Universidad Nacional Autónoma de México

Facultad de Ciencias

Estadística 3

Dra. Lizbeth Naranjo Albarrán
Axel Arellano Carmona
Diego Morales González
León Diego Rivera Hernández
Sergio Enrique Rodríguez

Proyecto Series de Tiempo

Índice

1. Introducción

Las series de tiempo tienen muchas aplicaciones para muchas áreas del conocimiento, puesto que es una herramienta muy fuerte para predecir y son de ayuda con distintos problemas. Un ejemplo puede ser el querer saber si una inversión en un activo generará ganancias o pérdidas más si dicha inversión es una de alto riesgo; por ello los estadistas han desarrollado herramientas con las cual podemos tratar de predecir el futuro y ver si nuestras inversiones futuras darán sus frutos o no.

En el siguiente escrito se explicará y tratará de estimar los rendimientos de una serie de tiempo utilizando los métodos más conocidos para modelar series de tiempo. También se mostrarán los problemas de la serie temporal y las limitantes de los modelos ajustados.

1.1. Modelos ARIMA

Uno de los principales modelos de series de tiempo es el modelo ARIMA (*autoregressive integrated moving average*) que junta la parte autorregresiva (AR(p)) que tienen las series de tiempo con la parte de promedio móviles (MA(q)) y el número de diferencias necesarias para llegar al ruido blanco (d) (ver ecuación (1)). Este modelo encuentra patrones usando polinomios de retrasos y determinando sus coeficientes. De esta forma, es que se tiene un modelo dinámico para muchos tipos de series de tiempo que tienen tendencia, ciclos y ruido blanco que usa los datos del pasado y no por variables independientes.

$$\left(1-\sum_{i=1}^{p}\Phi_i B^i\right)(1-B)^d X_t = \delta + \left(1+\sum_{1=i}^{q}\theta_i B^i \epsilon_t\right) \tag{1}$$

1.2. Modelos ARCH

En ocasiones no es posible estabilizar la varianza debido a factores que son muy difíciles de predecir y que causan comportamientos muy repentinos. Este tipo de comportamiento no es posible modelarlo con un modelo AR(p), MA(q) o ARIMA(p,d,q) ya que en este tipo de modelos clásicos una de las premisas principales es tener una varianza constante (condicional como incondicional) en la serie de tiempo.

Para ello existe otro tipo de modelos donde la volatilidad juega un papel primordial, uno de los más conocidos es el modelo con heterocedasticidad condicional autorregresiva, o por sus siglas en inglés ARCH.

En algunas series de tiempo que no son estacionarias y suelen tener algún tipo de tendencia se realizan determinadas diferenciaciones a la serie. Normalmente suelen ser una o dos para no diferenciar de más.Primero se obtiene el logaritmo de los datos, para evitar la volatilidad. Luego de esto, se toman diferencias a resultados obtenidos.

A lo anterior se le conoce como rendimiento o rendimiento logarítmico. Suele presentar distintas ventajas y desventajas, por ejemplo:

- Es un modelo sencillo.
- El modelo puede producir grupos de volatilidad.
- Es probable que los modelos ARCH pronostiquen en exceso la volatilidad al necesitar muchos parámetros para ajustarse y que algunos sean insignificantes.

Hay maneras de verificar que se trabaja con una serie de tiempo ARCH. Para detectar la presencia de efectos ARCH se pueden utilizar los estadísticos tradicionales de Ljung-Box en los correlogramas de los residuos de la ecuación de la media y de los residuos al cuadrado.

A pesar de su simpleza se suelen usar más parámetros para ajustar un modelo, lo cual en la mayoría de los casos suele llevar a predicciones erróneas. Lo que inclina a investigar el siguiente modelo.

1.3. Modelos GARCH

Los modelos GARCH pueden ser vistos como una generalización de los modelos ARCH. Estos permiten una representación con mayor parsimonia de la volatilidad; más aún, dependen de las observaciones pasadas, depende también de su propio pasado.

En éstos modelos la estructura de la varianza condicional depende, además, del cuadrado de los errores retrasados q períodos como en el modelo ARCH(q), de las varianzas condicionales retrasadas p períodos.

Definición 1. *Sea r_t una serie de log retornos y $a_t = r_t - \mu$ entonces el proceso $\{a_t\}$ es llamado GARCH(p,q) si:*

$$a_t = \sigma_t \eta_t,\ \eta_t \sim IIDN(0,1) \tag{2}$$

Donde

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{p} \alpha_i a_{t-i}^2,\ \alpha_0 > 0,\ \alpha_i \geqslant 0,\ \beta_j \geqslant 0 \tag{3}$$

$$\sum_{i=1}^{max\{p,q\}} (\alpha_i + \beta_i) \tag{4}$$

Lo anterior para garantizar que la varianza sea positiva y existan los momentos de orden superior. De (4) *obtenemos que la varianza marginal a_t es finita, mientras que su varianza condicionada σ_t^2 varía en el tiempo.*

Para verificar si el modelo que hay que ajustar es un ARCH o GARCH se debe ajustar un modelo ARIMA previamente. Si los residuales, su valor absoluto y su cuadrado presentan correlación, es posible empezar a intuir que se debería de ajustar en su lugar uno de los modelos ya antes mencionados.

Las pruebas no paramétricas y gráficas que se usaron en el siguiente trabajo fueron:

- Anderson Darling para normalidad.
- Breusch Pagan para varianza constante.
- Ljung-Box para no correlación.
- ARCH LM test para determinar si se puede representar el modelo con un GARCH o ARCH
- Comparamos en una recta los valores de la normal con los residuos.
- Las gráficas del ACF y PACF para verificar correlación.
- Los criterios de selección de modelos como el criterio de Información de Akaike (AIC), el criterio de Información Bayesiano (BIC) y finalmente el criterio de Información de la Devianza (DIC).

1.4. Datos

En el siguiente trabajo se consideró el conjunto de datos del precio mensual de la plata en mxn durante los últimos 25 años, consta de 256 mediciones realizadas mensualmente, que inician en enero de 1995 y finalizan en febrero de 2020.

Los datos presentan las características siguientes:

- Es una serie sin ciclos.
- Sin varianza constante.

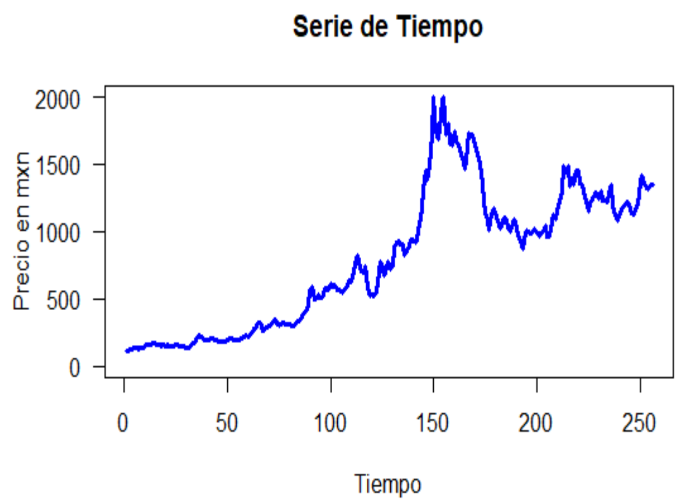

Figura 1: Representación gráfica de los datos.

- No es estacionaria.
- No hay datos faltantes.
- Se puede observar que existe volatilidad en la serie (ver figura 1).

Como se acostumbra en finanzas es más eficaz estimar los rendimientos aplicando el logaritmo para así tener una resta. Por lo tanto, los rendimientos se verán de la siguiente forma: $ln(p_t) - ln(p_{t-1})$ (ver figura 2).

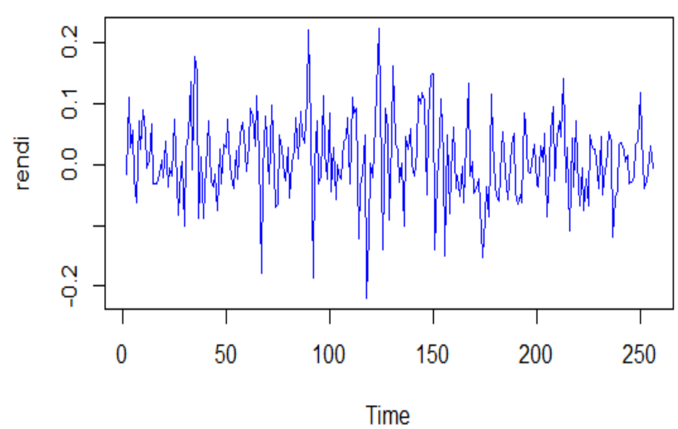

Figura 2: Representación gráfica de los rendimientos, después de haber transformado los datos aplicando el logaritmo.

De la figura 2 se puede observar que la varianza no es constante, que tampoco tiene ciclos y además que no es estacionaria.

1.5. Volatilidad en 2010

Se puede apreciar un crecimiento repentino cerca del dato 150,alrededor del año 2010.

Países de la Unión Europea con estrategias de economía pública poco planeada empezaron a crear una deuda incontrolable, causando que los países con economías más estables tuvieran que prestarles, devaluando el precio del euro. De esta forma, se empezó a tener mucha incertidumbre respecto a la volatilidad del euro y los inversionistas empezaron a interesarse más en *commodities* como el oro y la plata. Esto, junto con políticas problemáticas de China, y EU todavía recuperándose de su crisis del 2008 es que se pronosticó el aumento en el precio de la plata, pero también se pronosticó su caída en cuanto se estabilizara la situación[5].

2. Ajuste de Modelos

De lo anterior, es seguro concluir que ajustar algún modelo de tipo Holt-Winters no lineal o SARIMA no es la mejor opción.

Tratar de ajustar un modelo ARIMA tampoco resulta buena opción (ver figuras 3 y 4), debido a que se intentó estabilizar la varianza por tres métodos distintos (logaritmo, método Guerreo y Loglike) sin resultados satisfactorios. Por lo tanto, se debe de intentar diferentes técnicas para intentar controlar la varianza o predecirla, puesto que justamente con este tipo de activos es de interés saber el riesgo de cada inversión.

```
Coefficients:
          ar1      ar2     ma1     ma2   xmean
      -0.7995  -0.6578  1.0130  0.7747  0.0097
s.e.   0.1679   0.1110  0.1494  0.1105  0.0046

sigma^2 estimated as 0.004246:  log likelihood = 334.42,  aic = -656.83

$degrees_of_freedom
[1] 250

$ttable
      Estimate     SE t.value p.value
ar1    -0.7995 0.1679 -4.7621   0.000
ar2    -0.6578 0.1110 -5.9237   0.000
ma1     1.0130 0.1494  6.7822   0.000
ma2     0.7747 0.1105  7.0079   0.000
xmean   0.0097 0.0046  2.0858   0.038

$AIC
[1] -2.575808

$AICc
[1] -2.574863

$BIC
[1] -2.492484
```

Figura 3: *Summary* del ajuste del modelo *ARMA(2,2)* a los datos.

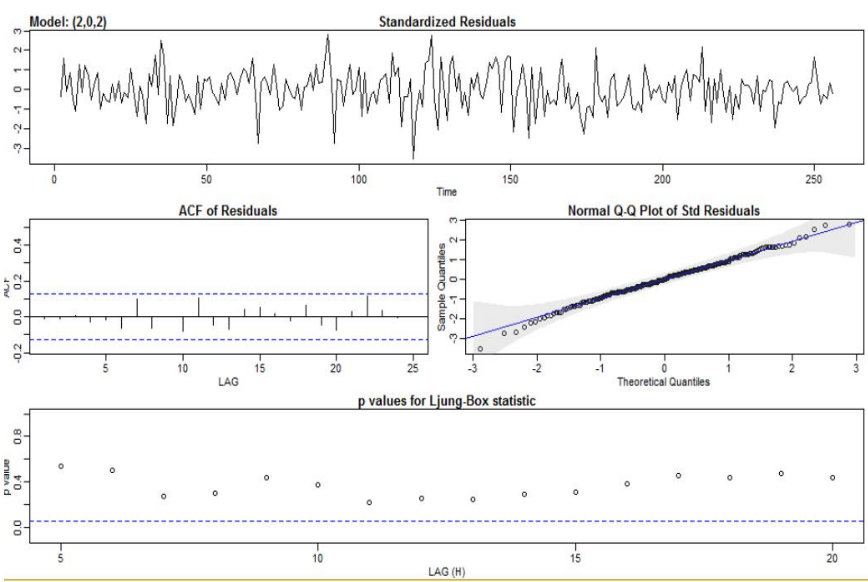

Figura 4: Varias gráficas de los residuales para el modelo *ARMA(2,2)*.

2.1. Buscando alternativas

En este caso se optó por un modelo GARCH, dado que los valores para los ACF y PACF (ver figuras 5 a 7)de los residuales, los residuales al cuadrado (ver figura 8) y el valor absoluto de los residuales (ver figura 9) se salen varios valores de las bandas de confianza.

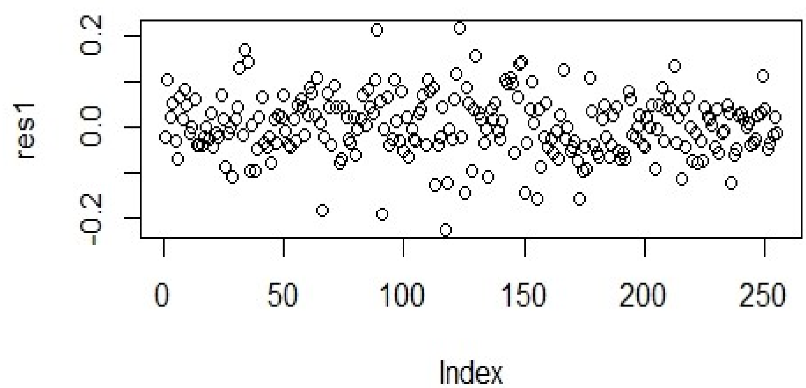

Figura 5: Gráfica de los residuales para el modelo *GARCH(2,2)*.

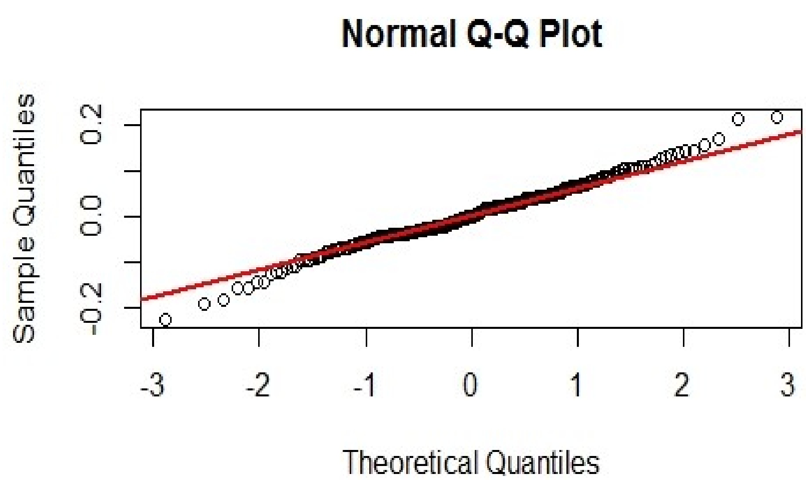

Figura 6: *QQplot* para los residuales del modelo *GARCH(2,2)*.

Ajustando varios modelos se puede observar que en su mayoría no pasan los supuestos del modelo, específicamente el supuesto donde los residuos no están correlacionados ya que en las gráficas del ACF Y PACF existen valores que salen de las bandas y revisándolas por separado usando el test de Box-Ljung se tiene el mismo resultado i.e. con los parámetros que usados no fue posible modelar toda la dependencia. A pesar de que los residuales se distribuyen normal.

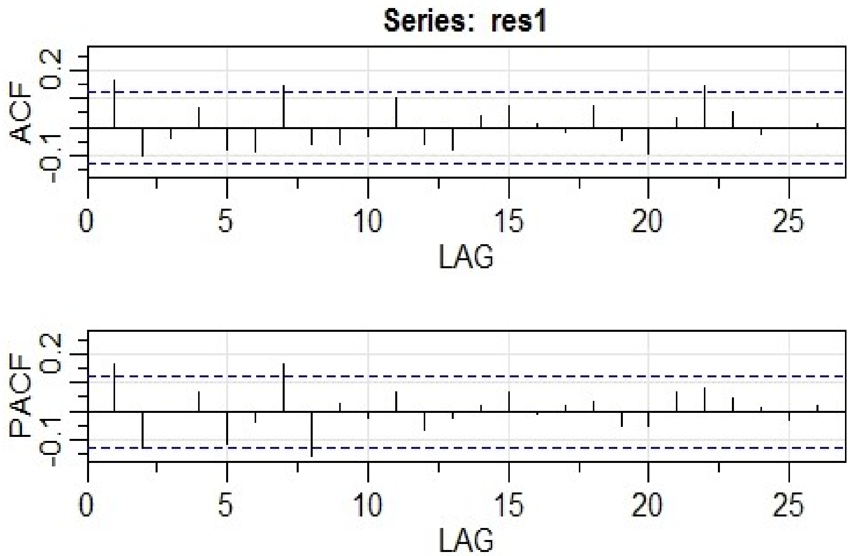

Figura 7: Gráficas de los ACF y PACF de los residuales del modelo *GARCH(2,2)*.

En conclusión, no se tiene un ruido blanco. Además, para explicar la correlación, la mayoría de los parámetros no son significativos.

2.2. Otro intento

Notando cuáles parámetros nos salían significativos y cuáles no es que intentamos encontrar el mejor. Nos fijamos que estábamos modelando los GARCH con media 0, pero justamente el comportamiento el comportamiento de la media va variando. Por lo tanto, usamos un GARCH pero no sólo eso, sino también ajustando la media con un ARMA encontramos mejores resultados.

Ahora tenemos parámetros significativos, también al probar supuestos vemos que no hay correlación en los residuales y su distribución también es normal. De aquí tenemos un ruido blanco gaussiano en los residuales. Ajustando un GARCH(1,2) con la distribución de su media siendo ARMA(2,2) llegamos a los siguientes resultados (ver cuadro 1). Obsérvese que tres de los parámetros no significativos y un valor AIC de -2.53 y un valor BIC de -2.4. Dados estos datos, se propuso que era posible encontrar un modelo aún mejor.

	Estimate	Std. Error	t value	Pr(>|t|)
\mu	0.006387	0.004704	1.3577	0.174572
ar1	-0.849210	0.173472	-4.8954	0.000001
ar2	-0.660387	0.128428	-5.1421	0.000000
ma1	1.037285	0.145327	7.1376	0.000000
ma2	0.780046	0.114013	6.8417	0.000000
omega	0.000761	0.000492	1.5458	0.122161
alpha1	0.127690	0.069568	1.8355	0.066437
beta1	0.702313	0.616533	1.1391	0.254648
beta2	0.000000	0.518539	0.518539	1.000000

Cuadro 1: Tabla con los valores de los coeficientes del modelos descrito.

Los residuales pasaron los supuestos de ruido blanco gaussiano. Notando que tenemos 3 parámetros no significativos y un AIC de -2.53 y BIC de -2.4 consideramos que podíamos encontrar un mejor modelo.

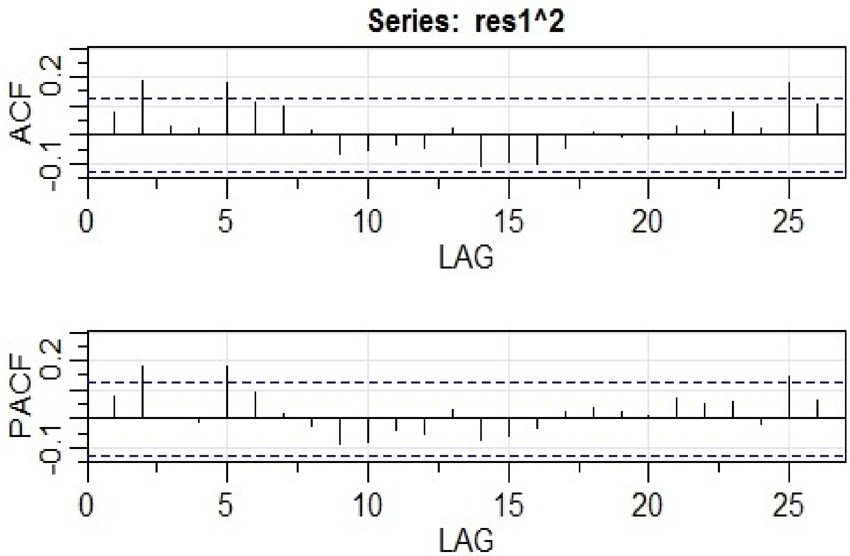

Figura 8: Gráficas de los ACF y PACF de los residuales al cuadrado del modelo *GARCH(2,2)*.

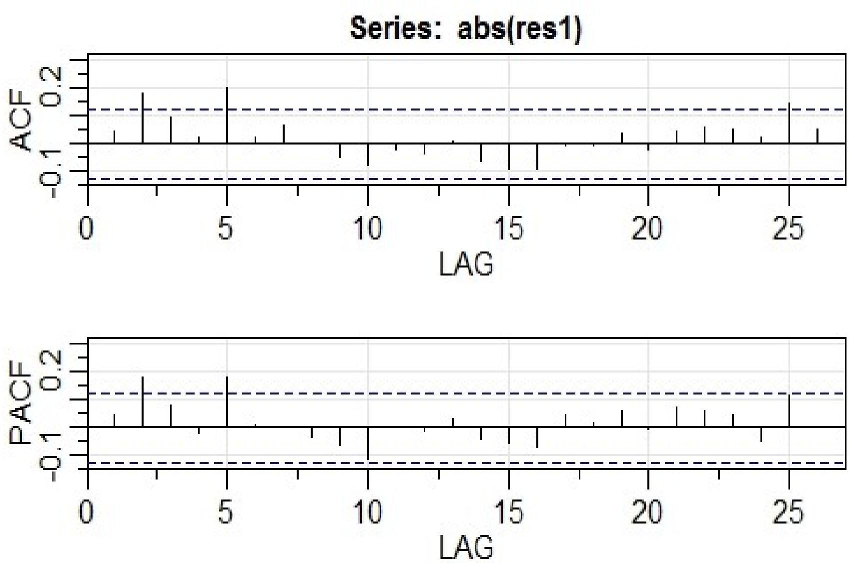

Figura 9: Gráficas de los ACF y PACF del valor absoluto de los residuales del modelo *GARCH(2,2)*.

3. Modelo Seleccionado

Considérese un modelo GARCH(1,1) con distribución de su media un ARMA(2,2); se tienen resultados más favorables. Con un AIC de -2.54 y un BIC de -2.43 siendo un poco mejores que el modelo anterior. La significancia de los parámetros fue lo que se consideró principalmente para la elección de este modelo.

	Estimate	Std. Error	t value	Pr(>|t|)
\mu	0.006387	0.004704	1.3577	0.174550
ar1	-0.849212	0.173465	-4.8956	0.000001
ar2	-0.660387	0.128054	-5.1571	0.000000
ma1	1.037286	0.145322	7.1379	0.000000
ma2	0.780046	0.113501	6.8726	0.000000
omega	0.000761	0.000375	2.0306	0.042300
alpha1	0.127690	0.067476	1.8924	0.058442
beta1	0.702313	0.104214	6.7391	0.000000

Cuadro 2: Tabla con los valores de los coeficientes del modelos definitivo seleccionado.

Teniendo solo un parámetro no significativo, es seguro afirmar que, en este modelo, todos los coeficientes utilizados son útiles.

4. Predicciones

Una vez encontrado el mejor modelo, se utilizará el mismo para predecir, habiendo removido los datos del último año, para observar qué tanto se acerca a los datos reales (ver figura 10). Determinando los intervalos de confianza y revisando el comportamiento de la varianza.

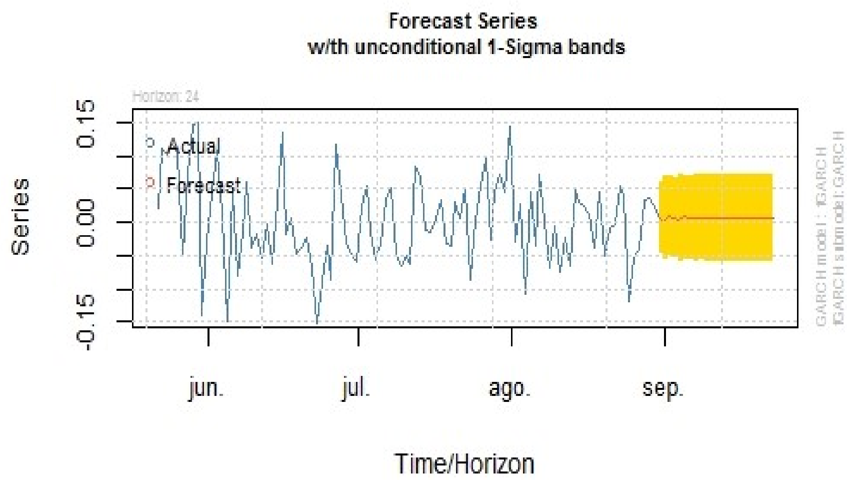

Figura 10: Gráfica de los datos con las predicciones realizadas en R con el modelo seleccionado.

El resultado de la predicción a primera vista parece ser erróneo puesto que se parece demasiado a la media, pero en verdad era lo esperado. Debe de estar equilibrada la probabilidad de ganar como de perder, de lo contrario se estaría en una situación de un mercado con arbitraje en el que la gente puede ganar más que la tasa libre de riesgo con poco riesgo. Es por esto que la estimación tiene que dar cerca del punto medio, porque lo que es de mayor interés encontrar el riesgo de la inversión a lo largo del tiempo, también conocido como la volatilidad. Por lo tanto, se grafica a continuación en la figura 11 la variabilidad a lo largo de los años.

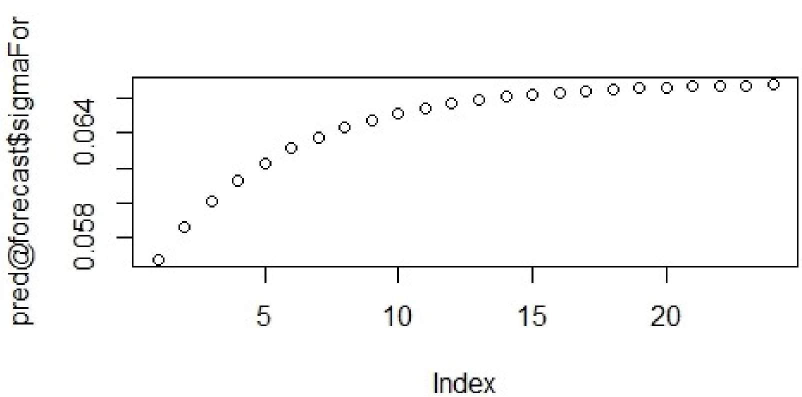

Figura 11: Gráfica de la variabilidad de los datos a lo largo de los años.

Se espera una varianza creciente los primeros meses, que se estabilizará después con un crecimiento muy paulatino. Finalmente, con el fin de verificar que tan eficaz es la predicción, se compara la serie sin los datos del último año contra la serie con los datos completos (ver figura 12).

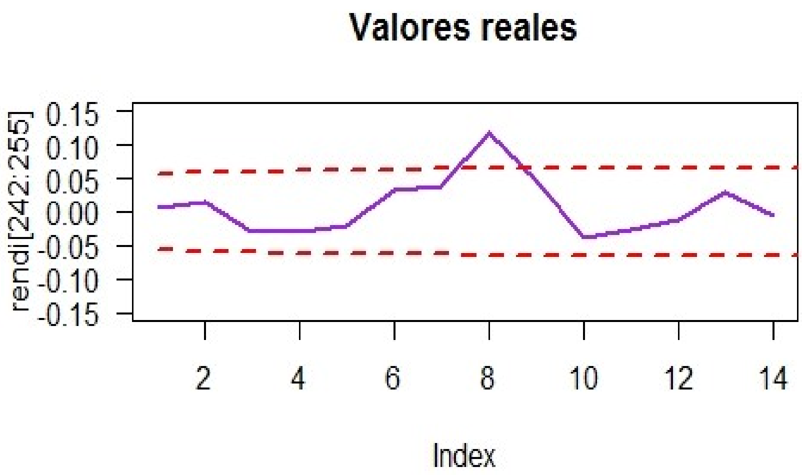

Figura 12: Comparación gráfica de lo datos predichos y los valores reales.

Solo un valor se sale de las bandas de confianza, por lo que si se estuviera interesado en invertir en plata podríamos darnos una idea casi segura de cuánto podríamos perder y ganar. De esta forma, es posible diversificar el portafolio de inversiones para que en casi ningún caso nos quedemos sin dinero y ganemos una cantidad deseable. Esto dependerá de qué tan adversos seamos al riesgo y entre más lo seamos, menores serán nuestras ganancias y pérdidas.

Al graficar la estimación con el modelo pasado GARCH con más parámetros no significativos nos salieron idénticas las gráficas por lo menos a la vista y al ver la diferencia entre los intervalos de confianza nos regresaron valores en milésimas. Por lo que parecería que para predecir no hay tanta diferencia, pero creemos que al largo plazo va a ser menos precisa la predicción por parte del modelo con parámetros no significativos ya que no tendrá suficientes coeficientes para expresar el comportamiento.

Es importante considerar que puede haber crisis como la que se está viviendo hoy en día que pueden afectar los rendimientos de una forma extrema ya sea para bien o mal y ese tipo de eventos, aunque poco probables, se saldrían por mucho de las bandas de confianza. El GARCH se ve limitado en el sentido que no puede estimar eventos tan radicales.

Referencias

[1] Ruey S. Tsay. *An introduction to analysis of financial data with R*, Wiley, Chicago, USA, 2013.

[2] Christian Francq. *GARCH models: structure, statistical inference and financial application*, Wiley, Paris, France, 2010.

[3] Cruz Oropeza R. *Modelo ARCH*, CIMAT, 2013.

[4] Quesada P. M. (2010) *Análisis de Series. Procesos Heterocedásticos.* Tesis de grado de maestría, Universidad de Granada.

[5] Mueller M. (2010) *El precio de la plata aumenta un 74% en 2010*. Recuperado el 4 de abril de 2020. https://www.oroyfinanzas.com/2010/12/el-precio-de-la-plata-aumenta-un-74-en-2010/