

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

ESCUELA DE POSGRADO

MAESTRÍA EN ESTADÍSTICA APLICADA



**“USO DE LOS MODELOS HETEROCEDÁSTICOS CON
BOOTSTRAP EN EL ANÁLISIS DEL ÍNDICE GENERAL DE LA
BOLSA DE VALORES DE LIMA”**

Presentada por:

JUAN CARLOS OROSCO GAVILÁN

**TESIS PARA OPTAR EL GRADO DE MAGISTER SCIENTIAE EN
ESTADÍSTICA APLICADA**

Lima - Perú

2019

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

ESCUELA DE POSGRADO

MAESTRÍA EN ESTADÍSTICA APLICADA

**“USO DE LOS MODELOS HETEROCEDÁSTICOS CON BOOTSTRAP EN
EL ANÁLISIS DEL ÍNDICE GENERAL DE LA BOLSA DE VALORES DE
LIMA”**

TESIS PARA OPTAR EL GRADO DE MAGISTER SCIENTIAE

Presentada por:

JUAN CARLOS OROSCO GAVILÁN

Sustentada y aprobada ante el siguiente jurado:

**Mg. Jesús Salinas Flores
PRESIDENTE**

**Mg.Sc. Grimaldo Febres Huamán
ASESOR**

**Mg.Sc. Jaime Porras Cerrón
MIEMBRO**

**Mg.Sc. Rino Sotomayor Ruiz
MIEMBRO**

ÍNDICE GENERAL

I. INTRODUCCIÓN	1
II. REVISIÓN DE LITERATURA	5
2.1. Antecedentes de la investigación	5
2.1.1. Investigaciones acerca del análisis de la Bolsa de Valores de Lima .	5
2.1.2. Investigaciones acerca de la metodología Bootstrap en series tempo- rales	7
2.2. Series temporales	10
2.2.1. Definición de series temporales	10
2.2.2. Objetivos del análisis de series temporales	10
2.2.3. Procesos estocásticos	10
2.2.4. Procesos estacionarios	11
2.2.5. Procesos de ruido blanco	12
2.2.6. Estimación de los momentos de procesos estacionarios	12
2.3. Modelos estacionarios lineales	13
2.3.1. Procesos autorregresivos (AR)	13
2.3.2. Procesos de medias móviles (MA)	13
2.3.3. Procesos autorregresivos de medias móviles (ARMA)	14
2.4. Modelos no estacionarios lineales	14
2.4.1. Procesos autorregresivos integrados de promedios móviles (ARIMA)	14
2.5. Modelos estacionarios de heterocedasticidad condicional	15
2.5.1. Distribución de los residuales en modelos heterocedásticos	16
2.5.2. Procesos ARCH	17
2.5.3. Procesos GARCH	18
2.5.4. Procesos EGARCH	20
2.5.5. Procesos TGARCH	21

2.5.6.	Metodología de modelamiento de procesos heterocedásticos	22
2.5.7.	Retornos	28
2.6.	Metodología Bootstrap	29
2.6.1.	Introducción	29
2.6.2.	Principio del Bootstrap	29
2.6.3.	Aplicaciones del Bootstrap	30
2.6.4.	Principales algoritmos Bootstrap en series de tiempo	34
2.6.5.	Metodología Bootstrap de predicción para modelos heterocedásticos	37
 III. MATERIALES Y MÉTODOS		41
3.1.	Materiales	41
3.2.	Metodología de la investigación	41
3.2.1.	Tipo de investigación	41
3.2.2.	Diseño de la investigación	41
3.2.3.	Formulación de la hipótesis para los datos aplicados	42
3.2.4.	Identificación de las variables	42
3.2.5.	Población	44
3.2.6.	Datos de entrenamiento	44
3.2.7.	Datos de prueba	44
3.2.8.	Metodología aplicada	45
 IV. RESULTADOS Y DISCUSIÓN		47
4.1.	Análisis exploratorio	47
4.2.	Análisis para el modelo GARCH	51
4.2.1.	Identificación	51
4.2.2.	Estimación	51
4.2.3.	Diagnóstico	54
4.2.4.	Comparación de intervalos de predicción	55
4.3.	Análisis para el modelo EGARCH	59
4.3.1.	Identificación	59
4.3.2.	Estimación	59
4.3.3.	Diagnóstico	60
4.3.4.	Comparación de intervalos de predicción	60

4.4.	Análisis para el modelo TGARCH	66
4.4.1.	Identificación	66
4.4.2.	Estimación	66
4.4.3.	Diagnóstico	67
4.4.4.	Comparación de intervalos de predicción	67
V.	CONCLUSIONES	73
VI.	RECOMENDACIONES	75
VII.	REFERENCIAS BIBLIOGRÁFICAS	76
VIII.	ANEXOS	81

ÍNDICE DE FIGURAS

1.	Esquema del algoritmo Bootstrap para la estimación del error estándar . . .	31
2.	(a) Evolución diaria del Índice General de precios de las acciones de la Bolsa de Valores de Lima, del 4 de enero de 2010 al 31 de diciembre de 2014 y (b) Evolución de los retornos diarios de los precios de las acciones del IGBVL, del 5 de enero de 2010 al 31 de diciembre de 2014.	48
3.	(a) Histograma de frecuencias y principales medidas descriptivas (b) Densidad normal estándar, t-Student y de los retornos diarios de los precios de las acciones del Índice General de la Bolsa de Valores de Lima. del 5 de enero de 2010 al 30 de octubre de 2014.	49
4.	(a) Función de Autocorrelación (fac) y (b) Función de Autocorrelación Parcial (facp), de los retornos del IGBVL (r_t).	50
5.	(a) Función de Autocorrelación (fac) de la serie filtrada b_t y (b) Función de Autocorrelación (fac) de los cuadrados de la serie filtrada b_t^2	52
6.	(a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim N(0, 1)$	53
7.	(a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim t(5)$	54
8.	Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim N(0, 1)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.	56

9.	Intervalos de predicción al nivel de confianza del 90 % (a) y 95 % (b) del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim t(5)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.	57
10.	Intervalos de predicción de la volatilidad a niveles de confianza del 90 % (a) y 95 % (b) del modelo ARMA(1,1)-GARCH(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 General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap para el mes de diciembre del 2014.	58
11.	(a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$	61
12.	(a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim t(5)$	62
13.	Intervalos de predicción al nivel de confianza del 90% (a) y 95 % (b) del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.	63
14.	Intervalos de predicción al nivel de confianza del 90 % (a) y 95 % (b) del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim t(5)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.	64
15.	Intervalos de predicción de la volatilidad a niveles 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 General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap para el mes de diciembre del 2014.	65
16.	(a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$	68

17.	(a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim t(5)$	69
18.	Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.	70
19.	Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim t(5)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.	71
20.	Intervalos de predicción de la volatilidad a niveles de confianza del 90% y 95% del modelo ARMA(1,1)-TGARCH(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 General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap para el mes de diciembre del 2014.	72

ÍNDICE DE ANEXOS

Anexo 1	Histograma de frecuencias, principales medidas descriptivas y evolución de los retornos diarios de los precios de las acciones del IGBVL	81
Anexo 2	Pruebas de Autocorrelación de Ljung-Box para las series r_t , b_t y b_2	82
Anexo 3	Criterios de selección AIC y BIC para los retornos del IGBVL, obtenidos de modelos con $\varepsilon_t \sim N(0, 1)$	82
Anexo 4	Criterios de selección AIC y BIC para los retornos del IGBVL, obtenidos de modelos con $\varepsilon_t \sim t(5)$	83
Anexo 5	Estimación de coeficientes de modelos para los retornos diarios del IGBVL con $\varepsilon_t \sim N(0, 1)$	84
Anexo 6	Estimación de coeficientes de modelos para los retornos diarios del IGBVL con $\varepsilon_t \sim t(5)$	85
Anexo 7	Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim N(0, 1)$	86
Anexo 8	Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim t(5)$	86
Anexo 9	Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$	87
Anexo 10	Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim t(5)$	87
Anexo 11	Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$	88
Anexo 12	Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim t(5)$	88

Anexo 13	Función realizado en el <i>software</i> R para el cálculo de las predicciones h pasos hacia adelante del modelo ARMA(1,1)-GARCH(1,1).	89
Anexo 14	Función realizado en el <i>software</i> R para el cálculo de las predicciones h pasos hacia adelante del modelo ARMA(1,1)-EGARCH(1,1).	92
Anexo 15	Función realizado en el <i>software</i> R para el cálculo de las predicciones h pasos hacia adelante del modelo ARMA(1,1)-TGARCH(1,1).	94

RESUMEN

La presente investigación es de naturaleza aplicada, y tiene el objetivo de analizar y evaluar la metodología Bootstrap en modelos heterocedásticos aplicados en la predicción del Índice General de la Bolsa de Valores de Lima (IGBVL), periodo 2010 - 2014. Se presenta sucintamente, los conceptos básicos de series temporales, los procesos seriales heterocedásticos, la metodología Bootstrap y sus aplicaciones a la inferencia estadística y a las series temporales, donde se presenta el algoritmo para procesos heterocedásticos GARCH propuesto por Pascual et al. (2006) y generalizados para los modelos EGARCH y TGARCH. Con los procedimientos mostrados fueron obtenidas las predicciones mediante la metodología paramétrica y metodología Bootstrap, que fueron comparados con valores reales y finalmente fueron evaluados los desempeños de ambas metodologías. Del estudio se obtuvo que los modelos que mejor ajustan a la serie son los modelos ARMA(1,1)-GARCH(1,1), ARMA(1,1)-EGARCH(1,1) y ARMA(1,1)-TGARCH(1,1) cada uno de ellos con el supuesto de distribución t de Student con 5 grados de libertad de los residuales, el estudio comparativo mostró que la aplicación de la metodología Bootstrap en la serie de los retornos del Índice General de la Bolsa de Valores de Lima, permite obtener intervalos de predicciones con mayores e iguales amplitudes en algunos horizontes hacia adelante en comparación con la metodología paramétrica, y también permitió construir con un buen desempeño los intervalos de predicción para las volatilidades, así siendo esta una alternativa para la construcción de intervalos de predicción en los modelos GARCH, EGARCH y TGARCH.

Palabras claves: Modelos heterocedásticos, intervalos de predicción, bootstrap, IGBVL.

ABSTRACT

The present research is from applied nature, and it has the objective of analyzing and evaluating the Bootstrap methodology in heterocedastic models, applied in the prediction of the Índice General de la Bolsa de Valores de Lima (IGBVL), period 2010 - 2014. It presents succinctly, the concepts basic temporal series, heteroskedastic serial processes, the Bootstrap methodology and its applications to statistical inference and time series, in this is show the algorithm for heteroscedastic processes GARCH proposed by Pascual et al. (2006) and generalized for models EGARCH and TGARCH. With the procedures shown, predictions were obtained using parametric methodology and Bootstrap methodology, which were compared with real values and finally the performances of both methodologies were evaluated in terms of their prediction. This study obtained that the models that best fit the series are the ARMA(1,1)-GARCH(1,1), ARMA(1,1)-EGARCH(1,1) and ARMA(1,1)-TGARCH(1,1) models each of with the assumption of t-Student distribution with 5 degrees of freedom of the residuals, the comparative analysis showed that the application of the Bootstrap methodology in the series of the returns of Índice General de la Bolsa de Valores de Lima, allow to obtain prediction intervals with greater and equal amplitudes in some forward horizons compared to the parametric methodology, and also allowed to construct with a good performance the prediction intervals for volatilities, thus being an alternative for the construction of prediction intervals in the GARCH, EGARCH and TGARCH models.

Keywords: Heterocedastic models, predictions intervals , bootstrap, IGBVL.

I. INTRODUCCIÓN

Una serie temporal es un conjunto de observaciones ordenadas en el tiempo, los datos que resultan de observaciones de esta naturaleza son comunes en diversas áreas como en: Economía, Biología, Física, entre otras, siendo la predicción de valores futuros uno de los principales objetivos de 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 va 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.

En 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 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 , entre otros. Con ciertas características inherentes de las serie que permiten mejorar el modelamiento de las volatilidades. Estos modelos son respaldados por la estadística clásica, ya que se basan en supuestos, principalmente en la normalidad de los errores.

Por otra parte, Morettin y Tolo (2011) mencionan que el modelamiento de series temporales financieras es un problema complejo. Esta complejidad no sólo es debido a la variedad de las series en uso (acciones, tasas de cambio, ect.), frecuencias de las observaciones (minuto, hora, día, etc.) y por la disponibilidad de grandes conjuntos de datos. La complejidad también es

debida a la existencia de propiedades características de la serie llamados hechos estilizados, que son comunes a un gran número de series financieras y que son difíciles de ser modeladas, consecuentemente estas hacen que las técnicas estadísticas tradicionales frecuentemente no sean las más adecuadas para modelar dichas series.

Efron y Tibshirani (1993) desarrollaron y publicaron el análisis formal del Bootstrap. Es entonces cuando realmente este enfoque cobra una importante fuerza teórica y capta el interés de toda la comunidad de estadísticos, quienes comienzan a explorarlo y utilizarlo para solucionar una amplia gama de problemas en probabilidades e inferencia. Este proceder ha sido considerado por la American Statistical Association como «el único gran descubrimiento en estadística desde 1970». (Miranda, 2003).

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 et al. (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 General de la Bolsa de Valores de Lima, 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 General de la Bolsa de valores de Lima (IGBVL).

Por otra parte, se observa en la literatura que los autores de los libros sobre el análisis de series temporales y predicción generalmente dedican poca importancia a la predicción por intervalo y dan poca orientación sobre como calcularlos. En general, los intervalos de predicción son calculados sobre la suposición de que el modelo es conocido, y los errores tienen distribución normal. Con estas suposiciones, intervalos de predicción pueden ser fácilmente obtenidos con la media y un múltiplo de la desviación estándar. Además en el campo de las series financieras existe poca literatura sobre el procedimiento para obtener intervalos de predicción para retornos y volatilidades. Además de eso, la distribución con colas más pesadas que la distribución normal y la presencia de asimetría son algunas características de las series de tiempo en el mercado financiero, estos hacen que, al obtener intervalos de predicción utilizando el enfoque anteriormente mencionado afecten la cobertura de los intervalos, haciendo las predicciones inapropiadas, llevando, generalmente, a un riesgo mayor de lo deseable. (Pascual et al., 2006).

Una alternativa para solucionar el problema dada las características propias de los retornos y volatilidades 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.

El método bootstrap, viene siendo frecuentemente utilizado en la literatura de series temporales y puede ser usado para construir intervalos de confianza para los parámetros del modelo, realizar pruebas de hipótesis, entre otras utilidades. Uno de los métodos bootstrap para obtener intervalos de predicción en modelos heterocedásticos que ha mostrado buenos resultados y que puede ser generalizable para otros modelos de familia GARCH es el propuesto por Pascual et al. (2006). En este método, propuesto para trabajar con modelos GARCH(1,1), se incorpora la incertidumbre en los intervalos de predicción debido a la estimación de parámetros del modelo en cada replica bootstrap que no depende de la distribución

de las innovaciones.

En Perú no existe un estudio acerca de la Bolsa de Valores de Lima para los retornos y volatilidades mediante la metodología bootstrap. Este estudio aporta sobre el análisis a los índices bursátiles de la Bolsa de Valores de Lima, aportando una alternativa de método al estudio de predicciones en el mercado financiero y una propuesta de extensión del algoritmo de Pascual et al. (2006), para la construcción de intervalos de predicción para los modelos GARCH, EGARCH y TGARCH.

Al fin de delinear metodológicamente el presente trabajo, se ha centrado la investigación y análisis en el objetivo de: Aplicar y evaluar la metodología bootstrap en la predicción de modelos GARCH, EGARCH y TGARCH en el Índice General de la Bolsa de Valores de Lima, periodo 2010 - 2014.

Y como objetivos específicos se tienen:

1. Encontrar los intervalos de predicción mediante la metodología bootstrap para modelos GARCH, EGARCH y TGARCH en el Índice General de la Bolsa de Valores de Lima, periodo 2010 - 2014.
2. Comparar los enfoques bootstrap y paramétrico de construcción de intervalos de predicción para modelos GARCH, EGARCH y TGARCH en el Índice General de la Bolsa de Valores de Lima, periodo 2010 - 2014.

II. REVISIÓN DE LITERATURA

Se presenta a continuación, algunos antecedentes relevantes de la investigación, una revisión de series temporales, modelos estacionarios lineales, modelos no estacionarios lineales, modelos de heterocedasticidad condicional y la metodología Bootstrap.

2.1. Antecedentes de la investigación

2.1.1. Investigaciones acerca del análisis de la Bolsa de Valores de Lima

Elescano y Agüero (2004) analizaron la familia de procesos estocásticos para la heterocedasticidad condicional autorregresiva - ARCH en el estudio titulado “Modelos ARCH: Una aplicación en el pronóstico de la volatilidad de acciones cotizadas en la Bolsa de Valores de Lima”. Obteniendo un ajuste al modelo ARMA(0,2)-GARCH(2,2) para pronosticar la volatilidad de las cotizaciones de las acciones de la empresa minera Atacocha (uno de los índices principales de la Bolsa de Valores de Lima), utilizando la serie de datos observados desde 1992 a 2003. Entre los principales resultados fueron; que los modelos ARCH estudiados bajo el supuesto de que el proceso ε_t sigue una distribución condicionalmente normal. Sin embargo, se obtuvo mejores resultados en el ajuste al utilizar distribuciones marginales no gaussianas tales como: distribución t de Student, mixtura de distribuciones Normal-Poisson, mixtura Normal-Lag Normal, distribución exponencial generalizada y mixtura de distribuciones normales serialmente dependientes o variables t-student, estimar directamente un modelo ARCH o GARCH sin eliminar la autocorrelación lineal en los, retornos puede dificultar la identificación y estimación de estos modelos, y además que la volatilidad (riesgo) forma parte de un conjunto de factores a tener presente al realizar inversiones en activos financieros, por lo tanto, el análisis de la volatilidad mediante modelos ARCH no puede realizarse independientemente de un análisis de la estructura de precios o retornos, de los fundamentos de las empresas (estados financieros) y del entorno macroeconómico y político del país.

Chambi (2017) en el estudio titulado “La volatilidad de los mercados financieros globalizados: Impacto en la Bolsa de Valores de Lima-Perú”, con el objetivo de determinar el tipo de volatilidad de la rentabilidad de la Bolsa de Valores de Lima, y las incidencias que recibe

de los otros mercados financieros; Para tal analizó la data diaria de la cotización en índices para luego a partir de las series, obtener la rentabilidad diaria, y mediante el procesamiento estadístico de los datos recogidos, y desarrollar las pruebas que se han utilizado para determinar el proceso autoregresivo y la heterocedasticidad condicionada mediante el modelado ARCH(1) y GARCH(1,1) y la prueba de raíz unitaria de Dickey – Fuller para calificar la estacionariedad de la serie y por otro lado en modo gráfico y mediante la matriz de correlación mostrar la incidencias que existen entre diversos mercados financieros globales.

Silupú y Calle (2007) estudiaron los factores relevantes que caracterizan el comportamiento de la rentabilidad de cada acción que cotiza en la Bolsa de Valores de Lima, principalmente de aquellas que se caracterizan por ser líquidas y de mayor negociación (acciones que pertenecen al Índice Selectivo de la Bolsa de Valores de Lima), en el periodo de 1999-2004, encontrando factores estacionales, autorregresivos, que caracterizan a cada acción logrando modelarla a través de modelos como TARARCH, GARCH que logran reflejar el comportamiento de la varianza condicional ante la presencia de conglomerados de volatilidad, determinada por las expectativas de los inversionistas frente a la previsión de sus carteras. Aunque este trabajo no logra llegar a realizar un buen pronóstico de los resultados, aún así se logra encontrar los factores relevantes para poder determinar que acontecimientos son los que generan mayor incertidumbre ante las decisiones de los inversionistas, que son características del tipo de acción y del sector al cual pertenecen. Entre los principales resultados se obtuvieron que; Todas las acciones analizadas se caracterizan porque su rentabilidad es estacionaria, es decir no se comportan como un camino aleatorio, lo que nos lleva a determinar que existe alguna manera de poder modelarlas en el corto plazo, además resulta que no cumplen con el supuesto de normalidad, ya que generalmente los datos se encuentran acumulados alrededor de la media (leptocúrticas) y presentan una asimetría a la derecha. Las acciones que pertenecen al sector Minero, se caracterizan por tener un beta mayor a uno lo que significa que es el sector más riesgoso a excepción de la acción MINSURI1. En las tres acciones Mineras se ha estimado un modelo TARARCH que logra determinar que las noticias negativas tienen un mayor efecto sobre las noticias positivas, existiendo un efecto asimétrico. Para el caso de Atacocha y Minsur existen componentes autorregresivos de primer orden AR(1) que se han incorporado en la modelación de la rentabilidad y es el caso de Atacocha que presenta una cierta estacionalidad en el mes de Marzo, (pago de impuestos) para la varianza condicional.

Para el caso del Subsector Telecomunicaciones (sector Servicios) la acción Telefónica se ha modelado a través de un TARCH(1,1) que logra capturar los efectos asimétricos de los acontecimientos sucedidos, observamos también que la acción se comporta de acuerdo al mercado (IGBVL), lo que ha llevado que como la rentabilidad del IGBVL se ha mostrado muy volátil por lo diferentes acontecimientos sucedidos en el periodo 2000 – 2004 también a afectado la volatilidad de la acción. Por último la acción CREDITC1 que pertenece al Sector Bancos y Finanzas se ha modelado a través de una GARCH (1,1) lo que muestra que no existen efectos asimétricos sobre el impacto de las noticias buenas o malas, Observamos que existen componentes autorregresivos que afectan la rentabilidad de la acción AR(2) siendo los acontecimientos pasados los que influyen en la rentabilidad de hoy.

Finalmente cabe destacar el trabajo de **Castillo (2014)** titulado “Modelación de la volatilidad del índice general de la Bolsa de Valores de Lima, periodo 2009-2011”, con el objetivo de describir los modelos de varianza condicional ARCH y GARCH junto con sus propiedades y demostraciones aplicado al valor diario del Índice General de la Bolsa de Valores de Lima (IGBVL). Al analizar los resultados de las estimaciones de los modelos que explicarían la volatilidad diaria de la Rentabilidad del Índice General de la Bolsa de Valores (RIGBVL), se concluyó que el modelo GARCH(1,1) es el adecuado, debido a que el modelo GARCH (1,1) tiene a diferencia de los demás modelos el menor valor tanto en el criterio de información de Akaike (AIC) como en el criterio de información de Schwarz. Previamente se modeló la media de la RIGVBL con el modelo AR(1).

2.1.2. Investigaciones acerca de la metodología Bootstrap en series temporales

Alvarez y Rubio (2000) presentaron la metodología Bootstrap como una alternativa para construir intervalos de predicción en series temporales, realizaron un estudio de los métodos bootstrap cuando las series analizadas son heterocedásticas, haciendo especial énfasis en la predicción de modelos heterocedásticos condicionalmente autorregresivos (ARCH). Presentaron un estudio empírico sobre el índice IBEX-35 con datos diarios desde 1992 hasta 1998 y se comparó la metodología bootstrap con los métodos tradicionales utilizados en la construcción de intervalos de predicción. En los modelos de rentabilidad-riesgo de tipo GARCH-M la predicción bootstrap mostró un excelente comportamiento en el corto y medio plazo.

Henrique (2003) estudió el tráfico aéreo en dos regiones de Portugal, con el objetivo de

obtener predicciones para el planeamiento y tomar decisiones anticipadamente, respecto al flujo aéreo. Aplicando la metodología Bootstrap por bloques en modelos de predicción por el método de Holt Winters y ARIMA, y realizando remuestras a los residuales de modelos obtenidos con anticipación, se obtuvieron predicciones por intervalos adecuados para el flujo aéreo.

Kim. et al. (2009) evaluaron el rendimiento de intervalos de predicción de series mensuales para el número de llegadas de turistas a Hong Kong y Australia, empleando el modelo autorregresivo (AR), el modelo AR usando el Bootstrap corregido por sesgo, los modelos estacionales ARIMA, las innovaciones de los modelos de estado-espacio para el suavizado exponencial, y modelos de series de tiempo estructurales de Harvey. Se encontraron que todos los modelos producen intervalos de predicción satisfactorios, excepto para el modelo autorregresivo (AR). En particular, los basados en el Bootstrap corregido por sesgo funcionan mejor en general, obteniendo intervalos con tasas de cobertura precisas, especialmente cuando el horizonte de pronóstico es largo.

Marques (2013) realizó un estudio de la metodología Bootstrap en métodos de alisamiento exponencial, principalmente, al método no paramétrico de Holt Winters para la predicción de variables ambientales de calidad de agua de superficie, de marzo de 2000 a diciembre de 2011. Las predicciones fueron comparadas con valores reales, utilizando el error cuadrático medio como medida para evaluar el desempeño de la metodología Bootstrap en cuanto a la calidad de las predicciones. El estudio comparativo mostró que las aplicaciones del Bootstrap en asociación al método de Holt Winters permitió obtener intervalos de confianza de las predicciones con menor amplitud a comparación de los obtenidos por los métodos usuales.

Stine (1987) propuso el primer método bootstrap para estimar el error cuadrático medio del predictor lineal clásico cuando el proceso es un modelo autorregresivo de orden p (AR(p)). El método se basó en la obtención de la función de distribución empírica de los errores centrados. A partir de ésta extraía muestras para reproducir un proceso AR(p) con los parámetros estimados inicialmente y fijando los primeros p valores de la serie. Dada la serie bootstrap se estiman los parámetros y se obtiene el predictor clásico para evaluar el error cuadrático medio de dicha predicción. Repitiendo un número elevado de veces este proceso podemos aproximar el error cuadrático medio de la predicción puntual.

Thombs y Schucany (1990) propusieron la construcción de intervalos de predicción en modelos AR(p), propusieron realizar una representación hacia atrás del proceso para posteriormente utilizar la representación habitual para obtener los valores futuros. Estos autores demostraron la validez asintótica de este método bootstrap y realizaron un estudio de simulación donde comparan con los intervalos clásicos para diferentes distribuciones y coberturas.

McCullough (1994) sugirió que el método percentil de obtener los intervalos de predicción no era el más adecuado cuando existe asimetría en el error de predicción y, por lo tanto, propuso un método de corrección de sesgo. Además, advirtió que la representación backward utilizada por Thombs y Schucany (1990) no asegura que los errores sean independientes e idénticamente distribuidas y utilizó un procedimiento más sofisticado para obtener dichos errores hacia atrás.

La validez asintótica del método de Thombs y Schucany (1990) para modelos AR fue demostrada por Jurado, Manteinga, y Sánchez (1995). La idea general se basa en obtener la serie diferenciada y aplicar el método TS para construir la réplica bootstrap de la serie que tiene la estructura autoregresiva. Posteriormente, se integra para tener réplicas bootstrap de la serie original.

Un inconveniente importante del método Thombs y Schucany (1990) consiste en la representación backward de la serie, no sólo por el gasto computacional sino porque no es fácil cuando los procesos no tienen distribución normal. Cao et al. (1997) proponen un método alternativo, que es consistente, para construir intervalos de predicción sin tener que realizar réplicas hacia atrás de la serie y demostraron la consistencia de este método tanto en el caso de utilizar la función de distribución empírica de los residuos directamente como en el de utilizar un bootstrap suavizado, es decir, cuando se extrae la muestra a partir de una suavización no paramétrica de la distribución.

Finalmente, cabe destacar que **Pascual, Romo, y Ruiz (1999)** propusieron un método que capta el error de predicción debido a la incertidumbre en la estimación pero que no es necesaria realizar una representación backward de la serie. Este método será presentado formalmente en capítulos posteriores

2.2. Series temporales

2.2.1. Definición de series temporales

Una serie de tiempo es una secuencia cronológica de observaciones de una variable particular. Se aplican series de tiempo en los distintos campos del saber: en economía, mercado, demografía, meteorología, ingeniería, etc. Muchos son los ejemplos de series de tiempo que podrían citarse, tales como: las ventas mensuales de una empresa en la última década, El número de automóviles producidos por año de determinada marca en período 1985-2000, La temperatura diaria promedio en los últimos 6 meses, etc.

2.2.2. Objetivos del análisis de series temporales

Según Uriel y Peiró (2005), son varios los objetivos por los cuales se puede querer analizar una serie de tiempo:

- Descripción: Cuando se tiene una serie de tiempo, el primer paso en el análisis es graficar los datos y obtener medidas descriptivas simples de las propiedades principales de la serie.
- Explicación: Cuando las observaciones son tomadas sobre dos o más variables, es posible usar la variación en una serie para explicar la variación en otras series.
- Predicción: Dada una serie de tiempo se puede querer predecir los valores futuros de la serie. Este es el objetivo más frecuente en el análisis de series de tiempo.
- Control: Cuando una serie de tiempo se genera por mediciones de calidad de un proceso, el objetivo del análisis puede ser el control de proceso.

2.2.3. Procesos estocásticos

Un proceso estocástico es una colección de variables aleatorias ordenadas en el tiempo. Sea C un conjunto arbitrario y una colección $y = \{y_t, t \in C\}$, tal que para cada $t \in C$, y_t es una variable aleatoria, es decir para cada valor de t del conjunto C está definida una variable aleatoria, y_t , y los valores observados de las variables aleatorias en distintos instantes forman una serie temporal (Uriel y Peiró, 2005).

Morettin y Toloi (2011) indicaron que el proceso queda caracterizado si se define la distribución de probabilidad conjunta de las variables aleatorias $(y_1, \dots, y_t, \dots, y_T)$, para cualquier valor de T . Estas distribuciones se denominan las *distribuciones finito-dimensionales* del proceso. Se dice que se conoce la estructura probabilística de un proceso estocástico cuando se conozcan estas distribuciones, que determinan la distribución de cualquier subconjunto de variables y , en particular, las distribuciones marginales de cada variable que tienen las siguientes propiedades:

- función de medias del proceso a $E(y_t) = \mu_t$,
- a la función de varianza del proceso a $Var(y_t) = \sigma_t^2$,
- a la función de autocovarianzas del proceso a

$$\gamma(t, t+j) = Cov(y_t, y_{t+j}) = E[(y_t - \mu_t)(y_{t+j} - \mu_{t+j})],$$

en particular $\gamma(t, t) = Var(y_t) = \sigma_t^2$, y

- al coeficiente de autocorrelación de orden $(t, t+j)$ a

$$\rho(t, t+j) = \frac{Cov(t, t+j)}{\sigma_t \sigma_{t+j}} = \frac{\gamma(t, t+j)}{\gamma^{1/2}(t, t) \gamma^{1/2}(t+j, t+j)}$$

2.2.4. Procesos estacionarios

Se dice que un proceso estocástico es estacionario en *sentido estricto* si:

- Las distribuciones marginales de todas las variables son idénticas,
- Las distribuciones finito-dimensionales de cualquier conjunto de variables sólo dependen de los retardos entre ellas.

Se dice que un proceso estocástico es estacionario en *sentido débil*, si para todo t :

- $\mu_t = \mu$,
- $\sigma_t^2 = \sigma^2$,
- $\gamma(t, t-k) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k$, $k = 0, \pm 1, \pm 2, \dots$

Las dos primeras condiciones indican que la media y la varianza son constantes. La tercera, que la covarianza entre dos variables depende sólo de su separación.

2.2.5. Procesos de ruido blanco

Se dice que $\{\varepsilon_t, t \in \mathbb{Z}\}$ es un ruido blanco si:

- $E(\varepsilon_t) = 0, t=1,2,\dots$
- $Var(\varepsilon_t) = \sigma^2, t=1,2,\dots$
- $Cov(y_t, y_{t-k}) = 0, k = 0, \pm 1, \pm 2, \dots$

La condición primera establece que la esperanza es siempre constante e igual a cero, la segunda condición que la varianza es constante y la tercera condición que las variables del proceso están incorreladas para todos los retardos.

2.2.6. Estimación de los momentos de procesos estacionarios

Sea un proceso estacionario con media μ , varianza σ^2 y covarianza γ_k del que se observa una realización (y_1, y_2, \dots, y_T) . A continuación se presentan los estimadores de la media, la varianza, las covarianzas y las autocorrelaciones del proceso a partir de esta única realización disponible.

- El estimador de la media poblacional es la media muestral $\bar{y} = \frac{\sum_{t=1}^T y_t}{T}$
- El estimador de la autocovarianza poblacional es dado por

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})$$

Donde la varianza del proceso se estima por $\hat{\gamma}_0$.

- La autocorrelación se estima por, $\hat{\rho} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$

Además de estudiar la correlación de una forma global, también es importante estudiar la correlación parcial entre $y(t)$ y $y(t+k)$, cuando se fijan las variables intermedias $y(t+1), y(t+2), \dots, y(t+k-1)$, esto es la correlación simple entre $y(t)$ y $y(t+k)$ después de eliminar el efecto que sobre estas producen las variables intermedias, esta es llamada *función de autocorrelación parcial (FACP)*.

Se representa la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP) dando los coeficientes de correlación muestral o estimados en función del

retardo. A esta representación se le denomina *correlograma*, o *función de autocorrelación muestral*.

2.3. Modelos estacionarios lineales

2.3.1. Procesos autorregresivos (AR)

Los procesos autorregresivos constituyen un tipo importante de procesos estacionarios. Son obtenidos al imponer una dependencia de tipo lineal entre el valor del proceso en un determinado instante y los valores del pasado. Se dice que y_t sigue un proceso estocástico autorregresivo de orden p , o $AR(p)$ si se escribe como:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

donde $\phi_i, i = 1, \dots, p$ son constantes fijas y ε_t es una secuencia de variables no correlacionadas con media cero y varianza σ^2 , esto es, un proceso de ruido blanco.

Definiendo el operador autorregresivo estacionario de orden p ,

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, entonces el proceso $AR(p)$ se puede escribir también como:

$$\phi(B)y_t = \varepsilon_t$$

2.3.2. Procesos de medias móviles (MA)

El proceso de medias móviles consiste en expresar y_t en términos de un proceso más simple, como es el proceso de ruido blanco. Así, un proceso de medias móviles de orden q se define en cada instante t , como la media ponderada de las $q + 1$ observaciones de un proceso de ruido blanco ε_t . Estos procesos son siempre estacionarios.

Un proceso se dice de medias móviles de orden q , o $MA(q)$, si:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

siendo $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ el operador de medias móviles de orden q , entonces el proceso $MA(q)$ se puede expresar también como:

$$y_t = \theta(B)\varepsilon_t$$

2.3.3. Procesos autorregresivos de medias móviles (ARMA)

Tanto los procesos autorregresivos como los procesos de medias móviles constituyen buenos modelos para las series temporales que se pueden observar en la “vida real”. Así es natural considerar procesos más generales que son construidos a partir de la combinación entre estos dos modelos (Morettin y Toloí, 2011).

Se dice que y_t sigue un proceso autorregresivo de medias móviles de orden (p, q) , o $ARMA(p, q)$, si satisface la ecuación:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

donde ε_t es un proceso de ruido blanco, si $\phi(B)$ y $\theta(B)$ son los operadores autorregresivos y de medias móviles respectivamente, entonces es posible también expresar en la forma siguiente:

$$\phi(B)y_t = \theta(B)\varepsilon_t$$

2.4. Modelos no estacionarios lineales

2.4.1. Procesos autorregresivos integrados de promedios móviles (ARIMA)

Los modelos presentados en los capítulos anteriores son apropiados para describir series estacionarias, sin embargo muchas series encontradas en la práctica no son estacionarias, si una serie de tiempo no es estacionaria en el sentido antes definido, se denomina serie de tiempo no estacionaria.

Para el tratamiento de la no estacionariedad en la media se propone la diferenciación sucesiva de la serie, aprovechando la propiedad de que gozan una gran parte de los procesos estocásticos, de convertirse en estacionarios al diferenciarlos cierto número de veces. Las primeras diferencias de los valores de la serie de tiempo (y_1, y_2, \dots, y_T) son:

$$\nabla y_t = y_t - y_{t-1} = (1 - B)y_t$$

para $t = 2, 3, \dots, T$

Las segundas diferencias dadas por:

$$\nabla^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = (1 - B)^2 y_t$$

para $t = 3, 4, \dots, T$

Generalizando para $d \geq 0$

$$\nabla^d Y_t = (1 - B)^d y_t$$

El objetivo es determinar la serie estacionaria obtenida por la menor diferenciación, una vez que la varianza aumenta con la diferenciación.

La diferenciación no elimina todos los casos de no estacionariedad. La estacionariedad puede ser conseguida a través de transformaciones que permiten estabilizar la varianza. Se puede entonces recurrir a transformaciones logarítmicas de la serie o por la transformación de Box y Cox, ver Murteira y Muller (1990).

Una metodología bastante utilizada en el análisis de modelos paramétricos es conocida como metodología ARIMA de Box y Jenkins . Tal metodología consiste en ajustar modelos autorregresivos integrados de medias móviles, $ARIMA(p, d, q)$, a un conjunto de datos.

Si $z_t = \nabla^d y_t$ fuese estacionaria, es posible representar z_t por un modelo $ARMA(p, q)$, es decir:

$$\phi(B)z_t = \theta(B)\varepsilon_t$$

Entonces, se dice que z_t sigue un modelo autorregresivo, integrado, de medias móviles, o modelo $ARIMA$ de orden (p, d, q) si:

$$\phi(B)\nabla^d y_t = \theta(B)\varepsilon_t$$

y se denota como $ARIMA(p, d, q)$, donde p y q son ordenes de $\phi(B)$ y $\theta(B)$, respectivamente y ε_t es un proceso de ruido blanco,

2.5. Modelos estacionarios de heterocedasticidad condicional

En esta sección se presentan brevemente los procesos estacionarios que son no lineales en la varianza, es decir, presentan heterocedasticidad condicional a lo largo del tiempo. En estos tipos de modelos la varianza condicional está en función de los cuadrados de las observaciones como de las varianzas condicionales pasadas.

Las series de tiempo financieras, como precios de acciones, tasa de cambio, tasa de inflación, etc. a menudo presentan el fenómeno de acumulación de la volatilidad, es decir, existen

lapsos en los que sus precios muestran amplias variaciones durante prolongados períodos y luego se dan intervalos de una calma relativa (Peña, 2010).

Conocer la volatilidad es muy importante en muchas áreas. Por ejemplo, existe una enorme cantidad de trabajo en econometría sobre la variabilidad de la inflación a lo largo del tiempo. Para algunas personas con poder de decisión, la inflación en sí misma quizá no sea dañina, pero no es deseable su variabilidad porque dificulta la planificación financiera. Sucede lo mismo con los importadores, exportadores y comerciantes que acuden a los mercados de cambio de divisas, pues la variabilidad de las tasas de cambio representa grandes pérdidas o ganancias. A los inversionistas de las casas de bolsa obviamente les interesa la volatilidad de los precios de las acciones, pues una gran volatilidad puede significar enormes pérdidas o ganancias y, en consecuencia, provocar mayor incertidumbre (Bodas, 2013).

Existen variedades de modelos no lineales disponibles en la literatura, sin embargo, para poder realizar este trabajo se da el enfoque solamente en los procesos ARCH (“Autorregresive Conditional Heterocedasticity”), GARCH (“Generalized ARCH”), EGARCH (“Exponential GARCH”) y TGARCH (“Threshold GARCH”).

2.5.1. Distribución de los residuales en modelos heterocedásticos

Originalmente, los modelos heterocedásticos fueron construidos en base al supuesto de que las series de tipo financieras siguen una distribución Normal, sin embargo evidencias sugieren que estos tipos de series no se comportan como una Normal, por la presencia de la forma leptocútica y presencia de colas pesadas en su estructura. Para resolver este problema Bollerslev (1987) propone trabajar con el supuesto de los residuales con una distribución de t de Student en el modelo ARCH, y consecuentemente puede ser generalizada a otros tipos de modelos heterocedásticos.

Así siguiendo el planteamiento de (Bollerslev, 1987), sea ε_t sigue una distribución t de Student con ν grados de libertad, es decir:

$$\varepsilon_t | \sigma_t^2 \sim \Gamma\left(\frac{\nu+1}{2}\right) \Gamma\left(\frac{\nu}{2}\right)^{-1} ((\nu-2)\sigma_t^2)^{-1/2} \left(1 + \frac{\varepsilon_t^2}{\sigma_t^2(\nu-2)}\right)^{-(\nu+1)/2}$$

Para un ν muy grande, se tiene que ε_t tiende a una Normal con media cero y varianza σ^2 .

Como se observó, el modelo heterocedástico con distribución t de Student de los residuales se introduce el parámetro adicional ν , que son los grados de libertad que permiten medir el grado de la leptocurtosis en los datos. Heracleous (2003) en su trabajo de doctorado evaluó la capacidad del coeficiente de curtosis para capturar con precisión los grados de libertad implícitos y la capacidad del modelo GARCH con t de Student para estimar con precisión el verdadero parámetro de grado de libertad, concluyendo que con un número de $\nu < 10$ es posible explicar el exceso de curtosis en la serie, obteniendo resultados satisfactorios con $\nu = 5$. Así en esta investigación se consideró trabajar los modelos heterocedásticos con residuales t de Student con $\nu = 5$ en cada modelo evaluado, además del ya conocido modelo heterocedástico con distribución de residuales Normales.

2.5.2. Procesos ARCH

Los modelos autorregresivos con heterocedasticidad condicional (ARCH) fueron introducidos por Engle (1982) con el objetivo de estimar la variancia de la inflación en el Reino Unido. El modelo ARCH plantea una estrategia de modelación dinámica de la variancia; en determinadas situaciones se desea justamente predecir la variancia condicional, es decir, determinar su patrón de comportamiento estadístico. Esto ocurre al analizar los mercados financieros; por lo cual el modelo ARCH, es el modelo básico para poder estimar la variancia condicional.

El planteamiento básico del modelo ARCH consiste en modelar la variancia de una variable dependiente en función de los valores pasados de la propia variable y de las variables independientes que se incluyan en el modelo. Así, un proceso ARCH simple viene definido por dos expresiones: en la primera se modela la esperanza condicionada de la serie en función de la variancia y de un término de error, y en la segunda ecuación es la variancia condicionada la que es modelada en función de sus propios valores pasados.

En la siguiente definición se presenta de manera formal el modelo ARCH.

Dado una serie y_t y sean $\mu_t = E(y_t|F_{t-1})$, $\sigma_t^2 = Var(y_t|F_{t-1})$ la media y variancia condicional de y_t , donde F_{t-1} es la información hasta el instante $t - 1$. Se supondrá que $\mu_t = 0$ de modo que $\sigma_t^2 = E(y_t^2|F_{t-1})$.

Un proceso y_t es llamado de proceso $ARCH(r)$ si satisface las siguientes ecuaciones

$$y_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \dots + \alpha_r y_{t-r}^2 = \alpha_0 + \sum_{i=1}^r \alpha_i y_{t-i}^2$$

donde: $\alpha_0 > 0$, $\alpha_i \geq 0$, $i > 0$, y ε_t es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza uno.

Con el supuesto de que $\varepsilon_t \sim N(0, 1)$ o $\varepsilon_t \sim t_\nu$ (distribución t de Student con ν grados de libertad).

2.5.3. Procesos GARCH

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 como el modelo ARMA puede ser más parsimonioso, en el sentido de presentar menos parámetros que un modelo autorregresivo o de medias móviles, el modelo GARCH puede ser usado para describir la volatilidad con menos parámetros que un modelo ARCH.

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 $\beta_j > 0$.

Herencia (1997) indica que, las condiciones $\omega > 0$, $\alpha_i > 0$ y $\beta_j > 0$ impuestas por Bollerslev (1986) fueron garantizadas para que la varianza condicional sea positiva, sin embargo diversos estudios demostraron que existen condiciones menos restrictivas para algunas especificaciones (p, q) , es decir la presencia de estimaciones negativas de los parámetros no significa que el modelo esté mal especificado o que las estimaciones sean malas, sin embargo es importante resaltar que para el caso del modelo GARCH(1,1) no es posible descuidar las condiciones de Bollerslev.

ε_t , es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza uno, e independiente de y_{t-i} para $i > 1$. En la práctica usualmente suponemos que $\varepsilon_t \sim N(0, 1)$ o $\varepsilon_t \sim t_\nu$ (distribución t de Student con ν grados de libertad).

Por otro lado, las condiciones para la estacionariedad del modelo GARCH(p,q) presentadas por Bollerslev (1986), es definida la varianza condicional como:

$$Var(y_t) = \frac{\omega}{1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j}, \text{ si y solamente si, } \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j < 1$$

Se tiene también que, σ_t y ε_t son dos procesos estacionarios independientes entre sí. Sea el proceso ε_t ruido blanco. El proceso σ_t es estacionario, pero tiene estructura dinámica, siendo su valor en t función del conjunto $\mathbf{y}_{t-1} = (y_{t-1}, \dots, y_1)$ de los valores de la serie previos a t (información hasta el instante $t - 1$). La condición de independenciam entre ε_t y σ_t garantiza que la serie y_t tenga media marginal igual a cero, ya que:

$$E(y_t) = E(\sigma_t)E(\varepsilon_t) = 0,$$

y también media condicional nula, ya que:

$$E(y_t | \mathbf{y}_{t-1}) = E(\sigma_t | \mathbf{y}_{t-1})E(\varepsilon_t) = 0.$$

Como el proceso y_t es estacionario, tendrá una varianza marginal constante, que llamaremos σ^2 . calculada mediante:

$$E(y_t^2) = E(\sigma_t^2)E(\varepsilon_t^2) = E(\sigma_t^2) = \sigma^2$$

y coincide con la varianza del proceso σ_t , ya que $E(\varepsilon_t^2) = 1$. Sin embargo, el proceso y_t tendrá una varianza condicionada que no es constante. En efecto:

$$Var(y_t^2 | \mathbf{y}_{t-1}) = E(\sigma_t^2 | \mathbf{y}_{t-1})E(\varepsilon_t^2) = \sigma_t^2$$

dado que $E(\varepsilon_t^2 | \mathbf{y}_{t-1}) = E(\varepsilon_t^2) = 1$. Por tanto, el proceso σ_t^2 representa la varianza condicionada de la serie en cada instante, que va variando en el tiempo con cierta estructura estacionaria.

La independenciam entre los procesos ε_t y σ_t garantiza que la serie y_t carece de autocorrelación y forma un proceso de ruido blanco. En efecto, las autocorvianzas de la serie son:

$$E(y_t y_{t-k}) = E(\sigma_t \varepsilon_t \sigma_{t-k}) = E(\varepsilon_t)E(\sigma_t \sigma_{t-k}) = 0,$$

ya que el proceso ε_t es independiente de su pasado y de los valores previos y actuales del proceso σ_t . Por tanto, el proceso y_t tendrá autocovarianzas nulas.

Para investigar algunas propiedades de los modelos GARCH, se considera el caso especial GARCH(1,1), que es un modelo bastante usado en la práctica, para lo cual la volatilidad o varianzas condicionales siguen la ecuación:

$$\sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

con $0 \leq \alpha_1, \beta_1 < 1, \alpha_1 + \beta_1 < 1$.

Bollerslev (1986) mostró condiciones necesarias y suficientes para la existencia de momentos de orden $2n$, así se tienen que:

$E(y_t^2) = \frac{\omega}{1 - \alpha_1 - \beta_1}$, y $E(y_t^4) = \frac{3\omega(1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)}$, por tanto su curtosis es igual a:

$$K = \frac{E(y_t^4)}{[E(y_t^2)]^2} = \frac{3(1 - (\alpha_1 + \beta_1)^2)}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2}$$

que es mayor que 3 para todo $\alpha \geq 0$. dado que el denominador es positivo, la cual demuestra que si y_t sigue un modelo GARCH, las colas de y_t serán más largas o pesadas en comparación a una distribución normal.

2.5.4. Procesos EGARCH

Los procesos GARCH asumen que los errores positivos y negativos tienen un efecto simétrico en la volatilidad. En la práctica este supuesto es frecuentemente violado, en particular en retornos financieros, en que la volatilidad es incrementada más por retornos negativos que por los positivos, este efecto es llamado de apalancamiento. El proceso EGARCH (Exponential GARCH), planteado por Nelson (1991) modela el efecto mencionado.

Una serie y_t es llamado de proceso $EGARCH(p, q)$ si satisface las siguientes ecuaciones:

$$y_t = \sigma_t \varepsilon_t$$

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^q g(\varepsilon_{t-i}) + \sum_{i=1}^p \beta_i \log(\sigma_{t-i}^2)$$

Donde:

$$g(\varepsilon_{t-i}) = \alpha_i \varepsilon_{t-i} + \gamma(|\varepsilon_{t-i}| - E(|\varepsilon_{t-i}|)), i = 1, \dots, q,$$

$\omega, \alpha_i, \beta_i, \gamma_i$ son números reales, y ε_t , es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza uno.

Este modelo permite 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 y Tolo, 2011).

Para el caso más simple, es decir, un modelo EGARCH(1,1) se tiene:

$$y_t = \sigma_t \varepsilon_t$$

$$\log(\sigma_t^2) = \omega + \alpha \varepsilon_t + \gamma[|\varepsilon_{t-1}| - E(|\varepsilon_{t-1}|)] + \beta \log(\sigma_{t-1}^2)$$

Cuando la varianza condicional cuando $y_{t-1} > 0$ es:

$$\log(\sigma_t^2) = \omega + (\alpha + \gamma)\varepsilon_{t-1} + \beta \log(\sigma_{t-1}^2)$$

Si la condición de estacionariedad es satisfecha se tiene que la varianza incondicional, cuando las innovaciones siguen una distribución normal estándar es dada por:

$$\sigma_t^2 = \exp\left(\frac{\omega}{1-\beta}\right)$$

Donde el modelo es estacionario y tiene curtosis finita si $|\beta| < 1$.

2.5.5. Procesos TGARCH

Otra variante de los modelos GARCH que consiste en incorporar el elemento threshold (umbrales), propuesto por Zakoian (1994), el cual introduce un modelo TGARCH (Threshold GARCH) capaz de captar asimetrías según el signo de los residuos. El modelo TGARCH permite evaluar si inflaciones por sobre lo esperado tienen un efecto sobre la incertidumbre distinto del que tendría una inflación menor a la esperada.

Una serie y_t es llamado de proceso $TGARCH(p, q)$ si satisface las siguientes ecuaciones:

$$y_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 + \sum_{i=1}^q \gamma_i \varepsilon_{t-i}^2 S_{t-i}$$

Donde:

$$S_{t-i} = \begin{cases} 1 & \text{si } \varepsilon_{t-i} < 0 \\ 0 & \text{si } \varepsilon_{t-i} \geq 0. \end{cases}$$

El coeficiente γ_i corresponde al parámetro de apalancamiento y cuando $\gamma_i = 0$, el modelo corresponde a un GARCH estándar. De acuerdo con este modelo, cuando se tiene un impacto positivo (buenas noticias) el efecto que se tiene en la volatilidad es igual a α_i . Asimismo,

cuando el impacto es negativo (malas noticias) el efecto en la volatilidad es $\alpha_i + \gamma_i$. Además ε_t , es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza uno, y ω , α_i , β_i , γ_i son no negativos satisfaciendo las condiciones semejantes a las de los modelos GARCH.

Para el caso más simple, es decir, un modelo TGARCH(1,1) se tiene:

$$y_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \omega + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{t-1}^2 + \gamma_i \varepsilon_{t-1}^2 S_{t-1}$$

Así, se tiene que la varianza condicional cuando $S_{t-1} < 0$ es:

$$\sigma_t^2 = \omega + (\alpha_1 + \gamma_1) \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

Y la varianza condicional cuando $S_{t-1} \geq 0$ es:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

Es decir, si el retorno es negativo se tendrá que el efecto sobre la varianza condicional es mayor, en comparación de que cuando el retorno es positivo cuando α_1 es positivo.

Si el proceso es estacionario la varianza condicional de y_t es dado por:

$$\sigma_t^2 = \frac{\omega}{1 - \alpha - \beta - \gamma \text{Prob}(\varepsilon_t < 0)}.$$

2.5.6. Metodología de modelamiento de procesos heterocedásticos

La metodología descrita por Box y Jenkins (1970), consiste en determinados procedimientos para poder obtener un proceso adecuado para la serie temporal objeto de estudio. Este procedimiento consta de: el análisis de exploración, la identificación de un modelo, la estimación de parámetros del modelo, el diagnóstico del modelo y finalmente la predicción.

A continuación se detalla cada paso a seguir de esta metodología para obtener un proceso heterocedásticos.

1. Análisis exploratorio

Consiste en el análisis del histograma de frecuencias, las principales medidas descriptivas, análisis de estacionariedad mediante gráficos (gráficos de las funciones de autocorrelación y autocorrelación parcial) y la prueba estadística de existencia de raíces unitarias de Dickey y Fuller (DF).

Prueba de raíces unitarias de Dickey Fuller (DF)

Una estrategia eficiente para probar la existencia de una raíz unitaria (no estacionariedad) en una serie, es probar la hipótesis nula de un proceso con raíz unitaria y una constante versus un proceso estacionario alrededor de una tendencia.

La prueba de Dickey y Fuller se basa en asumir que la serie se puede aproximar por un proceso AR(1). Inicialmente se asume que y_t sigue un modelo AR(1) y se procede a transformar el modelo de la siguiente manera:

$$\nabla y_t = \rho y_{t-1} + \varepsilon_t$$

Donde la existencia de raíz unitaria (no estacionariedad) equivale a $\rho = 0$. Por tanto se tiene las hipótesis:

$$H_0 : \rho = 0$$

$$H_1 : \rho < 0$$

El estadístico de prueba es denotado por τ y su distribución bajo H_0 permite calcular los valores críticos, de tal forma que el criterio de rechazo es $\hat{\tau} < \tau_\alpha$ donde $\hat{\tau}$ es el valor calculado del estadístico.

2. Identificación del modelo

Para la identificación de un modelo heterocedástico primeramente se tiene que tomar en cuenta si los retornos y_t son o no correlacionadas. Para tal caso es verificada mediante la prueba de Ljung-Box y/o el análisis gráfico de las funciones de autocorrelación (FAC).

Solamente en el caso de que y_t presente correlación se realiza un ajuste al modelo ARMA, para remover la correlación serial, por tanto:

$$\phi(B)y_t = \theta_0 + \theta(B)a_t$$

siendo a_t el residuo de la aplicación de un modelo ARMA, que viene a ser la serie a modelar de un proceso heterocedástico.

Para verificar si la serie presenta heterocedasticidad condicional, se usa la prueba de Ljung-Box para y_t^2 o en el caso de ser necesario modelar un ARMA se procede a la prueba de Ljung-Box a los cuadrados de los residuales a_t^2 del modelo ARMA.

Prueba de Ljung-Box

La prueba de Ljung y Box es una prueba para las autocorrelaciones de series de tiempo. Que a pesar de no detectar caídas específicas en el comportamiento de ruido blanco, puede indicar si esos valores son muy altos (Morettin y Toloï, 2011).

$$H_0 : \rho(1) = \rho(2) = \dots = \rho(k) = 0$$

H_1 : por lo menos un $\rho(k) \neq 0$.

con el estadístico de prueba:

$$Q(k) = n(n+2) \sum_{j=1}^k \frac{\hat{\rho}_j^2}{n-j}$$

Donde n es el tamaño de la serie y $\hat{\rho}_j$ es el estimador de la j -ésima autocorrelación. La estadística Q tendrá aproximadamente una distribución χ^2 con k grados de libertad, se rechaza la hipótesis nula si $Q(k) \geq \chi_k^2$.

Herencia (1997) menciona que, el proceso de identificación puede ser realizado aprovechando la estructura de autocorrelaciones y autocorrelaciones parciales de los cuadrados de las observaciones que muestran un comportamiento similar a los modelos ARMA(p,q). Sin embargo, en modelos heterocedásticos es frecuente la selección del modelo a través de criterios como AIC y BIC.

Otro aspecto observado empíricamente en los ajustes de modelos heterocedásticos en series financieras es que el valor de p y q son raramente mayores que dos y que la contribución de los términos de orden superior a uno son insignificantes. (Bollerslev, 1986)

De esta manera, la identificación de orden de un modelo a ser ajustado se realiza mediante la selección por los criterios de Akaike (AIC) y de Schwarz (BIC) de los modelos GARCH, EGARCH y TGARCH de orden (1,0), (0,1), (1,1), (1,2) (2,1) y (2,2).

Criterio de información de Akaike (AIC)

EL AIC es uno de los criterios comunmente usados para la selección de modelos, este es una medida estadística de verosimilitud del conjunto de modelos a incluir, penalizado por el número de parámetros de los modelos (Peña, 2010). El criterio de información de Akaike es definido como:

$$AIC = -2\log(L) + 2p$$

Donde L es la función de máxima verosimilitud del modelo ajustado y p es el número de

parámetros del modelo. El menor valor del AIC, significa el mejor modelo ajustado.

Criterio de Schwarz (BIC)

El término de penalización del AIC, $2p$, no depende del tamaño muestral “n” de la población considerada. Esto conduce al hecho de que un mismo número de parámetros comunes es seleccionado mediante el AIC, tanto para muestras pequeñas como para muestras grandes. Es por ello que el AIC no es un estimador consistente del número adecuado de parámetros (Peña, 2010). Schwarz (1978) sugirió que el AIC podría no ser asintóticamente justificable, y presentó un criterio de información alternativo a partir de un enfoque bayesiano, el BIC (Bayesian Information Criterion). Con este criterio, se penaliza el número de parámetros con $\ln(n)$, en lugar de 2. Así,

$$BIC = -2\log(L) + \log(n)p$$

Donde L es la función de máxima verosimilitud del modelo ajustado, p es el número de parámetros del modelo y “n” es el tamaño muestral. De acuerdo con este criterio, el modelo que indique el menor valor de BIC será el mejor modelo ajustado.

3. Estimación

Los estimadores de los parámetros del modelo GARCH, EGARCH y TGARCH son obtenidos por el método de máxima verosimilitud.

A continuación se define para el modelo ARCH(p), a partir de ésta pueden ser extendibles a otros modelos heterocedásticos.

Asumiendo que la serie y_t tiene condicionalmente una distribución normal, esto es

$$f(y_t|F_{t-1}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(y_t-u_t)^2}{\sigma^2}}$$

con $t = 1, 2, \dots, N$. Y su función de verosimilitud condicional es dada por

$$L(\alpha|y_1, y_2, \dots, y_N) = f(y_N|F_{N-1})f(y_{N-1}|F_{N-2})\dots f(y_{r+1}|F_r)f(y_1, \dots, y_r|\alpha)$$

Así, usando la la función de log-verosimilitud podemos escribir la función con los parámetros a estimar

$$l(\alpha|y_t) = -\frac{1}{2}\log\sigma^2 - \frac{1}{2}\frac{\varepsilon_t^2}{\sigma^2}$$

Así, para N observaciones se tiene que la función de verosimilitud

$$L(\alpha|y_1, y_2, \dots, y_N) = \frac{1}{N} \sum_{t=1}^T l(\alpha|y_t)$$

La función log verosimilitud puede ser maximizada respecto a los vectores de parámetros desconocidos, realizada a través de algoritmos de optimización numérica, tales como Newton-Raphson, Scoring, Gauss-Newton, etc.

En algunas aplicaciones es más apropiado asumir que los ε_t tienen una distribución t-Student estandarizado, esto es, $\varepsilon_t = \frac{t_v}{\sqrt{v/(v-2)}}$ con $t_v \sim$ t-student con v grados de libertad. Así se tiene

$$f(\varepsilon_t|\mathbf{V}) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\pi}} \left(1 + \frac{\varepsilon_t^2}{v-2}\right)^{-(v+1)/2}$$

Con $\Gamma(a) = \int_0^\infty y^{a-1} e^{-y} dy$

Obteniendo así la función de verosimilitud $L(\alpha|y_1, y_2, \dots, y_N)$ y log-verosimilitud $l(\alpha|y_t)$, y maximizada respecto a los vectores de parámetros desconocidos, realizada a través de algoritmos de optimización numérica, tales como Newton-Raphson, Scoring, Gauss-Newton, etc.

4. Verificación

Para un proceso heterocedástico, los residuos del modelo:

$$\varepsilon_t = \frac{y_t}{\sigma_t}$$

son variables aleatorias independientes e idénticamente distribuidas con distribución normal o t-Student.

Para verificar la adecuación del modelo se calcula la estadística Q de Ljung-Box para la secuencia ε_t , además de eso el cálculo del coeficiente de asimetría, curtosis y un qq-plot son utilizados para probar la validación de la distribución normal. Para verificar si aún existe heterocedasticidad condicional en los residuos del modelo se aplica la prueba de Ljung-Box para la secuencia ε_t^2 .

5. Predicción

Considerando un modelo GARCH(p,q), para una serie de retornos y_1, y_2, \dots, y_T , sus pre-

dicciones puntuales hacia un horizonte k son denotados por y_{T+k} para $k > 0$, y sus predicciones puntuales para la volatilidad σ_{T+k}^2 , k pasos al frente son dados por:

$$\hat{\sigma}_{t+k}^2 = \omega + \sum_{i=1}^p \alpha_i \hat{y}_{T+k-i}^2 + \sum_{j=1}^q \beta_j \hat{\sigma}_{T+k-j}^2$$

Donde $\hat{\sigma}_{T+k-j}^2 = y_{T+k-j}^2$, si $k - j \leq 0$

Considerando un modelo EGARCH(p,q), para una serie de retornos y_1, y_2, \dots, y_T , sus predicciones puntuales hacia un horizonte k son denotados por y_{T+k} para $k > 0$, y sus predicciones puntuales para la volatilidad σ_{T+k}^2 , k pasos al frente son dados por:

$$\log(\hat{\sigma}_t^2) = \omega + \sum_{i=1}^q g(\hat{\varepsilon}_{T+k-i}) + \sum_{i=1}^p \beta_i \log(\hat{\sigma}_{T+k-i}^2)$$

Donde:

$$g(\hat{\varepsilon}_{T+k-i}) = \alpha_i \hat{\varepsilon}_{T+k-i} + \gamma(|\hat{\varepsilon}_{T+k-i}| - E(|\hat{\varepsilon}_{T+k-i}|)), i = 1, \dots, q,$$

Considerando un modelo TGARCH(p,q), para una serie de retornos y_1, y_2, \dots, y_T , sus predicciones puntuales hacia un horizonte k son denotados por y_{T+k} para $k > 0$, y sus predicciones puntuales para la volatilidad σ_{T+k}^2 , k pasos al frente son dados por:

$$\hat{\sigma}_{T+k}^2 = \omega + \sum_{i=1}^q \alpha_i \hat{\varepsilon}_{T+k-i}^2 + \sum_{i=1}^p \beta_i \hat{\sigma}_{T+k-i}^2 + \sum_{i=1}^q \gamma_i \hat{\varepsilon}_{T+k-i}^2 S_{T+k-i}$$

Donde:

$$S_{t-i} = \begin{cases} 1 & \text{si } \hat{\varepsilon}_{T+k-i} < 0 \\ 0 & \text{si } \hat{\varepsilon}_{T+k-i} \geq 0. \end{cases}$$

Herencia (1997) menciona que a medida que aumenta el horizonte de predicción k , la información de torna cada vez menos importante en el cálculo de predicción. Muestra que a medida que aumenta k la predicción de la volatilidad converge a la varianza incondicional de la serie σ^2 , así recomienda, dependiendo de los valores de los parámetros estimados realizar predicciones de la serie de tipo financiera hasta 40 observaciones.

Uno de los métodos más frecuentemente utilizados para obtener intervalos de predicción es el método asintótico. Así los intervalos de predicción para $(1 - \gamma)\%$ de y_{T+1} son dados por $\pm z_{\gamma/2} \sigma_{T+1}$, donde $z_{\gamma/2}$ es el $\gamma/2$ -cuantil de la densidad normal estándar. Para las predicciones hacia adelante $k > 1$, la distribución de los errores de predicción no son normales aun si

ε_t es normal. Sin embargo la aproximación del intervalo de predicción para un $(1 - \gamma)\%$ de los retornos y_{T+k} para $k > 1$ es dado por:

$$\pm z_{\gamma/2} E_T(\sigma_{T+k})$$

Sin embargo, la presencia de distribución con colas más pesadas que la distribución normal y de asimetría (hechos estilizados de los retornos), hace que los intervalos para y_{T+1} presentadas no sean adecuadas (Pascual et al., 2006).

Con respecto a los valores futuros de predicción para las volatilidades son dados por $E_T(\sigma_{T+k}^2)$, sin embargo la distribución del error de predicción para la varianza condicional no ha sido derivado, de esta manera, sus intervalos de predicción no pueden ser obtenidos (Pascual et al., 2006).

2.5.7. Retornos

Una serie de retornos es definido 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 son 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 leptocurtica (distribución no normal).
- En muchos casos, la distribución condicional también presenta colas pesadas.

2.6. Metodología Bootstrap

2.6.1. Introducción

El primero en considerar la técnica del bootstrap de una manera sistemática fue Efron. El nombre “bootstrapping” es usado por Efron para seguir las ideas de Tukey (1985) quien usó el nombre de “Jackknifing” para describir su método de remuestreo. Desde entonces, se ha convertido rápidamente en una popular y poderosa herramienta estadística usada para problemas difíciles en el análisis estadístico.

El bootstrap constituye la técnica más versátil y conocida dentro del método de remuestreo, su importancia radica en que es un método para estimar la distribución de una estadística con muestras finitas. Fernández (1999) menciona que debido a que el sustento teórico matemático-estadístico del bootstrap es bastante complejo, hasta finales de la década del '80, la eficiencia del método era probada de manera empírica, es decir, en el terreno de la práctica. Este método es computacionalmente intenso, sin embargo, las modernas computadoras son más que suficientes para los requerimientos computacionales requeridos para éste método.

Los métodos bootstrap son robustos debido a que no asumen ninguna distribución particular de la población y han sido aplicados a diversos problemas de muestreo, sustituyendo la complejidad del análisis teórico por el poder de cálculo de los ordenadores.

La esencia del “bootstrapping” es la idea que, en la ausencia de cualquier otro conocimiento sobre la distribución de una población, la distribución de valores encontrada en una muestra aleatoria de tamaño n de la población es la mejor información para aproximar su distribución de probabilidades. (Alvarez y Rubio, 2000).

2.6.2. Principio del Bootstrap

En la formulación formal del Bootstrap inicialmente propuesto por Efron (1979), 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 μ y de varianza σ^2 finitos, pero desconocidos.

Además 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 vamos a extraer 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.

Así se comienza por construir la función de distribución empírica \hat{F}_n , estimador de F , definida por:

$$\hat{F}_n(x) = \frac{\#\{i : X_i \leq x\}}{n}$$

Esta función se obtiene a partir de los valores observados, siendo atribuido el peso $\frac{1}{n}$ a cada una de las observaciones. Con \hat{F}_n fijo, definimos una muestra Bootstrap $(x_1^*, x_2^*, \dots, x_n^*)$, donde el símbolo * indica que hubo remuestreo con reposición de la muestra inicial.

La metodología Bootstrap se basa exactamente en la idea de producir un grande número de muestras a partir de la muestra original (población Bootstrap) y calcular el valor de la estadística de interés para cada una de ellas, y así estimar la distribución muestral del estadístico. Los resultados pueden generalizarse para la población. La ventaja del bootstrap es que no necesitamos conocer el proceso generador de los datos.

La aproximación bootstrap es válida para la mayoría de las estadísticas las cuales son consistentes. Si $B \rightarrow \infty$ entonces las estimaciones del error estándar, del sesgo y del error cuadrático medio se igualan a las estimaciones de máxima verosimilitud. (Efron y Tibshirani, 1993).

Para el cálculo de las estimaciones bootstrap es necesario un valor de B muy grande, generalmente a partir de B=1000 genera buenos resultados. (Efron y Tibshirani, 1993).

2.6.3. Aplicaciones del Bootstrap

Como el resultado del proceso descrito en líneas arriba, se pueden derivar al menos cuatro importantes y más usadas aplicaciones prácticas, detalladas a continuación:

a. Estimación Bootstrap del Error Típico

El bootstrap fue introducido por Efron y Tibshirani (1993) como un método basado en cálculos intensivos mediante un ordenador para estimar el error muestral de un estadístico. Explícitamente, la estimación bootstrap del error de un estadístico es como sigue:

- Se extraen B muestras bootstrap independientes de la función de distribución empírica \hat{F} .

- Se computa el estadístico de interés en cada una de las B muestras, obteniendo $\hat{\theta}_b^*$.
- Se estima el error muestral de $\hat{\theta}$ mediante la desviación estándar de la función de distribución obtenida de los B $\hat{\theta}_i^*$, es decir a través de $\hat{F}^*(\hat{\theta}^*)$.

$$\hat{s}e_B = \left\{ \sum_{b=1}^B \frac{[\hat{\theta}_b^* - \hat{\theta}^*(\cdot)]^2}{(B-1)} \right\}^{1/2}$$

Donde: $\hat{\theta}^*(\cdot) = \sum_{b=1}^B \frac{\hat{\theta}_b^*}{B}$

Este procedimiento se puede observar en la Figura 1, donde se muestra el algoritmo bootstrap para la estimación del error estándar de un estadístico $\hat{\theta} = s(x)$; cada muestra bootstrap es una variable aleatoria independiente de tamaño n de \hat{F} , y B es el número de repeticiones bootstrap.

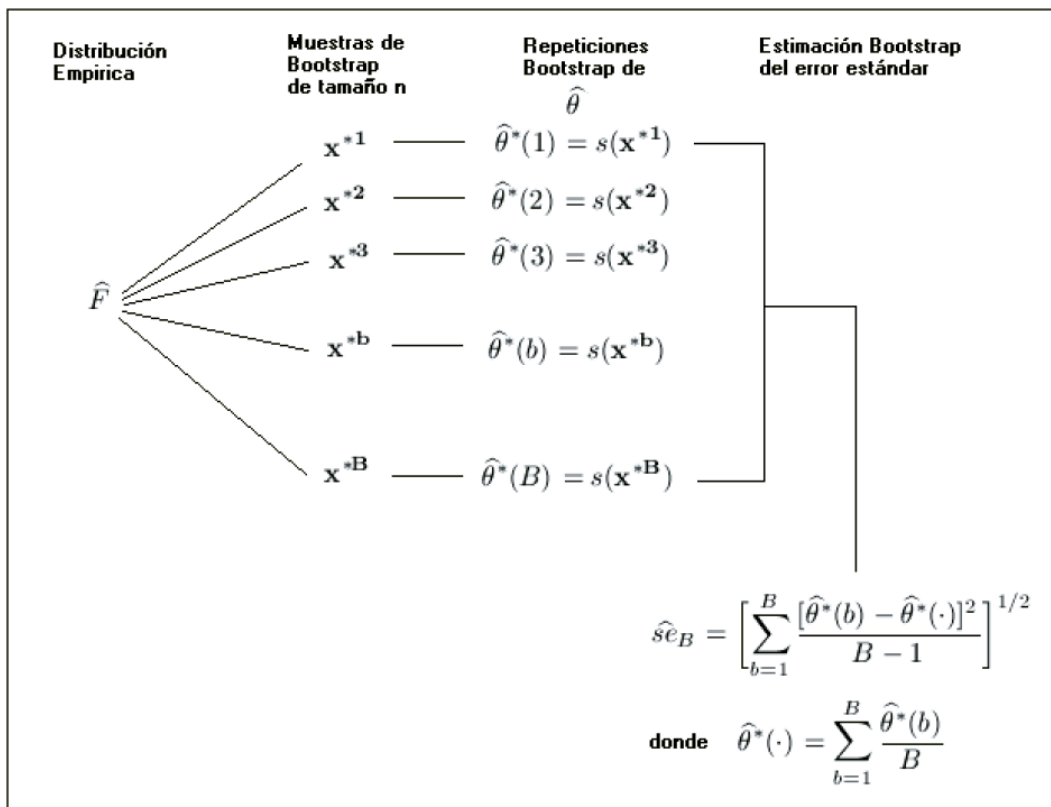


Figura 1: Esquema del algoritmo Bootstrap para la estimación del error estándar
Fuente: Efron y Tibshirani (1994).

b. Intervalo de Confianza Bootstrap

Miranda (2003), describe que existen 03 métodos frecuentemente utilizados a través de los cuales se pueden construir intervalos de confianza bootstrap y son:

- Método de aproximación normal
- Método de los percentiles
- Método del Bootstrapping estudentizado

El método de aproximación normal, utiliza la misma estructura de los procedimientos paramétricos de la construcción de intervalos de confianza. Si es posible asumir que el estadístico se distribuye según la curva normal pero el cálculo del error típico resulta analíticamente difícil o no existe fórmula para su cálculo, entonces podemos emplear la distribución muestral bootstrap para estimar el error típico o insertarlo en la correspondiente expresión del intervalo de confianza paramétrico. Así el intervalo de confianza bootstrap será de la forma

$$[\hat{\theta} - \hat{s}e_R z_{1-\frac{\alpha}{2}}; \hat{\theta} + \hat{s}e_R z_{1-\frac{\alpha}{2}}]$$

Donde: $\hat{s}e_R = \sqrt{\frac{\sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2}{B-1}}$, es el error estándar obtenido y $\hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B}$.

El método del percentil, hace uso literal de la idea básica del bootstrap, es decir $\hat{F}^*(\hat{\theta}^*)$ se aproxima a $F(\hat{\theta})$. La idea es muy simple: un intervalo con un nivel de confianza $1 - \alpha$ incluye todos los valores de $\hat{\theta}^*$ entre los percentiles $\frac{\alpha}{2}$ y $(1 - \frac{\alpha}{2})$ de la distribución de $\hat{F}^*(\hat{\theta}^*)$.

En este caso, el intervalo está dado de la forma:

$$[\hat{F}^{-1}(\alpha/2); \hat{F}^{-1}(1 - \alpha/2)]$$

Donde $\hat{F}^{-1}(\alpha/2)$ representa el percentil de $\alpha/2$ de la distribución del estimador $\hat{\theta}^*$, es decir un valor tal que la probabilidad acumulada hasta dicho valor sea $\alpha/2$, y $\hat{F}^{-1}(1 - \alpha/2)$ representa el percentil del $1 - \alpha/2$.

El método del percentil conserva la esencia no-paramétrica del enfoque bootstrap y libera al usuario de las asunciones de la estadística paramétrica.

El método del Bootstrapping estudentizado, La idea aquí es sustituir los percentiles de la distribución t por los percentiles de la distribución de los valores estudentizados del estadístico $\hat{\theta}$ en la muestras bootstrap. Es decir de los valores definidos por:

$$t^*(B) = \frac{\hat{\theta}^*(B) - \hat{\theta}}{\hat{s}e(\hat{\theta}^*(B))}$$

Donde $\hat{\theta}^*(B)$ y $\hat{s}e(\hat{\theta}^*(B))$ son los valores del estimador y valor estimado del error del estimador en la B-ésima muestra bootstrap respectivamente; $\hat{\theta}$ es el estimado en la muestra original.

Luego, el intervalo de confianza del $100(1 - \alpha)\%$ por bootstrapping estudentizado estará dado por:

$$\left(\hat{\theta} + \hat{F}_t^{-1}\left(\frac{\alpha}{2}\right)\hat{s}e; \hat{\theta} + \hat{F}_t^{-1}\left(1 - \frac{\alpha}{2}\right)\hat{s}e\right)$$

Donde $\hat{F}_t^{-1}\left(\frac{\alpha}{2}\right)$ representa el percentil de $\frac{\alpha}{2}$ de la distribución de t^* , es decir un valor tal que la probabilidad acumulada hasta dicho valor sea $\frac{\alpha}{2}$ y $\hat{F}_t^{-1}\left(1 - \frac{\alpha}{2}\right)$ representa el percentil del $1 - \frac{\alpha}{2}$.

c. Prueba de Hipótesis Bootstrap

La prueba de hipótesis para el caso bilateral es de la siguiente manera. Sea B el número de remuestras independientes de tamaño n de \hat{F} . Para cada remuestra b , $b = 1, \dots, B$ es calculado el valor de la estadística de prueba t_b^* rechazando H_0 cuando $|t_b^*| \geq c$ entonces el p-valor bootstrap utilizado puede ser aproximado por:

$$p_{boot} = \frac{1 + \#\{|t_b^*| \geq |t_{obs}|\}}{B + 1}$$

Utilizando este procedimiento para obtener el p-valor bootstrap, diferentes pruebas de hipótesis pueden ser realizados, como por ejemplo probar la hipótesis de diferencias de medias, probar la hipótesis de distribuciones. Entonces para probar esta hipótesis utilizando el procedimiento bootstrap descrito encima, tenemos que seguir los siguientes pasos:

1. Obtener B remuestras bootstrap de tamaño $n + m$, las primeras n observaciones son llamadas de z^* y las siguientes de y^*
2. Calcular el estadístico de prueba para cada una de las B remuestras

$$t_b^* = \bar{z}^* + \bar{y}^*, b = 1, \dots, B.$$

3. Obtener el p-valor bootstrap.

$$p_{boot} = \frac{1 + \#\{|t_b^*| \geq |t_{obs}|\}}{B + 1}$$

donde t_{obs} es el estadístico de prueba obtenida de la muestra original.

d. Modelo de Regresión Bootstrap

El análisis de modelos de regresión ocupa un papel muy importante en el análisis estadístico. Se estudian los efectos de variables explicativas o covariables sobre una variable respuesta. El modelo de regresión lineal es descrito en notación vectorial de la siguiente manera:

$$y_i = x_i^t \beta + \varepsilon_i, \quad i = 1, 2, \dots, n$$

con $x_i^t = (1, x_{i1}, x_{i2}, \dots, x_{i(p-1)})$ y $\beta^t = (\beta_0, \beta_1, \dots, \beta_{p-1})$

Los pasos a seguir para estimar el coeficientes de regresión β bootstrap son:

1. Calcular los estimadores MCO de β , $\hat{\beta}$, y estimar los residuales del modelo $\hat{\varepsilon}_i = y_i - x_i^t \hat{\beta}$.
2. Muestrear con reemplazo n errores $\hat{\varepsilon}_i$ desde $\hat{\varepsilon}_i$ y formar las n observaciones muestrales bootstrap desde $y_i^* = x_i^t \hat{\beta} + \hat{\varepsilon}_i^*$, para $i = 1, \dots, n$.
3. A partir de y_i^* y obtenidas desde la muestra (y^*, x) , calcular y guardar los estimadores MCO, $\hat{\beta}_{(b)}^*$.
4. Repetir los pasos 2 y 3, B veces, obteniendo $\hat{\beta}_{(1)}^*, \hat{\beta}_{(2)}^*, \dots, \hat{\beta}_{(B)}^*$

De tal forma que ahora ya se puede proseguir con la estimación de las estadísticas y estadísticas de prueba que se deseen.

2.6.4. Principales algoritmos Bootstrap en series de tiempo

A continuación se presentan las principales propuestas de las aplicaciones de la metodología Bootstrap en modelos para series temporales.

Stine (1987) propuso el primer método bootstrap para estimar el error cuadrático medio del predictor lineal clásico cuando el proceso es un modelo autorregresivo de orden p (AR(p)). El método se basaba en la obtención de la función de distribución empírica de los errores centrados. A partir de ésta extraía muestras para reproducir un proceso AR(p) con los parámetros estimados inicialmente y fijando los primeros p valores de la serie. Dada la serie bootstrap se estiman los parámetros y se obtiene el predictor clásico para evaluar el error cuadrático medio de dicha predicción. Repitiendo un número elevado de veces este proceso es posible aproximar el error cuadrático medio de la predicción puntual.

Thombs y Schucany (1990) propusieron la construcción de intervalos de predicción en modelos AR(p), propusieron realizar una representación hacia atrás del proceso para posteriormente utilizar la representación habitual para obtener los valores futuros. El esquema de dicho método es el siguiente

1. Construir los residuales de la representación backward y calcular la función de distribución empírica de dichos errores centrados y reescalados según propuso Stine (1987).

2. Extraer los errores bootstrap, ε_i^* a partir de dicha distribución.

3. Se define la serie backward de la serie fijando los últimos p valores observados

$$Y_i^* = \hat{\phi}_1 Y_{i+1}^* + \dots + \hat{\phi}_p Y_{i+p}^* + \varepsilon_i^*, i = t - p, t - p - 1, \dots, 1$$

$$Y_i^* = Y_i, i = t - p + 1, \dots, t$$

4. Calcular los estimadores de la serie bootstrap.

5. Construir los residuales hacia adelante de la serie y obtener la función de distribución empírica de dichos errores centrados y reescalados.

6. Extraer los errores bootstrap, a_i^* , a partir de dicha distribución.

7. Definir las réplicas bootstrap de los valores futuros

$$Y_{t+i}^* = \hat{\phi}_1 Y_{t+i-1}^* + \dots + \hat{\phi}_p Y_{t+i-p}^* + a_i^*, i = 1, \dots, s$$

$$Y_{t+i}^* = Y_{t+i}, \text{ si } i \leq 0$$

8. Repetir un número B elevado de veces los pasos 2-7 para aproximar la distribución de Y_{t+s}^* . Los cuantiles correspondientes a esa distribución serán los límites del intervalo de predicción.

Estos autores demostraron la validez asintótica de este método bootstrap y realizaron un estudio de simulación donde comparan con los intervalos clásicos para diferentes distribuciones y coberturas.

McCullough (1994) sugirió que el método percentil de obtener los intervalos de predicción no era el más adecuado cuando existe asimetría en el error de predicción y, por lo tanto, propuso un método de corrección de sesgo. Además, advirtió que la representación backward utilizada por Thombs y Schucany (1990) no asegura que los errores sean independientes e

idénticamente distribuidas y utilizó un procedimiento más sofisticado para obtener dichos errores hacia atrás.

Un inconveniente importante del método Thombs y Schucany (1990) consiste en la representación backward de la serie, no sólo por el gasto computacional sino porque no es fácil cuando los procesos no tienen distribución normal.

Cao et al. (1997) proponen un método alternativo, que es consistente, para construir intervalos de predicción sin tener que realizar réplicas hacia atrás de la serie. El método se puede describir en los siguientes pasos

1. Obtener los residuos forward centrados y reescalados para construir la función de distribución empírica.
2. Extraer una muestra aleatoria, a_i^* , a partir de dicha distribución.
3. Construir los valores futuros bootstrap utilizando como errores la muestra anterior y las estimaciones iniciales de los parámetros.

$$Y_{t+i}^* = \hat{\phi}_1 Y_{t+i-1}^* + \dots + \hat{\phi}_p Y_{t+i-p}^* + a_i^*, i = 1, \dots, s$$

$$Y_{t+i}^* = Y_{t+i}, \text{ si } i \leq 0$$

4. Repetir las etapas anteriores un número elevado de veces para aproximar la distribución bootstrap del valor Y_{t+s}^* .

Cao et al. (1997) demostraron la consistencia de este método tanto en el caso de utilizar la función de distribución empírica de los residuos directamente como en el de utilizar un bootstrap suavizado, es decir, cuando en la etapa 2 se extrae la muestra a partir de una suavización no paramétrica de la distribución.

Finalmente, cabe destacar que Pascual et al. (1999) propusieron un método que capta el error de predicción debido a la incertidumbre en la estimación pero que no es necesaria realizar una representación backward de la serie. El esquema de remuestreo para un modelo AR(p) se describe bajo las siguientes etapas

1. Obtener los residuos forward centrados y reescalados para construir la función de distribución empírica.
2. Extraer una muestra aleatoria, a_i^* , a partir de dicha distribución.

3. Generar una serie con la misma estructura AR(p) utilizando los errores bootstrap y estimar los parámetros bootstrap de esta serie.
4. Construir los valores bootstrap futuros utilizando como errores una muestra obtenida como en la etapa 2 y las estimaciones bootstrap de los parámetros. En esta etapa se fijan los últimos p valores de la serie original para construir los valores bootstrap futuros.
5. Repetir las etapas anteriores un número elevado de veces para aproximar la distribución bootstrap del valor Y_{t+s}^* .

2.6.5. Metodología Bootstrap de predicción para modelos heterocedásticos

Pascual et al. (2006) propone intervalos de predicción para los modelos GARCH mediante la metodología bootstrap. El algoritmo propuesto por Pascual et al. (2006) incorpora una componente de incerteza debido a la estimación de los parámetros para obtener los intervalos de predicción resultantes para los retornos y volatilidades de los modelos GARCH.

Se presenta a continuación el procedimiento bootstrap utilizado en series de tiempo por el algoritmo propuesto por Pascual et al. (2006) para modelos $GARCH(1, 1)$, con el supuesto de que la especificación del modelo es conocida. Y las extensiones para los modelo EGARCH y TGARCH. Con el objetivo es estimar directamente la distribución de y_{T+k} y σ_{T+k} condicionalmente de los datos existentes, la construcción de los intervalos de predicción.

a. Intervalo de predicción Bootstrap para GARCH

Dado $\{y_1, y_2, \dots, y_T\}$ una secuencia de T observaciones generadas por un proceso $GARCH(1, 1)$.

Los parámetros desconocidos $\theta = (w, \alpha, \beta)$ son estimados por máxima verosimilitud $\hat{\theta}_T = (\hat{w}, \hat{\alpha}, \hat{\beta})$, los residuales son calculados por $\hat{\varepsilon}_t = \frac{y_t}{\hat{\sigma}_t}$, para $t = 1, 2, \dots, T$ donde

$$\hat{\sigma}_t^2 = \hat{w} + \hat{\alpha}y_{t-1}^2 + \hat{\beta}\hat{\sigma}_{t-1}^2$$

Son las varianzas condicionales estimadas para $t=2,3,\dots,T$, y $\hat{\sigma}_1^2 = \frac{\hat{w}}{1 - \hat{\alpha} - \hat{\beta}}$ es la varianza marginal estimada.

Para implemetar la técnica bootstrap, es necesario obtener replicas bootstrap $Y_T^* = \{y_1^*, \dots, y_T^*\}$ que simula la estructura original de la serie. Estas réplicas son obtenidas de la siguiente

ecuaciones:

$$\hat{\sigma}_t^{*2} = \hat{w} + \hat{\alpha} y_{t-1}^{*2} + \hat{\beta} \hat{\sigma}_{t-1}^{*2},$$

$$y_t^* = \varepsilon_t^* \hat{\sigma}_t^*$$

para $t=1,2,\dots,T$, donde ε_t^* son muestras aleatorias con reemplazo de la distribución empírica \hat{F}_T de los residuales centrados, y $\hat{\sigma}_1^{*2} = \hat{\sigma}_1^2$.

Una vez estimados los parámetros de la serie bootstrap $\hat{\theta}_T^* = (\hat{w}^*, \hat{\alpha}^*, \hat{\beta}^*)$, mediante el método de máxima verosimilitud mostrada en la sección 2.5.6, las predicciones bootstrap de valores futuros son obtenidos mediante:

$$\hat{\sigma}_{T+k}^{*2} = \hat{w}^* + \hat{\alpha}^* y_{T+k-1}^{*2} + \hat{\beta}^* \hat{\sigma}_{T+k-1}^{*2},$$

$$y_{T+k}^* = \varepsilon_{T+k}^* \hat{\sigma}_{T+k}^*$$

para $k = 1, 2, \dots$, con ε_{T+k}^* muestras aleatorias con reemplazo de \hat{F}_T , $y_T^* = y_T$ y

$$\hat{\sigma}_T^{*2} = \frac{\hat{w}^*}{1 - \hat{\alpha}^* - \hat{\beta}^*} + \hat{\alpha}^* \sum_{j=0}^{T-2} \hat{\beta}^{*j} (y_{T-j-1}^2 - \frac{\hat{w}^*}{1 - \hat{\alpha}^* - \hat{\beta}^*})$$

Aunque $\hat{\sigma}_T^{*2}$ es diferente para toda replica bootstrap, el valor es obtenido usando los parámetros estimados bootstrap.

Una vez obtenido el conjunto de B replicas bootstrap $(y_{T+k}^{*(1)}, \dots, y_{T+k}^{*(B)})$ para y_{T+k} , los límites son definidos como los cuantiles de la función de distribución bootstrap de y_{T+k}^* . Específicamente, si $G_y^*(h) = Pr(y_{T+k}^* \leq h)$ es la función de distribución de y_{T+k}^* y su estimación es dado por $G_{y,B}^*(h) = \#(y_{T+k}^{*b} \leq h) / B$. Entonces el intervalo de predicción al $100(1-\gamma)\%$ para y_{T+k}^* está dado por:

$$[L_{y,B}^*(y), U_{y,B}^*(y)] = [Q_{y,B}^*(\frac{\gamma}{2}), Q_{y,B}^*(1 - \frac{\gamma}{2})]$$

Donde $Q_{y,B}^* = G_{y,B}^{*-1}$

De igual manera se obtienen los intervalos de predicción para la volatilidad k períodos en el futuro. Dado un conjunto $(\hat{\sigma}_{T+k}^{*(1)}, \dots, \hat{\sigma}_{T+k}^{*(B)})$ de replicas bootstrap de la volatilidad para algún horizonte k, usando como límites de predicción los cuantiles de la función de distribución bootstrap de $\hat{\sigma}_{T+k}^*$. En este caso, si $G_{\sigma}^*(h) = Pr(\hat{\sigma}_{T+k}^* \leq h)$ es la función de

distribución de $\hat{\sigma}_{T+k}^*$ y su estimación es dado por $G_{\sigma,B}^*(h) = \#(\hat{\sigma}_{T+k}^{*b} \leq h)/B$, el intervalo de predicción al $100(1-\gamma)\%$ para $\hat{\sigma}_{T+k}^*$ está dado por:

$$[L_{\sigma,B}^*(\sigma), U_{\sigma,B}^*(\sigma)] = [Q_{\sigma,B}^*(\frac{\gamma}{2}), Q_{\sigma,B}^*(1 - \frac{\gamma}{2})]$$

Donde $Q_{\sigma,B}^* = G_{\sigma,B}^{*-1}$.

Resumiendo, los pasos para obtener **intervalos de predicción bootstrap para el proceso GARCH(1,1)** son:

- a) Calcular las estimaciones de los parámetros generados por el proceso GARCH $\theta = (w, \alpha, \beta)$ dadas por: $\hat{\theta} = (\hat{w}, \hat{\alpha}, \hat{\beta})$ y los residuales centrados $\hat{\epsilon}_t$, denotando \hat{F}_T como su función de distribución empírica.
- b) A partir de $\hat{\sigma}_t^{*2}$ y y_t^* , generar la serie bootstrap y_t^* , ajustarla y obtener estimaciones bootstrap $\hat{\theta}^* = (\hat{w}^*, \hat{\alpha}^*, \hat{\beta}^*)$, mediante el método de máxima verosimilitud mostrada en la sección 2.5.6.
- c) Calcular los valores futuros bootstrap de los retornos y de las volatilidades para un horizonte k mediante $\hat{\sigma}_{T+k}^{*2}$ y y_{T+k}^* .
- d) Repetir los pasos 2 y 3, B veces.
- e) Finalmente calcular B replicas bootstrap $(y_{T+k}^{*(1)}, \dots, y_{T+k}^{*(B)})$ y $(\sigma_{T+k}^{*(1)}, \dots, \sigma_{T+k}^{*(B)})$ para y_{T+k} y σ_{T+k} respectivamente. Los intervalos de predicción bootstrap son calculados por los cuantiles de $G_{y,B}^*(h)$ y $G_{\sigma,B}^*(h)$ mediante las funciones de distribución bootstrap de y_{T+k}^* y $\hat{\sigma}_{T+k}^*$ respectivamente.

b. Intervalo de predicción Bootstrap para EGARCH(1,1)

los pasos para obtener intervalos de predicción bootstrap para procesos EGARCH son:

- a) Calcular las estimaciones de los parámetros generados por el proceso EGARCH $\theta = (w, \alpha, \beta, \gamma)$ dadas por: $\hat{\theta} = (\hat{w}, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$ y los residuales centrados $\hat{\epsilon}_t$, denotando \hat{F}_T como su función de distribución empírica.
- b) A partir de $\log(\hat{\sigma}_t^{*2})$ y y_t^* , generar la serie bootstrap y_t^* , ajustarla y obtener estimaciones bootstrap $\hat{\theta}^* = (\hat{w}^*, \hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^*)$, mediante el método de máxima verosimilitud mostrada en la sección 2.5.6.

- c) Calcular los valores futuros bootstrap de los retornos y de las volatilidades para un horizonte k mediante $\log(\hat{\sigma}_{T+k}^{*2})$ y y_{T+k}^* .
- d) Repetir los pasos 2 y 3, B veces.
- e) Finalmente calcular B replicas bootstrap $(y_{T+k}^{*(1)}, \dots, y_{T+k}^{*(B)})$ y $(\log(\sigma_{T+k}^{*(1)}), \dots, \log(\sigma_{T+k}^{*(B)}))$ para y_{T+k} y $\log(\sigma_{T+k})$ respectivamente. Los intervalos de prediccion bootstrap son calculados por los cuantiles de $G_{y,B}^*(h)$ y $G_{\sigma,B}^*(h)$ mediante las funciones de distribución bootstrap de y_{T+k}^* y $\log(\hat{\sigma}_{T+k}^*)$ respectivamente.

c. Intervalo de predicción Bootstrap para TGARCH(1,1)

los pasos para obtener intervalos de predicción bootstrap para procesos TGARCH son:

- a) Calcular las estimaciones de los parámetros generados por el proceso GARCH $\theta = (w, \alpha, \beta, \gamma)$ dadas por: $\hat{\theta} = (\hat{w}, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$ y los residuales centrados $\hat{\epsilon}_t$, denotando \hat{F}_T como su función de distribución empírica.
- b) A partir de $\hat{\sigma}_t^{*2}$ y y_t^* , generar la serie bootstrap y_t^* , ajustarla y obtener estimaciones bootstrap $\hat{\theta}^* = (\hat{w}^*, \hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^*)$, mediante el método de máxima verosimilitud mostrada en la sección 2.5.6.
- c) Calcular los valores futuros bootstrap de los retornos y de las volatilidades para un horizonte k mediante $\hat{\sigma}_{T+k}^{*2}$ y y_{T+k}^* .
- d) Repetir los pasos 2 y 3, B veces.
- e) Finalmente calcular B replicas bootstrap $(y_{T+k}^{*(1)}, \dots, y_{T+k}^{*(B)})$ y $(\sigma_{T+k}^{*(1)}, \dots, \sigma_{T+k}^{*(B)})$ para y_{T+k} y σ_{T+k} respectivamente. Los intervalos de prediccion bootstrap son calculados por los cuantiles de $G_{y,B}^*(h)$ y $G_{\sigma,B}^*(h)$ mediante las funciones de distribución bootstrap de y_{T+k}^* y $\hat{\sigma}_{T+k}^*$ respectivamente.

III. MATERIALES Y MÉTODOS

3.1. Materiales

Los materiales y equipos de los cuales se hizo uso en la presenta tesis son los siguientes:

- Textos de consulta, monografías y artículos especializados.
- Una computadora de escritorio, con un procesador *Intel Core i5*, 3.30 GHz, con memoria *RAM* de 8GB y con un sistema operativo 8.1 de 64 bits.
- *Microsoft Excel* 2013.
- RStudio versión 0.98.873 con sus paquetes *forecast*, *tseries*, *timeSeries* y *rugarch*.
- Eviews versión 8.
- Procesador de texto LaTeX.

3.2. Metodología de la investigación

3.2.1. Tipo de investigación

El tipo de investigación es descriptivo de corte longitudinal retrospectivo, ya que se inferirán los retornos de los índices de la Bolsa de Valores de Lima, mediante intervalos de predicción en el período 2010 al 2014.

3.2.2. Diseño de la investigación

Es una investigación no experimental con diseño longitudinal, debido a que se obtuvieron los datos del mercado bursátil de los índices históricos de la página web de la Bolsa de Valores de Lima, *www.bvl.com.pe*. en el período 2010 al 2014.

3.2.3. Formulación de la hipótesis para los datos aplicados

- El modelo con el que se obtiene un intervalo de predicción adecuado de los retornos del Índice General de la Bolsa de Valores de Lima en el periodo 2010 al 2014 es el modelo GARCH con $\varepsilon_t \sim t(5)$.
- Las predicciones por intervalos bootstrap para el mes de diciembre del 2014 de los retornos del Índice General de la Bolsa de Valores de Lima en el periodo enero del 2010 a noviembre del 2014, presentan mejores comportamientos en cobertura y amplitud que los intervalos de predicción paramétricos, en cada modelo.

3.2.4. Identificación de las variables

Las variables con las que se contaron en la investigación fueron:

- **Índice General de la Bolsa de Valores de Lima (IGBVL)**

El Índice General de la Bolsa de Valores de Lima (IGBVL), 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 Lima. Integrada por un conjunto de valores que concentra el 80% de la negociación del mercado.

La cartera del IGBVL vigente a partir del año 2010 al 2015 está formada por las siguientes empresas:

1. Volcan B
2. Rio Alto Mining
3. Ferreycorp
4. Cerro Verde
5. ADR Buenaventura
6. Maple Energy
7. Graña Montero
8. Relapasa
9. Minsur Inv.

10. Credicorp
11. Alicorp
12. Trevali Mining
13. Casagrande
14. Bco. Continental
15. Intergr. Financ Services
16. Southern
17. Unión and Cementos
18. Corp. Aceros Areq. Inv.
19. El Brocal
20. Milpo
21. SIDERPERU
22. Cementos Pacasmayo
23. Luz del Sur
24. Atacocha B
25. Panoro Minerales
26. Candente
27. Austral Group
28. EDEGEL
29. Scotiabank
30. Telefónica B
31. Sulliden Gold
32. EDELNOR
33. Luna Gold
34. Cemento Andino B
35. Agro Ind Pomalca

El cálculo del IGBVL es realizado mediante la ecuación:

$$IGBVL = \sum_{i=1}^{35} \left[\left(\frac{P_{it}}{P_{i0}} \right) w_i F_i \right]$$

Donde:

i : índice de cada cartera $i=1,2,\dots,35$

P_{it} : Precio del valor de la empresa i en la fecha t .

P_{i0} : Precio del valor de la empresa i en la fecha 0.

w_i : Ponderación del valor de la cartera i .

F_i : Factor de ajuste por eventos aplicables a la cotización del valor de la empresa i .

■ Retornos del Índice General de la Bolsa de Valores de Lima (IGBVL)

Variable de tipo cuantitativa continua, que representa una transformación del Índice General de la Bolsa de Valores de Lima. ver sección 2.4.6.

3.2.5. Población

Está conformada por los precios diarios de cierre de activos del mercado bursátil de la bolsa de valores de Lima, del 4 de enero del 2010 al 31 de diciembre del 2014, con un total de 1303 observaciones. Se consideró este periodo pues a partir del 2015 la metodología de cálculo de la cartera del Índice General de la Bolsa de Valores no es la misma, siendo sustituido por el nuevo índice SP/BVL Perú General Index (SPBLPGPT).

3.2.6. Datos de entrenamiento

Está conformada por los precios diarios de cierre de activos del mercado bursátil de la bolsa de valores de Lima, del 4 de enero del 2010 al 30 de noviembre del 2014, con un total de 1283 observaciones.

3.2.7. Datos de prueba

Está conformada por los precios diarios de cierre de activos del mercado bursátil de la bolsa de valores de Lima, del 1 de diciembre del 2014 al 31 de diciembre del 2014, con un total de 20 observaciones. Considerando lo mencionado en la sección 2.5.6 acerca de las limitaciones de predicciones en series de tipo financieras.

3.2.8. Metodología aplicada

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

Los pasos que se realizaron para poder contrastar la hipótesis de la presente investigación para los retornos de los índices general de la bolsa de valores de Lima fueron los siguientes:

1. Análisis exploratorio

Se realizaron los siguientes pasos:

- Representación gráfica de la serie
- Histograma de frecuencias
- Test de estacionariedad de Dickey y Fuller

2. Identificación de modelos

- Se analizaron las funciones de autocorrelación (FAC) y funciones de autocorrelación parcial (FACP).
- Se obtuvo el mejor modelo de la serie de retornos mediante la comparación del criterio de Akaike (AIC) y criterio de Schwarz (BIC). Se obtuvieron los parámetros de los modelos heterocedásticos GARCH(p,q), EGARCH(p,q) y TGARCH(p,q).

3. Estimación

- Se realizó la estimación de los parámetros de los modelos mediante el método de máxima verosimilitud.

4. Diagnóstico de modelos

Para la adecuación de los modelos y los datos se analizó los residuales de cada modelo sugerido, se observó en que medida los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco. Para esto se procedió a analizar lo siguiente:

- Funciones de autocorrelación (FAC) y funciones de autocorrelación parcial (FACP)
- Test de Ljung-Box (Q)
- Histograma de frecuencia de los residuales

- Test de normalidad de Jarque y Bera

5. Predicción y comparación de modelos

Para realizar la comparación de intervalos de predicción en los modelos GARCH, EGARCH y TGARCH, se tuvieron en consideración lo siguiente:

- Las predicciones para los retornos y las volatilidades fueron calculados para el mes de diciembre del 2014 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 ε_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 2014 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.
- En los Anexos 13, 14 y 15 son mostrados los programas para el cálculo de las predicciones de los retornos y volatilidades del proceso ARMA(1,1)-GARCH(1,1), ARMA(1,1)-EGARCH(1,1) y ARMA(1,1)-TGARCH(1,1) desarrollado con el *software RStudio*.

IV. RESULTADOS Y DISCUSIÓN

4.1. Análisis exploratorio

Inicialmente se observó la presencia de una observación atípica, tal como se muestra en el Anexo 1, tal observación es referida al día 06 de junio del 2011, que refiere al cierre inesperado de la Bolsa de Valores, este acontecimiento se debe a la caída de la Bolsa de Valores de Lima, considerada la peor caída en la historia del mercado bursátil, esto de acuerdo al historial del mercado bursátil de Lima. Por lo tanto, para poder capturar este efecto se creó una variable dummy (D_t), la cual toma el valor de 1 el día referido y cero en el resto de casos. Así, para cada análisis posterior se consideró la presencia de la variable dummy (D_t).

De la figura 2 (a) de los índices de precios, se observaron la presencia de alzas significativas durante los últimos meses del año 2010, caídas inesperadas en el 2011, caídas significativas durante los primeros meses del año 2013 y a partir de mediados del 2013 se mantuvieron bajos los precios de cierre de las acciones a comparación de los años anteriores. En la evolución de los retornos de la figura 2 (b) 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 el prueba de raíz unitaria de Dickey y Fuller con valor del estadístico $\tau = -31,813$ y de probabilidad $p = 0,00$, rechazando la hipótesis de existencia de raíces unitarias en la serie, es decir la serie de retornos r_t es estacionaria.

De la figura 3, 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 (Skewness) con valor $s = -0,1492$, indicando que las distribuciones de los retornos de la media está sesgada ligeramente hacia la izquierda, el coeficiente de curtosis (Kurtosis) con valor $k = 8,5640 > 3$, indicando que los retornos siguen una distribución leptocúrtica. Por la prueba de normalidad de Jarque-Bera se observó el valor del estadístico $JB=1685.635$ y valor de probabilidad $p=0.000$, rechazando la hipótesis de normalidad de los

retornos. Realizando la comparación de la densidad de los retornos con la densidad normal, se observó la además la presencia de colas largas en su distribución. Así, concluyendo que los retornos no se distribuyen normalmente.

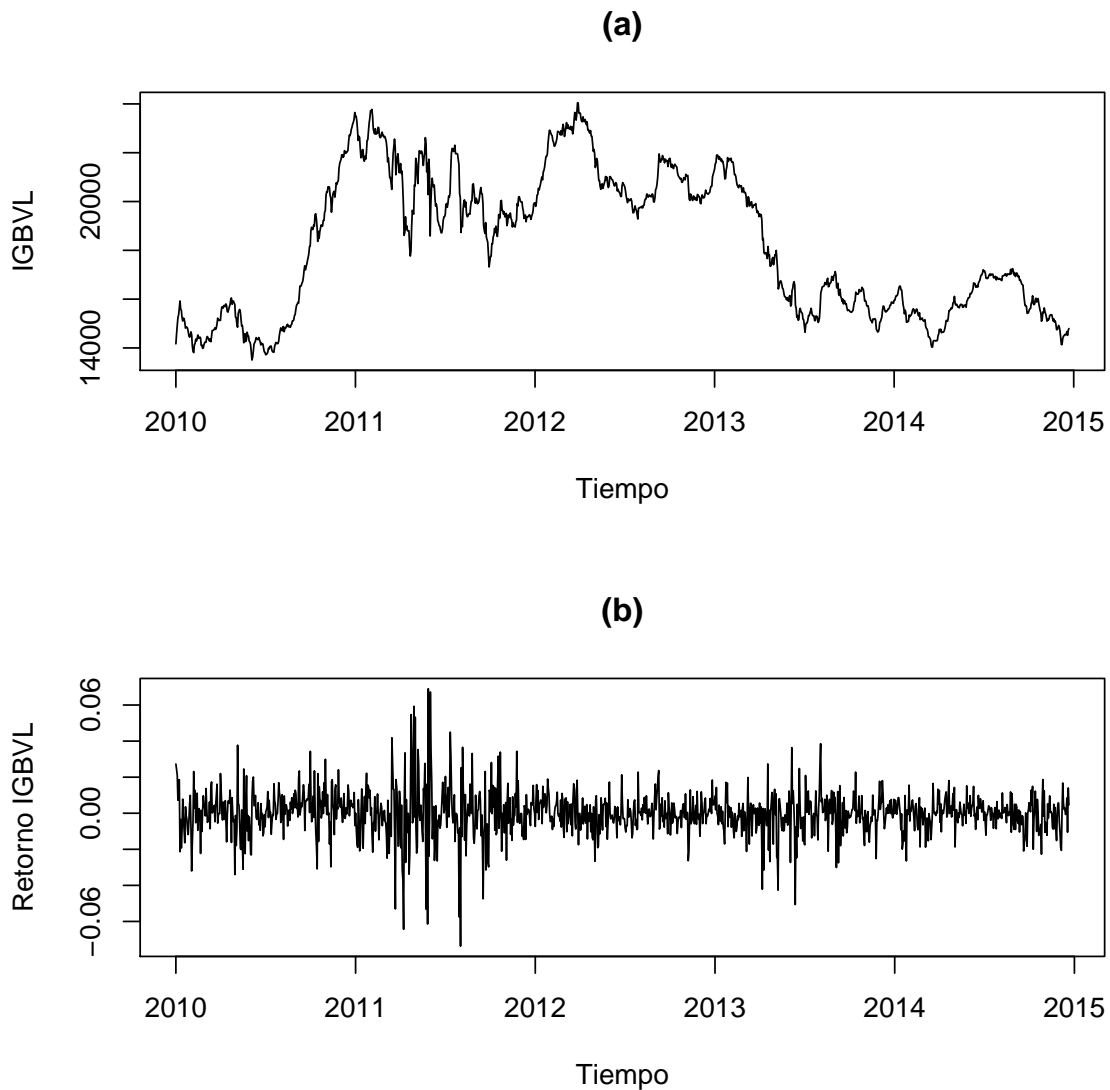


Figura 2: (a) Evolución diaria del Índice General de precios de las acciones de la Bolsa de Valores de Lima, del 4 de enero de 2010 al 31 de diciembre de 2014 y (b) Evolución de los retornos diarios de los precios de las acciones del IGBVL, del 5 de enero de 2010 al 31 de diciembre de 2014.

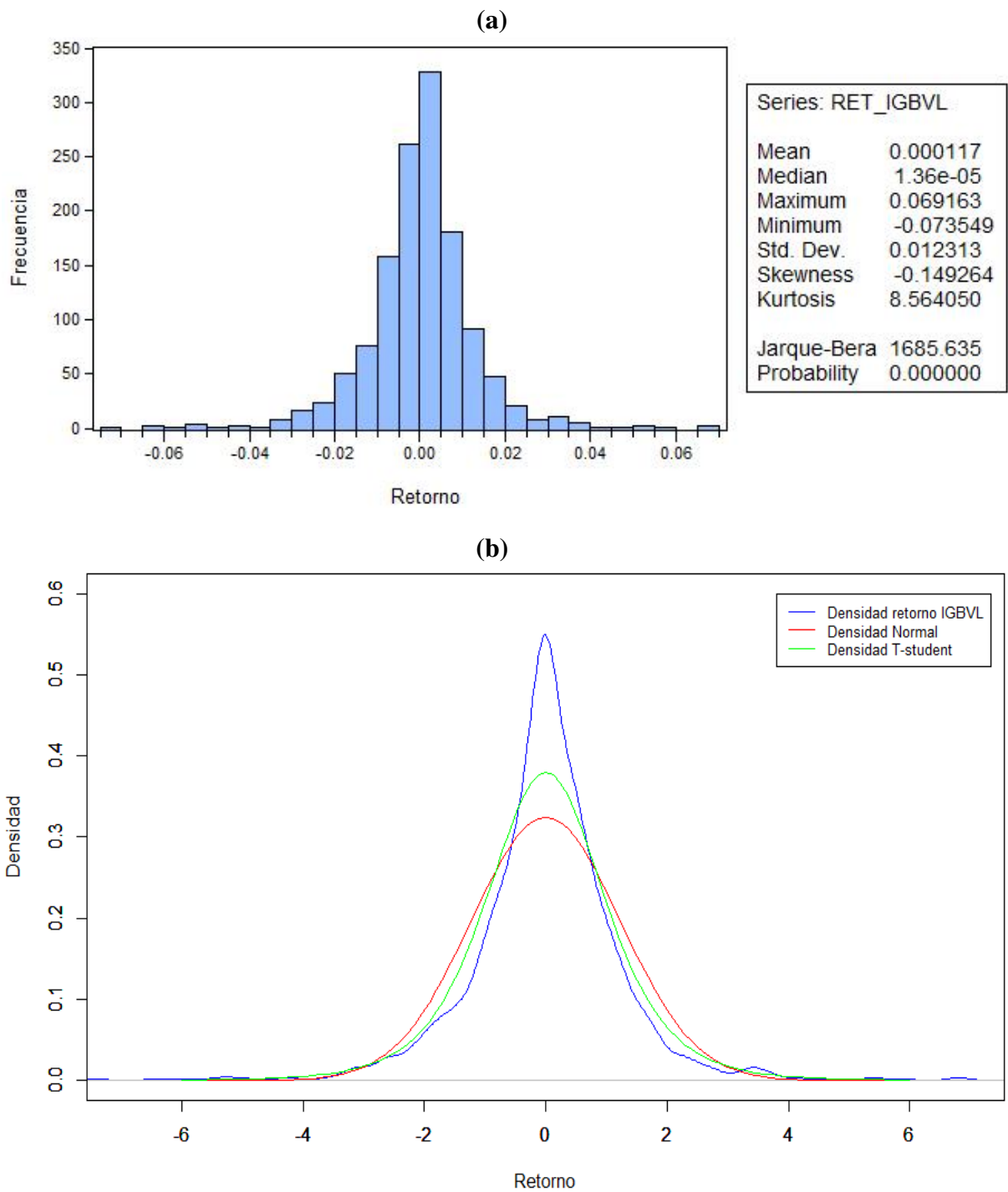


Figura 3: (a) Histograma de frecuencias y principales medidas descriptivas (b) Densidad normal estándar, t-Student y de los retornos diarios de los precios de las acciones del Índice General de la Bolsa de Valores de Lima. del 5 de enero de 2010 al 30 de octubre de 2014.

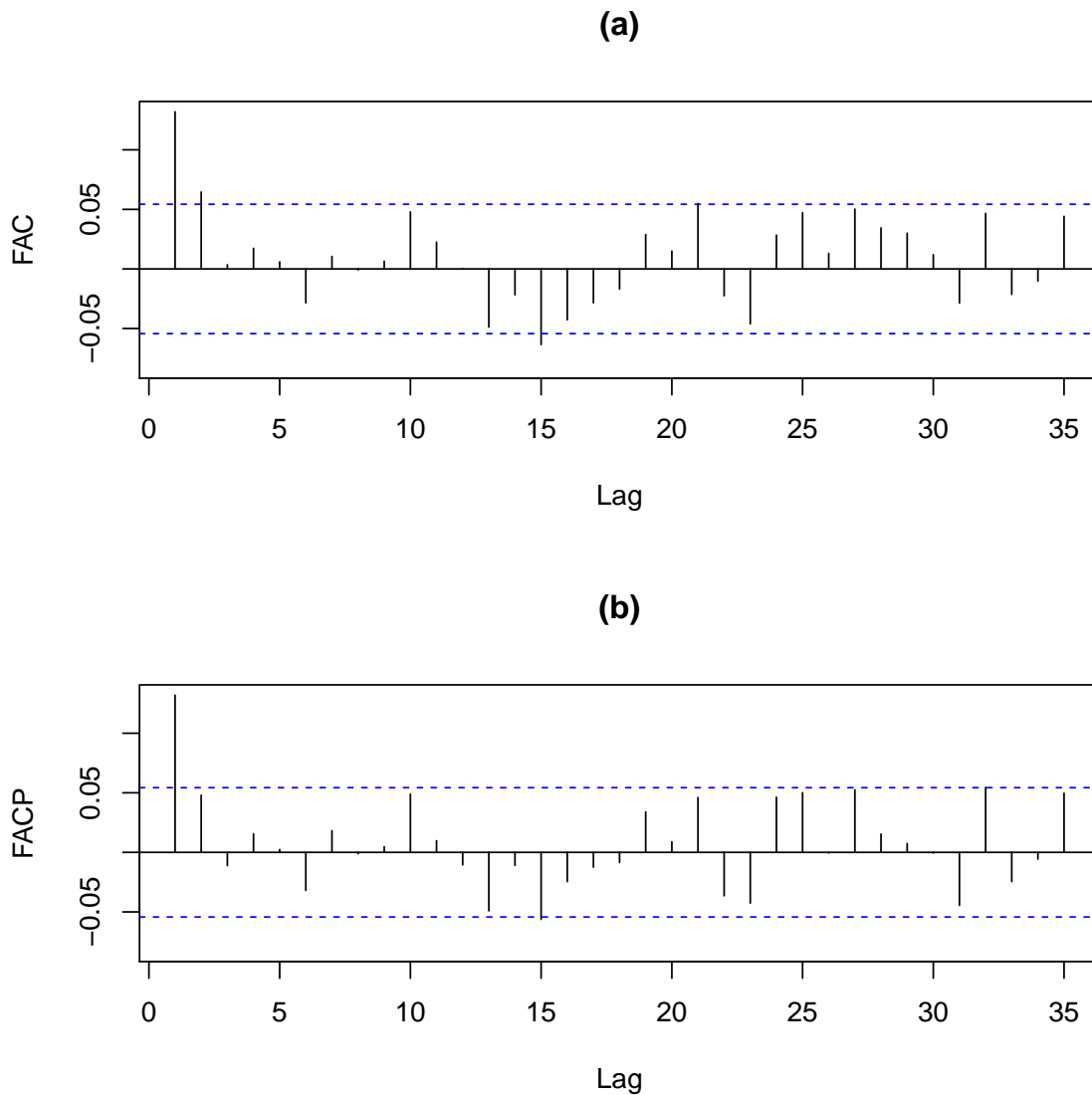


Figura 4: (a) Función de Autocorrelación (fac) y (b) Función de Autocorrelación Parcial (facp), de los retornos del IGBVL (r_t).

Se observó mediante la prueba de autocorrelación de Ljung y Box, mostrados en el Anexo 2, que la serie de retornos r_t con valores de estadístico Q grandes y de probabilidad menores a 0.000 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, dado que los valores de sus funciones de autocorrelación y autocorrelación parcial presentan valores alternados positivos y negativos. De la Figura 4, correspondientes a sus fac y pacf de r_t es posible que el modelo fuese ajustado a un $AR(1)$ o incluso a un $MA(1)$, sin embargo, con el fin de no perder información pues con la serie ajustada se evaluará la

volatilidad, se obtuvo que el modelo apropiado para r_t es un ARMA(1,1), definido por:

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

Donde: D_t es la variable dummy y b_t es el residuo del modelo.

Por la prueba de autocorrelación de Ljung y Box para b_t , ver Anexo 2, se tiene 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, ver Anexo 2, 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 rechaza la hipótesis de existencia de autocorrelación para b_t^2 , es decir se observó la presencia de heterocedasticidad condicional en b_t . Evidenciado también por sus fac y pacf de b_t^2 , mostrado en la Figura 5.

4.2. Análisis para el modelo GARCH

4.2.1. Identificación

Mediante los criterios AIC y BIC, mostrados en el Anexo 3, se determinó que los parámetros de un modelo GARCH que presenta un mejor ajuste para la serie de retornos, es el modelo ARMA(1,1)-GARCH(1,1) con valores del AIC=-6.248787 y BIC=-6.228926 y considerando el modelo más parsimonioso. Siendo representado por las siguientes ecuaciones.

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

$$\sigma_t^2 = \omega_0 + \alpha b_{t-1}^2 + \beta \sigma_{t-1}^2$$

Con residuales $\varepsilon_t = \frac{b_t}{\sigma_t}$ considerando su distribución como una normal estandar o una t-student con 5 grados de libertad.

4.2.2. Estimación

a. **Suponiendo que $\varepsilon_t \sim N(0, 1)$**

Se observó que en el modelo ajustado, todos los parámetros estimados son altamente significativos, tal como se muestran en el Anexos 5.

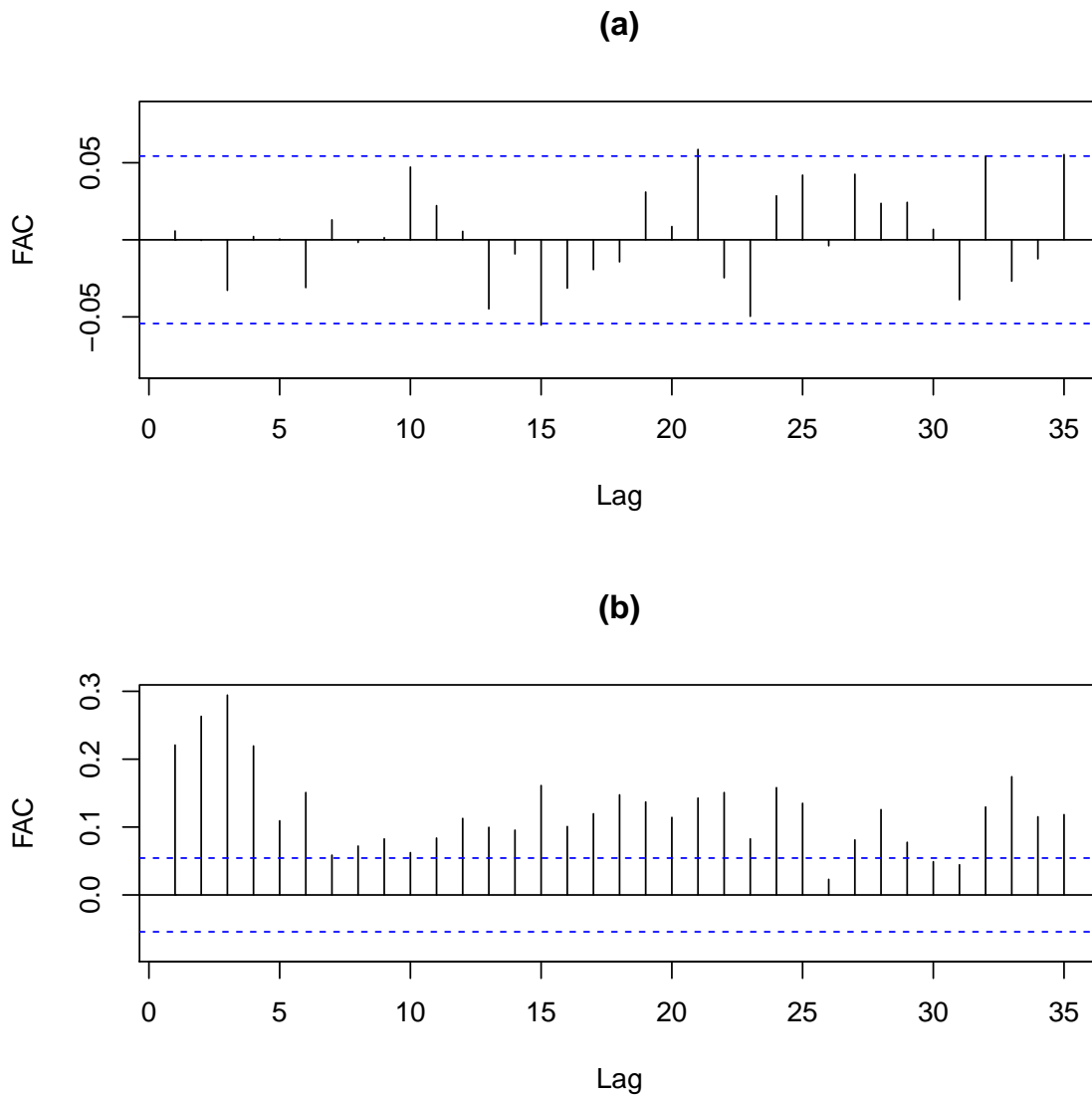


Figura 5: (a) Función de Autocorrelación (fac) de la serie filtrada b_t y (b) Función de Autocorrelación (fac) de los cuadrados de la serie filtrada b_t^2 .

Los estimados de los parámetros del modelo ARMA(1,1)-GARCH(1,1) ajustado para la serie de retornos y su volatilidad son representados por las siguientes ecuaciones.

$$r_t = -0,146952D + 0,454218r_{t-1} - 0,281201b_{t-1} + b_t$$

$$\sigma_t^2 = 0,0000 + 0,069655^2_{t-1} + 0,916419\sigma_{t-1}^2$$

b. **Suponiendo que $\varepsilon_t \sim t(5)$**

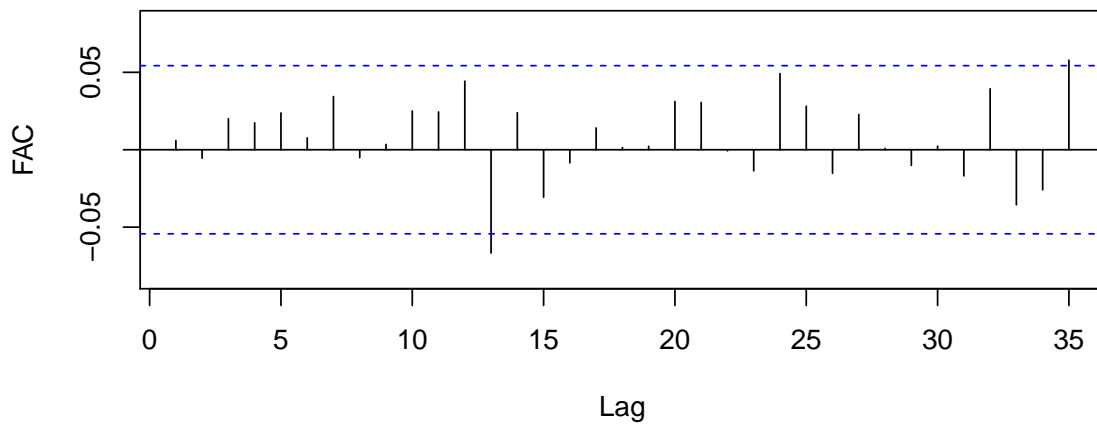
Se observó que en el modelo ajustado, todos los parámetros estimados son significativos a un nivel de confianza del 90 %, tal como se muestran en el Anexos 6.

Las siguientes ecuaciones muestran los estimados del modelo ARMA(1,1)-GARCH(1,1) para la serie de retornos y su respectiva volatilidad.

$$r_t = -0,138151D + 0,656124r_{t-1} - 0,546733b_{t-1} + b_t$$

$$\sigma_t^2 = 0,0000 + 0,095305b_{t-1}^2 + 0,882074\sigma_{t-1}^2$$

(a)



(b)

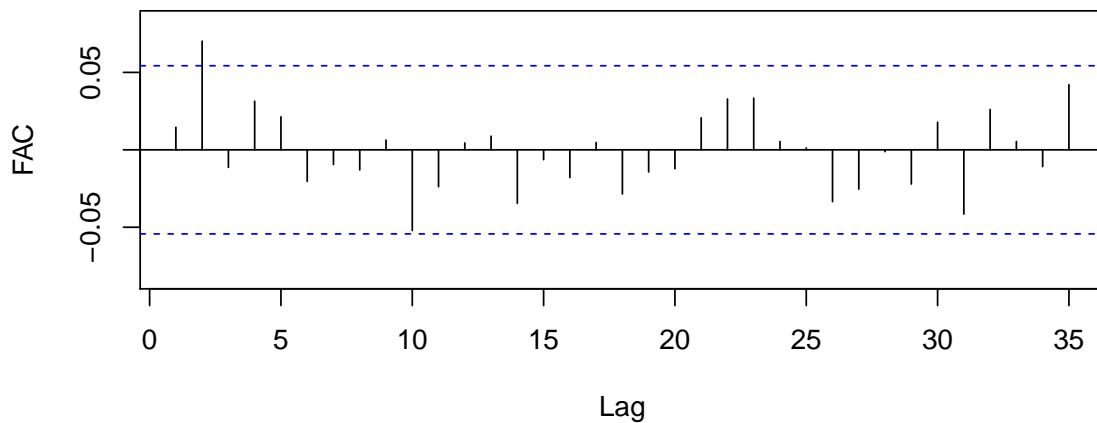


Figura 6: (a) Función de Autocorrelación (fac) de los residuos y (b) Función de Auto-correlación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim N(0,1)$.

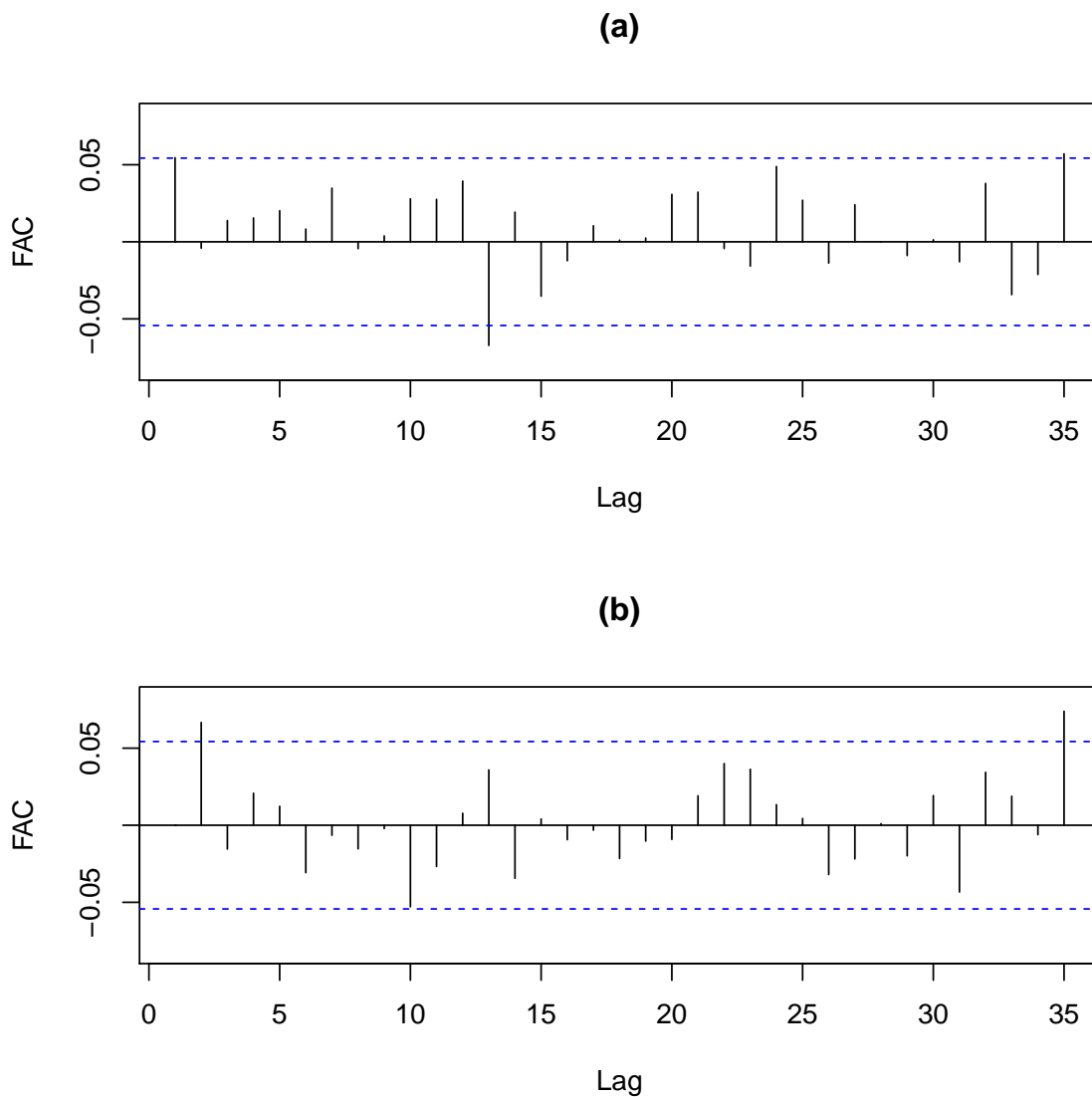


Figura 7: (a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim t(5)$.

4.2.3. Diagnóstico

Evaluando los residuales del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim N(0, 1)$ y $\varepsilon_t \sim t(5)$ se observaron de las Figuras 6 y 7 que no presentan valores altamente significativos en sus funciones de autocorrelación y autocorrelaciones al cuadrado, respectivamente, es decir la no correlación serial en los residuos y residuos al cuadrado, indicando que los modelos para la media y la varianza son adecuados. De esta manera se determinó que los residuos del modelo ARMA(1,1)-GARCH(1,1), se comportan como un ruido blanco, por tanto se mostró

que no hay evidencia de una especificación incorrecta del modelo propuesto.

Así también, Del Anexo 7 y 8, se observó que los residuales del modelo presentan un coeficiente de curtosis mayor a tres, en ambos supuestos de los residuales. Comparando ambos modelos se determinó que el modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim t(5)$ presenta un mejor comportamiento, respecto a sus criterios de información Akaike y bayesiano.

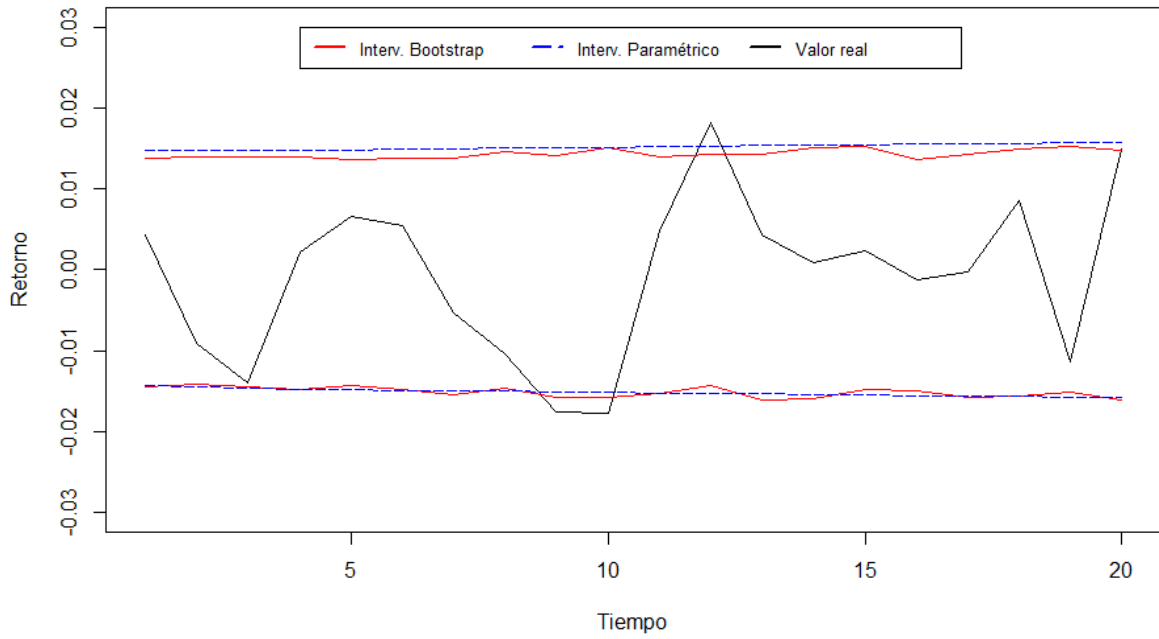
4.2.4. Comparación de intervalos de predicción

En la Figuras 8 y Figura 9 se muestran los intervalos de predicción bootstrap y paramétrico para el modelo ARMA(1,1)-GARCH(1,1), considerando a los residuales con distribución normal y t-student, respectivamente. 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 $h=9,10,12$ y también mediante la metodología bootstrap se observaron que presentan similares amplitudes en la mayoría de los horizontes, respecto a los intervalos paramétricos en ambos supuestos.

A un nivel de confianza del 95 %, se observó que los límites de predicción para los retornos con ambas metodologías cubren todos los valores reales y también mediante la metodología paramétrico se observaron que presentan menores amplitudes en la mayoría de los horizontes, respecto a los intervalos bootstrap en ambos supuestos de los residuales.

En la Figura 10 se observó que los intervalos de predicción bootstrap para las volatilidades del modelo ARMA(1,1)-GARCH(1,1) a niveles de confianza del 90 % y 95 % cubren todos los valores reales de las volatilidades, y en los modelos con $\varepsilon_t \sim t(5)$ se observaron que presentan una mayor amplitud de intervalo de predicción para las volatilidades.

(a)



(b)

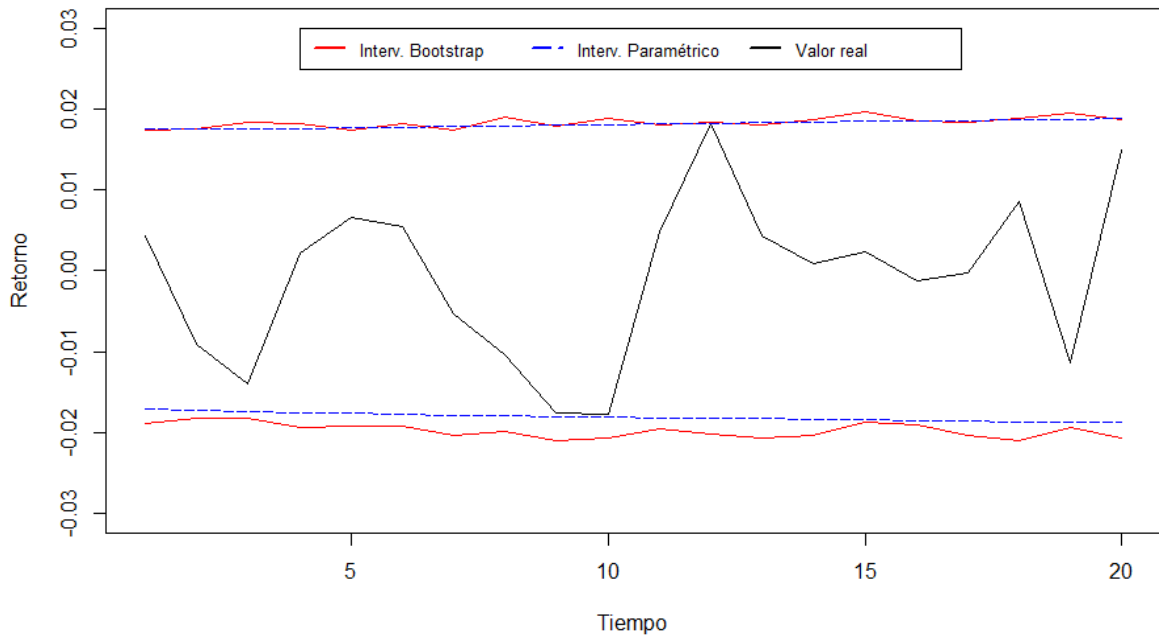
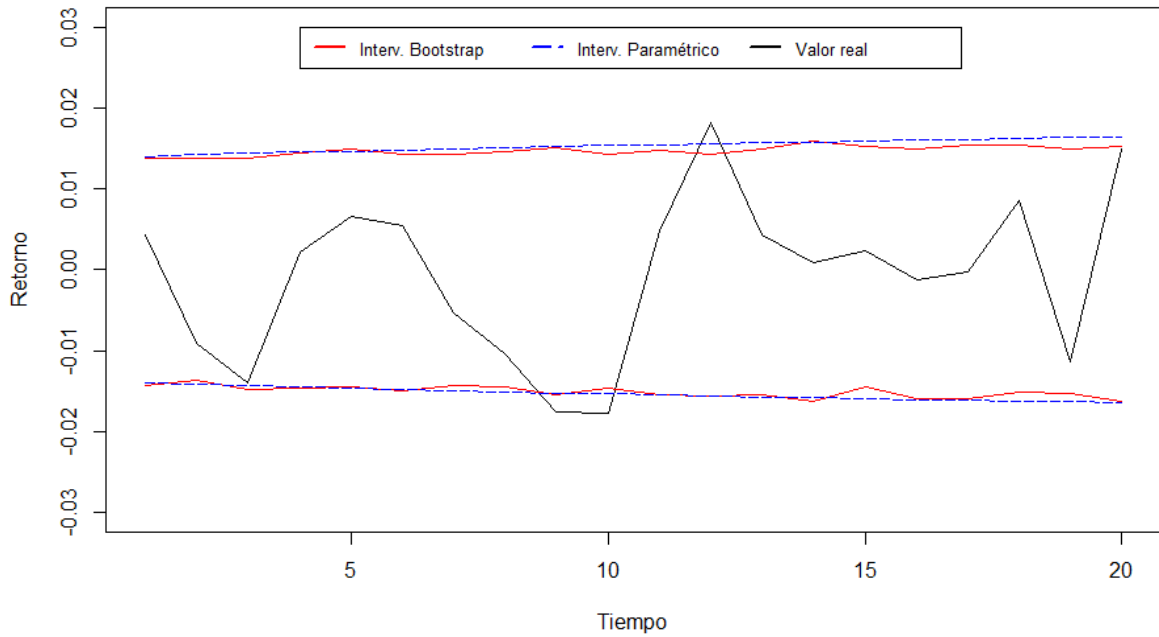


Figura 8: Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim N(0,1)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.

(a)



(b)

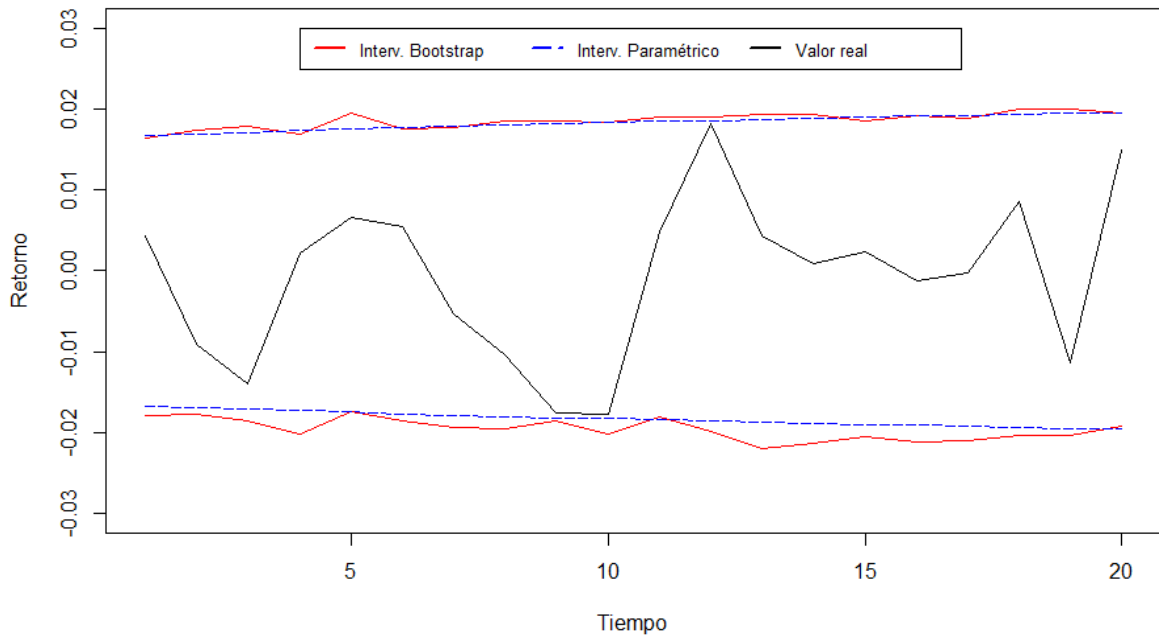
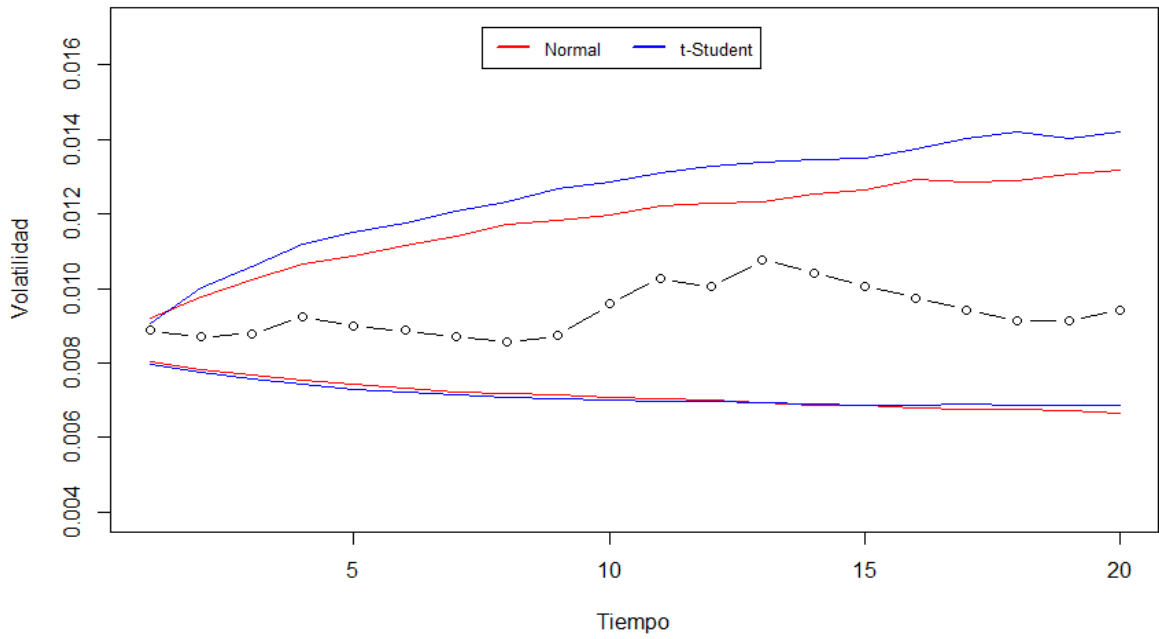


Figura 9: Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim t(5)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.

(a)



(b)

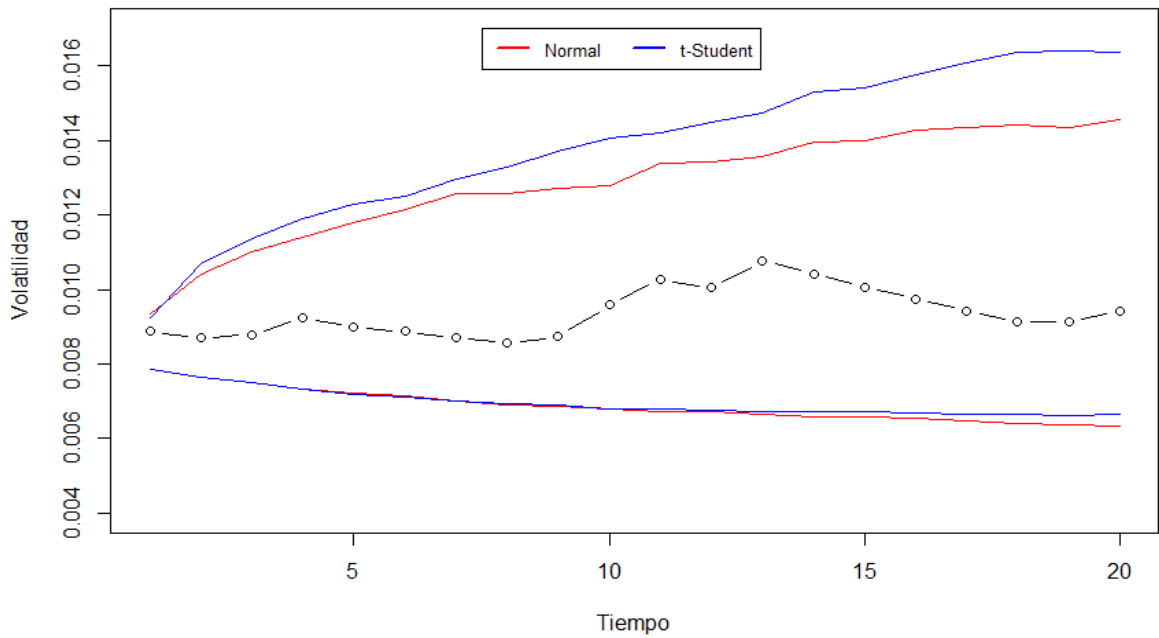


Figura 10: Intervalos de predicción de la volatilidad a niveles de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim N(0, 1)$ y $\varepsilon_t \sim t(5)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap para el mes de diciembre del 2014.

4.3. Análisis para el modelo EGARCH

4.3.1. Identificación

Por los criterios de información de akaike (AIC) y bayesiano (BIC), mostrados en el Anexo 3, se determinó que los parámetros del modelo EGARCH que presenta un mejor ajuste para la serie de retornos, es el modelo ARMA(1,1)-EGARCH(1,1) con valores del AIC=-6.249163 y BIC=-6.225330, y considerando el modelo más parsimonioso. representado por las siguientes ecuaciones.

$$r_t = dD_t + \phi r_{t-1} + \theta b_{t-1} + b_t$$
$$\log(\sigma_t^2) = \omega + \alpha \varepsilon_{t-1} + \gamma[|\varepsilon_{t-1}| - E(|\varepsilon_{t-1}|)] + \beta \log(\sigma_{t-1}^2)$$

Con residuales $\varepsilon_t = \frac{b_t}{\sigma_t}$ considerando su distribución como una normal o una t-student con 5 grados de libertad.

4.3.2. Estimación

a. Suponiendo que $\varepsilon_t \sim N(0, 1)$

Se observó que en el modelo ajustado, todos los parámetros estimados son altamente significativos, tal como se muestran en el Anexo 5. 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.

$$r_t = -0,147232D_t + 0,491914r_{t-1} - 0,302002b_{t-1} + b_t$$
$$\log(\sigma_t^2) = -0,36581 + 0,18659\varepsilon_{t-1} - 0,048435[|\varepsilon_{t-1}| - E(|\varepsilon_{t-1}|)] + 0,97489\log(\sigma_{t-1}^2)$$

b. Suponiendo que $\varepsilon_t \sim t(5)$

Se observó que en el modelo ajustado, todos los parámetros estimados son significativos a un nivel de confianza del 95 %, tal como se muestran en el Anexos 6.

El modelo ARMA(1,1)-EGARCH(1,1) ajustado a la serie de retornos y su respectiva volatilidad son representados por las siguientes ecuaciones.

$$r_t = -0,137337D_t + 0,677284r_{t-1} - 0,562092b_{t-1} + b_t$$

$$\log(\sigma_t^2) = -0,45567 + 0,20689\varepsilon_{t-1} - 0,04978[|\varepsilon_{t-1}| - E(|\varepsilon_{t-1}|)] + 0,96588\log(\sigma_{t-1}^2)$$

4.3.3. 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 observaron de las Figuras 11 y 12 que a pesar de existir rezagos fuera de las bandas en sus funciones de autocorrelación y autocorrelaciones al cuadrado, estas no representan no son altamente significativos, así determinando 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, De los Anexo 9 y 10 , 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 estandar o t-student, sin embargo, es posible observar 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 AIC y BIC mostrados en los Anexos 3 y 4,

4.3.4. Comparación de intervalos de predicción

En la Figuras 13 y Figura 14 se muestran los intervalos de predicción bootstrap y paramétrico para el modelo ARMA(1,1)-EGARCH(1,1) de los retornos del IGBVL, considerando a los residuales con distribución normal y t-student, respectivamente. 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=9,10,12, y también mediante la metodología bootstrap se observó que presentan similares amplitudes de intervalos en algunos horizontes los intervalos de predicción bootstrap resultaron de menor amplitud, 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 cubren todos los valores reales 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 supuestos de los residuales.

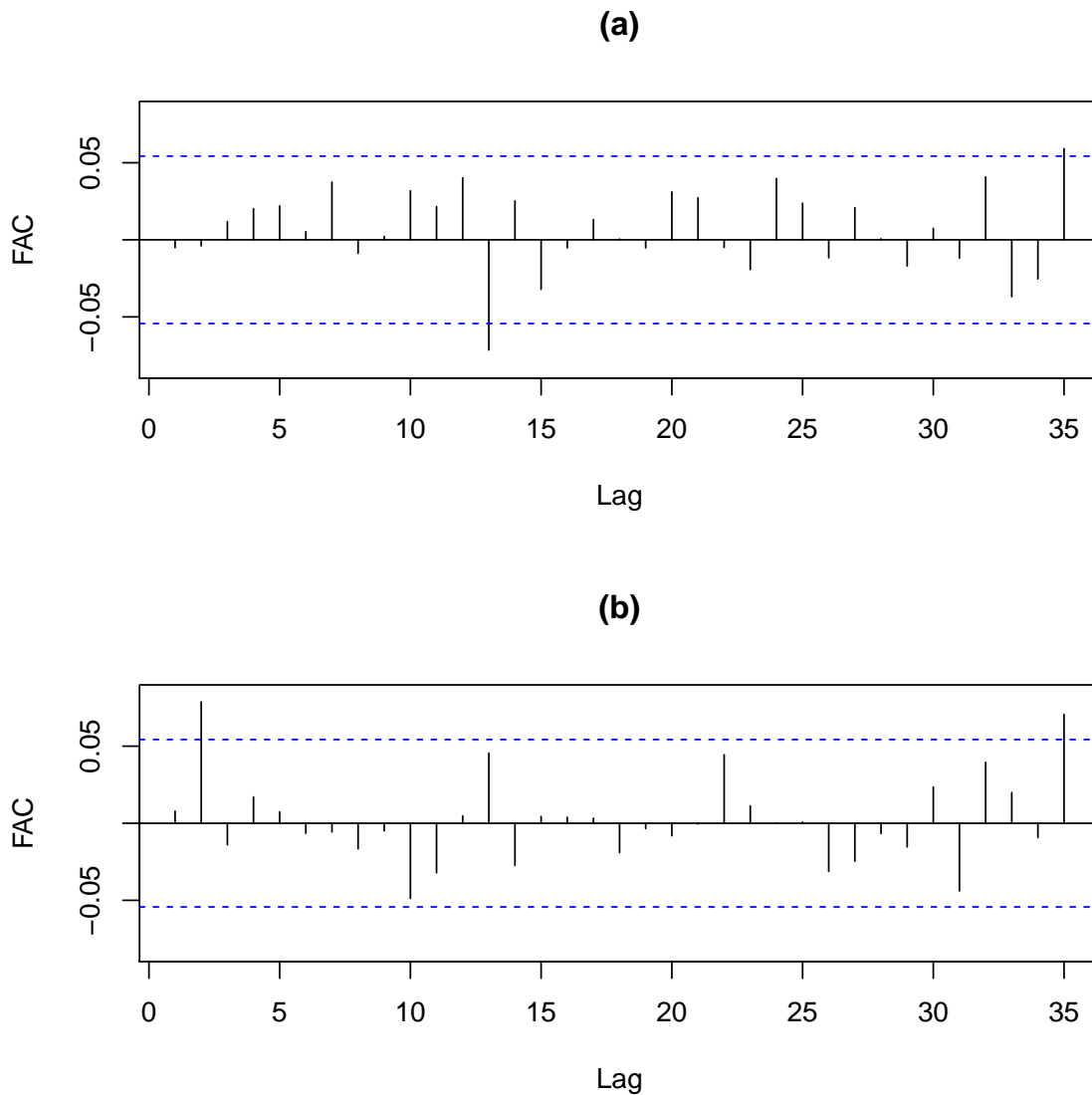


Figura 11: (a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$.

En la Figura 15 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 $\varepsilon_t \sim N(0, 1)$ se observó que presentan mayores amplitudes en la mayoría de los horizontes de intervalo de predicción para las volatilidades de los retornos en la mayoría de los horizontes, a comparación del modelo considerando $\varepsilon_t \sim t(5)$.

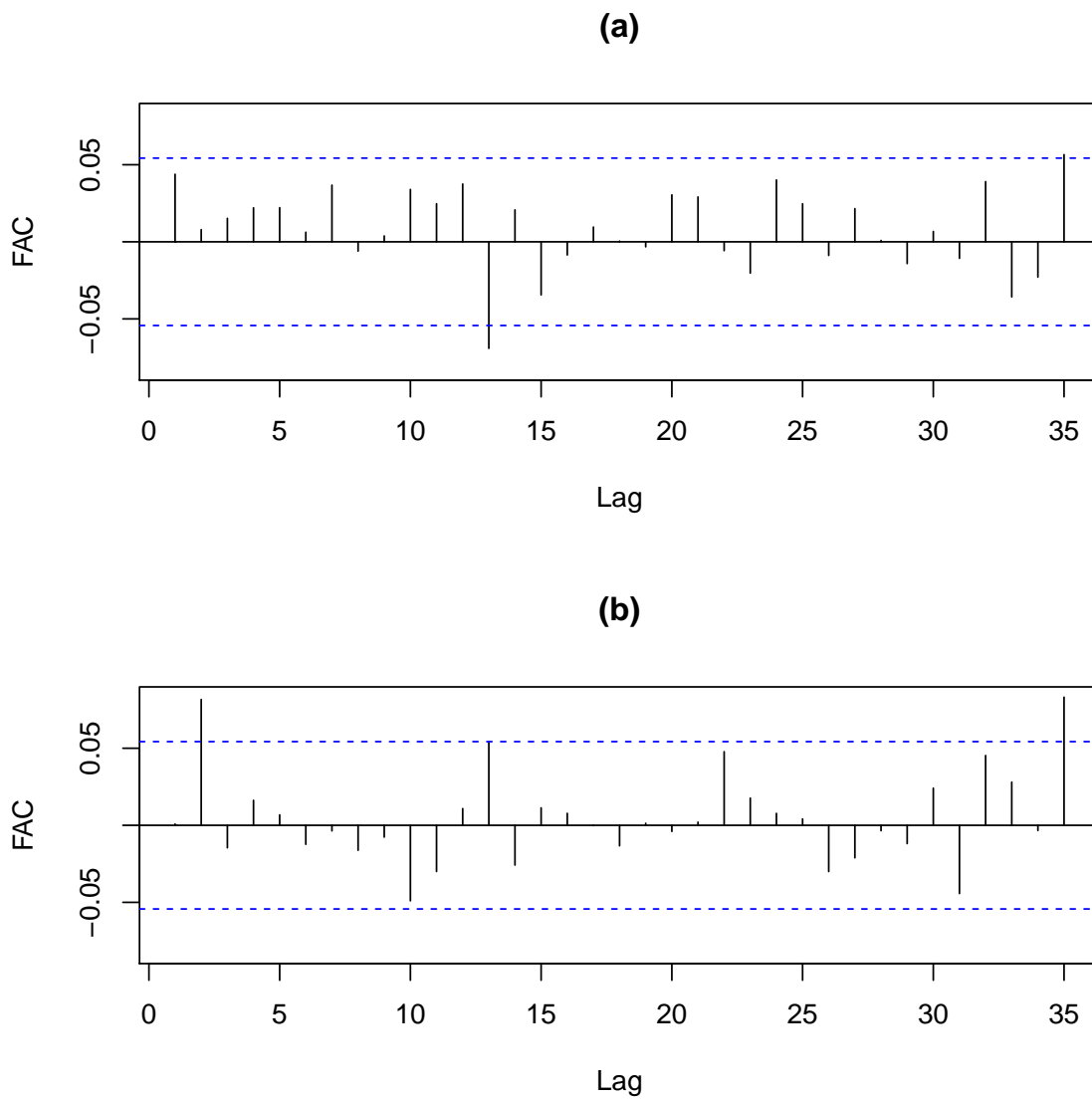
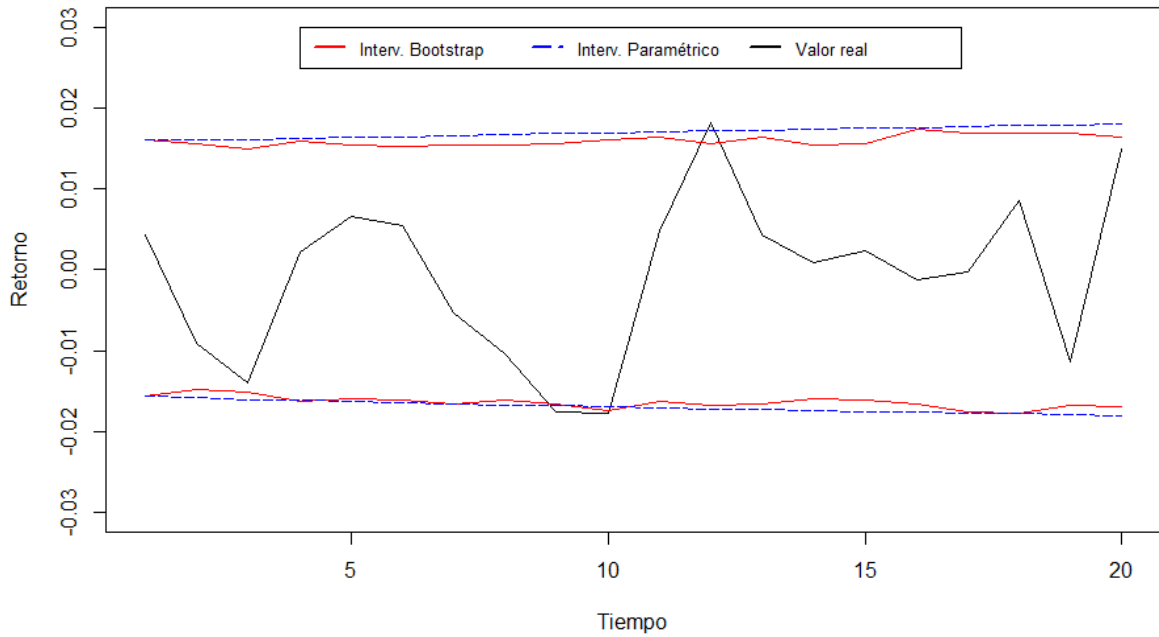


Figura 12: (a) Función de Autocorrelación (fac) de los residuos y (b) Función de Auto-correlación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim t(5)$.

(a)



(b)

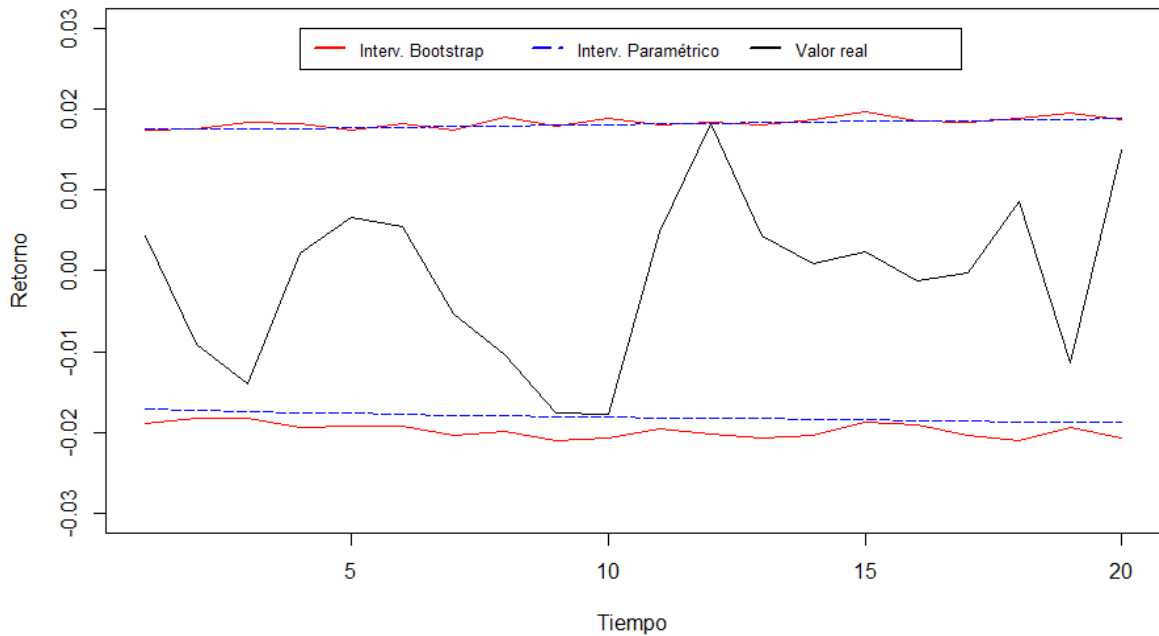
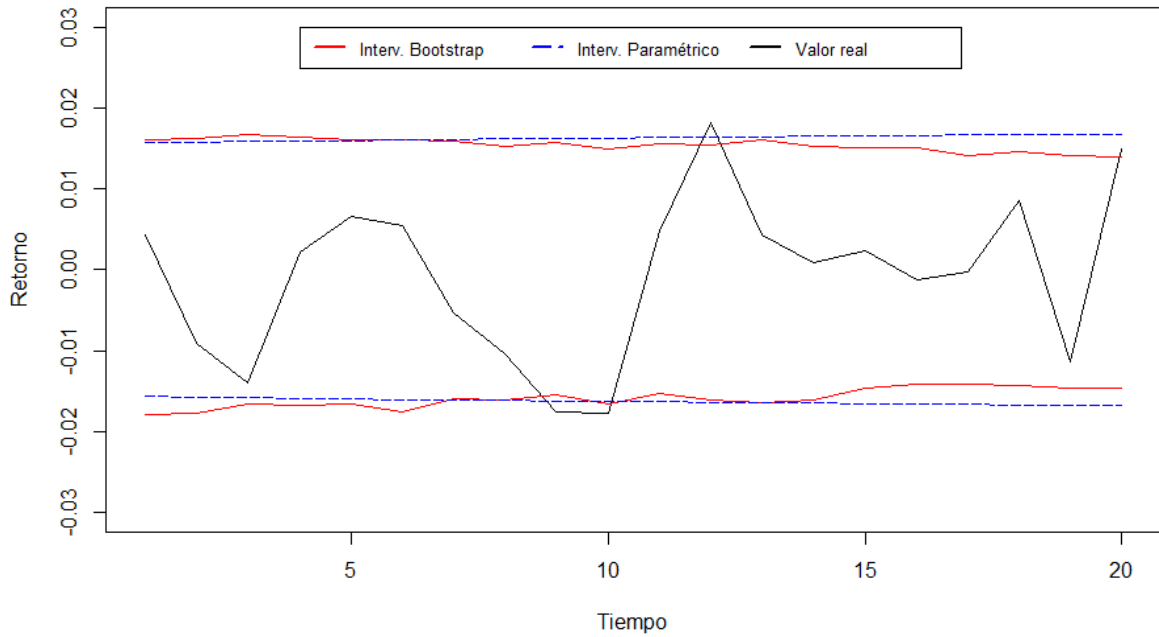


Figura 13: Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.

(a)



(b)

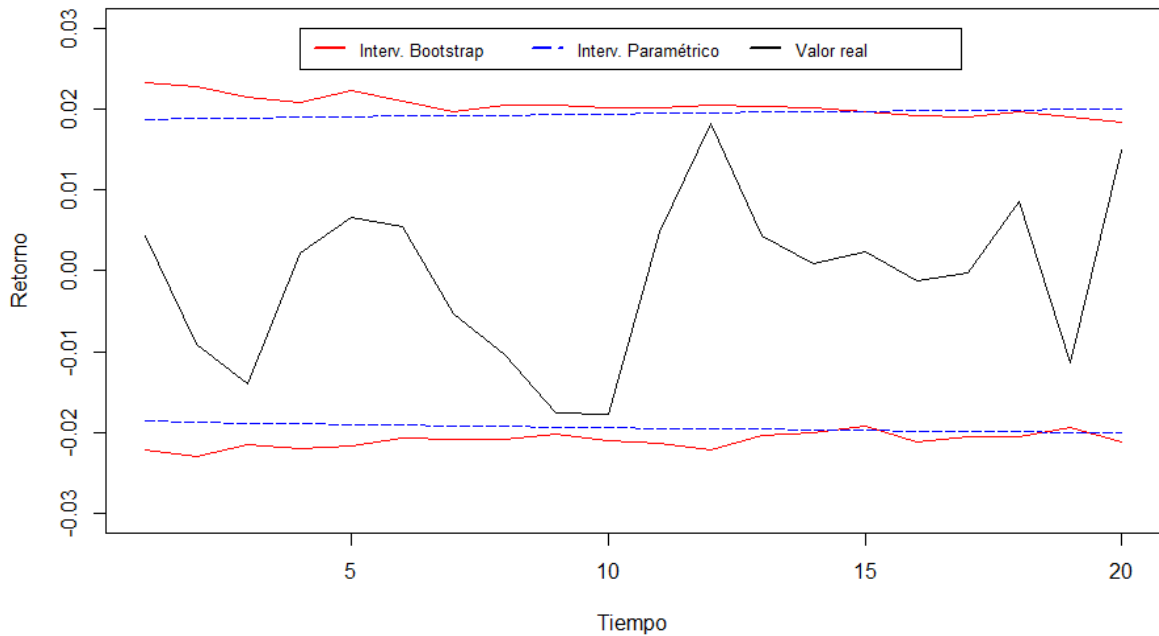
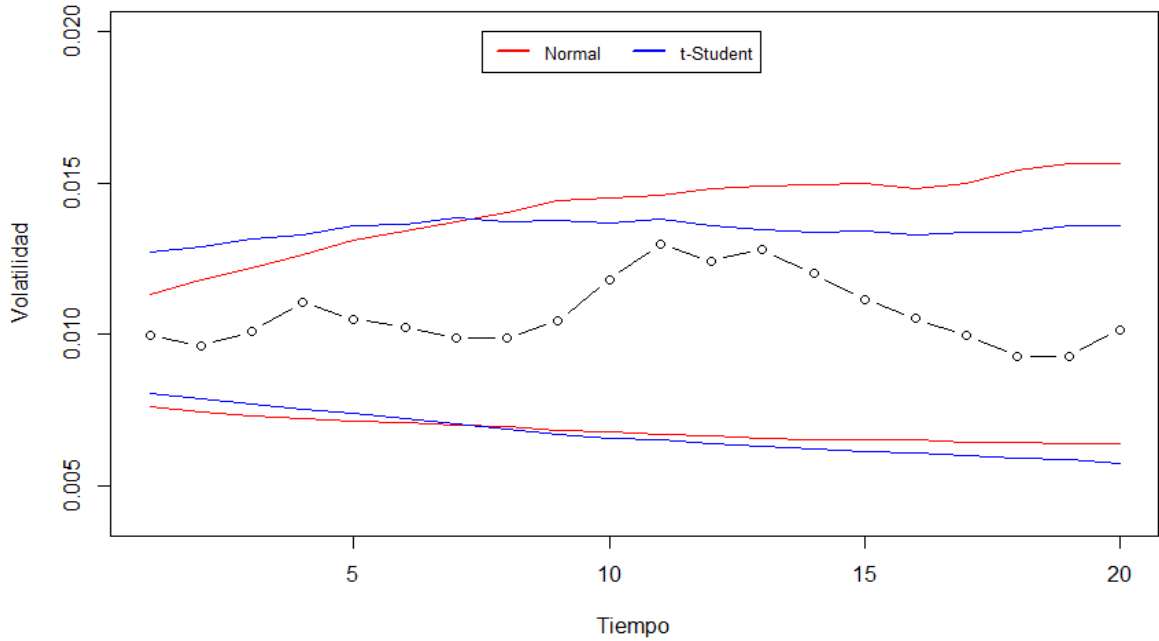


Figura 14: Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim t(5)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.

(a)



(b)

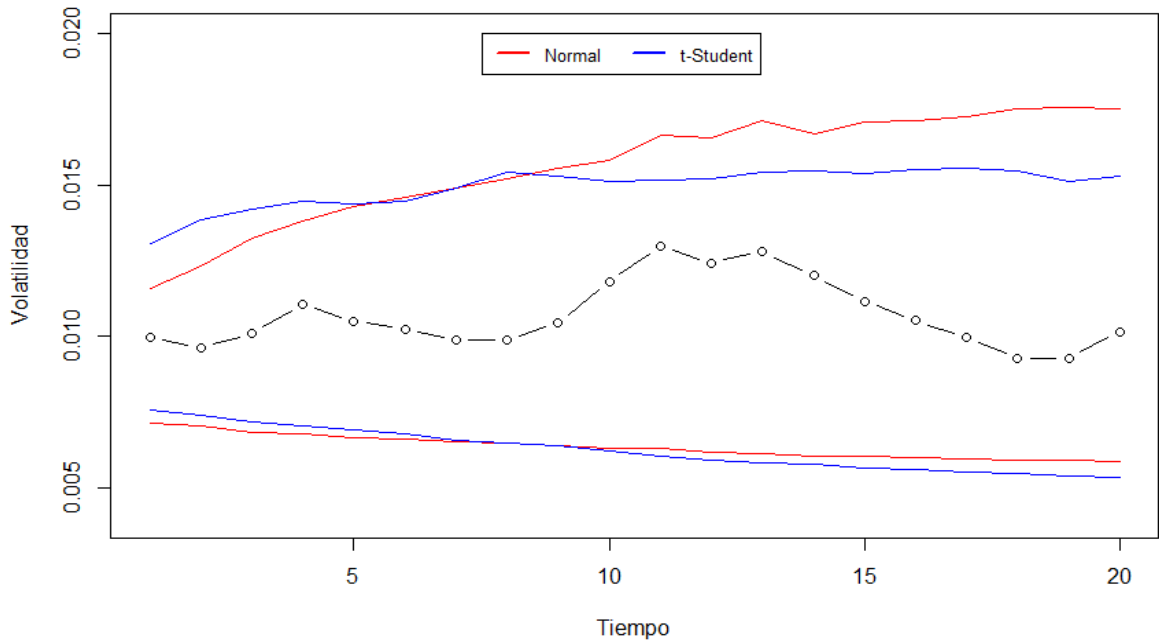


Figura 15: Intervalos de predicción de la volatilidad a niveles de confianza del 90 % (a) y 95 % (b) del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$ y $\varepsilon_t \sim t(5)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap para el mes de diciembre del 2014.

4.4. Análisis para el modelo TGARCH

4.4.1. Identificación

Para capturar el efecto asimétrico de la volatilidad se ajustaron los retornos del IGBVL a un modelo TGARCH. Así, mediante los criterios AIC y BIC, mostrados en el Anexo 3, se determinó que los parámetros de un modelo TGARCH que presenta un mejor ajuste para la serie de retornos del IGBVL es el modelo ARMA(1,1)-TGARCH(1,1) con valores del AIC=-6.254528 y BIC=-6.230696. A pesar de que el modelo ARMA(1,1)-TGARCH(2,2) presentó su AIC y BIC ligeramente mayor, por parsimonia seleccionamos el modelo con menos parámetros, así representado por las siguientes ecuaciones.

$$r_t = dD_t + \phi r_{t-1} + \theta b_{t-1} + b_t$$
$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma \varepsilon_{t-1}^2 S_{t-1}$$

Donde:

$$S_{t-1} = \begin{cases} 1 & \text{si } \varepsilon_{t-1} < 0 \\ 0 & \text{si } \varepsilon_{t-1} \geq 0. \end{cases}$$

Además, los residuales $\varepsilon_t = \frac{b_t}{\sigma_t}$ considerados su distribución como una normal o una t-student con 5 grados de libertad.

4.4.2. Estimación

a. **Suponiendo que $\varepsilon_t \sim N(0, 1)$**

Se observó que en el modelo ajustado, todos los parámetros estimados son altamente significativos, tal como se muestran en el Anexos 5.

La estimación del modelo ARMA(1,1)-TGARCH(1,1) ajustado a la serie de retornos y su volatilidad son representados por:

$$r_t = -0,147146D_t + 0,492248r_{t-1} - 0,313823b_{t-1} + b_t$$
$$\sigma_t^2 = 0,000 + 0,041587\varepsilon_{t-1}^2 + 0,912014\sigma_{t-1}^2 + 0,057960\varepsilon_{t-1}^2 S_{t-1}$$

Donde:

$$S_{t-1} = \begin{cases} 1 & \text{si } \varepsilon_{t-1} < 0 \\ 0 & \text{si } \varepsilon_{t-1} \geq 0 \end{cases}$$

b. **Suponiendo que $\varepsilon_t \sim t(5)$**

Se observó que en el modelo ajustado, todos los parámetros estimados son significativos a un nivel de confianza del 90 %, tal como se muestran en el Anexos 6.

El modelo ARMA(1,1)-TGARCH(1,1) ajustado para la serie de retornos y su volatilidad son representados por las siguientes ecuaciones.

$$r_t = -0,137347D_t + 0,685095r_{t-1} - 0,572117b_{t-1} + b_t$$
$$\sigma_t^2 = 0,000 + 0,052028\varepsilon_{t-1}^2 + 0,870747\sigma_{t-1}^2 + 0,092097\varepsilon_{t-1}^2 S_{t-1}$$

Donde:

$$S_{t-1} = \begin{cases} 1 & \text{si } \varepsilon_{t-1} < 0 \\ 0 & \text{si } \varepsilon_{t-1} \geq 0. \end{cases}$$

4.4.3. Diagnóstico

Evaluando los residuales de los modelos ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$ y $\varepsilon_t \sim t(5)$ se observaron de las Figuras 16 y 17 que no presentan valores altamente significativos en sus funciones de autocorrelación y autocorrelaciones al cuadrado, respectivamente, es decir la no correlación serial en los residuos y residuos al cuadrado, indicando que los modelos para la media y la varianza describen adecuadamente los retornos y las volatilidades. Así también, De los Anexo 11 y 12, se observó que los residuales del modelo presentan un coeficiente de curtosis mayor a tres y la presencia de colas pesadas, diferenciando así de un comportamiento normal estándar o t-student.

4.4.4. Comparación de intervalos de predicción

En la Figuras 18 y Figura 19 se muestran los intervalos de predicción bootstrap y paramétrico para el modelo ARMA(1,1)-TGARCH(1,1), considerando a los residuales con distribución normal y t-student, respectivamente. A un nivel de confianza del 90 %, se observaron que los límites de predicción para los retornos con ambas metodologías no cubren los horizontes $h=9,10,12$ y que presentan similares amplitudes de intervalos de predicción, en ambos supuestos de los residuales .

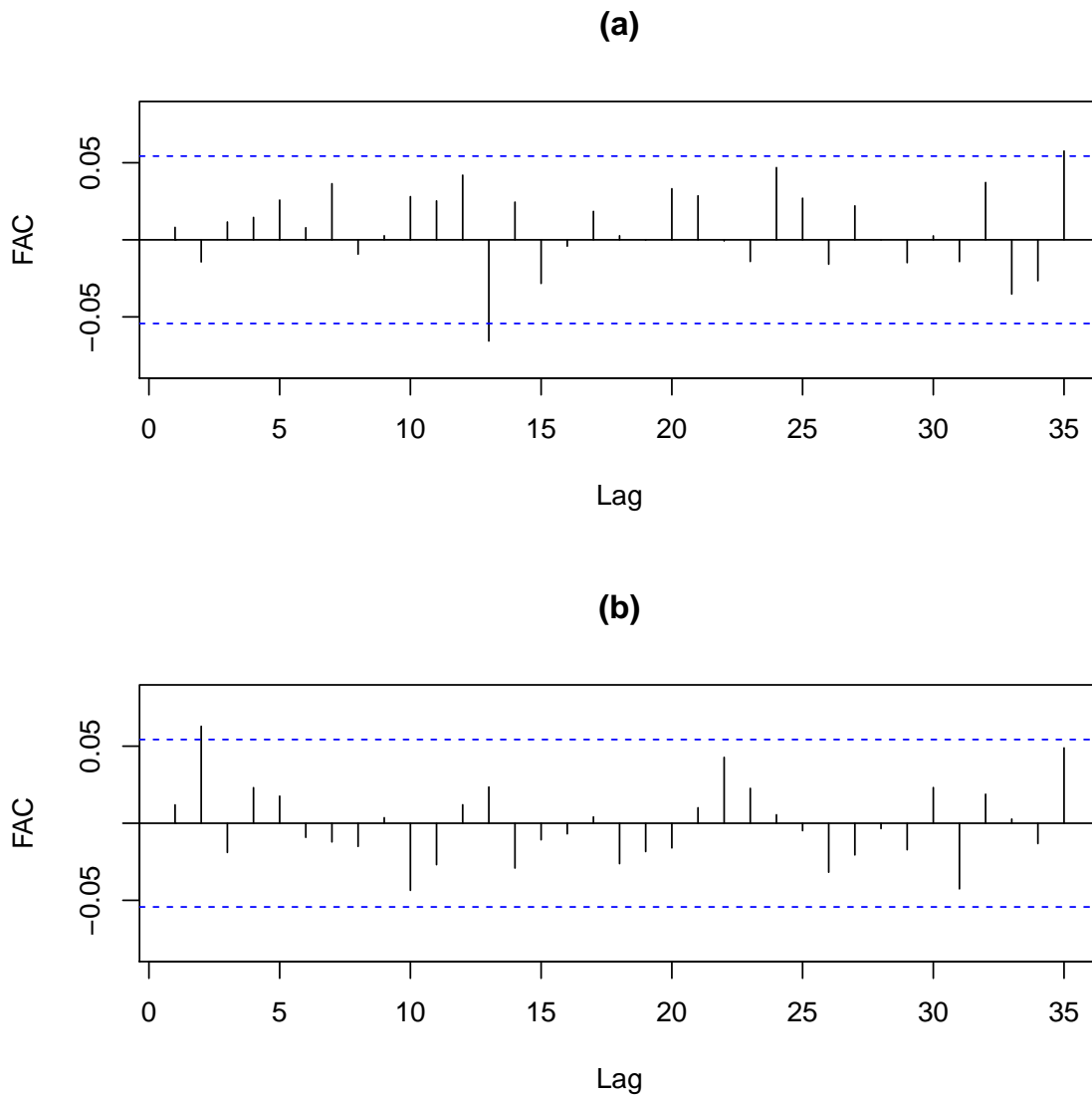


Figura 16: (a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo $ARMA(1,1)$ -TGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$.

A un nivel de confianza del 95%, se observó que los límites de predicción para los retornos con ambas metodologías cubren todos los valores reales y mediante la metodología paramétrica se observaron que presentan menores amplitudes en todos los horizontes, respecto a los intervalos bootstrap en ambos supuestos de residuales.

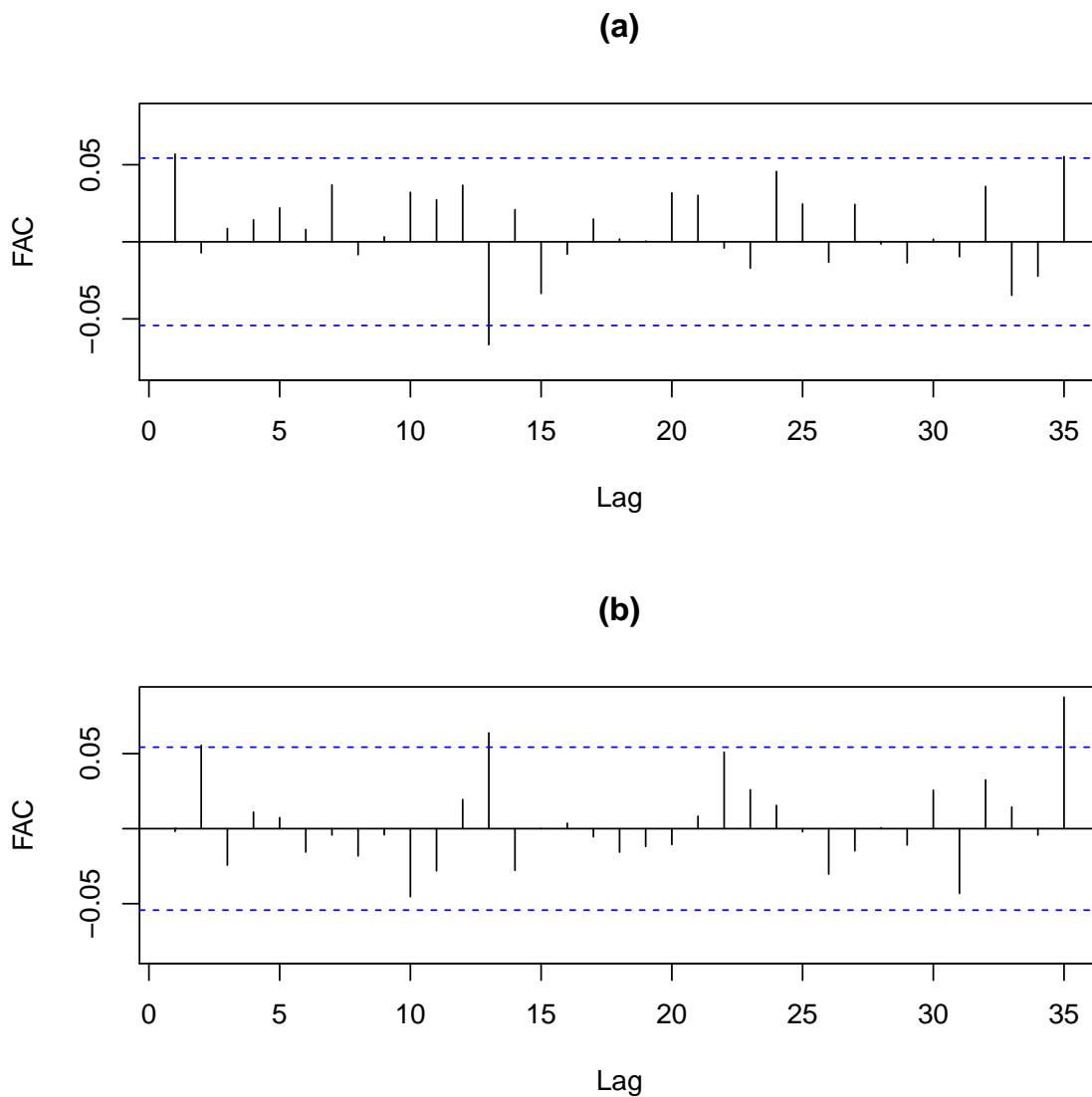
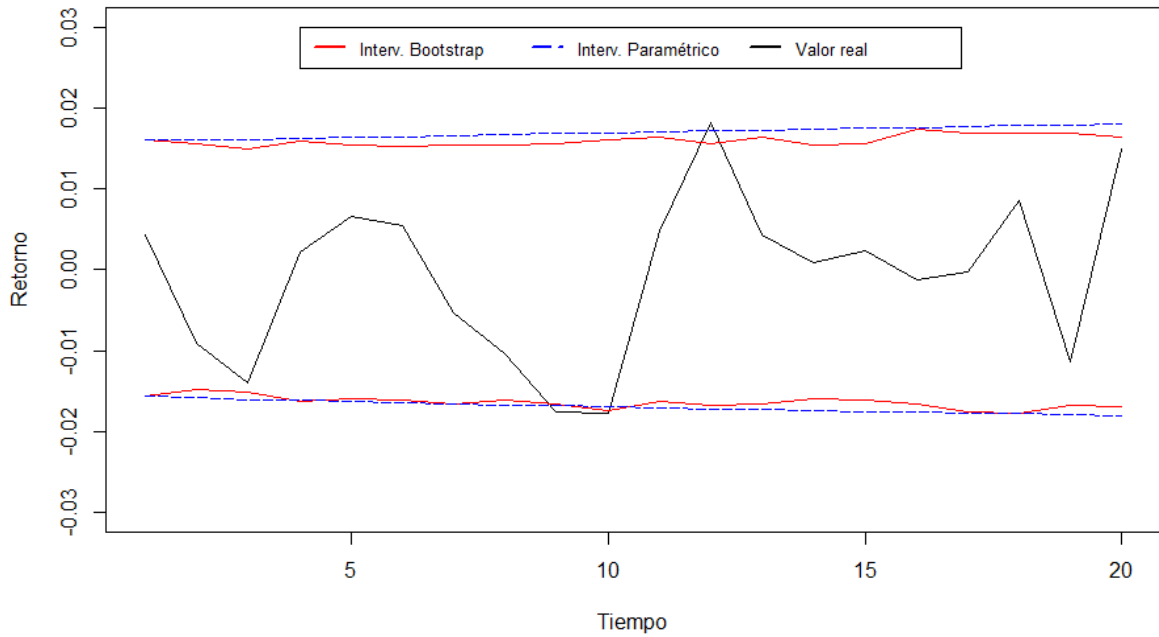


Figura 17: (a) Función de Autocorrelación (fac) de los residuos y (b) Función de Autocorrelación (fac) de los residuos al cuadrado, del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim t(5)$.

En la Figura 20 se observó que los intervalos de predicción para las volatilidades del modelo ARMA(1,1)-TGARCH(1,1) a niveles de confianza del 90% y 95% cubren todos los valores reales de las volatilidades, y con los modelos donde $\varepsilon_t \sim t(5)$ se observaron que presentan mayores amplitudes de intervalo de predicción para las volatilidades de los retornos.

(a)



(b)

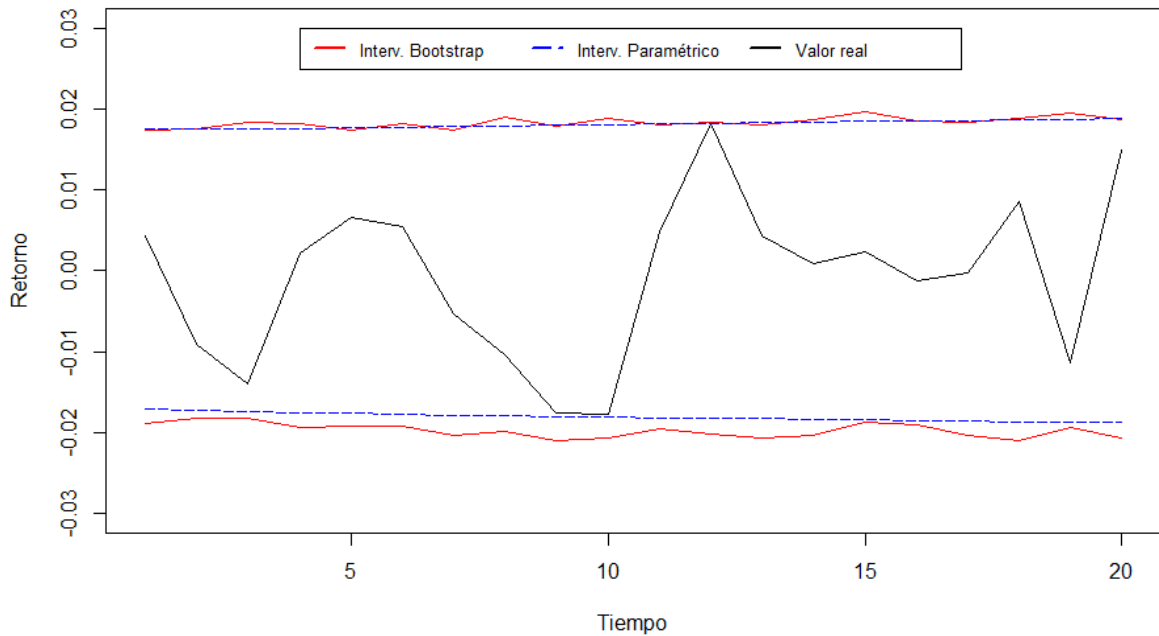
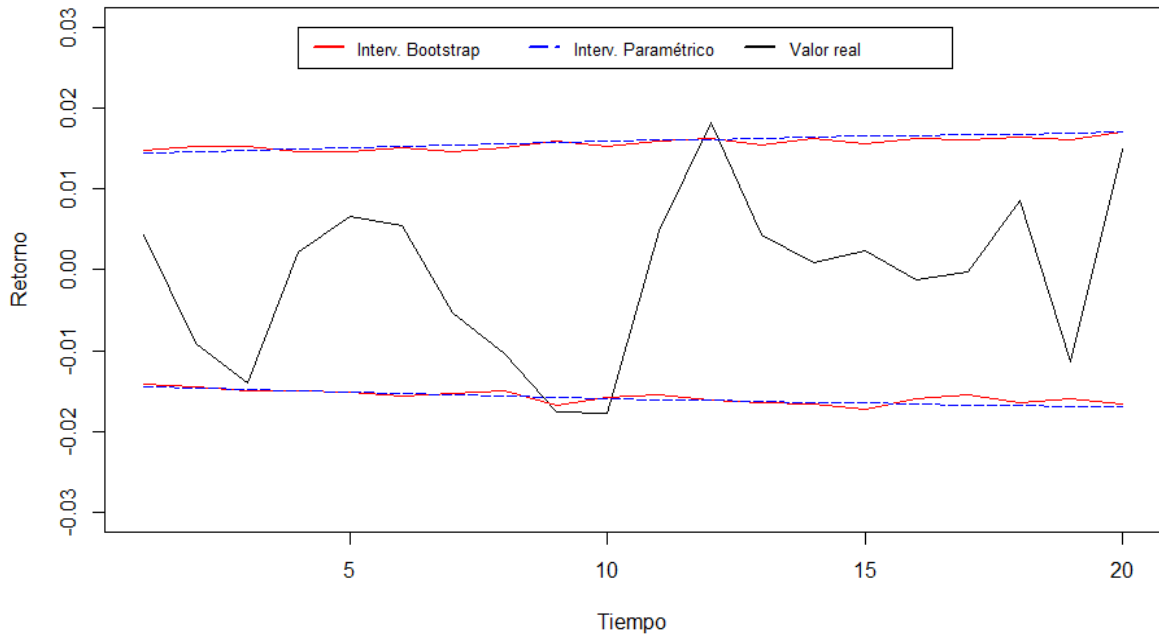


Figura 18: Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.

(a)



(b)

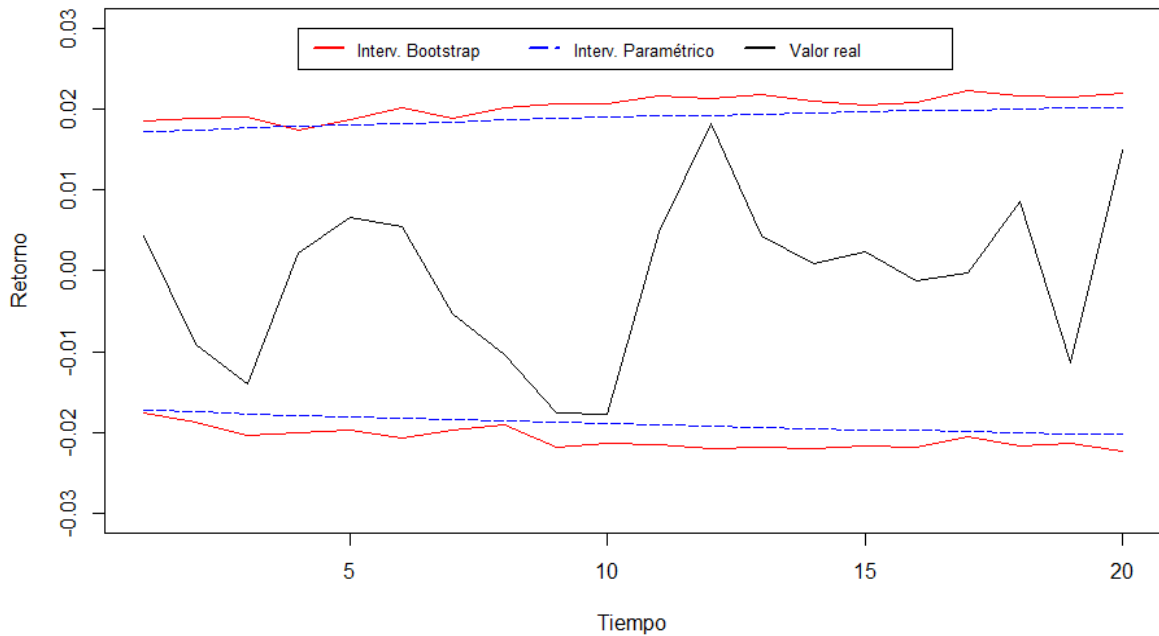
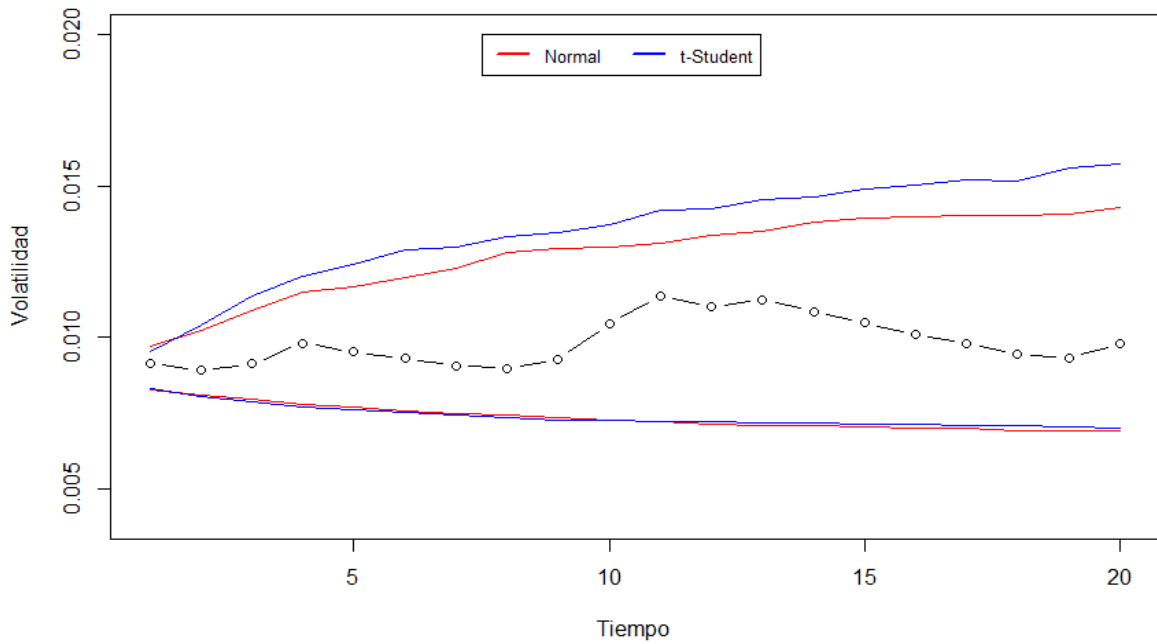


Figura 19: Intervalos de predicción al nivel de confianza del 90% (a) y 95% (b) del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim t(5)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap y paramétrico para el mes de diciembre del 2014.

(a)



(b)

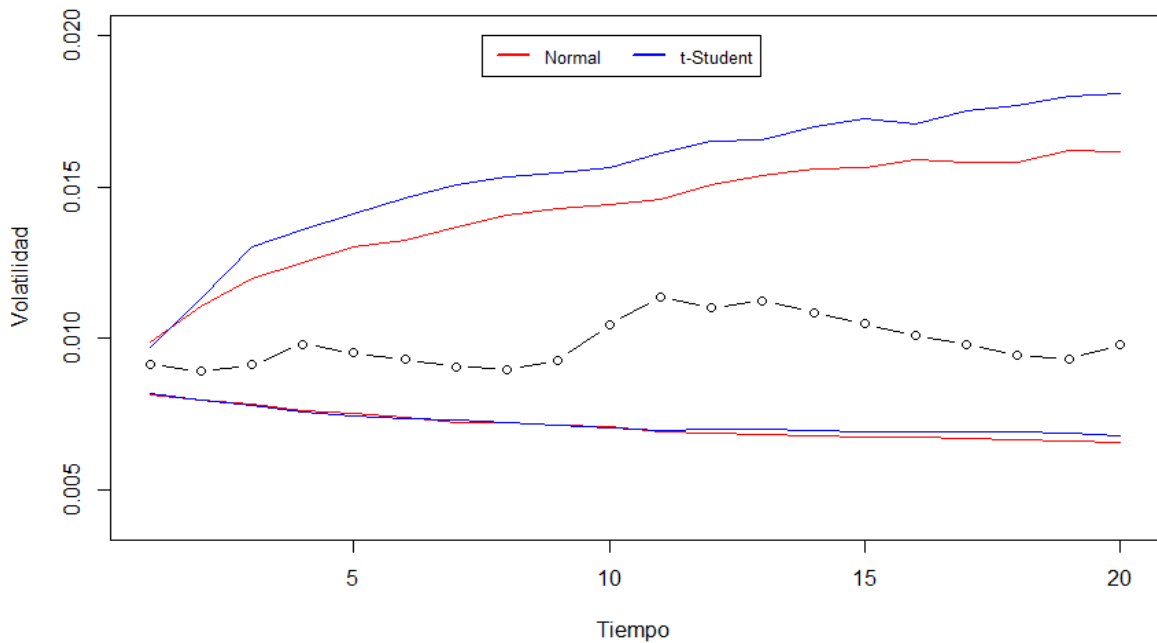


Figura 20: Intervalos de predicción de la volatilidad a niveles de confianza del 90% y 95% del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$ y $\varepsilon_t \sim t(5)$, de los retornos de los precios de las acciones del Índice General de la Bolsa de Valores de Lima, mediante las metodologías bootstrap para el mes de diciembre del 2014.

V. CONCLUSIONES

Del análisis realizado en la investigación se han llegado a las siguientes conclusiones:

1. El orden de los modelos de predicción que se ajustan adecuadamente a los retornos diarios de los precios del índice general de la bolsa de valores de Lima fueron: ARMA(1,1)-GARCH(1,1), ARMA(1,1)-EGARCH(1,1) y ARMA(1,1)-TGARCH(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.
2. 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 ARMA(1,1)-GARCH(1,1), ARMA(1,1)-EGARCH(1,1) y ARMA(1,1)-TGARCH(1,1) 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.
3. 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 para los modelos ARMA(1,1)-GARCH(1,1), ARMA(1,1)-EGARCH(1,1) y ARMA(1,1)-TGARCH(1,1) 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 General de la Bolsa de Valores de Lima.

4. 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, en todos los casos presentaron sus valores reales de las volatilidades de los retornos dentro de los márgenes de confianza. Los intervalos de predicción bootstrap de los modelos ARMA(1,1)-GARCH(1,1) y ARMA(1,1)-TGARCH(1,1) con $\varepsilon \sim t(5)$ presentaron mejores resultados, en cuanto a su amplitud de intervalo.

VI. RECOMENDACIONES

Al finalizar la presente investigación se llegaron a las siguientes recomendaciones:

1. Se recomienda analizar otros tipos de series de retornos de los Índices de la Bolsa de Valores de Lima, mediante la metodología bootstrap.
2. Se sugiere realizar simulaciones para cada tipo de modelo heterocedástico mediante la metodología Bootstrap para la construcción de intervalos de predicción.
3. Evaluar los resultados con la metodología Bootstrap considerando diferentes valores de tamaño de remuestra B .
4. Para un trabajo posterior, se sugiere analizar la construcción de intervalos de predicción considerando múltiples variables.

VII. REFERENCIAS BIBLIOGRÁFICAS

Alpuim, T. 1998. Séries Temporáis . 2 ed. Lisboa, Associação dos Estudantes da Faculdade de Ciências.

Alvarez, A. M., y Rubio, P. O. 2000. Metodología bootstrap en series heterocedásticas. una aplicación al ibex-35. XIV Reunión ASEPELT. Anales de Economía Aplicada(9), 84-99.

Alvarez, A. M., y Rubio, P. O. 2002. Avances recientes en métodos bootstrap para procesos arch. una aplicación en el mercado español de valores. Estudios de economía Aplicada, 20(2), 487-498.

Bodas, D. 2013. Una técnica para la optimización de los parámetros de indicadores técnicos bursátiles mediante algoritmos evolutivos multiobjetivo. Tesis Phd. Universidad Complutense de Madrid, Departamento de Arquitectura de Computadores y Automática.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307-327.

Bollerslev, T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. The Review of Econometrics and Statistics, 69, 542-547.

Box, G., y Jenkins, G. 1970. Time Series Analysis: Forecasting and Control. San Francisco:Holden-Day.

Cao, R., M., F., Manteinga, G., y Sánchez, P. 1997. Saving computer time in constructing consistent bootstrap prediction intervals for autoregressive processes. Communications in Statistics-Simulation and Computation, 26, 961-978.

Carnero, F. M. 2003. Heterocedasticidad condicional, atípicos y cambios de nivel en series temporales financieras . Tesis Doc. Universidad Carlos III de Madrid.

Castillo, E. 2014. Modelación de la volatilidad del índice general de la Bolsa de Valores de Lima, periodo 2009-2011. Tesis grad. Universidad Nacional Agraria La Molina.

Chambi, P. 2017. La volatilidad de los mercados financieros globalizados: Impacto en la bolsa de valores de lima-perú. Revista de la facultad de Ciencias Contables - UNMSM, 25(47), 103-111.

Connolly, R. A. 1989. An examination of the robustness of the weekend effect. Financial and Quantitative Analysis, 24, 133-169.

Davison, A., y Hinkley, D. 1997. Bootstrap methods and their application. Cambridge: Cambridge University Press.

Efron, B. 1979. Bootstrap methods: Another look at the jackknife. Annals of Statistics, 7, 1-26.

Efron, B., y Tibshirani, R. 1993. An Introduction to the Bootstrap. New York, Chapman and Hall.

Elescano, A., y Agüero, Y. 2004. Modelos arch: Una aplicación en el pronóstico de la volatilidad de acciones cotizadas en la bolsa de valores de lima. Revista de la facultad de CC. MM. UNMSM, 7(1), 64-79.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. The Econometric Society, 50, 987-1007.

Fan, J., y Yao, Q. 1994. Nonlinear time series. New York: Springer.

Fernández, J. 1999. El método Jackknife para la estimación de varianzas de estimadores. Tesis Lic. Universidad Nacional Mayor de San Marcos.

Henrique, M. 2003. Modelos de previsão em séries temporais. Aplicação da meto-

dologia Bootstrap. Tese Ms. Universidade nova de Lisboa, Faculdade de Ciências y Tecnologia.

Heracleous, M. S. 2003. Volatility Modeling Using the Student's t Distribution. Dissertation Doc. of Philosophy in Economics, Faculty of the Virginia Polytechnic Institute, University of Virginia.

Herencia, M. E. 1997. Volatilidade nos modelos ARCH e variância estocástica: um estudo comparativo. Tese Ms. Instituto de Matemática e computação Científica-UNICAMP. Campinas, Brasil.

Jurado, G., Manteinga, G., y Sánchez, P. 1995. Predicting using box-jenkins, nonparametric and bootstrap techniques. *Technometrics*, 37, 303-310.

Kim., J., Song, H., Wong, K., Athanasopoulos, G., y Liu, S. 2009. Beyond point forecasting: Evaluation of alternative prediction intervals for tourist arrivals. Monash University. Department of Econometrics and Business Statistics, 30, 1-19.

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..

Marques, J. I. 2013. A metodologia Bootstrap associada ao método de Holt-Winters na previsão de séries temporais . Tese Mes. Universidade de Minho, Escola de Ciências, Brasil.

McCullough, B. 1994. Bootstrapping forecast intervals: An application to $ar(p)$ models. *Journal of Forecasting*, 13, 51-66.

Miranda, A. 2003. El método de remuestreo y su aplicación en la investigación biomédica. Esp. Bio. Ciudad de la Habana, Escuela Nacional de Salud Pública Carlos J. Finlay.

Morettin, P. A., y Toloi, C. M. C. 2011. Análise de Séries Temporais. São Paulo: Bloucher.

- Murteira, B., y Muller, D. 1990. Análise de sucessão cronológicas. Lisboa: McGraw-Hill.
- Nelson, D. B. 1991. Conditional heteroskedasticity in asset returns. *The Econometric Society*, 52, 347-370.
- Pascual, L., Romo, J., y Ruiz, E. 1999. Bootstrap predictive inference for arima processes. *Working Papers. Statistics and Econometrics*, 40(4), 86-98.
- Pascual, L., Romo, J., y Ruiz, E. 2006. Bootstrap prediction for return and volatilities in garch models. *Computational Statistics and Data Analysis*, 50(9), 2293-2312.
- Peña, D. 2010. Análisis de series temporales . Madrid, ES: Alianza editorial.
- Peña, D., Tiao, G., y Tsay, R. 2001. A course in time series analysis. New York: J Wiley.
- Raffo, L. E., y Quispe, A. C. 2012. Aplicación de la metodología garch al precio de cierre en la bolsa de valores de lima. *Revista de la facultad de Ingeniería Industrial UNMSM*, 15(2), 96-105.
- Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Shimizu, K. 2009. Bootstrapping stationary ARMA-GARCH models. Wiesbaden: Springer Science.
- Silupú, B., y Calle, S. 2007. Análisis de la eficiencia en la Bolsa de Valores de Lima. Periodo: 1999-2004. Tesis grad. Universidad Nacional de Piura.
- Stine, R. 1987. Estimating properties of autoregressive forecasts. *American Statistical Association*, 82, 1072-1078.
- Thombs, L., y Schucany, W. 1990. Bootstrap prediction intervals for autoregression. *American Statistical Association*, 85, 486-492.

Tineo, F. 2005. Estimación de Kaplan Meier bootstrap de la curva de supervivencia. Tesis Lic. Universidad Nacional Mayor de San Marcos.

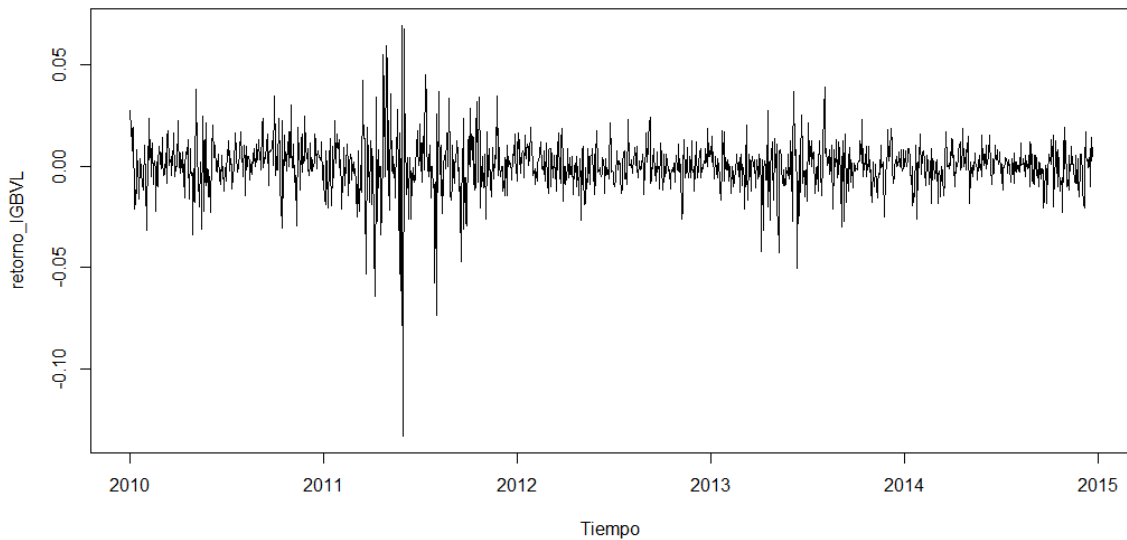
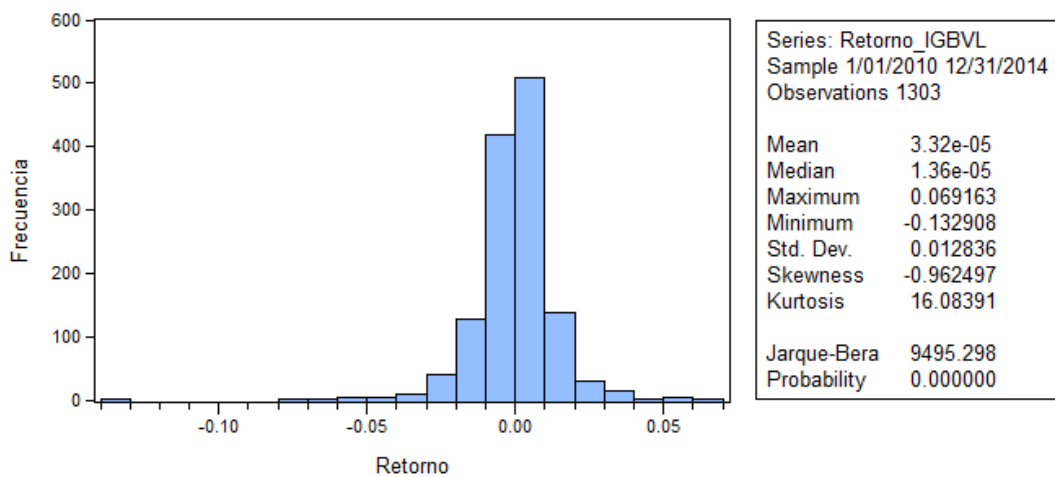
Tukey, J. W. 1958. Bias and confidence in not-quite large sample. *Annals of Mathematical Statistics*, 29, 614.

Uriel, E., y Peiró, A. 2005. Introducción al análisis de series temporales. España: Alfa centauro.

Zakoian, M. 1994. Threshold heteroscedastic models. *Journal of Economic Dynamics and Control*, 18, 931-955.

VIII. ANEXOS

ANEXO 1: Histograma de frecuencias, principales medidas descriptivas y evolución de los retornos diarios de los precios de las acciones del IGBVL.



ANEXO 2: Pruebas de Autocorrelación de Ljung-Box para las series r_t , b_t y b_t^2

Serie	Estadístico de Ljung - Box					
	Q(1)	Q(2)	Q(3)	Q(5)	Q(10)	Q(20)
r_t	22.739	28.182	28.199	28.637	32.927	47.956
(prob.)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
b_t	0.125	3.299	3.426	3.731	7.649	18.597
(prob.)	(0.723)	(0.192)	(0.330)	(0.589)	(0.663)	(0.548)
b_t^2	69.910	172.30	275.80	352.48	398.89	579.14
(prob.)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

ANEXO 3: Criterios de selección AIC y BIC para los retornos del IGBVL, obtenidos de modelos con $\varepsilon_t \sim N(0, 1)$.

Serie	Criterio	Orden del modelo GARCH			
		(1,1)	(1,2)	(2,1)	(2,2)
\$b.t\$	AIC	-6.248787	-6.247489	-6.248527	-6.246495
	BIC	-6.228926	-6.223656	-6.224695	-6.218691
		Orden del modelo EGARCH			
		(1,1)	(1,2)	(2,1)	(2,2)
\$b.t\$	AIC	-6.249163	-6.247888	-6.247820	-6.247090
	BIC	-6.225330	-6.220084	-6.220015	-6.215314
		Orden del modelo TARCH			
		(1,1)	(1,2)	(2,1)	(2,2)
\$b.t\$	AIC	-6.257056	-6.254528	-6.253474	-6.257767
	BIC	-6.230696	-6.229252	-6.225670	-6.235991

ANEXO 4: Criterios de selección AIC y BIC para los retornos del IGBVL, obtenidos de modelos con $\varepsilon_t \sim t(5)$.

Serie	Criterio	Orden del modelo GARCH			
		(1,1)	(1,2)	(2,1)	(2,2)
\$b.t\$	AIC	-6.312620	-6.311110	-6.311177	-6.311382
	BIC	-6.288787	-6.283305	-6.283373	-6.279606
		Orden del modelo EGARCH			
		(1,1)	(1,2)	(2,1)	(2,2)
\$b.t\$	AIC	-6.312932	-6.311748	-6.311401	-6.310287
	BIC	-6.285127	-6.279971	-6.279625	-6.274538
		Orden del modelo TARCH			
		(1,1)	(1,2)	(2,1)	(2,2)
\$b.t\$	AIC	-6.316680	-6.315656	-6.315208	-6.298250
	BIC	-6.288876	-6.283880	-6.283432	-6.262502

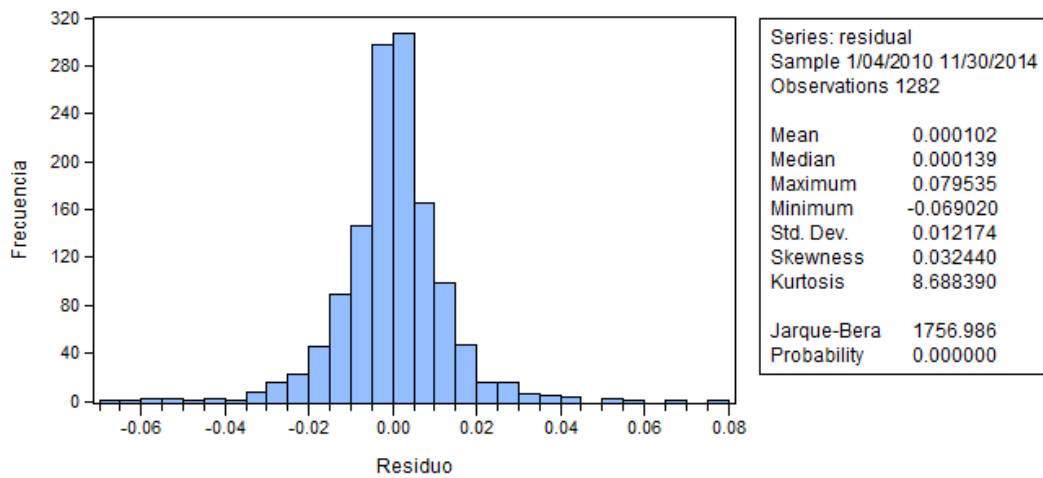
ANEXO 5: Estimación de coeficientes de modelos para los retornos diarios del IGBVL con $\varepsilon \sim N(0, 1)$.

ARMA(1,1)-GARCH(1,1)				
Variable	Coeficiente	Std. Error	z-Stat.	Prob.
d	-0.146952	0.052378	-2.805635	0.0050
ϕ	0.454218	0.095658	4.748346	0.0000
θ	-0.281201	0.107157	-2.624203	0.0087
ω	1.91E-06	4.23E-07	4.523480	0.0000
α	0.069655	0.008231	8.462138	0.0000
β	0.916419	0.008217	111.5255	0.0000
ARMA(1,1)-EGARCH(1,1)				
Variable	Coeficiente	Std. Error	z-Stat.	Prob.
d	-0.147232	0.040421	-3.642461	0.0003
ϕ	0.491914	0.082257	5.980200	0.0000
θ	-0.302002	0.094383	-3.199738	0.0014
ω	-0.365813	0.051987	-7.036606	0.0000
α	0.186597	0.018412	10.13471	0.0000
γ	-0.048435	0.011306	-4.284092	0.0000
β	0.974896	0.005076	192.0627	0.0000
ARMA(1,1)-TGARCH(1,1)				
Variable	Coeficiente	Std. Error	z-Stat.	Prob.
d	-0.147146	0.051003	-2.885012	0.0039
ϕ	0.492248	0.086226	5.708822	0.0000
θ	-0.313823	0.099054	-3.168205	0.0015
ω	2.30E-06	4.40E-07	5.226987	0.0000
α	0.041587	0.010018	4.151415	0.0000
γ	0.057960	0.013404	4.324107	0.0000
β	0.912014	0.008457	107.8370	0.0000

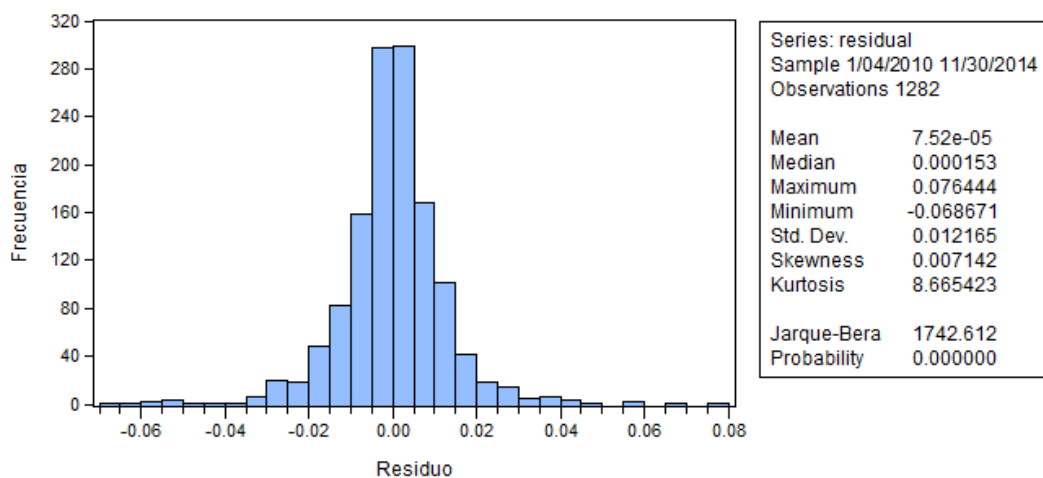
ANEXO 6: Estimación de coeficientes de modelos para los retornos diarios del IGBVL con $\varepsilon \sim t(5)$.

ARMA(1,1)-GARCH(1,1)				
Variable	Coficiente	Std. Error	z-Stat.	Prob.
d	-0.138151	0.079813	-1.730920	0.0835
ϕ	0.656124	0.080008	8.200673	0.0000
θ	-0.546733	0.091896	-5.949472	0.0000
ω	3.93E-06	1.42E-06	2.765234	0.0057
α	0.095305	0.021814	4.368997	0.0000
β	0.882074	0.022764	38.74891	0.0000
ARMA(1,1)-EGARCH(1,1)				
Variable	Coficiente	Std. Error	z-Stat.	Prob.
d	-0.137337	0.059158	-2.321516	0.0203
ϕ	0.677284	0.074143	9.134824	0.0000
θ	-0.562092	0.086322	-6.511549	0.0000
ω	-0.455677	0.112493	-4.050720	0.0001
α	0.206899	0.037182	5.564509	0.0000
γ	-0.049787	0.021223	-2.345955	0.0190
β	0.965885	0.010954	88.17664	0.0000
ARMA(1,1)-TGARCH(1,1)				
Variable	Coficiente	Std. Error	z-Stat.	Prob.
d	-0.137347	0.085234	-1.611409	0.1071
ϕ	0.685095	0.072335	9.471125	0.0000
θ	-0.572117	0.084853	-6.742474	0.0000
ω	4.90E-06	1.55E-06	3.158926	0.0016
α	0.052028	0.022486	2.313814	0.0207
γ	0.092097	0.033568	2.743592	0.0061
β	0.870747	0.024116	36.10639	0.0000

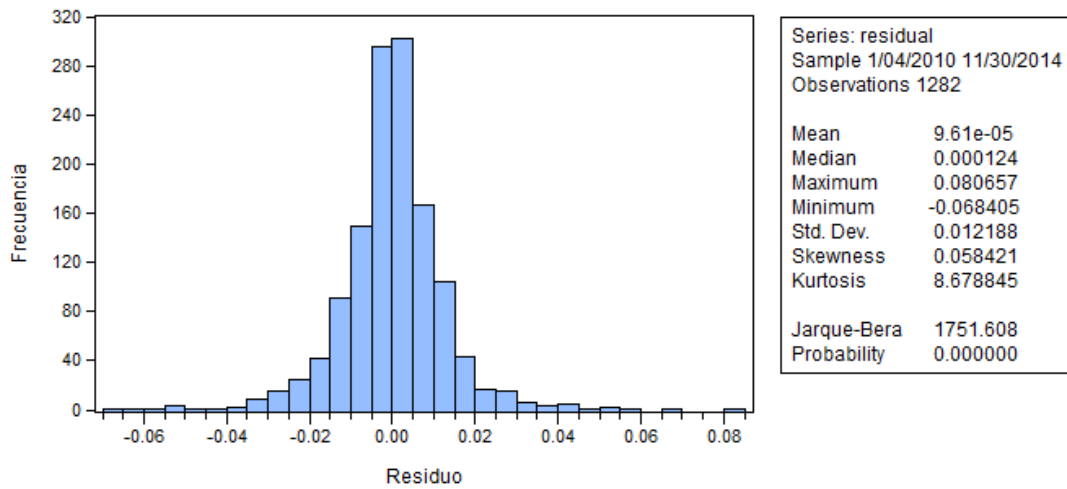
ANEXO 7: Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim N(0, 1)$.



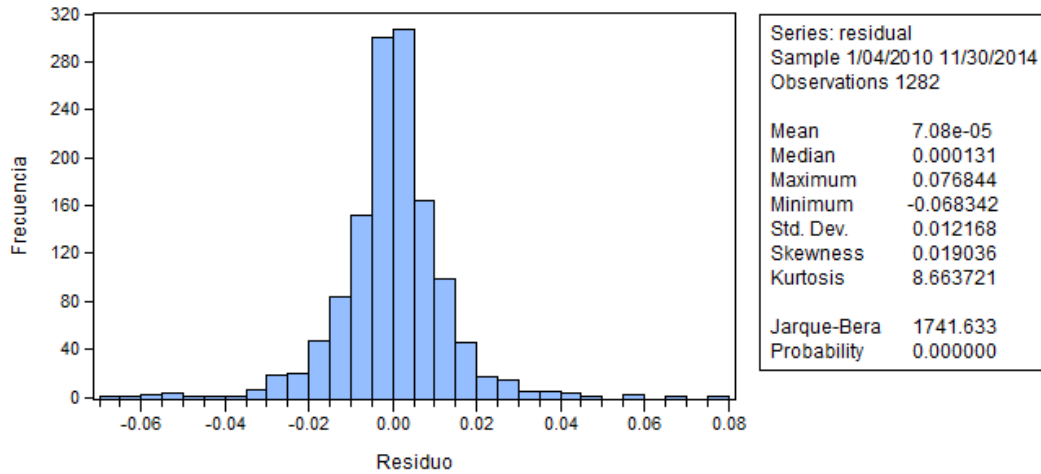
ANEXO 8: Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-GARCH(1,1) con $\varepsilon_t \sim t(5)$.



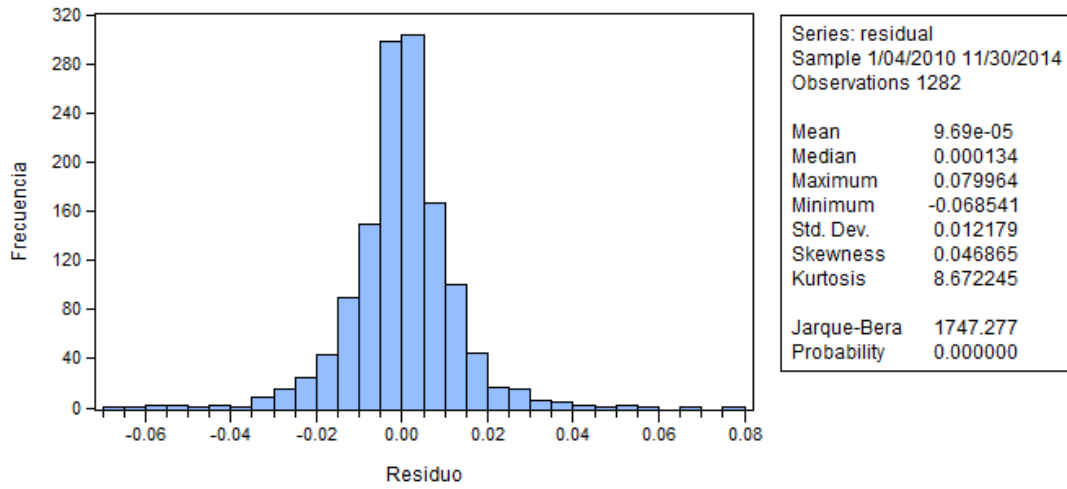
ANEXO 9: Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$.



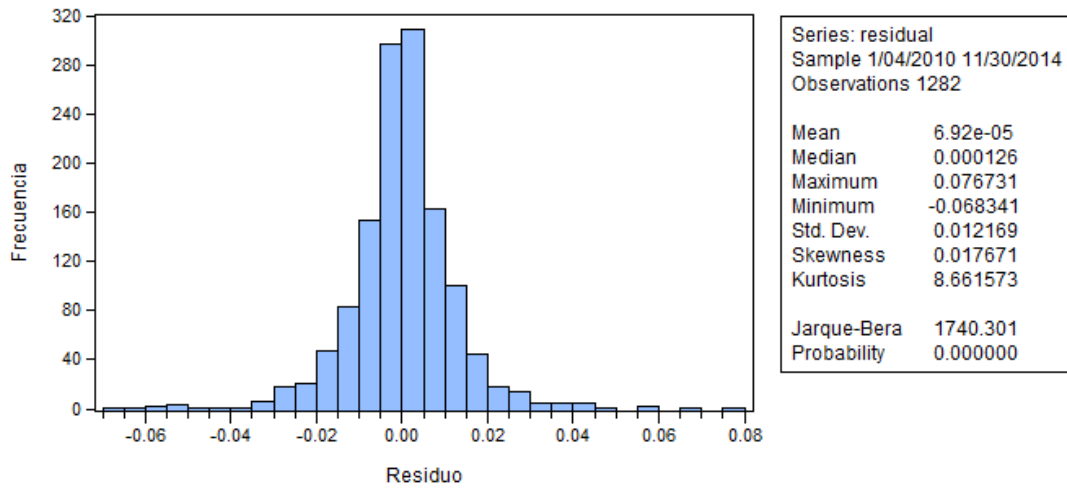
ANEXO 10: Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-EGARCH(1,1) con $\varepsilon_t \sim t(5)$.



ANEXO 11: Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim N(0, 1)$.



ANEXO 12: Histograma de frecuencias y principales medidas descriptivas de los residuales del modelo ARMA(1,1)-TGARCH(1,1) con $\varepsilon_t \sim t(5)$.



ANEXO 13: Función realizado en el *software* R para el cálculo de las predicciones h pasos hacia adelante del modelo ARMA(1,1)-GARCH(1,1).

```
boot_garch<-function(data,B,h,dummy){
  N<-dim(data)[1]
  datos<-ts(as.matrix(data[1:(N-h)]),start=2010,frequency=262))
  n<-dim(datos)[1]
  spec<-ugarchspec(variance.model=list(model="sGARCH",garchOrder=c(1,1))
    ,mean.model=list(armaOrder=c(1,1),include.mean=FALSE,
    external.regressors=matrix(dummy)),
    distribution.model="norm")
  model<-ugarchfit(spec=spec,data=datos,
    fit.control=list(stationarity=1), solver="hybrid")
  coef<-model@fit$coef
  res_centr<-datos/as.numeric(sigma(model))
  residual<-(res_centr-mean(res_centr))/sd(res_centr)
  y_boot<-matrix(0,n,B)
  var_boot<-matrix(0,n,B)
  var2<-matrix(0,n,B)
  coef1<-matrix(0,B,length(coef))
  s2<-yb<-c()
  yp<-s2p<-matrix(NA,ncol=h,nrow=B)
  for (i in 1:B){
    res<-sample(residual,n,replace=T)
    var1<-as.numeric(coef[4]/(1-coef[5]-coef[6]))
    for(j in 2:n){
      var_boot[1,i]<-var1
      y_boot[1,i]<-res[1]*sqrt(var_boot[1])
      var_boot[j,i]<-coef[4]+coef[5]*y_boot[j-1,i]^2+
        coef[6]*var_boot[j-1,i]
      y_boot[j,]<-res[j]*sqrt(var_boot[j,i])
    }
  }
```

```

restric<-ugarchfit(spec=spec, data=y_boot[,i],
                  fit.control=list(stationarity=1),
                  solver="hybrid")@fit$coef
while((restric[5]+restric[6])>=1 || length(restric)!=6){
  res<-sample(residual,n,replace=T)
  var1<-as.numeric(coef[4]/(1-coef[5]-coef[6]))
  for(j in 2:n){
    var_boot[1,i]<-var1
    y_boot[1,i]<-res[1]*sqrt(var_boot[1,i])
    var_boot[j,i]<-coef[4]+coef[5]*y_boot[j-1,i]^2+
      coef[6]*var_boot[j-1,i]
    y_boot[j,i]<-res[j]*sqrt(var_boot[j,i])
  }
  restric<-ugarchfit(spec=spec, data=y_boot[,i],
                    fit.control=list(stationarity=1),
                    solver="hybrid")@fit$coef
}
coef1[i,]<-restric
res_pred = sample(residual, h, replace=TRUE)
suma = 0
for (k in 0:(n-2)){
  suma = suma+ (coef1[i,6]^(k))*(datos[n-k-1]^2-
              coef1[i,4]/(1-coef1[i,5]-coef1[i,6]))
}
var_cond = coef1[i,4]/(1-coef1[i,5]-coef1[i,6]) + coef1[i,5]*suma
for (k in 2:h){
  s2[1] = coef1[i,4]+coef1[i,5]*datos[n]^2+coef1[i,6]*var_cond
  yb[1] = res_pred[1]*sqrt(s2[1])
  s2[k] = coef1[i,4]+coef1[i,5]*yb[k-1]^2+coef1[i,6]*s2[k-1]
  yb[k] = res_pred[k]*sqrt(s2[k])
}
yp[i,]<-yb

```

```
s2p[i,]<-s2
}
OUT= list(yp=yp,s2p=s2p)
return(OUT)
}
```

ANEXO 14: Función realizado en el *software* R para el cálculo de las predicciones h pasos hacia adelante del modelo ARMA(1,1)-EGARCH(1,1).

```
boot_egarch<-function(data,B,h,dummy){
  N<-dim(data)[1]
  datos<-ts(as.matrix(data[1:(N-h)],start=2010,frequency=262))
  n<-dim(datos)[1]
  spec<-ugarchspec(variance.model=list(model="eGARCH",garchOrder=c(1,1))
    ,mean.model=list(armaOrder=c(1,1),include.mean=FALSE,
    external.regressors=matrix(dummy)),
    distribution.model="norm")
  model<-ugarchfit(spec=spec,data=datos,
    fit.control=list(stationarity=1), solver="hybrid")
  coef<-model@fit$coef
  res_centr<-datos/as.numeric(sigma(model))
  residual<-(res_centr-mean(res_centr))/sd(res_centr)
  M<-sqrt(2/pi)
  y_boot<-var_boot<-var_boot2<-var2<-matrix(0,n,B)
  coef1<-coef2<-matrix(0,B,length(coef))
  s2<-yb<-c()
  yp<-s2p<-matrix(NA,ncol=h,nrow=B)
  for (i in 1:B){
    res<-sample(residual,n,replace=T)
    var1<-as.numeric(coef[4]/(1-coef[6]))
    for(j in 2:n){
      var_boot[1,i]<-var1
      y_boot[1,i]<-res[1]*sqrt(exp(var_boot[1,i]))
      var_boot[j,i]<-coef[4]+coef[5]*res[j-1]+
        coef[6]*var_boot[j-1,i]+ coef[7]*(abs(res[j-1])-M)
      y_boot[j,i]<-res[j]*sqrt(exp(var_boot[j,i]))
    }
    coef2[i,]<-ugarchfit(spec=spec, data=y_boot[,i],
```



```

        fit.control=list(stationarity=1),
        solver="hybrid")@fit$coef
for(l in 2:n){
  var_boot2[1,i]<-coef2[i,4]/(1-coef2[i,6])
  var_boot2[l,i]<-coef2[i,4]+coef2[i,5]*res_centr[l-1]+
    coef2[i,6]*var_boot2[l-1,i]+
    coef2[i,7]*(abs(res_centr[l-1])-M)
}
res_pred = sample(residual, h, replace=TRUE)
for (k in 2:h){
  s2[1]<-coef2[i,4]+coef2[i,5]*datos[n]/sqrt(exp(var_boot2[n,i]))+
    coef2[i,6]*var_boot2[n,i]+coef2[i,7]*
    (abs(datos[n]/sqrt(exp(var_boot2[n,i])))-M)
  yb[1] = res_pred[1]*sqrt(exp(s2[1]))
  s2[k] = coef2[i,4]+coef2[i,5]*res_pred[k-1]+coef2[i,6]*s2[k-1]+
    coef2[i,7]*(abs(res_pred[k-1])-M)
  yb[k] = res_pred[k]*sqrt(exp(s2[k]))
}
yp[i,]<-yb
s2p[i,]<-exp(s2)
}
OUT= list(yp=yp,s2p=s2p)
return(OUT)
}

```

ANEXO 15: Función realizado en el *software* R para el cálculo de las predicciones h pasos hacia adelante del modelo ARMA(1,1)-TGARCH(1,1).

```
boot_tgarch<-function(data,B,h,dummy){
  N<-dim(data)[1]
  datos<-ts(as.matrix(data[1:(N-h)]),start=2010,frequency=262))
  n<-dim(datos)[1]
  spec<-ugarchspec(variance.model=list(model="gjrGARCH",
                                       garchOrder=c(1,1)),mean.model=list(armaOrder=c(1,1),
                                       include.mean=FALSE,external.regressors=matrix(dummy)),
                   distribution.model="norm")
  model<-ugarchfit(spec=spec,data=datos,
                  fit.control=list(stationarity=1))
  coef<-model@fit$coef
  res_centr<-datos/as.numeric(sigma(model))
  residual<-(res_centr-mean(res_centr))/sd(res_centr)
  y_boot<-var_boot<-var_boot2<-var2<-matrix(0,n,B)
  coef1<-coef2<-matrix(0,B,length(coef))
  s2<-yb<-c()
  yp<-s2p<-matrix(NA,ncol=h,nrow=B)
  for (i in 1:B){
    res<-sample(residual,n,replace=T)
    var1<-as.numeric(coef[4]/(1-coef[5]-coef[6]-coef[7]/2))
    for(j in 2:n){
      var_boot[1,i]<-var1
      y_boot[1,i]<-res[1]*sqrt(var_boot[1,i])
      var_boot[j,i]<-coef[4]+coef[5]*y_boot[j-1,i]^2+
        coef[6]*var_boot[j-1,i]+
        coef[7]*y_boot[j-1,i]^2*ifelse(y_boot[j-1,i]<0,1,0)
      y_boot[j,i]<-res[j]*sqrt(var_boot[j,i])
    }
    restric<-ugarchfit(spec=spec, data=y_boot[,i],
```

```

fit.control =list(stationarity = 1))@fit$coef
while (length(restric)!=7|| (restric[5]+restric[6]+restric[7]/2)>=1||
      (restric[5]+restric[7])<=0) {
  res<-sample(residual,n,replace=T)
  var1<-as.numeric(coef[4]/(1-coef[5]-coef[6]-coef[7]/2))
  for(j in 2:n){
    var_boot[1,i]<-var1
    y_boot[1,i]<-res[1]*sqrt(var_boot[1,i])
    var_boot[j,i]<-coef[4]+coef[5]*y_boot[j-1,i]^2+
      coef[6]*var_boot[j-1,i]+
      coef[7]*y_boot[j-1,i]^2*ifelse(y_boot[j-1,i]<0,1,0)
    y_boot[j,i]<-res[j]*sqrt(var_boot[j,i])
  }
  restric<-ugarchfit(spec=spec, data=y_boot[,i],
                    fit.control =list(stationarity = 1))@fit$coef
}
coef1[i,]<-restric
res_pred<-sample(residual,h, replace=TRUE)
for ( l in 2:n){
  var_boot2[1,i]<-coef1[i,4]/(1-coef1[i,5]-coef1[i,6]-coef1[i,7]/2)
  var_boot2[l,i]<-coef1[i,4]+coef1[i,5]*datos[l-1]^2+
    coef1[i,6]*var_boot2[l-1,i]+
    coef1[i,7]*datos[l-1]^2*ifelse(datos[l-1]<0,1,0)
}
for (k in 2:h){
  s2[1]<-coef1[i,4]+coef1[i,5]*datos[n]^2+
    coef1[i,6]*var_boot2[n,i]+
    coef1[i,7]*datos[n]^2*ifelse(datos[n]<0,1,0)
  yb[1] = res_pred[1]*sqrt(s2[1])
  s2[k] = coef1[i,4]+coef1[i,5]*yb[k-1]^2+coef1[i,6]*s2[k-1]+
    coef1[i,7]*yb[k-1]^2*ifelse(yb[k-1]<0,1,0)
  yb[k] = res_pred[k]*sqrt(s2[k])
}

```

```
}  
yp[i,]<-yb  
s2p[i,]<-s2  
}  
OUT= list(yp=yp,s2p=s2p)  
return(OUT)  
}
```