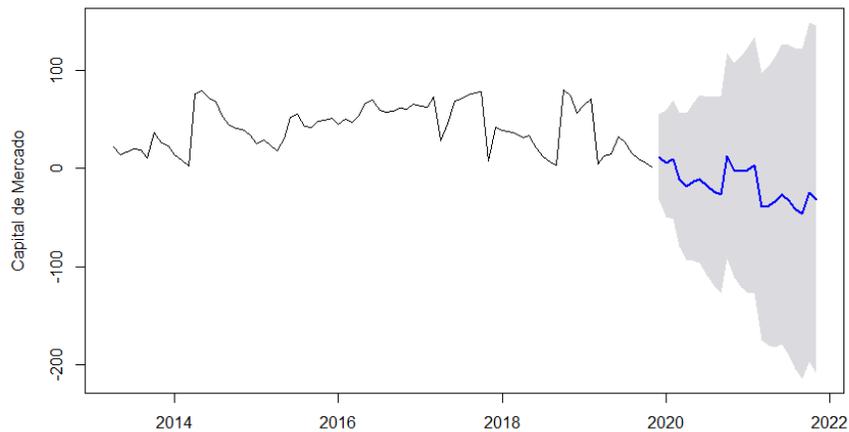
plazo siguen siendo interesantes, *dash* mantiene un comportamiento bastante homogéneo al de *ethereum* con caídas más dispersas.

Tabla XLVI. **Resumen del modelo VaR *dash***

Modelo estudiado	ARIMA estacional
Metadata	1264 registros
$\alpha$	1 %
VaR esperado	12.6
VaR excedido realmente	9
% de VaR	0.7 %
Cobertura incondicional (Kupiec)	$p - value = 0.325$
Independencia de las coberturas (Christoffersen)	$p - value = 0.614$

Fuente: elaboración propia.

Figura 81. **Pronóstico con modelo ARIMA para *litecoin***



Fuente: elaboración propia.

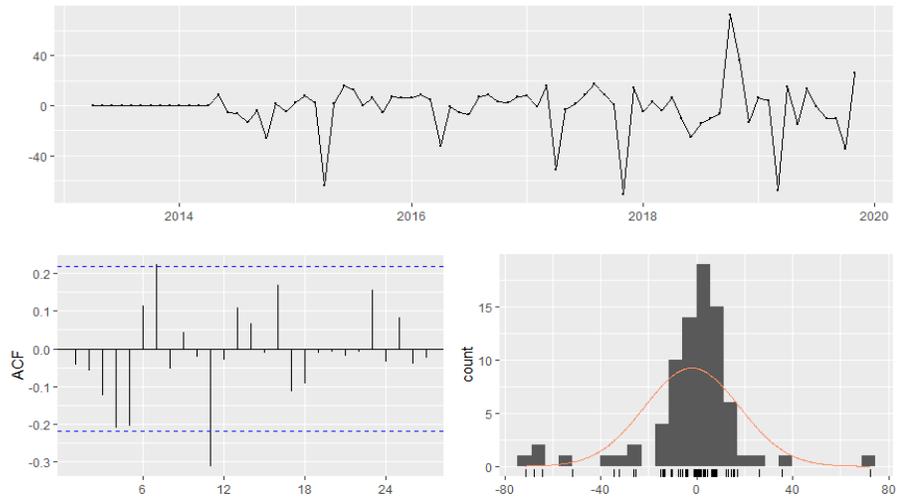
Tabla XLVII. **Prueba de normalidad para *litecoin***

<b>Modelo</b>	Residuos <i>litecoin</i>
<b>D</b>	0.129
<b>Valor-P</b>	0.2698

Fuente: elaboración propia.

En la figura 81 se puede apreciar gráficamente el pronóstico para la divisa de *litecoin*, en donde se aprecia que el modelo es bastante eficiente en poder captar los atrasos de la serie y reflejarlas en el futuro dado comportamientos pasados adaptados al volumen de capital, sin embargo la serie denota un comportamiento singular, una tendencia visual hacia la baja, lo cual denota que podría tener intervalos más amplios sobre valores negativos en sus predicciones, sin embargo, en cuanto a los residuos del modelo, en el extremo inferior de la figura 82, se aprecia un comportamiento adecuado en el gráfico de autocorrelación total, además de un histograma de residuos satisfactorio en el extremo inferior derecho, además la prueba de normalidad en la tabla XLVII confirma este supuesto, validando la eficiencia del modelo.

Figura 82. **Análisis de residuos para predicciones de *litecoin***



Fuente: elaboración propia.

Tabla XLVIII. **Pronóstico a 24 meses de *litecoin* con cifras en millones de dólares**

Periodo pronosticado	Valor más probable	IC. Límite inferior	IC. Límite superior
1	USD 11.24	-USD 32.37	USD 54.86
2	USD 5.17	-USD 49.25	USD 59.58
3	USD 8.99	-USD 51.60	USD 69.59
4	-USD 11.82	-USD 79.59	USD 55.95
5	-USD 18.43	-USD 92.79	USD 55.92
6	-USD 13.77	-USD 93.87	USD 66.34
7	-USD 11.17	-USD 96.68	USD 74.35
8	-USD 17.81	-USD 108.46	USD 72.84
9	-USD 23.35	-USD 118.84	USD 72.13
10	-USD 26.93	-USD 127.02	USD 73.15
11	USD 12.37	-USD 92.12	USD 116.86

Continuación tabla XLVIII.

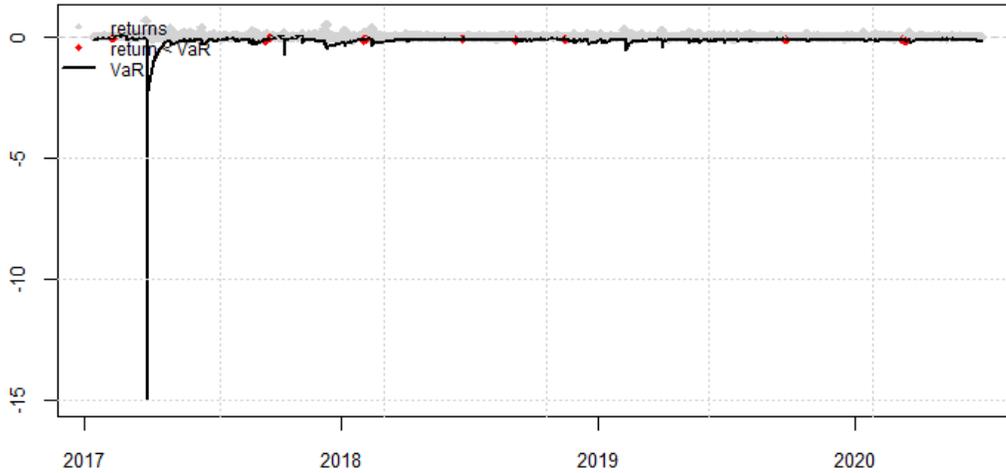
12	-USD	2.40	-USD	111.12	USD	106.31
13	-USD	3.41	-USD	120.84	USD	114.01
14	-USD	1.79	-USD	126.07	USD	122.49
15	USD	3.00	-USD	127.19	USD	133.20
16	-USD	39.36	-USD	175.56	USD	96.84
17	-USD	37.91	-USD	179.89	USD	104.07
18	-USD	34.51	-USD	181.96	USD	112.95
19	-USD	26.46	-USD	179.21	USD	126.30
20	-USD	32.72	-USD	190.60	USD	125.16
21	-USD	41.05	-USD	203.90	USD	121.79
22	-USD	46.35	-USD	214.00	USD	121.31
23	-USD	24.34	-USD	196.68	USD	148.00
24	-USD	32.18	-USD	209.07	USD	144.72

Fuente: elaboración propia.

#### **3.3.5.4. Análisis del valor en riesgo para *litecoin***

Como se mencionaba en el apartado anterior, dada la tendencia visual que conlleva la predicción del modelo hacia la serie de datos, existirán meses en donde el riesgo es alto, tal como la predicción para el mes 3, en donde el valor medio esperado está muy distante de un escenario pesimista dado el intervalo de confianza, para validar este comportamiento se realiza un análisis del valor en riesgo para *litecoin*.

Figura 83. **VaR *litecoin***



Fuente: elaboración propia.

Los retornos de los datos alcanzan el 1 % es decir el VaR (puntos rojos) 11 veces en comparación con las 12.6 veces esperadas, además, como se evidencia en la tabla XLIX las superaciones son correctas e independientes. Esto implica que se espera que una pérdida del 0.87 % se supere solo el 1 % del tiempo en función de la inversión inicial, sin embargo, dado el comportamiento característico, y su cercanía con una subestimación del valor en riesgo, no es un activo en extremo redituable dado su cercanía con la cota superior y su bajo valor para el horizonte temporal estudiado.

Tabla XLIX. **Resumen del Modelo VaR *litecoin***

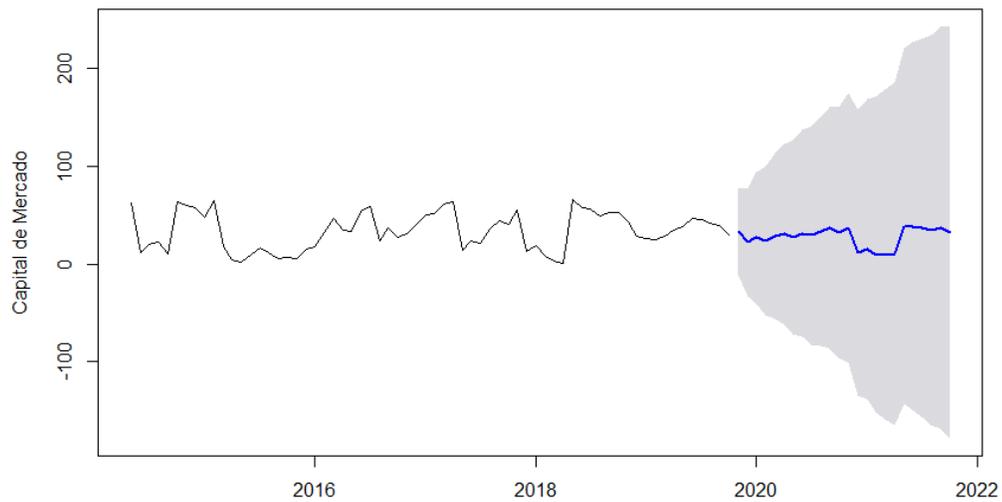
Modelo estudiado	ARIMA Estacional
Metadata	1264 registros
$\alpha$	1 %
VaR esperado	12.6
VaR excedido realmente	11

Continuación tabla XLIX.

% de VaR	0.9 %
Cobertura incondicional (Kupiec)	$p - value = 0.635$
Independencia de las coberturas (Christoffersen)	$p - value = 0811$

Fuente: elaboración propia.

Figura 84. **Pronóstico del modelo ARIMA para *monero***



Fuente: elaboración propia.

Tabla L. **Prueba de normalidad de *monero***

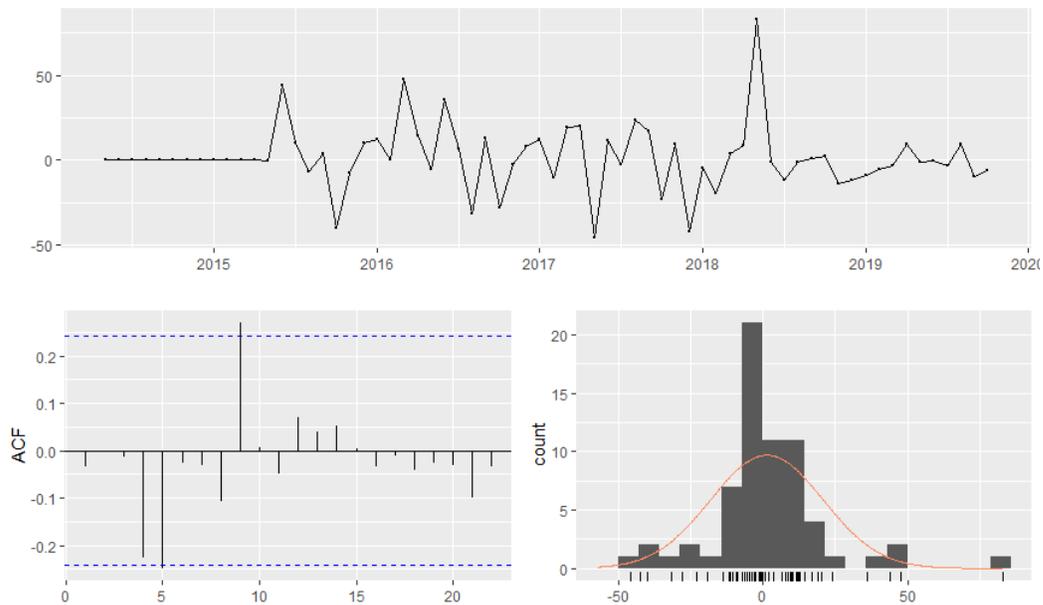
Modelo	Residuos <i>monero</i>
D	0.1790
Valor-P	0.3188

Fuente: elaboración propia.

En la figura 84 se aprecia de forma visual el pronóstico para la divisa *monero*, en dónde se aprecia un comportamiento bastante estable, además, se aprecia a que posterior al primer año, el modelo muestra una ligera caída de valor en donde es probable no sea conveniente realizar inversiones.

Adicionalmente, en la figura 85 en el extremo inferior izquierdo se observa que los atrasos del modelo se encuentran controlados por las bandas del gráfico de autocorrelación, además, la prueba de normalidad en la tabla L además del histograma de residuos en la figura 85 evidencian normalidad en la distribución de residuos, lo cual denota que el modelo se comporta de manera satisfactoria y es eficiente para realizar predicciones sobre la serie de datos.

Figura 85. **Análisis de residuos para predicciones de *monero***



Fuente: elaboración propia.

Tabla LI. **Pronóstico a 24 meses de *monero* con cifras en millones de dólares**

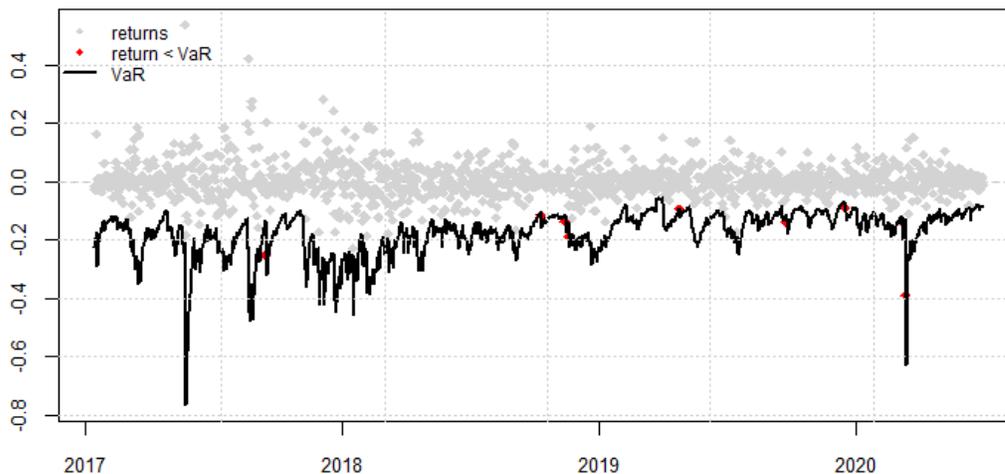
Periodo pronosticado	Valor más probable	IC. Límite inferior	IC. Límite superior
1	USD 33.08	-USD 10.88	USD 77.05
2	USD 21.84	-USD 33.64	USD 77.33
3	USD 26.66	-USD 40.29	USD 93.61
4	USD 23.98	-USD 52.20	USD 100.16
5	USD 28.30	-USD 56.26	USD 112.86
6	USD 30.24	-USD 61.90	USD 122.37
7	USD 26.80	-USD 72.35	USD 125.94
8	USD 31.26	-USD 74.44	USD 136.95
9	USD 29.51	-USD 82.35	USD 141.36
10	USD 33.25	-USD 84.44	USD 150.95
11	USD 36.76	-USD 86.50	USD 160.02
12	USD 31.93	-USD 96.65	USD 160.51
13	USD 37.10	-USD 101.06	USD 175.26
14	USD 11.66	-USD 134.38	USD 157.70
15	USD 15.27	-USD 138.62	USD 169.15
16	USD 9.42	-USD 151.81	USD 170.66
17	USD 9.65	-USD 158.65	USD 177.94
18	USD 10.66	-USD 164.40	USD 185.72
19	USD 38.59	-USD 142.99	USD 220.17
20	USD 38.39	-USD 149.48	USD 226.27
21	USD 36.90	-USD 157.06	USD 230.86
22	USD 34.32	-USD 165.55	USD 234.18
23	USD 36.39	-USD 169.20	USD 241.99
24	USD 32.16	-USD 179.01	USD 243.34

Fuente: elaboración propia.

### 3.3.5.5. Análisis del valor en riesgo para *monero*

Como se apreciaba de forma gráfica, la tabla LI, muestra que en los periodos 15 al 18 no es conveniente realizar inversiones dada la caída brusca que supone el modelo desde el pronóstico 14, sin embargo, tiende a crecer a partir del mes 19 del pronóstico, lo cual es una clara referencia que el saber cuándo vender y cuándo comprar es crucial en la compra de esta clase de activos bursátiles, para validar estas apreciaciones se realiza un análisis de valor en riesgo.

Figura 86. VaR *monero*



Fuente: elaboración propia.

Los retornos de los datos alcanzan el 1 % es decir el VaR (puntos rojos) 9 veces en comparación con las 12.6 veces esperadas, además, como se evidencia en la tabla LII, las superaciones son correctas e independientes. Esto implica que se espera que una pérdida del 0.71 % se supere solo el 1 % del

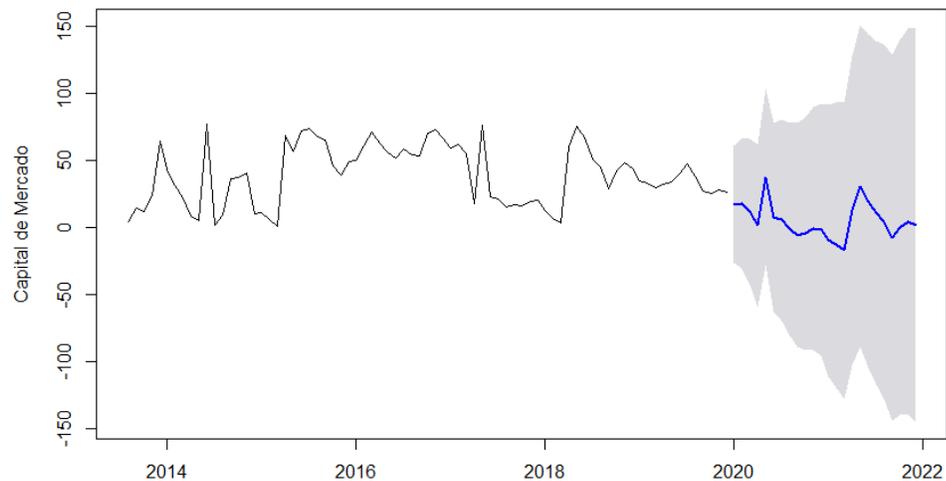
tiempo en función de la inversión inicial, por ende, refleja un comportamiento más equilibrado en proporción de riesgo que *litecoin* a pesar de su baja apreciación.

Tabla LII. **Resumen del modelo VaR *monero***

Modelo estudiado	ARIMA estacional
Metadata	1264 registros
$\alpha$	1 %
VaR esperado	12.6
VaR excedido realmente	9
% de VaR	0.7 %
Cobertura incondicional (Kupiec)	$p - value = 0.278$
Independencia de las coberturas (Christoffersen)	$p - value = 0.52$

Fuente: elaboración propia.

Figura 87. **Pronóstico del modelo ARIMA para *ripple***



Fuente: elaboración propia.

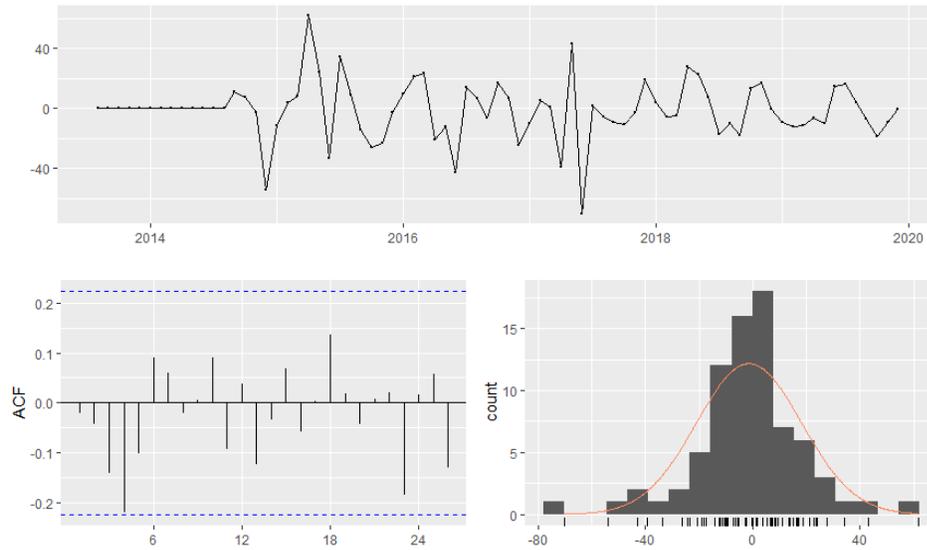
Tabla LIII. Prueba de normalidad de *ripple*

Modelo	Residuos <i>ripple</i>
D	0.047
Valor-P	0.217

Fuente: elaboración propia.

Como se aprecia gráficamente en la figura 87, existen patrones claros de alza y baja para la divisa *ripple* de forma visual, como se comentaba para la divisa *monero*, es probable que existan alzas significativas y bajas significativas en periodos cortos de tiempo, por lo que las sugerencias deben ser eficientes, por otro lado, el modelo tiene una distribución de residuos adecuada, en dónde en la figura 88 se aprecia como en el extremo inferior izquierdo en el gráfico de autocorrelación los atrasos se distribuyen de forma adecuada en los límites, además el histograma de residuos en el extremo inferior derecho de la figura, y la tabla LIV donde se presentan los resultados de la prueba de normalidad aplicada muestran resultados satisfactorios para este supuesto, por lo que se concluye que el modelo generalizado es eficiente para realizar pronósticos de todas las divisas individuales involucradas y la metodología empleada fue satisfactoria.

Figura 88. **Análisis de residuos de predicciones de *ripple***



Fuente: elaboración propia.

Tabla LIV. **Pronóstico a 24 meses de *ripple* con cifras en millones de dólares**

Periodo pronosticado	Valor más probable	IC. Límite inferior	IC. Límite superior
1	USD 17.18	-USD 26.24	USD 60.60
2	USD 17.47	-USD 30.90	USD 65.85
3	USD 12.19	-USD 42.07	USD 66.46
4	USD 1.41	-USD 59.20	USD 62.02
5	USD 37.52	-USD 28.11	USD 103.15
6	USD 7.57	-USD 62.90	USD 78.03
7	USD 5.47	-USD 69.55	USD 80.49
8	-USD 1.30	-USD 80.58	USD 77.97
9	-USD 5.78	-USD 89.10	USD 77.54
10	-USD 4.62	-USD 91.80	USD 82.57

Continuación tabla LIV.

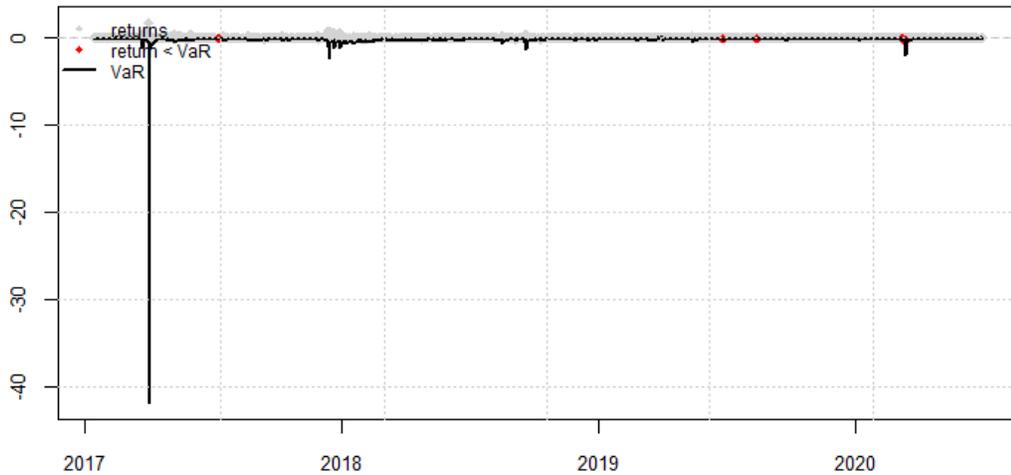
11	-USD	1.10	-USD	91.98	USD	89.78
12	-USD	1.74	-USD	96.17	USD	92.69
13	-USD	9.56	-USD	110.94	USD	91.82
14	-USD	13.24	-USD	119.30	USD	92.83
15	-USD	17.40	-USD	128.19	USD	93.40
16	USD	12.57	-USD	102.93	USD	128.07
17	USD	30.78	-USD	89.11	USD	150.68
18	USD	19.52	-USD	104.65	USD	143.69
19	USD	11.09	-USD	117.21	USD	139.40
20	USD	4.37	-USD	127.93	USD	136.68
21	-USD	7.78	-USD	143.97	USD	128.41
22	-USD	0.44	-USD	140.40	USD	139.53
23	USD	4.40	-USD	139.24	USD	148.05
24	USD	1.62	-USD	145.61	USD	148.85

Fuente: elaboración propia.

### 3.3.5.6. Análisis del valor en riesgo para *ripple*

Se aprecia que las alzas y bajas son un factor significativo en esta divisa, como se muestra en la tabla LIV existen horizontes temporales en dónde la moneda no crea valor y es conveniente comprar por ejemplo del periodo 8 al 15 del pronóstico, además se evidencia que es la divisa de menor valor bursátil evaluado en cuanto a los pronósticos realizados., para validar estos supuestos se realiza un análisis del valor en riesgo para *ripple*.

Figura 89. **VaR ripple**



Fuente: elaboración propia.

Los retornos de los datos alcanzan el 1 % es decir el VaR (puntos rojos) 5 veces en comparación con las 12.6 veces esperadas, además, como se evidencia en la tabla LV, las superaciones son correctas e independientes. Esto implica que se espera que una pérdida del 0.3955 % se supere solo el 1 % del tiempo en función de la inversión inicial, hasta ahora es el comportamiento más equilibrado en cuanto a los valores por debajo del umbral de riesgo permisible en pérdidas y ganancias, ahondado a que es la divisa de menor apreciación sugiere que su comportamiento es bastante uniforme.

Tabla LV. **Resumen del modelo VaR ripple**

Modelo estudiado	ARIMA estacional
Metadata	1264 registros
$\alpha$	1 %
VaR esperado	12.6
VaR excedido realmente	5

Continuación tabla LV.

% de VaR	0.4 %
Cobertura incondicional (Kupiec)	$p - value = 0.014$
Independencia de las coberturas (Christoffersen)	$p - value = 0.048$

Fuente: elaboración propia.

## 4. DISCUSIÓN DE RESULTADOS

En este capítulo se detallan diversas premisas relacionadas a los resultados obtenidos en las secciones anteriores, así también la proyección estimada para futuras inversiones en las criptodivisas estudiadas.

### 4.1. Descripción de la tendencia de las criptodivisas

Analizando descriptivamente a cada divisa en particular se observó que el comportamiento singular de las divisas es altamente asimétrico, es decir, la amplitud en los intervalos de confianza tiende a crecer o decrecer de forma significativa en periodos cortos de tiempo, en las figuras 2 a 13 se puede evidenciar de forma gráfica este fenómeno dado el comportamiento estadístico característico.

Debido a este comportamiento no es claro apreciar visualmente si en efecto existe una tendencia en la serie de datos analizada, o simplemente se trata de un periodo de inflación estacional, para tal fin se sometió cada serie de datos a una prueba de tendencias de Mann-Kendall en donde se deseaba validar en principio si existía una tendencia para cada divisa, además de ver hacia dónde se proyectaba dicha tendencia, de forma gráfica se puede apreciar en las figuras 14 a 19 la proyección de estas curvas.

A su vez la prueba concluyo que sí existe una tendencia en todas las divisas estudiadas (Tabla X a la XV) y no solo eso, también probó que las tendencias para todas las divisas son crecientes, dado el coeficiente de Kendall  $\tau$  era positivo, lo que denota una tendencia positiva para todas las series.

Esto resulta en una conclusión interesante, dado a que se proyecta en posibles rendimientos eficientes para todas las divisas, siendo *dash* la del coeficiente más positivo con un tau de Kendall de 0.8110 (tabla XI), esto puede apreciarse en la figura 78, en dónde tiende a tener un incremento en su valor bursátil consecutivo además de estabilidad en relación a la traza histórica.

Cabe destacar que la variable utilizada fue el valor del volumen de capital para todas las divisas, como mencionan los autores Salwach y Perz (2020) el volumen de capital bursátil al ser la variable que colecta de mejor forma tanto el volumen de las aperturas bursátiles generadas como para los cierres totales, por lo que es la variable que mejor describe al menos de forma bursátil a las divisas, por otro lado, en el estudio funcionaron de forma óptima tanto en el comportamiento de residuales de los modelos, como para obtener análisis de las divisas.

#### **4.1.1. Análisis de correlación de las divisas**

Por medio de un análisis de correlación, se determinó que las divisas si se encuentran correlacionadas proporcionalmente a baja escala esto puede apreciarse en la tabla XVI en dónde se evidencia que la mayor correlación teórica la presenta *litecoin* con *bitcoin* con un coeficiente de correlación de Pearson de -0.31, además en la tabla XLIX se puede apreciar todo el resumen de coeficientes por divisa.

Este análisis parte de los hallazgos de los autores Castro y Forero (2018) en donde se evidencia que al menos un punto porcentual en cuanto a la correlación con la criptomoneda de interés, es decir la de mayor valor, refleja un potencial retorno sobre la inversión con una tendencia similar a la divisa de mayor valor.

Al estar alejadas en valor bursátil las divisas se entiendo el por qué los bajos valores de correlación, además como se presenta más adelante si se refleja en la realidad, las divisas que mayor correlación presentaban en este estudio con *bitcoin* son las que mejores rendimientos presentan.

Las mejores asociaciones en relación a cada divisa se presentan a continuación, en función al alza o baja bursátil, en función del coeficiente de correlación de cada divisa.

Tabla LVI. **Asociaciones de correlación por divisa**

	Mayor correlación bursátil alta	Mayor correlación bursátil baja
<b><i>Bitcoin</i></b>	0.15 <i>Ethereum</i>	-0.31 <b><i>Litecoin</i></b>
<b><i>Dash</i></b>	0.12 <i>Bitcoin</i>	0.08 <b><i>Monero</i></b>
<b><i>Ethereum</i></b>	0.15 <i>Bitcoin</i>	-0.05 <b><i>Monero</i></b>
<b><i>Litecoin</i></b>	-0.31 <i>Bitcoin</i>	-0.11 <b><i>Monero</i></b>
<b><i>Monero</i></b>	0.08 <i>Dash</i>	-0.11 <b><i>Litecoin</i></b>
<b><i>Ripple</i></b>	0.01 <i>Bitcoin</i>	0.09 <b><i>Monero</i></b>

Fuente: elaboración propia.

Además, gracias al análisis de correlación se pudo probar que la relación de las divisas es no lineal, por lo tanto un modelo lineal es poco eficiente para modelar el fenómeno, para ello se graficó una matriz de dispersiones el cual se puede observar en la figura 20, en donde la dispersión de puntos denota concentraciones no lineales, además posteriormente se pretende probar los supuestos de un modelo lineal generado por uno de los análisis de varianza asociados a probar supuestos en el estudio, el modelo lineal en definitiva no es adecuado para modelar a los datos.

Además, no cumple los supuestos de normalidad en los residuos en dónde en la tabla XXI se puede apreciar un p-valor de  $1.29 * 10^{-9}$  además de una gráfica de normalidad de residuos en la figura 31 que denotan la poca eficiencia del modelo para simular el fenómeno.

Por lo que, gracias al análisis de tendencia, se infiere que es probable esperar potenciales rendimientos futuros positivos, dado el valor positivo de la prueba de Mann Kendall para todas las divisas, además, se probó que a pesar de ser baja sí existe mejor relación entre *bitcoin* y alguna de las divisas, por otro lado, se validó el supuesto que la relación de los datos es no lineal.

#### **4.1.1.1. Análisis del modelo de vector autorregresivo para las seis divisas**

En el apartado anterior se comparó el comportamiento de las divisas por medio de la correlación lineal histórica que presentaban, si bien es cierto el análisis presenta evidencia contundente en el contraste del comportamiento entre divisas, presenta ciertas limitantes., tanto en la imposibilidad de reflejar relaciones no lineales que existen entre las mismas, así como el no poder evidenciar conclusiones de muchos a muchos.

Por ende, para analizar cómo está constituida la influencia que tiene tanto la divisa de mayor apreciación del estudio, es decir *bitcoin* respecto al resto y la de menor apreciación, en este caso *ripple*, se optó por diseñar un modelo multivariante de vector autorregresivo, entre la multitud de bondades que ofrece el modelo, la que más interesa a nivel de este análisis, es su capacidad de extrapolar parámetros de las distancias entre vectores que lo conforman para generar inferencias generalizadas de los mismos.

En otras palabras, como un impacto en un arreglo de datos del vector, que en este caso caracteriza a una divisa en particular, influye en las inferencias del resto de divisas, para ello se toma de referencia un análisis ortogonal de impulso respuesta, el cual no es más que el ajuste de las series estacionales ajustadas a un corte de referencia, y como un impacto particular de un vector tiende a desestabilizar a cada una de las divisas, además, si el quiebre rompe el equilibrio estacional de forma significativa, se puede analizar qué tendencia tiene el mismo para identificar si tiene un impacto positivo o negativo.

Como se aprecia en la figura 24, *bitcoin* es la divisa que tiene una influencia contundente en el resto de divisas, un impacto en *bitcoin* supone una variación evidente en el resto de muestras indiferente a si esta es positiva o negativa, por ende, se deben de tener contemplados potenciales fluctuaciones bursátiles fuertes en *bitcoin* para un diseño de inversión eficiente.

Para complementar el análisis, se deseaba observar qué proporción de la varianza explicada del modelo estaba representada por cada divisa, se evidencia en la figura 30 que *bitcoin* es la divisa que compone gran parte de la varianza estructural por modelo para cada vector, lo que reafirma que si bien no es un mercado del todo especulativo, y como se concluía en el apartado anterior gracias al análisis de correlación, la divisa de mayor valor influye en el resto de la muestra del mercado, entonces aunque no sea una relación lineal 1:1c siempre debe tomarse como referencia cambios estructurales de valor en BTC.

Para complementar la afirmación anterior, se realizó el mismo análisis de impulso-respuesta, el cual puede apreciarse en las figuras 25 a 29, para el resto de divisas, con el fin de analizar como diversificar un portafolio de inversión, que supondría un impacto en divisas de menor apreciación respecto al resto, si bien es cierto en algunas no supone un quiebre estacional significativo para ninguna,

otras tienden a apreciarse a pesar de un impacto en la misma divisa, como es el caso de *ripple*, lo cual podría sugerir evidencia de momentos oportuno de venta acorde a impactos en el resto.

## **4.2. Comparativa del precio bursátil de las divisas**

Uno de los objetivos de investigación es poder validar por medio del análisis de varianza los supuestos asociados a las criptodivisas de los distintos mercados de capital, principalmente el probar que no existe diferencia significativa en su precio en los distintos mercados financieros (descentralización de precios), además, probar que el criterio de relación bursátil (el cual se expresa en la tabla XXIII) y el criterio divisa se cumplen, es decir si existe una diferencia significativa en las divisas estudiadas, así también si existe variabilidad en criterio de su comportamiento histórico diario.

### **4.2.1. ANOVA para factor de divisas y para el criterio de relación bursátil**

Para probar que efectivamente existe diferencia significativa entre los valores de capital de mercado medio en las distintas divisas se agrupo por medio del valor medio del volumen de capital en los distintos mercados financieros considerados, por medio de un análisis de varianza múltiple el cual está descrito en la tabla XXII, se obtuvo un resultado significativo en el p-valor de 0 el cual denota que efectivamente existe diferencia significativa en al menos una de las divisas.

Dada esta conclusión, se realizó una prueba post-hoc en dónde se obtienen los resultados de asociación para las divisas por medio de una prueba de Tukey (tabla XX), la cual denota a *bitcoin* en un grupo aparte, al igual que individualiza

a *ethereum* y *ripple*, sin embargo, *litecoin*, *dash* y *monero* se encuentran agrupados en un mismo grupo, esto es interesante dado a que *litecoin* es la moneda que tiene el mejor coeficiente de correlación con *bitcoin* -0.31 entonces se convierte en una de las divisas de valor más bajo, pero más interesantes en el estudio.

En cuanto al criterio de relación bursátil, se realizó de igual forma un análisis de varianza múltiple en donde los factores a analizar fueron la relación bursátil y las criptodivisas involucradas, el resumen de este análisis se puede apreciar en la tabla XXIII, en donde se obtiene un valor p de 0 para el criterio de relación bursátil siendo un resultado significativo, lo cual denota que indistinto del mercado de capital que se esté trabajando, si existen diferencias significativas entre los precios y su relación la divisa de forma diaria, esto se puede apreciar en la historia de las divisas, la cual no es simétrica, el resumen de las series temporales de forma gráfica puede apreciarse en las figuras 35 a 37.

#### **4.2.2. ANOVA para distintos mercados de capital**

Una de las principales premisas de las criptodivisas es que son activos descentralizados, que no están sujetos a ninguna legislación y su valor bursátil no está sujeto a ningún criterio de economía clásica, por lo que para probar tal fin, se seleccionó una muestra de las tres agrupaciones que se tenían de las divisas en el apartado anterior, siendo estas *bitcoin*, *dash* y *ripple*, en donde se utilizaron los datos de distintos mercados de capital, los cuales son el mercado de divisas de Londres (LSE), el mercado de divisas de Nueva York (NYSE) y el mercado de divisas de China (SSE).

Para las tres divisas seleccionadas, las cuales corresponden a los tres grupos de valor medio de volumen de capital, se seleccionaron periodos

homogéneos de los distintos mercados de capital, desde el 01 de enero de 2014 al 30 de diciembre del año 2019, con ello utilizando un diseño factorial fue posible realizar el análisis de varianza el cual tiene resultados satisfactorios en relación a las premisas asociadas a cada divisa de referencia, para validar la hipótesis de igual de medias se utiliza el valor p se tiene que para *bitcoin* un valor p de 0.879, para *dash* un p valor de 0.702, y para *ripple* un p valor de 0.850, en dónde para todos los casos se puede afirmar que no existe una diferencia estadísticamente significativa entre los precios medios del volumen de capital de mercado para las divisas consideradas, indiferente a la moneda de referencia que se utilice en el mercado financiero.

Este último punto es algo sumamente positivo, dado a que en efecto las inversiones pueden realizarse desde cualquier origen y obtener resultados equivalentes, lo cual vuelve reproducibles las recomendaciones de esta investigación, al menos dentro del horizonte temporal considerado.

Por otra parte, se puede validar que efectivamente el estimar que *bitcoin* es la moneda de referencia para el resto dado su alto valor bursátil es correcto, (ver figura 33 y 34) en dónde se denota que, para las diferencias de valor medio, efectivamente ningún intervalo contiene al 0, todos se encuentran al lado izquierdo, lo que concluye que los valores de *bitcoin* son significativamente más altos que los del resto de las divisas.

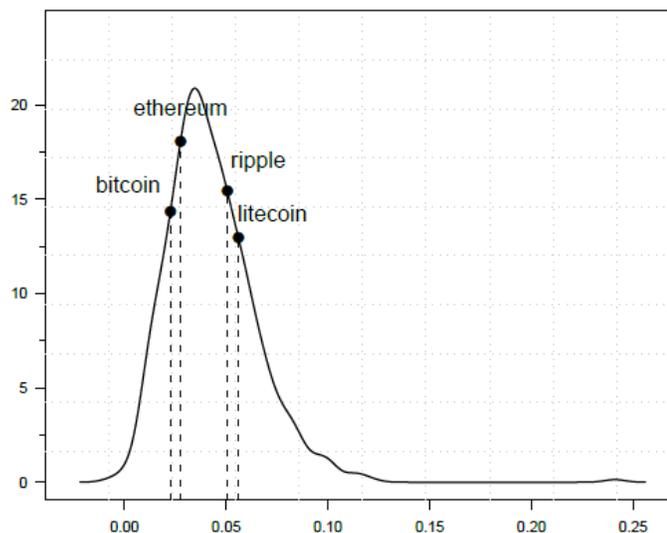
#### **4.3. Modelo matemático que describe a las criptodivisas**

Para modelar el fenómeno se tomó de referencia los métodos propuestos bajo la generalización de análisis de datos masivos de los autores Cryer y Chan (2008) para no generar una aproximación distinta acorde a cada serie de datos, se trabajó por medio de una serie de datos generalizada o serie de datos de

modelización, está última parte del supuesto de que las divisas tendrán un comportamiento relacional y autorregresivo de forma histórica, esto se pudo probar de forma eficiente (ver tabla XXX) en dónde las series sí disponen de una tendencia histórica, además de una componente estacional y autorregresiva

Los supuestos de esta serie de datos de modelización parten de que se asignará un peso menor a la serie de referencia y gradualmente se asignan pesos mayores a las series de valor mínimo entre el conjunto de datos, para tal fin en el estudio se puede observar en la tabla XXXI que se asignan pesos de referencia desde un 5 % de participación de *bitcoin* hasta un 25 % de participación en relación a *ripple* o *monero*. Probabilísticamente se expone un comportamiento con aristas como la figura a continuación.

Figura 90. **Distribución de pesos para la serie de modelación**



Fuente: elaboración propia.

Con ello lo que se pretende es poder generar aproximaciones con datos de todas las divisas con un modelo aplicable posteriormente a cada divisa de forma individual, como se puede apreciar en la figura 66, las distribuciones están comprendidas con las divisas de mayor valor que ocupan gran parte del área de la curva, tomando siempre valores de *bitcoin*, la serie generalizada de tiempo se puede apreciar en la figura 63, en dónde si se compara con la descomposición por componentes de *bitcoin* en la figura 38 tienen un parecido visual más que evidente, lo cual denota que al final, las divisas tienden a comportarse de forma homogénea con las divisas de alto valor de mercado, por ende, sí se trata de un mercado con una componente especulativa.

#### **4.3.1. Modelo de series de tiempo generado**

El modelo elegido está basado en metodologías de series de tiempo y modelos autorregresivos integrados de medias móviles, se decidió formular dos modelos para decidir cuál era la alternativa más eficiente para poder caracterizar la serie de tiempo de modelación y a la vez poder generar predicciones en relación a la misma, el primer modelo es un modelo planteado de forma teórica, contemplando las diferencias estacionales optimas de forma iterativa para la serie y con la ayuda de un árbol Bayesiano (ver figura 62) la cual es una herramienta visual que ayuda a entender los atrasos del modelo, que tan recursivos son durante la frecuencia del modelo y qué tan concentrados están a lo largo de toda la serie (Cryer y Chan, 2008).

Dado a que se tenía que incluso el último atraso era recurrente en la serie se optó por plantear un modelo ARIMA (2,1,0) (2,1,0), es decir un modelo con componente autorregresivo, integrado sin medias móviles y adicional estacional, esto para una frecuencia de 12 periodos, dado a que la frecuencia de las

observaciones es mensual, la ecuación general del modelo generado es la siguiente, la cual es equivalente a la expansión realizada en la ecuación 56:

$$\Delta x_t = c + \phi_1 \Delta x_{t-1} + \phi_2 \Delta x_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t \quad (\text{Ec. 58})$$

El modelo ARIMA teórico planteado obtuvo resultados satisfactorios, tanto para los criterios de penalización de la información en la cual se tiene de referencia un criterio de información de Akaike de 603.19, esta y las demás métricas de entrenamiento del modelo se pueden ver en la tabla XXXII. Además, el modelo presenta normalidad en toda la distribución de los residuos (ver figura 65 y tabla XXXII) lo cual lo vuelven un modelo viable mediante el cual se pueden generar predicciones que simulen el fenómeno de forma eficiente.

Por otra parte, se utilizó otro acercamiento para poder brindar solución al planteamiento de un modelo de comparación, esto fue por medio de un algoritmo de aprendizaje de máquina, el seleccionado fue el algoritmo multivariante de árboles de clasificación, el cual a pesar de tener la posibilidad de utilizar a todas las variables bursátiles como predictores, lo cual probablemente hubiese vuelto más robusto el modelo, únicamente fue entrenado con la variable de volumen de capital bursátil como predictor, utilizando el criterio de información de Akaike como criterio de optimización, luego de la convergencia del modelo en el valor más eficiente se obtienen métricas de entrenamiento bastante buenas, en dónde destaca un criterio de información de Akaike de 673.51 planteando un modelo ARIMA (1,0,0), es decir, un modelo con componente autorregresiva, pero sin diferenciación, medias móviles o parte estacional ver tabla XXXV.

Sin embargo, el modelo no puede ser considerado como un modelo válido, dado a que no presenta normalidad en la distribución los residuos como se evidencia en la figura 70 y en la tabla XXXVI, fuera de que el modelo no puede

ser contemplado como un modelo predictivo válido, de igual forma se realizó una comparativa de métricas de estimación, en dónde el modelo planteado de forma teórica posee más de un 30 % de eficiencia sobre el modelo planteado por computadora, este ejercicio es interesante, dado a que muchos software de análisis estadístico automatizado suelen plantear modelos ARIMA poco eficientes basados únicamente en la optimización de criterios de información indiferente a algunas características singulares de la serie de datos.

Por lo tanto, el modelo seleccionado es el modelo ARIMA (2,1,0) (2,1,0), el cual se aplicó de forma individualizada a cada divisa, lo cual era el objetivo central de haber distribuido de esa forma el modelado, dando resultados satisfactorios con cada serie individualizada de datos (ver sección 3.4.3 de la investigación) en dónde muestra que para cada divisa los atrasos del modelo se comportan de forma regular, además todas las predicciones asociadas a divisas presentan una distribución normal en los residuos del modelo validado por medio de una prueba de Kolmogorov Smirnov, lo cual denota que la metodología empleada es adecuada y se diseñó un modelo que tiene capacidad predictiva para cada una de las seis divisas consideradas.

#### **4.4. Modelo de eficiencia financiera**

Para poder analizar y generar conclusiones las potenciales pérdidas y ganancias que hubiese obtenido un inversionista en caso de colocar recurso financiero en uno o varios de los activos virtuales estudiados, se planteó un modelo de eficiencia financiera teórico, el cual toma como insumo las inferencias generadas por el modelo de series de tiempo planteado, la cual con la variable de precio, tiene la capacidad de proyectar potenciales valores futuros para cada divisa y con ello poder calcular rendimientos teóricos con métricas financieras de retorno sobre la inversión.

A su vez, para cuantificar el factor riesgo, se analiza el valor en riesgo (VaR) para cada horizonte temporal acorde a las inferencias y un nivel de significancia situado en el 1% en dónde además de una ratio de devolución, se puede cuantificar los puntos críticos de pérdidas en el flujo.

#### **4.4.1. Rentabilidad esperada por predicción**

En la sección 3.4.3 se utilizó el modelo ARIMA planteado para generar predicciones relacionadas a cada divisa, con las cuales se obtuvo el valor medio estimado para cada divisa, así como un intervalo de confianza asociado al valor medio de predicción, con ello lo que se pretende es realizar recomendaciones acordes a los valores predichos antes de realizar un cálculo más exhaustivo por medio de indicadores financieros.

Adicional a ello se realizó un análisis de valor en riesgo por medio de un modelo (VaR) para las pérdidas y ganancias que brinda el modelo ARIMA estacional.

Se estimó como un techo proporcional significativo que si al menos un 1 % de los valores superaba el umbral de riesgo estimado, se habría subestimado el riesgo de inversión y por ende, son potenciales las pérdidas que supondría el activo, dada la homologación de fechas, se disponen de 1264 registros para las seis divisas, entonces, si al menos 12.6 observaciones superan la cota de riesgo, supondría que se ha subestimado el riesgo de la inversión. Un análisis gráfico y resumen tabular puede apreciarse en la figura 74 y XL para *bitcoin* y así sucesivamente para cada divisa.

El análisis acota a que no se ha subestimado el VaR en ninguna de las divisas estudiadas, esto se debe a que gracias a las condiciones del horizonte

temporal para las divisas, las apreciaciones son mucho más significativas que las depreciaciones, esto se denota de mejor forma en el apartado posterior, sin embargo, el modelo VaR si evidencia qué divisas son las más cercanas a haber subestimado el riesgo de inversión, lo cual se ve correlacionado en su intervalo de confianza de predicción como es el caso de *monero* XMR.

#### **4.4.1.1. Rendimiento *bitcoin***

Como se aprecia en la tabla XXXIX, las mejores inversiones para *bitcoin* comienzan a partir del mes 12, es decir al final del primer año, hasta el mes 21, en dónde el mes más adecuado para vender la divisa es el mes 14, en dónde se tiene el siguiente intervalo de confianza en miles de millones de dólares americanos:

$$(-39.88 \leq 86.18 \leq 212.25)$$

Con un rendimiento teórico esperado de 90.13 % sobre la compra de divisas realizada de forma teórica en el mes 1.

#### **4.4.1.2. Rendimiento *ethereum***

Para *ethereum* los periodos de inversión más eficientes están comprendidos en los meses del 9 al mes 11 del primer año proyectado (tabla XLI) en dónde el mes más adecuado para vender dados sus posibles rendimientos es el mes 10, el cual cuenta con el siguiente intervalo de confianza en miles de millones de dólares americanos:

$$(-6.64 \leq 55.91 \leq 118.47)$$

Con un rendimiento teórico esperado de 75.09 % sobre la compra de divisas realizada de forma teórica en el mes 1.

#### **4.4.1.3. Rendimiento *dash***

En cuanto a *dash* los meses más eficientes para poder realizar inversiones están comprendidos entre los meses 12 al 18 dado que son los meses de mayor estabilidad de valor (tabla XLV) en donde el mes que se sugiere realizar la venta de divisas es en el mes 18, bajo la premisa del siguiente intervalo de confianza en miles de millones de dólares americanos:

$$(-84.04 \leq 80.32 \leq 244.68)$$

Para un rendimiento teórico esperado de 111.39 % sobre la compra de divisas realizada de forma teórica en el mes 1.

#### **4.4.1.4. Rendimiento *litecoin***

*Litecoin* por su parte es la divisa que presenta los valores menos eficientes de predicción financieramente hablando durante el primer año, los meses más eficientes de inversión están comprendidos a lo largo del segundo periodo del mes 16 al mes 22, siendo el mes 22 el mejor en relación a la venta de la divisa, este tiene el siguiente intervalo de confianza en millones de dólares americanos:

$$(-214 \leq 46.35 \leq 121.31)$$

Para un rendimiento teórico esperado de 96.21 % sobre la compra de divisas realizada de forma teórica en el mes 1.

#### 4.4.1.5. Rendimiento *monero*

En el caso de *monero* se tiene un fenómeno similar al de *litecoin*, en dónde los rendimientos del primer año proyectado no son muy buenos financieramente hablando, sin embargo, en el segundo año en los meses 19 al 24 es el mejor momento para realizar inversiones (tabla LI), siendo el mes 19 el adecuado para vender la divisa, el cual cuenta con el siguiente intervalo de confianza en millones de dólares americanos:

$$(-142.99 \leq 38.39 \leq 220.17)$$

Para un rendimiento teórico esperado de 174.14 % sobre la compra de divisas realizada de forma teórica en el mes 1.

#### 4.4.1.6. Rendimiento *ripple*

En cuanto a *ripple* los meses más eficientes para realizar inversiones son los meses del 15 al 22 (tabla XLIV) siendo el mes 18 el mes dónde se podrían obtener mejores rendimientos sobre venta, el cual tiene el siguiente intervalo de confianza en millones de dólares americanos:

$$(-104.65 \leq 19.52 \leq 143.69)$$

Para un rendimiento teórico esperado de 199.62 % sobre la compra de divisas realizada de forma teórica en el mes 1.

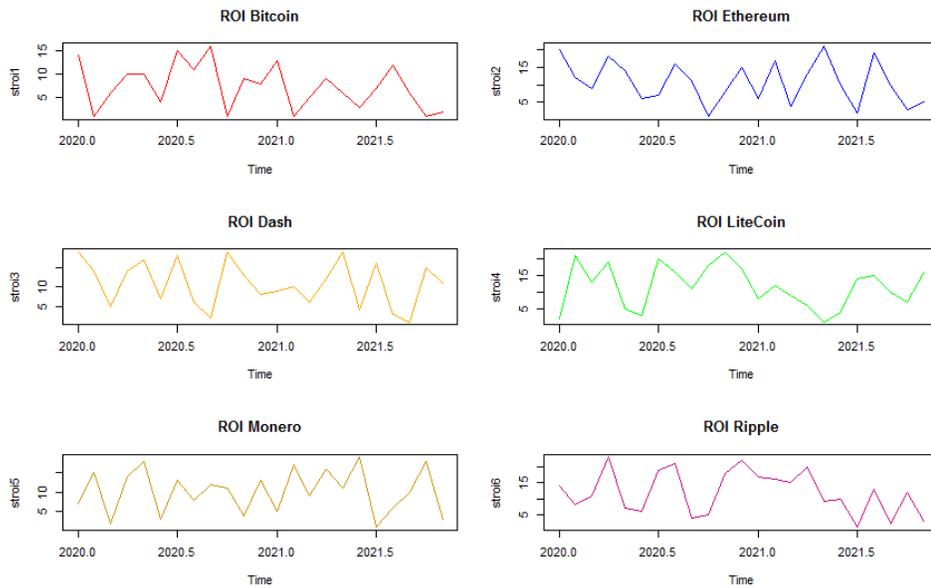
#### 4.4.2. Modelo de retorno sobre la inversión

La rentabilidad financiera, es decir lo que un inversionista puede llegar a percibir como beneficio posterior a una inversión de un activo suele ser cuantificado con el ratio financiero del retorno sobre la inversión (ROI) por sus siglas en inglés. Esta ratio mide la eficiencia del retorno de un activo, en ese caso cada criptomoneda, la cual denotaremos con la letra  $c$ , el ratio es considerado en periodos de tiempo homogéneos es decir periodos  $t$  para todas las inversiones Salwach y Paerz (2020) mencionan que dada la volatilidad de los cryptoactivos periodos fluctuantes son bastante frecuentes, por lo que a pesar de que sean volátiles, periodos de eficiencia financiera positiva son un buen indicador de retornos financieros optimistas.

A continuación, se presenta la forma algebraica que representa al ratio, en donde  $i$  hace referencia al periodo específico del cual se está hablando. Por lo que la medición de la eficiencia financiera por moneda desde el día 1 de la serie considerada está definido cómo:

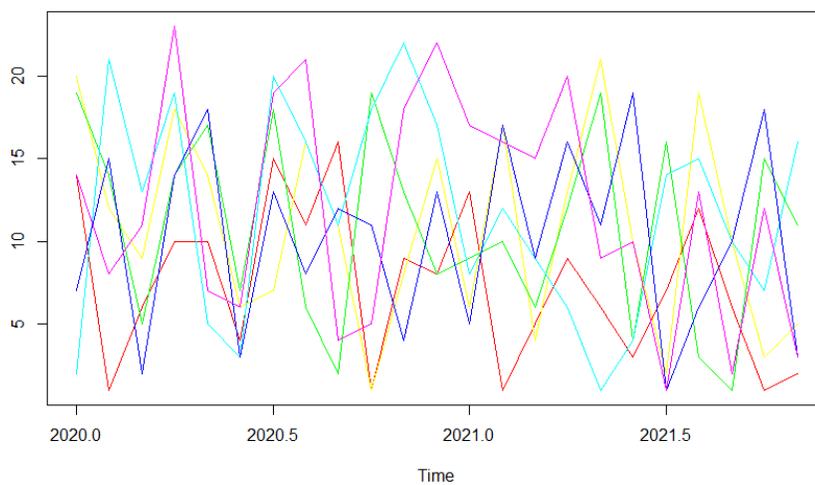
$$ROI(c, t_i) = \frac{\text{precio}(c, t_i) - \text{precio}(c, t_i - 1)}{\text{precio}(c, t_i - 1)} \quad (\text{Ec. 59})$$

Figura 91. **Serie de tiempo de ROI para cada criptodivisa**



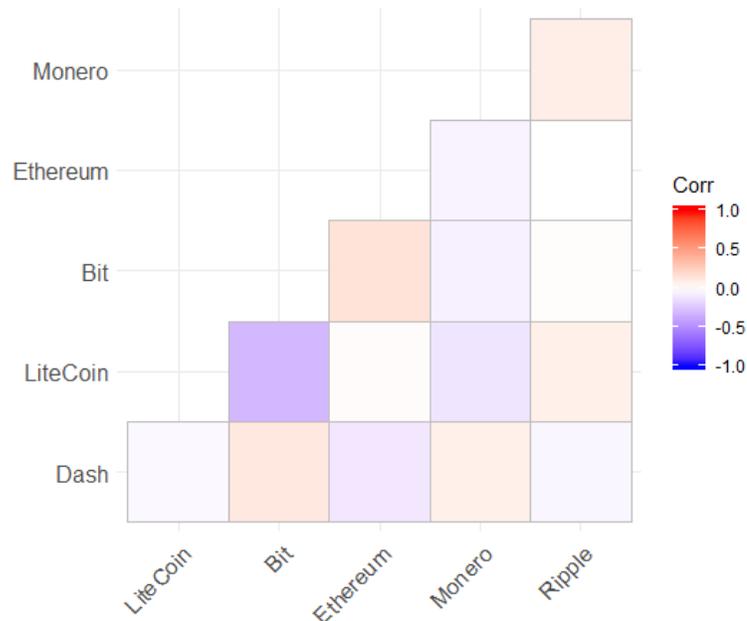
Fuente: elaboración propia.

Figura 92. **Sobreposición de todas las criptodivisas en función del ROI**



Fuente: elaboración propia.

Figura 93. **Correlación policorica de los ROI de cada criptodivisa**



Fuente: elaboración propia.

Se puede observar en la figura 93, que incluso en cuanto a su eficiencia financiera este tipo de activos tienen un comportamiento volátil, es decir cuando alguna serie original de datos presenta un cambio en la orientación de su tendencia, existen momentos adecuados en los cuales vender la divisa supondría un retorno financiero eficiente como se contempló en la sección anterior, en dónde bajo la compra de divisas a un precio adecuado se pueden generar rendimientos de hasta un 199 % como es el caso de *ripple*, y esto se ve evidenciado con una métrica financiera más robusta como lo es el ROI.

Por otra parte, se aprecia que los supuestos considerados para la correlación de divisas se cumplen significativamente, en dónde dada la existencia de relación porcentual en el comportamiento del volumen de capital bursátil por

divisa, se vería reflejado de forma efectiva sobre el retorno de la inversión, como se aprecia en la figura 93, *bitcoin* tiene la mayor relación en cuanto a retorno con *litecoin*, tal como refleja la tabla (XLVIII), además en cuanto a las divisas de valor intermedio *ethereum* es la que mayor relación tiene con *bitcoin*, además de *monero* y *ripple* son las que presentan los mejores rendimientos en cuanto a divisas de bajo valor bursátil.

Como puede observarse en la figura 91, el ratio del retorno sobre la inversión de igual forma refleja que sí existen puntos clave en los que el rendimiento financiero puede ser muy eficiente (bastante por encima del precio de compra pasado), sin embargo existen momentos en los que esa divisa puede caer por debajo del umbral positivo, y en lugar de generar valor es una pérdida neta, por lo que observar el comportamiento del modelo es clave para poder definir momentos de venta ideales, aunque en términos generales *ripple* y *monero* representan las divisas que a menor precio durante el año 2019, generan resultados eficientes al mediano plazo.

## CONCLUSIONES

1. El comportamiento estadístico de las criptomonedas evaluadas tiende mayoritariamente al alza en relación a su volumen de capital financiero, un modelo de series de tiempo ARIMA estacional puede captar de forma adecuada el comportamiento histórico, presentando un coeficiente de determinación  $R^2$  significativo, y una correcta bondad de ajuste a la normal en sus residuos, según la prueba de Kolmogorov Smirnov. Este modelo parte de una serie de tiempo generalizada por pesos probabilísticos, el cual brinda intervalos de confianza predictivos que demuestran el comportamiento inflacionario de las divisas.
2. Existe una tendencia positiva para todas las divisas estudiadas, a partir de la prueba de Mann Kendall se estima que todas las divisas tienen una proyección futura positiva, a su vez, existe un factor de equivalencia en la correlación del volumen de capital de las divisas hacia *bitcoin*, y el valor esperado de retorno sobre la inversión, es decir a mayor homogeneidad en las variaciones bursátiles respecto a *bitcoin*, es más probable un rendimiento financiero óptimo.
3. Se realizó una comparación de mercados de capital financiero por medio de un análisis de varianza para todas las criptomonedas, este refleja resultados no significativos para ninguna de las interacciones entre el valor medio de cada divisa y su mercado, lo cual demuestra que, indiferente al mercado financiero o moneda de referencia, todas las criptodivisas tienen un valor equivalente, lo cual valida la premisa de descentralización de las mismas.

4. Un modelo ARIMA (2,1,0)(2,1,0) planteado de forma teórica es el que mejor describe el comportamiento estadístico de todas las criptodivisas estudiadas, y presenta resultados satisfactorios y válidos en relación a la distribución de sus residuos, este modelo fue utilizado para generar un pronóstico de 24 meses para todas las divisas, los cuales reflejan los potenciales rendimientos financieros futuros de las divisas y también momentos de potenciales pérdidas.
  
5. Los resultados predictivos para los seis activos virtuales, muestran intervalos de confianza dónde se reflejan los momentos idóneos para la inversión y venta, esperando rendimientos estimados entre un 75 al 199 %, sobre monto de inversión en periodos de tiempo adecuados, *ethereum* y *litecoin* son las divisas que reflejan el potencial más alto de apreciación, por otro lado, las divisas de menor valor que reflejan correlación y rendimientos futuros eficientes son *monero* y *ripple*.

## RECOMENDACIONES

En relación a los resultados obtenidos y a las predicciones aproximadas con el modelo planteado, se sugieren los siguientes puntos para continuar y nutrir la investigación, además de mejorar los mecanismos para realizar un posible modelo más robusto para poder incluir criptomonedas adicionales.

1. Introducir variables bursátiles adicionales al estudio, que agreguen valor al análisis de varianza y con ello tener la posibilidad de brindar conclusiones adicionales del tipo comercial y financiero al estudio.
2. Analizar el efecto directo que supondría agregar más criptomonedas de baja apreciación al estudio y disminuir el peso de *bitcoin* sobre la generalización, con el fin de describir en mejor medida a monedas de bajo valor bursátil.
3. Utilizar de nuevo un algoritmo de aprendizaje de máquina para estimar el mejor modelo autorregresivo, pero incluyendo todas las covariables disponibles de los mercados bursátiles y su interacción, para probar si las conclusiones cambian o se obtiene un modelo más eficiente.
4. Utilizar un acercamiento predictivo distinto, directamente sobre los rendimientos financieros futuros por divisa, utilizando algoritmos robustos como redes neuronales para regresión.



## REFERENCIAS

1. Ammous, K. (2016). *Can Cryptocurrencies Fulfil the Functions of Money?* SSRN Electronic Journal. doi:10.2139/ssrn.2832769.
2. Arango, C. (2018). *Criptoactivos. Documento Técnico Criptoactivos*, 1-47. Recuperado Julio/Agosto, 2020, de: <https://www.banrep.gov.co/sites/default/files/publicaciones/archivos/documento-tecnico-criptomonedas.pdf>.
3. Bakar, N., y Rosbi, S. (2017). *Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction. International Journal of Advanced Engineering Research and Science*, 4(11), 130-137. doi:10.22161/ijaers.4.11.20.
4. Bakar, N., Rosbi, S., y Uzaki, K. (2017). *Cryptocurrency Framework Diagnostics from Islamic Finance Perspective: A New Insight of Bitcoin System Transaction. International Journal of Management Science and Business Administration*, 4(1), 19-28. doi:10.18775/ijmsba.1849-5664-5419.2014.41.1003.
5. Bianchi, D., y Dickerson, A. (2018). *Trading Volume in Cryptocurrency Markets*. SSRN Electronic Journal. doi:10.2139/ssrn.3239670.

6. Castro, J. y Forero, E. (2018). *Comparando la rentabilidad y el riesgo de los commodities y el bitcoin ¿qué opción es más rentable y de menor riesgo?*. (Tesis de postgrado). Corporación Universitaria de Dios, Bogotá.
7. Cocco, L., Concas, G., y Marchesi, M. (2015). *Using an artificial financial market for studying a cryptocurrency market*. *Journal of Economic Interaction and Coordination*, 12(2), 345-365. doi:10.1007/s11403-015-0168-2.
8. CoinMarketCap. (2019). *Learn about Cryptocurrency*. Consultado en agosto 6, 2020, de: <https://coinmarketcap.com/alexandria/#blog>.
9. Conrad, C., Custovic, A., y Ghysels, E. (2018). *Long- and Short-Term Cryptocurrency Volatility Components: A GARCH-MIDAS Analysis*. *SSRN Electronic Journal*. doi:10.2139/ssrn.3161264.
10. Cryer, J., y Chan, K. (2008). *Time series analysis with applications in R*. New York: Springer.
11. Dodge, Y. (2010). *The concise encyclopedia of statistics*. New York, NY: Springer.
12. Francés, C. (2018). *The cryptocurrency market: A network analysis*. *ESIC Market Economics and Business Journal*, 49(3), 569-606. doi:10.7200/esicm.161.0493.4.

13. Gandal, N., y Halaburda, H. (2016). *Can We Predict the Winner in a Market with Network Effects? Competition in Cryptocurrency Market*. SSRN Electronic Journal. doi:10.2139/ssrn.2832836.
14. Hayes, A. (2015). *Cryptocurrency Value Formation: An Empirical Analysis Leading to a Cost of Production Model for Valuing Bitcoin*. SSRN Electronic Journal. doi:10.2139/ssrn.2648366.
15. Hileman, G., y Rauchs, M. (2017). *2017 Global Cryptocurrency Benchmarking Study*. SSRN Electronic Journal. doi:10.2139/ssrn.2965436.
16. Klenke, A. (2014). *Probability theory: A comprehensive course*. London: Springer.
17. Kuhn, M., y Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY: Springer New York.
18. Liu, Y., y Tsyvinski, A. (2018). *Risks and Returns of Cryptocurrency*. doi:10.3386/w24877.
19. Mendenhall, W., Scheaffer, R. L., y Wackerly, D. D. (2010). *Estadística matemática con aplicaciones*. México: Grupo Editorial Iberoamérica.
20. Perz, P., y Gemzik-Salwach, A. (2020). *Cryptocurrencies as A Subject of Financial Investments. Risk Analysis and Potential Benefits On the Example of Bitcoin*. Acta Scientiarum Polonorum. Oeconomia, 19(2), 15-23. doi:10.22630/aspe.2020.19.2.13.

21. Rachlin, R. (2019). *How to Use Ratios to Monitor ROI Performance*. Return on Investment Manual, 38-48. doi:10.4324/9781315503813-3.
22. Radovanov, B., Marcikić, A., y Gvozdenović, N. (2018). *A Time Series Analysis Of Four Major Cryptocurrencies*. *Facta Universitatis, Series: Economics and Organization*, 271. doi:10.22190/fueo1803271r.
23. Rauchs, M. (2017). *Global cryptocurrency benchmarking study*. United Kingdom. Recuperado de: <https://www.jbs.cam.ac.uk/wp-content/uploads/2020/08/2017-04-20-global-cryptocurrency-benchmarking-study.pdf>.
24. Rincon, J., y Bareño, E. (2018). *Comparando la Rentabilidad y el Riesgo de los Commodities y el Bitcoin ¿Qué opción es más rentable y de menor Riesgo? (Master's thesis, Universidad Minuto de Dios, 2018)*. Bogotá: Corporación Universitaria Minuto de Dios.
25. Salwach, A. y Perz, P. (2020). *Cryptocurrencies as a subject of financial investments. Risk analysis and potential benefits on the example of bitcoin*. Recuperado de: [https://www.researchgate.net/publication/342844857\\_CRYPTOCURRENCIES\\_AS\\_A\\_SUBJECT\\_OF\\_FINANCIAL\\_INVESTMENTS\\_RISK\\_ANALYSIS\\_AND\\_POTENTIAL\\_BENEFITS\\_ON\\_THE\\_EXAMPLE\\_OF\\_BITCOIN](https://www.researchgate.net/publication/342844857_CRYPTOCURRENCIES_AS_A_SUBJECT_OF_FINANCIAL_INVESTMENTS_RISK_ANALYSIS_AND_POTENTIAL_BENEFITS_ON_THE_EXAMPLE_OF_BITCOIN).

26. Suárez, J., y Bautista, J. (2016). *Criptodivisas en el entorno global y su incidencia en Colombia*. *Revista Le Bret*, 0(8), 151. doi:10.15332/rl.v0i8.1691.
27. Sovbetov, Y. (2018). *Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, LiteCoin, and Monero*. *Journal of Economics and Financial Analysis*, , 1-27.
28. Tsyvinski, A. (2018). *Risks and returns of cryptocurrency*. Recuperado de: [https://www.nber.org/system/files/working\\_papers/w24877/w24877.pdf](https://www.nber.org/system/files/working_papers/w24877/w24877.pdf).