MODELACIÓN DE LA SERIE DE RETORNOS DIARIOS DE LA ACCIÓN DE ECOPETROL EN EL PERIODO: 27/NOVIEMBRE/2007-25/NOVIEMBRE/2013

Kuri S. Katherine A.[*]; Ojeda E. Cesar A.[**]; Ovalle M. Diana P.[***]

[*]*Universidad del Valle, Facultad de Ingeniería, Escuela de Estadistica*

e-mail: *katherine.kuri@correounivalle.edu.co*

[**]*Universidad del Valle, Facultad de Ingeniería, Escueal de Estadística*

e-mail: *cesar.ojeda@correounivalle.edu.co*

[***]*Universidad del Valle, Facultad de Ingeniería, Escueal de Estadística*

e-mail: *diana.ovalle@correounivalle.edu.co*

Abstract: The objective is to model the series of daily action of the Colombian oil company Ecopetrol. Yields or returns are calculated based on the closing price of the stock and are modeled using time series methodology involving the effect of conditional heteroscedastic in explaining volatility. The work shows that yields Ecopetrol action can be modeled with GARCH (1,1) fit.

Keywords: Volatility Models, GARCH, TGARCH, Returns, Stylized Facts

Resumen: El objetivo es modelar la serie de retornos diarios de la empresa Colombiana de petróleos Ecopetrol. Los rendimientos o retornos se calculan de acuerdo al precio del cierre diario de la acción y se modelan por medio de la metodología de series temporales involucrando el efecto de heterocedasticidad condicional para explicar la volatilidad. El trabajo realizado muestra que los rendimientos de la acción de Ecopetrol obedecen a un modelo GARCH(1,1).

Palabras claves: Modelo de Volatilidad, GARCH, TGARCH, Retornos, Hechos Estilizados.

---

1. INTRODUCCIÓN

La empresa Colombiana de Petróleos, Ecopetrol inicia su historia en el año 1951 como una compañía del estado. A lo largo de su operación Ecopetrol se ha consagrado como una empresa estable en el campo financiero logrando ubicarse como una de las cuarenta petroleras más grandes en el mundo y una de las cuatro más importantes en Latinoamérica según el reporte integrado de gestión sostenible 2011 de Ecopetrol.

Con el fin de crecer e internacionalizase Ecopetrol lanza al mercado la venta de paquetes de acciones que cualquier colombiano puede adquirir. Es inquietante para un inversor conocer que tan bueno son los rendimientos de las acciones y si vale la pena invertir.

La econometría permite ajustar modelos a las series de rendimientos de cualquier activo involucrando la teoría de series temporales con el fin de identificar la estructura condicional con respecto al tiempo que se encuentra detrás de los retornos. Según [Tsay, 2005] para los inversionistas el análisis de los rendimientos de un activo presenta un resumen completo de la oportunidad de invertir sin que la escala de medición de los precios influya, por esta razón, el objetivo de este artículo es modelar los rendimientos diarios de la acción de la empresa Ecopetrol durante el período 27 noviembre 2007 - 25 noviembre 2013. Cabe resaltar que este período de tiempo ha sido escogido por la disponibilidad de los datos.

Este trabajo es el resultado del proyecto final de econometría del programa académico de estadística desarrollado durante el segundo semestre del año 2013. Su importancia radica en que permitirá a los inversionistas conocer el comportamiento de la volatilidad condicional de los rendimientos de la acción diaria de Ecopetrol facilitando la toma de decisiones a la hora de invertir.

2. METODOLOGÍA

Para la modelación de los retornos de la acción diaria de Ecopetrol se usó el software estadístico R [Core Team, 2012] y las librerías utilizadas son FinTs desarrollada por [Graves, 2008] adecuada para el análisis de datos financieros y Rugarch implementada por [Ghalanos, 2013] cuyo objetivo es el ajuste de modelos GARCH univariados. En total se tomó el precio de la acción diaria de Ecopetrol de 1462 días obtenidos de la página web: *http://www.ecopetrol.com.co/indicadores_historico.aspx?indID=5*

La modelación de la serie de retornos de la acción Ecopetrol inicia con un análisis descriptivo cuyos resultados se resumen en la sección 3 de Resultados y Discusión, donde se observa algunas características de interés de los rendimientos a través de los Log-Retornos de esta acción, teniendo en cuenta que según [Tsay, 2005] define el rendimiento de los activos como se muestra en la ecuación 1 donde P_{t-1} es el precio del activo en el tiempo $t-1$ y P_t es el precio en el tiempo t (hoy).

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1)$$

De la ecuación anterior se derivan los Log-retornos calculados como se muestra en la ecuación (2)

$$r_t = ln(1 + R_t) \quad (2)$$

En la modelación de series temporales es importante identificar si el fenómeno analizado es estacionario o no. De manera formal, [Tsay, 2005] define un proceso débilmente estacionario de orden n si todos los momentos conjuntos hasta de orden n existen y son invariantes en el tiempo. De esta manera, un proceso débilmente estacionario de orden 2 presenta media y varianza constante con covarianzas y funciones de autocorrelación que no dependen del tiempo.

Así, para el ajuste del modelo se hace uso de la Función de Autocorrelación (ACF) y la Función de Autocorrelación Parcial (PACF) de los rendimientos con el fin de conocer si existe un modelo en media condicional, también se hace uso de la prueba Ljung-Box donde se contrasta el siguiente planteamiento de hipótesis:

$$H_o : \rho_1 = \rho_2 = \cdots = \rho_k = 0$$
$$H_a : \rho \neq 0 \; para \; algún \; i = 1, 2, \cdots, k$$

Donde k es el número de rezagos a probar, si existe alguna correlación significativa indica la posible existencia de un modelo en media condicional, luego de hallar el modelo en media condicional si lo hay se procede a hallar el modelo en varianza condicional, en este artículo se inicia con un modelo general TGARCH(p,q) ([Glosten et al., 1993]; [Zakoian, 1994]) , las especificaciones de este modelo se presentan en la ecuación 3 donde el parámetro λ captura el efecto de asimetría [Nelson, 1991] que puede presentarse en las series financieras, llamado efecto leverage donde algunas series financieras tienden a disminuir la volatilidad cuando los retornos aumentan y a aumentarla cuando los retornos caen. Cabe resaltar que la volatilidad es definida por [Tsay, 2005] como la desviación estándar condicional de los rendimientos de un activo. Los parámetros α_i y β_j involucran el efecto

* Estadística, Universidad del Valle.

** Profesor Escuela de Estadística, Universidad del Valle.

*** Estadística, Universidad del Valle.

de la volatilidad condicional.

El parámetro C está asociado al efecto del premio al riesgo dando lugar a los modelos GARCH-M [Bollerslev, 1986] implicando que los retornos se encuentran correlacionados serialmente con la heterocedasticidad condicional. Los a_t son shocks aleatorios que dependen de la desviación estándar condicional y de la distribución de los errores ε_t.

$$\sigma^2 = \alpha_0 + \sum_{i=1}^{p}(\alpha_1 + \gamma_i N_{t-i})a_{t-1}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2 \quad (3)$$

Donde $r_t = u + c\sigma_t^2 + a_t$, $a_t = \sigma_t\varepsilon_t$
$\varepsilon \sim iid(GED(V))$

$$N_{t-1} = \begin{cases} 1; \; si\ a_{t-i} < 0 \\ 0; \; si\ a_{t-i} \geq 0 \end{cases}$$

Las series de los retornos financieros poseen unas regularidades empíricas llamadas hechos estilizados, entre ellos se encuentra las colas pesadas de la distribución de los retornos donde esta distribución suele ser más leptocurtica que la distribución normal, por esta razón se propone el uso la distribución GED ya que gracias a que las colas de esta distribución son más pesadas que la distribución Normal permite tener una mayor probabilidad de obtener valores extremos.

Al probar la significancia de los parámetros se observa si el parámetro de asimetría es significativo, si resulta no significativo el modelo se puede reducir a un modelo más sencillo GARCH(p,q) [Bollerslev, 1986]especificado en la ecuación 4. Este modelo tiene la particularidad que no tiene en cuenta el efecto de asimetría en la serie, este tipo de modelos responden de la misma manera frente a shocks negativos y positivos.

$$\sigma^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i a_{t-1}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2 \quad (4)$$

Donde $r_t = a_t$, $a_t = \sigma_t\varepsilon_t$, $\varepsilon \sim iid(GED(V))$

Luego de proponer un modelo en media condicional (si existe) y un modelo en varianza condicional se procede a la validación de supuestos de los posibles modelos, utilizando los shocks estandarizados $a_t^s = \dfrac{\hat{a}_t}{\hat{\sigma}_t}$

Las pruebas para comprobar los supuestos son las siguientes:

- Los estadísticos de Ljung-Box de a_t^s para probar la validez de la ecuación en media condicional.
- Los estadísticos de Ljung-Box de $(a_t^s)^2$ para probar la validez de la ecuación de la volatilidad condicional.
- La Curtosis, la asimetría y gráficos de probabilidad de a_t^s para probar la validez de la distribución usada.
- La estabilidad de los parámetros tanto individual como grupal para identificar si los coeficientes estimados varían o no con el tiempo.

Tabla 1. Descriptivas Porcentuales

Núm. Observaciones	1462
Media	0.058587
Desviación Estándar	1.687459
Coeficiente Asimetría	0.8060464
Exceso Curtosis	14.84405
Mínimo	-9.378296
Máximo	19.06693

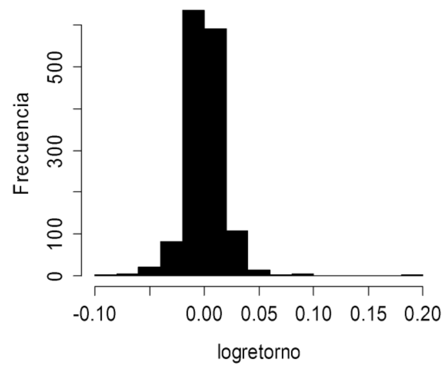

Figura 1. Histograma Log-Retorno

3. RESULTADOS Y DISCUSIÓN

El precio de la acción diaria de Ecopetrol en el período evaluado cuenta con un total de 1463 observaciones, por el procedimiento de cálculo, los log-retornos cuentan con una observación menos obteniendo una serie de 1462 datos.

La serie de Log-retornos y de los precios se observa en la figura 2, en la cual se evidencia la volatilidad representada por los fuertes cambios con respecto a la magnitud de los retornos. El histograma de la figura 1 muestra que la densidad incondicional de los log-retornos es simétrica con un centramiento cercano a cero y se caracteriza por una curtosis mayor a la presentada en la distribución normal.

Las estadísticas descriptivas resumidas en la tabla 1 se encuentran en términos porcentuales evidenciando que los log-retornos en promedio son del 0.05858778 %. Con respecto a la curtosis de la distribución normal, los datos presentan un exceso de 14.84405 y un coeficiente de asimetría de 0.8060464.

En la figura 3 se presenta la ACF y la PACF

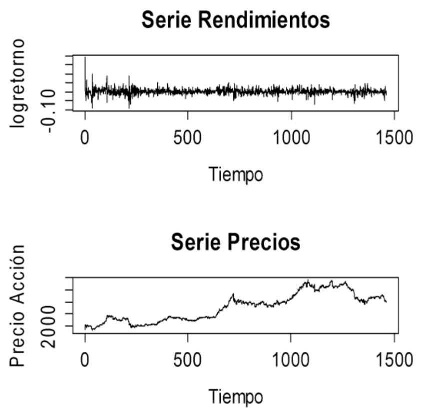

Figura 2. Serie de Rendimientos y Serie del Precio de la acción respectivamente

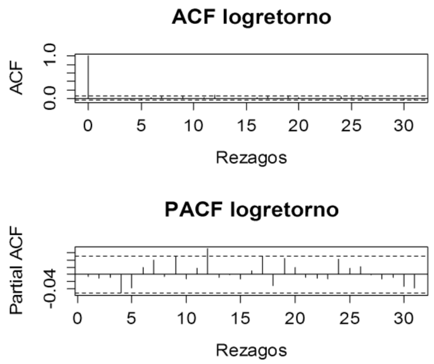

Figura 3. ACF y PACF Log-retorno

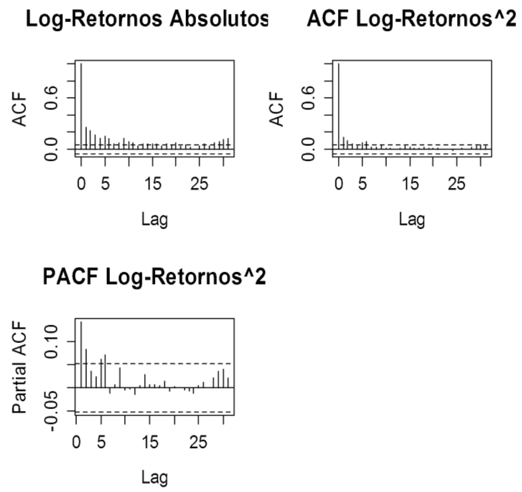

Figura 4. ACF y PACF Log-retorno

Tabla 2. Modelo TGARCH-M(2,2)

Parámetro	Estimador	Error Estándar	Valor P
Media	0.000452	0.000708	0.5225
C(ARCH-M)	-0.480642	2.772726	0.8623
α_0	0.000028	0.000018	0.1280
α_1	0.119905	0.060438	0.0472
α_2	0.024476	0.033513	0.4651
β_1	0.618355	0.448189	0.1676
β_2	0.081630	0.297755	0.7839
γ_1	0.184296	0.120400	0.1258
γ_2	-0.091170	0.164695	0.5798
Shape(V)	1.165074	0.058991	0.0000

de los rendimientos, se observa que no existe valores significativos individualmente debido a la no significancia estadística de la ACF a lo largo de los rezagos, se procedió a aplicar el test de Ljung-Box para probar la significancia de la autocorrelaciones a hasta el rezago 20.

El valor p asociado al estadístico de prueba es 0.1037 el cual permite concluir que los retornos no presentan autocorrelación y por tanto no hay presencia de autocorrelación serial.

La figura 4 permite concluir que se observan autocorrelaciones significativitas tanto en los log-retornos absolutos como en los log-retornos al cuadrado, indicando que no existe independencia serial en los log-retornos.

Con el fin de identificar un modelo apropiado, se inició con un modelo general TARCH-M(2,2) para evaluar la significancia de los parámetros estimados. El modelo ajustado inicial se muestra en el ecuación 5.

$$\sigma^2 = 0.000028 + \sum_{i=1}^{2}(\alpha_1 + \gamma_i N_{t-i}) + \sum_{j=1}^{2}\beta_j\sigma_{t-j}^2 \qquad (5)$$

Donde $r_t = u - 0.4806\sigma_t^2 + a_t$, $a_t = \sigma_t\varepsilon_t$
$\varepsilon \sim iid(GED(V))$

$$N_{t-1} = \begin{cases} 1;\ si\ a_{t-i} < 0 \\ 0;\ si\ a_{t-i} \geq 0 \end{cases}$$

Al ajustar el modelo TGARCH-M(2,2) usando como distribución de los errores GED (Nelson 1991) donde su parámetro V es otro componente a estimar en el ajuste, se encontraron los resultados mostrados en la tabla 2 los cuales indican que los parámetros asociados al premio al riesgo ARCH-M(C), al modelo GARCH(2,2), los parámetros de asimetría y la media no son significativos, por lo tanto se procedió a estimar un modelo GARCH(1,1), especificado en la ecuación 6.

$$\sigma^2 = 0.000031 + 0.195549a_{t-1}^2 + 0.682121\sigma_{t-1}^2 \qquad (6)$$

Donde $r_t = a_t$, $a_t = \sigma_t\varepsilon_t$
$\varepsilon \sim iid(GED(V))$

En la tabla 3 se muestran los valores p asociados a la significancia de los parámetros, en ella se observa que todos los parámetros son significativos.

Tabla 3. Modelo GARCH(1,1)

Parámetro	Estimador	Error Estándar	Valor P
α_0	0.000031	0.000010	0.1280
α_1	0.195549	0.043967	0.000009
β_1	0.682121	0.068735	0.000000
Shape(V)	1.142147	0.058621	0.000000

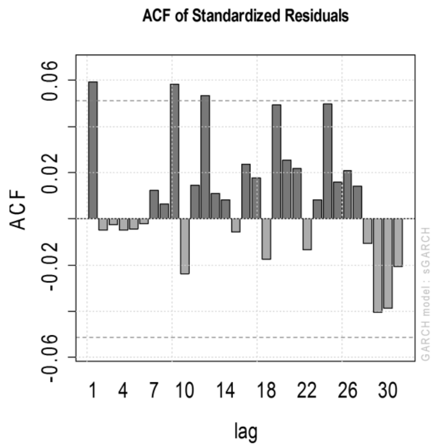

Figura 5. ACF Residuales Estandarizados

En este trabajo no se considera un modelo en media condicional ya que como se mencionó antes las funciones de autocorrelación de los retornos no son significativas; sin embargo en la figura 5 se observa que hay autocorrelaciones que se pensarían significativas con respecto a los residuales estandarizados, por lo tanto se procedió a aplicar la prueba Ljung-Box para evaluar formalmente la significancia de dichas correlaciones hasta el rezago 20.

El valor p resultante es 0.31 el cual indica que no existe suficiente evidencia para rechazar la hipótesis nula y por lo tanto, las autocorrelaciones conjuntas hasta el rezago 20 de los residuales estandarizados no son significativas indicando que el modelo en media condicional se encuentra correctamente especificado.

El modelo ajustado para explicar la volatilidad condicional es un GARCH(1,1) especificado en la ecuación 6. Según la figura 6 la funció n de autocorrelación parcial de los residuales estandarizados al cuadrado no son significativas, por lo tanto el modelo de heterocedasticidad condicional está bien especificado; sin embargo se aplicó el test de Ljung-Box para verificar este resultado.

El valor p resultante es 0.999 el cual indica que no existe suficiente evidencia para rechazar la hipótesis nula y por lo tanto, las autocorrelaciones conjuntas hasta el rezago 20 de los residuales estandarizados al cuadrado no son significativas indicando que el modelo en varianza se encuentra adecuadamente especificado.

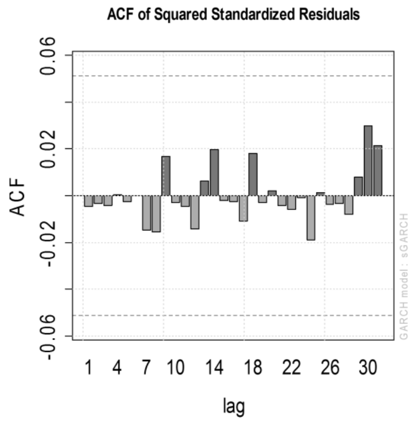

Figura 6. ACF Residuales Estandarizados al Cuadrado

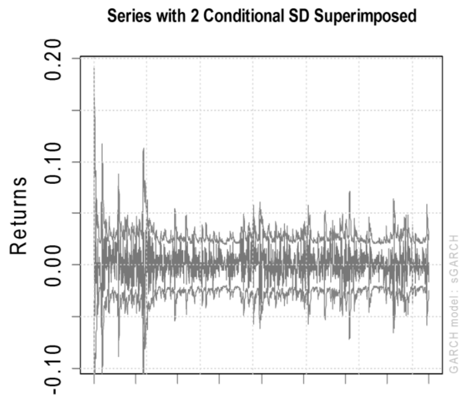

Figura 7. Serie de Rendimientos con dos desviaciones estándar superpuestas

La figura 7 muestra que el modelo en heterocedasticidad condicional logra capturar la volatilidad de los retornos a dos desviaciones estándar confirmando el buen ajuste del modelo.

Un preocupante resultado que se obtuvo es la estabilidad de los parámetros dentro del modelo ya que no se cumple de manera estadística, es decir que la estimación puede variar con el tiempo según lo menciona [Hansen, 1992]. Este supuesto de estabilidad se prueba con el siguiente planteamiento de hipótesis:

Ho: Las estimaciones son estables.

Ha: Al menos una de las estimaciones no son estables.

El estadístico de prueba conjunto a un nivel de significancia $\alpha = 0.05$ es 1.2826 y el valor critico 1.24, debido a que el estadístico es mayor a 1.24 se concluye que existe suficiente evidencia para rechazar la hipótesis nula y por tanto al menos una de las estimaciones no es estable.

Ahora interesa conocer cuáles son las estimaciones que están perjudicando la estabilidad del ajuste. El valor

Tabla 4. Estabilidad Individual

Parámetro	Estadístico de Prueba
α_0	0.3498
α_1	0.6040
β_1	0.4301
Shape(V)	0.7273

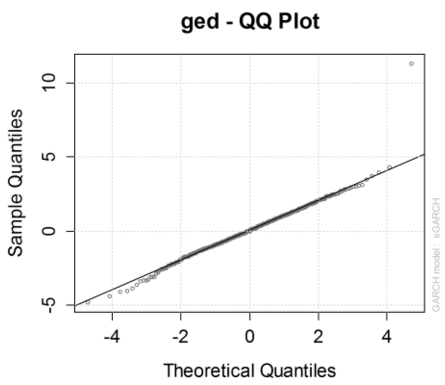

Figura 8. QQ-Plot Distribución Error Generalizada GED

crítico para la prueba individual asumiendo un nivel de significancia $\alpha = 0.05$ es 0.47 y los estadísticos de prueba se resumen en la Tabla 4. Los resultados indican que los parámetros α_1 y V(GED) no son estadísticamente estables.

Por otro lado, la verificación de la especificación correcta de la distribución de los errores como GED de parámetro V se realizó de manera gráfica y se presentan en las figuras 8 y 9 observando que los residuales estandarizados se ajustan correctamente a dicha distribución.

El parámetro de la distribución GED estimado es igual a 1.1421247 indicando que la distribución de los errores

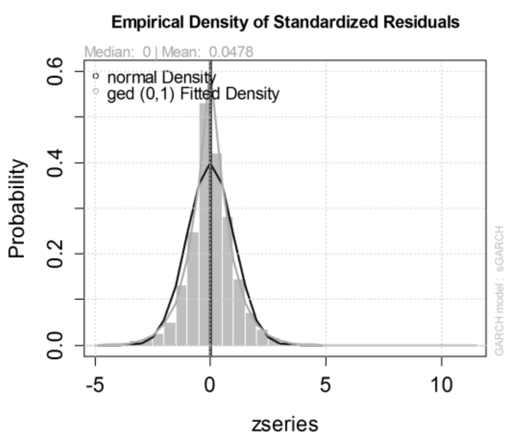

Figura 9. Comparación Distribución Normal y GED de los Residuales Estandarizados

Tabla 5. Criterios Selección de Modelos

Modelo	AIC	BIC
TGARCH(2,2)	-5.7053	-5.6691
TGARCH(2,2)	-5.7053	-5.6691
GARCH(1,1)	-5.7061	-5.6917

estandarizados es leptocúrtica con colas más pesadas que las referidas en la distribución normal.

Con respecto a la selección de modelos en la tabla 5 se observa que según los criterios Akaike AIC [Akaike, 1973] y Bayesiano BIC [Schwarz et al., 1978] el mejor modelo es el GARCH(1,1), resultado que respalda la idea de pasar de un modelo TGARCH(2,2) a un modelo reducido GARCH(1,1).

En la anterior figura 9, se observa un mejor ajuste por parte de la distribución GED con sus colas pesadas y su forma leptocurtica que la distribución Normal.

4. CONCLUSIONES

- El modelo GARCH(1,1) es adecuado para ajustar la heterocedasticidad condicional de los log-retornos de los precios de la acción de Ecopetrol.
- La no estabilidad en las estimaciones a lo largo del tiempo puede deberse a la existencia de posibles cambios estructurales o variables exógenas que para un análisis futuro se involucren en el modelo como los efectos calendario.
- El número de observaciones también puede ser un limitante en el ajuste del modelo impidiendo una mejor captura de la volatilidad en los rendimientos de la acción diaria de Ecopetrol.
- En la distribución de los errores estandarizados GED se observa un quantil atípico mayor a 10, por lo tanto para un análisis futuro se recomienda identificar la observación correspondiente e indagar la situación causante del valor tan alto del residual.
- El modelo ajustado resultante es de gran ayuda para los inversionistas, pues da una idea del rendimiento medio de los retornos de la acción diaria de Ecopetrol facilitando la toma de decisiones a la hora de invertir.

REFERENCIAS

[Akaike, 1973] Akaike, H. (1973). Information theory and an extension of the maximun likelihood principle. Second International Symposium on Information Theory. 267-281.

[Bollerslev, 1986] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

[Core Team, 2012] Core Team, R. (2012). R: A language and environment for statistical computing. *Vienna, Austria: R Foundation for Statistical Computing*. http://www.R-project.org/.

[Ghalanos, 2013] Ghalanos, A. (2013). *rugarch: Univariate GARCH models.* R package version 1.2-7.

[Glosten et al., 1993] Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801.

[Graves, 2008] Graves, S. (2008). Package fints, r versión 0.3-3. http://www.r-project.org.

[Hansen, 1992] Hansen, B. E. (1992). Tests for parameter instability in regressions with i (1) processes. *Journal of Business & Economic Statistics*, pages 321–335.

[Nelson, 1991] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370.

[Schwarz et al., 1978] Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

[Tsay, 2005] Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.

[Zakoian, 1994] Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5):931–955.