

# METODOLOGÍA BOOTSTRAP EN LA PREDICCIÓN DEL ÍNDICE DE LA BOLSA DE VALORES DE LIMA, PERÚ (S&P/BVL)

Juan Carlos Orosco Gavilán<sup>1\*</sup>

<sup>1</sup> Instituto Científico del Pacífico, Lima, Perú.

Recibido noviembre 2020; aceptado diciembre 2020

## Abstract

Bootstrap methodology is evaluated in heteroscedastic models applied in the prediction of the Lima Stock Exchange Index (S & P / BVL), period 2017 - 2019. Predictions were obtained using the parametric methodology from the algorithm of heteroscedastic GARCH processes with Bootstrap proposed by Pascual (2006) being this generalized for other types of heteroscedastic models and were compared with real values. The performance of both methodologies was evaluated. The models that best fit the series are the ARMA (1,1)-EGARCH (1,1) models with assumptions of the residuals with Normal and t-Student distributions with 5 degrees of freedom. The comparative study showed that the application of the Bootstrap methodology in the series of returns of the Lima Stock Exchange Index, allows prediction intervals with greater and equal amplitudes in some horizons compared to the parametric methodology. The construction of prediction intervals for volatilities were also obtained with good performance, thus, this is an alternative for their construction in heteroscedastic models.

## Resumen

La presente investigación tiene el objetivo de analizar y evaluar la metodología Bootstrap en modelos heterocedásticos aplicados en la predicción del Índice de la Bolsa de Valores de Lima (S&P/BVL), periodo 2017 – 2019. Las predicciones se obtuvieron mediante la metodología paramétrica y la metodología Bootstrap, y fueron comparados con valores reales, partir del algoritmo de procesos heterocedásticos GARCH con Bootstrap propuesto por Pascual (2006), siendo éste generalizado para otros tipos de modelos heterocedásticos. Los desempeños de ambas metodologías fueron evaluados. Del estudio se obtuvo que los modelos que mejor ajustan a la serie son los modelos ARMA(1,1)-EGARCH(1,1) con supuestos de los residuales con distribuciones Normal y t-Student con 5 grados de libertad. El estudio comparativo mostró que la aplicación de la metodología Bootstrap en la serie de los retornos del Índice de la Bolsa de Valores de Lima, permite obtener intervalos de predicciones con mayores e iguales amplitudes en algunos horizontes en comparación con la metodología paramétrica. También se construyeron los intervalos de predicción para las volatilidades, siendo ésta una alternativa para su construcción en modelos heterocedásticos.

**Key words:** COVID-19, Bootstrap, Stock Exchange, time series, S&P/BVL

**Palabras clave:** Bootstrap, Bolsa de valores, series temporales, S&P/BVL

## I. INTRODUCCIÓN

La predicción es uno de los principales objetivos en la aplicación de los modelos de series temporales, y es de interés en muchas áreas de conocimiento. En los mercados financieros, la predicción es un gran desafío para la estadística, pues la respuesta correcta a preguntas como, por ejemplo, si las acciones de la bolsa de valores van a subir, o si el valor del dólar va a caer, pueden significar un gran lucro o perjuicio para los inversores individuales o empresariales. Así también el comportamiento de los mercados bursátiles y su evolución ha despertado mucho interés en estos últimos tiempos entre los agentes económicos, ello se debe a que constituyen un medio de financiamiento de la actividad productiva de las empresas a partir del ahorro de los inversionistas.

La gran mayoría de las series de tipo financieras el supuesto de homocedasticidad resulta inadecuado, ya que a lo largo del tiempo su comportamiento no es estacionario ni en media ni en varianza, generando

---

\* Autor para correspondencia: juanoroscog@icip.edu.pe

volatilidades altas para unos períodos y bajas para otros, siendo frecuente este tipo de cambios. Para esto se han desarrollado varios modelos, con el fin de capturar el efecto de la volatilidad, entre ellos se encuentran principalmente los modelos ARCH y GARCH, y sus extensiones como EGARCH, PGARCH, TGARCH (Horacio, 2009).

Por otra parte, las series de tipo financieras o retornos presentan características propias, llamados comúnmente hechos estilizados, una de estas y las más importante es el comportamiento no Normal de dichas series, la cual genera un problema para la estadística clásica o paramétrica. Una alternativa para solucionar el problema dada las características propias de los retornos es obtener intervalos de predicción utilizando procedimientos bootstrap, los cuales no requieren suponer una distribución a priori para los errores. Los métodos estadísticos clásicos se apoyan en modelos matemáticos de naturaleza estocástica, de tal forma que los resultados que de ellos se derivan, requieren en muchas ocasiones, desarrollos analíticos complejos, lo que ha supuesto un obstáculo para su utilización comprensiva en muchas áreas científicas, básicamente para utilizar aquellos modelos estadísticos que se ajusten a características de los datos de cada disciplina. En términos más precisos, los modelos estadísticos incorporan distintos supuestos que establecen restricciones sobre las variables aleatorias analizadas.

En la actualidad los procedimientos de remuestreo son ampliamente conocidos y utilizados en diversos campos de la investigación y se han ido incorporando como técnicas bastante habituales en el análisis e interpretación de cierto tipo de datos. El método bootstrap o técnica del remuestreo, reemplaza las derivaciones teóricas del enfoque ya dicho por la evaluación de los estadísticos en submuestras obtenidas a partir de los datos originales, y mediante estos valores se obtienen estimadores de las medidas de exactitud o de la distribución muestral del estadístico. Pascual (2006) propusieron un método bootstrap para el cálculo de intervalos de predicción en modelos GARCH, obteniendo buenos resultados y siendo una alternativa para el cálculo de intervalos de predicción paramétrico, siendo este método generalizable a otros modelos heterocedásticos.

En este trabajo se realizó una aplicación de la metodología bootstrap a una serie temporal de tipo financiera que es el retorno del Índice de la Bolsa de Valores de Lima (S&P/BVL), comparando los resultados con la metodología paramétrica de predicción de intervalos. El índice trabajado es el de mayor representatividad dentro del mercado bursátil, que incluyen las empresas más importantes del país que cotizan en el Índice de la Bolsa de Valores de Lima.

En el Perú no existe un estudio acerca de la Bolsa de Valores de Lima para los retornos y volatilidades mediante la metodología bootstrap y a nivel mundial no es común el uso de esta metodología en la predicción de los índices bursátiles. Este estudio pretende aportar al análisis de los índices bursátiles de la Bolsa de Valores de Lima y de cualquier otra, dando una alternativa de método al estudio de predicciones en el mercado financiero.

## II. MATERIALES Y MÉTODOS

### 2.1 Principio de Bootstrap

En la formulación formal del bootstrap inicialmente propuesto por Efron (1993) se considera  $X_1, X_2, \dots, X_n$  una secuencia de variables aleatorias i.i.d con distribución común  $F$  desconocida y de valor medio  $\mu$  y de varianza  $\sigma^2$  finitos, pero desconocidos.

Así, sea  $(x_1, x_2, \dots, x_n)$  una realización de  $(X_1, X_2, \dots, X_n)$ . Una muestra  $(x_1, x_2, \dots, x_n)$  que va a desempeñar el papel de la “población Bootstrap”, de esta se extraerán con reposición, un grande número,  $B$ , de muestras, que designamos de “muestras Bootstrap” o “remuestras Bootstrap”. El objetivo es estimar la función de distribución de la población  $F$  o algún aspecto particular de esta.

### 2.2 Procesos Heterocedásticos

Los procesos heterocedásticos consisten en además de modelar la media, modela la varianza de una variable dependiente en función de los valores pasados de la propia variable; el proceso más simple de heterocedasticidad es el ARCH que viene definido por dos expresiones: en la primera se modela la esperanza condicionada de la serie en función de la varianza y de un término de error, y en la segunda ecuación es la varianza condicionada la que es modelada en función de sus propios valores pasados.

Una generalización natural de los modelos ARCH, donde la varianza condicional es también función de sus propios valores desfasados, es el modelo GARCH (generalizado ARCH) sugerida por Bollerslev (1986). Así una serie  $y_t$  es llamado de proceso GARCH(p,q) si satisface las siguientes ecuaciones:

$$y_t = \sigma_t \varepsilon_t$$

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

Donde:  $\omega > 0$ ,  $\alpha_i > 0$  y  $\beta_j > 0$ .  $\varepsilon_t$ , es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza uno, e independiente de  $y_{t-1}$  para  $i > 1$ . En la práctica usualmente se supone que  $\varepsilon_t \sim N(0,1)$  o  $\varepsilon_t \sim t_v$  (distribución t de Student con v grados de libertad).

A partir de los procesos ARCH y GARCH se planearon otros tipos de modelos con características propias como el proceso EGARCH y TGARCH que asumen que los errores positivos y negativos tienen un efecto simétrico en la volatilidad, estos modelos permiten que la volatilidad sea más afectada por retornos pasados negativos que de los positivos. Una ventaja en este modelo para garantizar la varianza positiva es la inexistencia de la restricción de que los parámetros sean positivos (Morettin, 2011).

En la presente investigación se obtuvo los modelos heterocedásticos de los retornos del Índice de la Bolsa de Valores de Lima (S&P/BVL), que es un indicador que mide el comportamiento del mercado bursátil y sirve para establecer comparaciones respecto de los rendimientos alcanzados por los diversos sectores participantes en la Bolsa de Valores Lima, y está conformada por un conjunto de valores que concentra el 80% de la negociación del mercado.

Una serie de retornos es definida como la diferencia de los logaritmos de una serie de tiempo financiera. Para  $t = 1, 2, \dots, T$  donde T es el total número de observaciones. Los retornos diarios del índice son denotados como:

$$r_t = \ln \left( \frac{p_t}{p_{t-1}} \right)$$

y la varianza del retorno es referido como la volatilidad de  $r_t$ .

Kosapattarapim (2013) menciona que los hechos estilizados de los retornos son en general propiedades de retornos financieros que son aceptados como verdaderos. Estudiados empíricamente muestran que estas propiedades son consistentes.

Los hechos estilizados relativos a retornos financieros pueden ser resumidos como:

- Los retornos son no correlacionados serialmente, o presentan en general pequeña autocorrelación.
- Los cuadrados de los retornos están autocorrelacionados.
- Series de retornos presentan conglomerados de volatilidades a lo largo del tiempo.
- La distribución (incondicional) de los retornos presenta colas más pesadas que una distribución normal, y la distribución es en general leptocúrtica (distribución no normal).
- En muchos casos, la distribución condicional también presenta colas pesadas.

### 2.3 Metodología aplicada al Índice S&P/BVL

Los datos fueron obtenidos del mercado bursátil de los índices históricos de la Bolsa de Valores de Lima, del período 2017 al 2019. La población estuvo conformada por los precios diarios de cierre de activos del mercado bursátil de la bolsa de valores de Lima, del 1 de enero del 2017 al 31 de diciembre del 2019, con un total de 781 observaciones. Las muestras fueron divididas en muestra de entrenamiento y la muestra de prueba, la primera conformada por los precios diarios de cierre de activos del mercado bursátil de la bolsa de valores de Lima, del 1 de enero del 2017 al 30 de noviembre del 2019, con un total de 761 observaciones; y la segunda conformada por los precios diarios de cierre de activos del mercado bursátil de la bolsa de valores de Lima, del 1 de diciembre del 2019 al 31 de diciembre del 2019, con un total de 20 observaciones.

Cada modelo obtenido fue calculado a partir de los datos de entrenamiento, y para poder realizar las evaluaciones respecto a la predicción en cada modelo, se realizó la comparación con los datos de prueba.

Los pasos que se realizaron para poder cumplir con los objetivos de la presente investigación fueron los propuestos por Box y Jenkins (1970), que consiste en el análisis exploratorio, identificación de modelos, estimación, diagnóstico del modelo, y finalmente la predicción. Una vez determinado el mejor modelo se realizaron las predicciones paramétricas y bootstrap. Este proceso se detallada de manera resumida en los siguientes pasos:

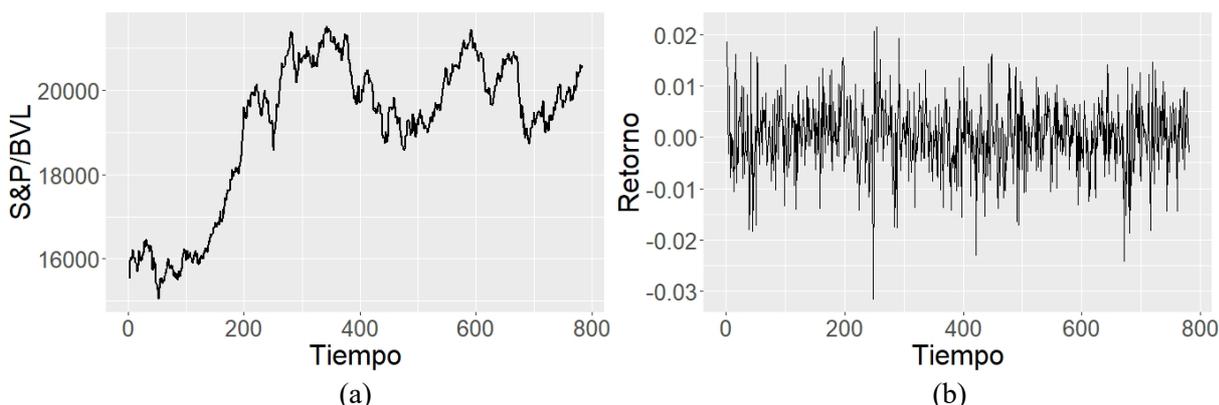
- Las predicciones para los retornos y las volatilidades fueron calculadas para el mes de diciembre del 2019 con un total de 20 observaciones,  $y_{t+k}$ , con  $k = 1, 2, \dots, 20$ .
- Los intervalos de predicción fueron calculados en los niveles de confianza del 90% y 95%.
- Para los intervalos de predicción bootstrap se consideraron muestras bootstrap de tamaño  $B=3000$ .
- Para los intervalos de predicción paramétricos se asumió que los residuales  $\varepsilon_t$  provienen de una distribución Normal con media cero y varianza 1, y distribución t de Student con 5 grados de libertad.
- Para evaluar los intervalos de predicción bootstrap e intervalos de predicción paramétrico para los retornos, fueron evaluados sus coberturas respecto a sus valores verdaderos del mes de diciembre 2019 y sus amplitudes de intervalos. Estas evaluaciones fueron realizadas a partir de los gráficos de series y sus respectivos intervalos de predicción.
- Para los intervalos de predicción de las volatilidades, se consideraron como las volatilidades reales a las volatilidades estimadas por los modelos.

Se uso el *software R* con los paquetes *rugarch*, *forecast*, *tseries*, *timeSeries* y *ggplot2* para obtener los pronósticos y la realización de las gráficas.

### III. RESULTADOS Y DISCUSIÓN

#### 3.1 Análisis exploratorio

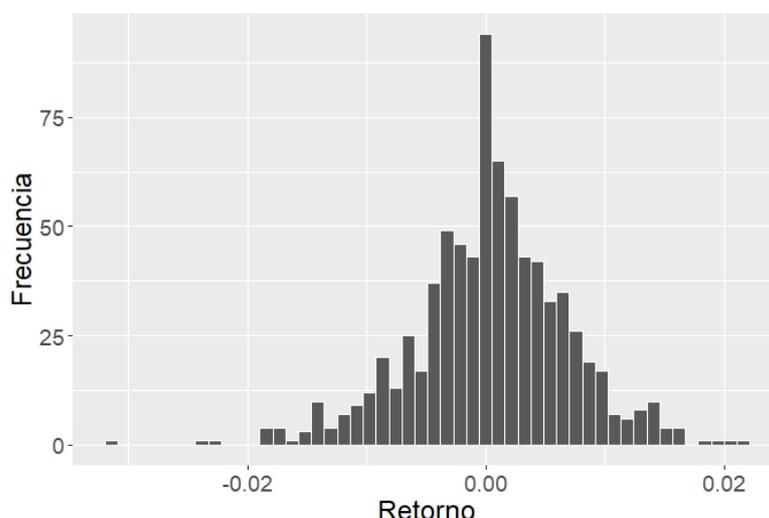
De la Figura 1(a) de los índices de precios, se observaron la presencia de alzas significativas durante los últimos seis meses del año 2017 y caídas y alzas inesperadas en el 2018 y 2019. En la evolución de los retornos de la figura 1(b) la evolución de los retornos  $r_t$  se observaron agrupamientos de las volatilidades a lo largo del tiempo, las cuales hacen referencia a las oscilaciones que presentan sus cotizaciones, y también que no presenta evidencia de no estacionariedad, por la prueba de raíz unitaria de Dickey y Fuller con valor del estadístico  $\tau = -24.563$  y de probabilidad  $p = 0,000$ , rechazando la hipótesis de existencia de raíces unitarias en la serie, es decir, la serie de retornos  $r_t$  es estacionaria.



**Figura 1.** (a) Evolución diaria de los precios de cierre y (b) Evolución de los retornos diarios de los precios de cierre, del Índice S&P/BVL del 1 de enero de 2017 al 31 de diciembre de 2019.

De la figura 2, se observó que los retornos se encuentran alrededor de cero, y la presencia de valores apartados de la parte central de las distribuciones, es decir, colas pesadas o largas. El coeficiente de asimetría con valor  $s = -0.3246$ , indicando que la distribución de los retornos de la media está sesgada ligeramente

hacia la izquierda, el coeficiente de curtosis con valor  $k = 4.3024 > 3$ , indicando que los retornos siguen una distribución leptocúrtica, dando indicios de los hechos estilizados presentes en el retorno del Índice S&P/GVL.



**Figura 2.** Histograma de frecuencias de los retornos diarios de los precios de cierre del S&P/BVL de 1 de enero de 2017 al 31 de diciembre de 2019.

Se observó mediante la prueba de autocorrelación de Ljung y Box, que la serie de retornos  $r_t$  con valores de estadístico  $Q$  grandes y de probabilidad menores a cero para todos los rezagos y a un nivel de significación de 0,05, la existencia de correlación serial, por tanto fue necesario la construcción de un modelo ARMA(p,q) para remover dicha correlación, se obtuvo que el modelo apropiado para  $r_t$  es un ARMA(1,0), definido por:

$$r_t = \phi_1 r_{t-1} + b_t$$

Donde:  $b_t$  es el residuo del modelo.

Por la prueba de autocorrelación de Ljung y Box para  $b_t$ , se obtuvo que con valores del estadístico  $Q$  y de probabilidad  $p$  los niveles de rezagos 1, 2, 3, 5, 10 y 20, con un nivel de significación de 0,01, no se rechaza la hipótesis nula de no correlación de los residuos, de este modo se eliminó la correlación serial para  $r_t$ . Para observar si  $r_t$  presenta heterocedasticidad condicional se analizó la serie  $b_t^2$ , mediante la prueba de autocorrelación para  $b_t^2$  de Ljung y Box, con valores del estadístico  $Q$  y de probabilidad  $p$  en todos los niveles de rezagos 1, 2, 3, 5, 10 y 20, con un nivel de significación del 0,01 se rechazó la hipótesis de existencia de autocorrelación para  $b_t^2$ , es decir se observó la presencia de heterocedasticidad condicional en  $b_t$ , justificando así el modelamiento por un proceso heterocedástico.

### 3.2 Identificación de modelos

Se determinó que el mejor ajuste para la serie viene dado por el modelo ARMA(1,0)-EGARCH(1,1) con distribuciones para sus residuales Normal y t-Student, con valores de criterios de informaciones menores a los demás modelos, y considerando el modelo más parsimonioso.

### 3.3 Estimación

Se observó que en el modelo ajustado, todos los parámetros estimados son significativos a un nivel de confianza del 90%, los valores de la estimación del modelo ARMA(1,1)-EGARCH(1,1) a la serie de retornos y su volatilidad son presentados por las siguientes ecuaciones.

Suponiendo que  $\varepsilon_t \sim N(0,1)$  se tiene

$$r_t = -0,147232D + 0,4919148r_{t-1} - 0,302002b_{t-1} + b_t$$

$$\log(\sigma_t^2) = -0,3658 + 0,1865\varepsilon_{t-1} - 0,0484[|\varepsilon_{t-1}| - E(|\varepsilon_{t-1}|)] + 0,9748\log(\sigma_{t-1}^2)$$

Suponiendo que  $\varepsilon_t \sim t(5)$  se tiene

$$r_t = -0,1373D + 0,6772r_{t-1} - 0,5620b_{t-1} + b_t$$

$$\log(\sigma_t^2) = -0,4556 + 0,2068\varepsilon_{t-1} - 0,0497[|\varepsilon_{t-1}| - E(|\varepsilon_{t-1}|)] + 0,9658\log(\sigma_{t-1}^2)$$

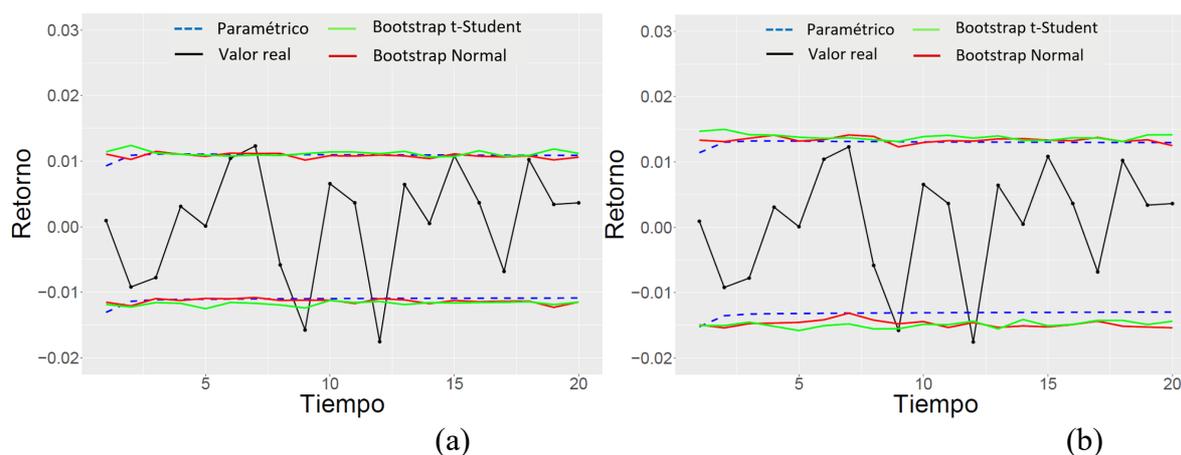
### 3.4 Diagnóstico

Evaluando los residuales de los modelos ARMA(1,1)-EGARCH(1,1) con  $\varepsilon_t \sim N(0,1)$  y  $\varepsilon_t \sim t(5)$  se determinó la no correlación serial en los residuos y residuos al cuadrado, indicando que los modelos para la media y la varianza son adecuados. Así también, se observó que los residuales del modelo presentan un coeficiente de curtosis mayor a tres y la presencia de colas pesadas diferenciándose con un comportamiento normal estándar o t-student, sin embargo, se observó también que considerando a los residuales como una distribución t-Student con 5 grados de libertad es posible capturar el efecto de colas pesadas en el modelo. Así se determinó que el modelo ARMA(1,1)-EGARCH(1,1) con  $\varepsilon_t \sim t(5)$  presenta mejor comportamiento en comparación de ARMA(1,1)-EGARCH(1,1) con  $\varepsilon_t \sim N(0,1)$ . Esto de acuerdo a sus criterios de información calculados (AIC y BIC).

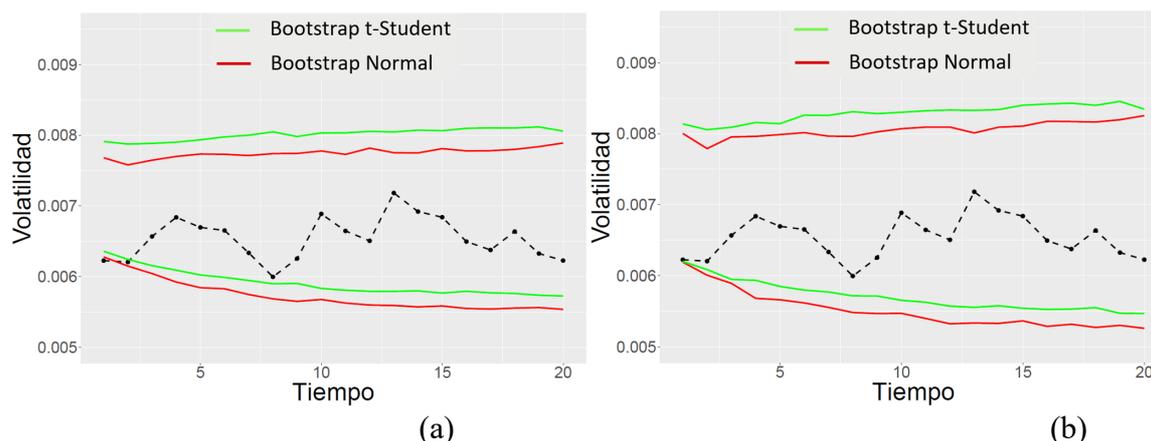
### 3.5 Predicción

En la Figura 3 se muestran los intervalos de predicción bootstrap y paramétrico para el modelo ARMA(1,1)-EGARCH(1,1) de los retornos del Índice S&P/BVL, a un nivel de confianza del 90% se observaron que los límites de predicción para los retornos con ambas metodologías no cubren todos los valores reales, específicamente en los horizontes  $h=7,9$  y  $12$ , y también mediante la metodología bootstrap se observó que presentan similares amplitudes y mayores amplitudes en algunos horizontes respecto a los intervalos paramétricos en ambos supuestos de los residuales. A un nivel de confianza del 95%, se observó que los límites de predicción para los retornos con ambas metodologías no cubren todos los valores reales en los horizontes  $h=9$  y  $12$  y mediante la metodología bootstrap se observaron que presentan mayores amplitudes en la mayoría de los horizontes, respecto a los intervalos bootstrap, esto en cada supuesto de los residuales.

En la Figura 4 se observó que los intervalos de predicción para las volatilidades del modelo ARMA(1,1)-EGARCH(1,1) a niveles de confianza del 90% y 95% cubren todos los valores reales de las volatilidades, y con el modelo con  $\varepsilon_t \sim t(5)$  se observó que presentan ligeramente mayores amplitudes en la mayoría de los horizontes de intervalo de predicción para las volatilidades de los retornos.



**Figura 3.** Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-EGARCH(1,1) con  $\varepsilon_t \sim N(0,1)$  y  $\varepsilon_t \sim t(5)$ , de los retornos de los precios de las acciones del Índice S&P/BVL, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2019.



**Figura 4.** Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-EGARCH(1,1) con  $\varepsilon_t \sim N(0,1)$  y  $\varepsilon_t \sim t(5)$ , de las volatilidades de los retornos de los precios de las acciones del Índice S&P/BVL, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2019.

#### IV. CONCLUSIONES

El orden de los modelos de predicción que se ajustan adecuadamente a los retornos diarios de los precios del Índice S&P/BVL fue el modelo ARMA(1,1)-EGARCH(1,1), Obteniendo mejores resultados con el supuesto de distribución de tipo t-student con 5 grados de libertad de los residuales en cada modelo.

Comparando sus coberturas de intervalos de predicción, se obtuvieron que los intervalos de predicción con niveles de confianza del 90% y 95% de los modelos para los retornos con residuales normales estándar y t-student con 5 grados de libertad, mediante las metodologías bootstrap no difieren sustancialmente de la metodología paramétrica. Son adecuadas en el sentido que presentaron un conjunto de valores entorno a sus verdaderos valores.

Comparando sus amplitudes en intervalos de predicción para los retornos con niveles de confianza del 90% y 95%, se determinaron que los intervalos mediante la metodología bootstrap se desempeñan adecuadamente, obteniendo intervalos de predicción de amplitudes mayores en algunos horizontes respecto a los intervalos paramétricos. De esta manera estableciendo una alternativa para el cálculo de los intervalos de predicción para los retornos del Índice S&P/BVL.

Considerando que no existe un método paramétrico para el cálculo de intervalos de predicción para las volatilidades de los retornos. Los intervalos de predicción mediante la metodología bootstrap para las volatilidades de los retornos presentaron comportamientos satisfactorios, con respecto a que presentaron sus valores reales de las volatilidades de los retornos dentro de los márgenes de confianza.

#### V. REFERENCIAS

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
- Bolsa de Valores de Lima*. Obtenido de <http://www.bvl.com.pe/>
- Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day,.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall,.
- Horacio, F. (2009). Egarch: un modelo asimétrico para estimar la volatilidad de series financieras. *Revista Ingenierías Universidad de Medellín*, 9, 49-60.
- Kosapattarapim, C. (2013). *Improving volatility forecasting of garch models: applications to daily returns in emerging stock markets*. Doc. Ph. School of Mathematics and Applied Statistics, University of Wollongong.
- Morettin, P. A., & Toloi, C. M. (2011). *Análise de séries temporais*. São Paulo, Brasil: Bloucher.
- Pascual, L., Romo, J., & Ruiz, E. (2006). Bootstrap prediction for return and volatilities in garch models. *Computational Statistics and Data Analysis*, 50(9), 2293-2312.