



ugr

Universidad
de Granada



Escuela
Internacional
de Posgrado

Series Temporales Avanzadas - Modelos Heterocedásticos

Una aplicación en la estimación de rendimientos financieros para
acciones, utilizando el software estadístico R.

Jorge Andrés Paguay Ortiz

Universidad de Granada

Departamento de Estadística e Investigación Operativa

Granada, España

2017

Series Temporales Avanzadas - Modelos Heterocedásticos

Una aplicación en la estimación de rendimientos financieros para acciones, utilizando el software estadístico R.

Jorge Andrés Paguay Ortiz

Trabajo presentado como requisito parcial para optar al título de:

Máster en Estadística Aplicada

Director:

Dr. Francisco Javier Alonso Morales

Línea de Investigación:

Análisis de series de tiempo, Aplicaciones a riesgos financieros, Inversiones

Universidad de Granada

Departamento de Estadística e Investigación Operativa

Granada, España

2017

Dedicatoria

A mi familia: Adriel y Yannira, por inspirarme a no desmayar.

A mis padres, por el regalo de la vida.

A mi hermano, por su entusiasmo.

A la UGR, por permitirme cursar este enriquecedor programa de estudios.

Agradecimientos

Quiero dar gracias a todas las personas que estuvieron pendientes de mi progreso en esta etapa de mi vida.

Quiero agradecer a los profesores del Máster en Estadística Aplicada de la UGR, por su dedicación a esta rama de la ciencia. En especial a Javier Alonso y Andrés González.

CONTENIDO

RESUMEN	8
1. MODELOS HETEROCEDÁSTICOS CONDICIONALES	9
1.1. MODELOS ARCH	10
1.1.1. MODELO ARCH(1)	11
1.1.2. MODELO ARCH(R)	13
1.2. MODELOS GARCH	14
1.2.1. MODELO GARCH(1,1)	16
1.2.2. MODELO IGARCH.....	17
1.2.3. MODELO EGARCH.....	18
1.3. CONSTRUCCIÓN DE MODELOS.....	19
1.3.1. IDENTIFICACIÓN DEL MODELOS	19
1.3.2. ESTIMACIÓN DEL MODELO	20
1.3.2.1. ESTIMACIÓN DE LOS PARÁMETROS, MODELO ARCH	20
1.3.2.2. ESTIMACIÓN DE LOS PARÁMETROS, MODELO GARCH	22
1.3.3. VALIDACIÓN DEL MODELO	23
2. R, COMO HERRAMIENTA PARA LA ESTIMACIÓN DE MODELOS	
HETEROCEDÁSTICOS CONDICIONALES.....	24
2.1. LECTURA DE DATOS	24
2.2. TRANSFORMACIÓN DE LOS DATOS EN UN OBJETO SERIE DE TIEMPO	26
2.3. ESTADÍSTICA DESCRIPTIVA DE LA SERIE DE TIEMPO	27
2.4. CALCULO DE RETORNOS	29
2.5. PRUEBAS PARA DETECTAR EFECTOS HETORECEDÁSTICOS	30
2.5.1. PRUEBA LJUNG-BOX, PARA INDEPENDENCIA	30

2.5.2.	PRUEBA MULTIPLICADOR DE LAGRANGE, PARA EFECTOS ARCH, GARCH.....	31
2.6.	FUNCIÓN DE AUTO-CORRELACIONES.....	33
2.7.	FUNCIÓN DE AUTO-CORRELACIONES PARCIALES	34
2.8.	MODELOS ARIMA(P,D,Q)	36
2.9.	MODELOS GARCH(R,S).....	38
2.10.	CRITERIO DE ELECCIÓN ENTRE MODELOS	39
3.	ESTIMACIÓN DE UN MODELO ARIMA-GARCH, PARA UNA SERIE DE	
	RENDIMIENTOS FINANCIEROS.	42
3.1.	ANÁLISIS DE LA SERIE DE RENDIMIENTOS, PRUEBA DE RAÍCES UNITARIAS.	44
3.2.	ESTIMACIÓN DEL MODELO ARIMA.....	45
3.3.	ANÁLISIS DE LA SERIE DE RESIDUOS DEL MODELO ARIMA, EFECTOS GARCH.....	51
3.4.	ESTIMACIÓN DEL MODELO GARCH	52
3.5.	PRESENTACIÓN MODELO ARIMA-GARCH ESTIMADO	56
	BIBLIOGRAFÍA.....	58
	ANEXO 1: CÓDIGO R.....	59
	ANEXO 2: COMPLEMENTOS TEÓRICOS.....	66
1.1.	SERIES DE TIEMPO Y PROCESOS ESTOCÁSTICOS	66
1.1.1.	FUNCIÓN DE MEDIAS.....	66
1.1.2.	FUNCIÓN DE AUTO-COVARIANZAS	66
1.1.3.	FUNCIÓN DE AUTO-CORRELACIONES.....	67
1.1.4.	FUNCIÓN DE AUTO-CORRELACIONES PARCIALES	67
1.1.5.	COMPONENTES DE UNA SERIE DE TIEMPO	68
1.2.	PROCESOS ESTOCÁSTICOS ESTACIONARIOS	69
1.2.1.	PROCESOS ESTOCÁSTICOS Estrictamente Estacionarios	69

1.2.2.	PROCESOS ESTOCÁSTICOS DÉBILMENTE ESTACIONARIOS	69
1.3.	PROCESOS LINEALES	69
1.3.1.	PROCESO LINEAL INVERSIBLE	70
1.3.2.	PROCESOS AUTOREGRESIVOS, $AR(p)$	70
1.3.3.	PROCESOS MEDIAS MÓVILES, $MA(q)$	71
1.3.4.	PROCESOS MEDIAS MÓVILES - AUTORREGRESIVOS, $ARMA(p,q)$	71
1.3.5.	PROCESOS MEDIAS MÓVILES - AUTORREGRESIVOS INTEGRADOS, $ARIMA(p,d,q)$	72

Resumen

En el presente trabajo se hace una recopilación de la teoría de series de tiempo, con énfasis en el estudio de variancia condicional. Para esto se analiza el caso de series temporales asociadas al retorno de activos financieros, donde, es habitual ver que la varianza cambie su magnitud con el paso del tiempo y por tanto los modelos de estimación de media condicional tengan sesgos al momento de ser aplicados.

Como respuesta a esta situación nacen los modelos que estiman varianza condicional para recoger el comportamiento fluctuante del error, que resulta de la estimación de media condicional. Y de esta forma tener un correcto ajuste de las series estudiadas.

Una vez que se ha comprendido los conceptos teóricos, revisamos el uso de varias funciones en R, que permiten hacer los cálculos con rapidez y confiabilidad. Es decir, nos familiarizamos con el lenguaje R para la estimación de modelos que estiman media condicional y varianza condicional, más comúnmente conocidos como modelos ARIMA y GARCH respectivamente.

Además estudiamos las pruebas de hipótesis necesaria para garantizar el cumplimiento de los requerimientos teóricos, antes de la estimación de cada modelo.

Finalmente, aplicamos la teoría de series de tiempo para estimación de media y varianza condicional en una serie de retornos real, mediante el empleo de funciones R.

Palabras claves.- Series de tiempo, media condicional, varianza condicional, retornos financieros, R.

1. Modelos Heterocedásticos Condicionales

En este capítulo se realiza una introducción formal al estudio de series de tiempo con varianza heterocedástica y a los modelos que permiten su estimación.

La varianza heterocedástica es una característica común en los activos financieros. Por ejemplo, las acciones mantienen un precio relativamente estable por periodos de tiempo y ante una noticia relevante, el precio comienza a fluctuar fuera del patrón en el que se encontraba, para finalmente estabilizarse en un nuevo patrón.

Es claro que este comportamiento genera cambios en la varianza de la serie a medida que el tiempo pasa. Los modelos de series de tiempo clásicos (Box-Jenkins) no están diseñados para recoger este tipo de comportamientos, ya que estos modelos asumen que la serie de tiempo siempre tiene varianza constante.

Ante este hecho aparecen los modelos de varianza condicional, estos modelos permiten explicar el comportamiento de series de tiempo con varianza no constante o heterocedástica, un comportamiento común en las series de retornos financieros. A continuación listamos estos modelos:

- El modelo ARCH planteado por Engle en 1982, de forma muy general podemos decir que, en este modelo la varianza condicionada se explica utilizando el cuadrado de los errores estandarizados.
- El modelo GARCH planteado por Bollerslev en 1986, en el cual la varianza condicional se explica por el cuadrado de los errores estandarizados y sus varianzas condicionadas pasadas.
- De los modelos anteriores se desprenden, modelos como: EGRACH, TGARCH, entre otros. Mismo que responden a variaciones de los modelos anteriores.

1.1. Modelos ARCH

Los modelos ARCH se utilizan para modelar series de tiempo heterocedásticas, como las series de retornos de activos financieros.

Podemos definir el retorno de un activo financiero como la variación porcentual del precio actual con respecto al precio del periodo anterior, de la siguiente forma:

$$r_t = \log(P_t) - \log(P_{t-1})$$

De forma que la formulación básica de este modelo corresponde a:

$$r_t = \sigma_t \varepsilon_t$$

Donde, ε_t es un proceso de ruido blanco formado por variables aleatorias normales independientes de media cero y varianza 1 y σ_t factor de volatilidad son procesos estacionarios independientes entre sí.

La condición de independencia en σ_t y ε_t , garantiza que la serie r_t tenga media igual a cero:

$$E(r_t) = E(\sigma_t \varepsilon_t) = E(\sigma_t)E(\varepsilon_t) = 0$$

De igual forma sucede con la media condicionada que es nula:

$$E(r_t | r_{t-1}) = E(\sigma_t | r_{t-1})E(\varepsilon_t) = 0$$

Al ser r_t un proceso estacionario, tenemos una varianza marginal constante σ^2 .

$$E(r_t^2) = E(\sigma_t^2 \varepsilon_t^2) = E(\sigma_t^2)E(\varepsilon_t^2) = \sigma^2 * 1 = \sigma^2$$

No así la varianza condicionada.

$$Var(r_t^2 | r_{t-1}) = E(\sigma_t^2 | r_{t-1})E(\varepsilon_t^2) = \sigma_t^2$$

Se puede ver que σ_t^2 representa la varianza condicionada de la serie en cada instante de tiempo, en adelante nos referiremos a ella como volatilidad.

También tenemos que:

$$E(\varepsilon_t^2 | r_{t-1}) = E(\varepsilon_t^2) = 1$$

Por tanto, σ_t^2 representa la varianza condicionada de la serie en cada instante de tiempo, que va variando con cierta estructura estacionaria.

La condición de independencia entre σ_t y ε_t , además de garantizar que la serie r_t tenga media marginal cero, carezca de autocorrelación y forme un ruido blanco. Sin embargo las variables de la serie nos son independientes.

1.1.1.1. Modelo ARCH(1)

Partiremos de una serie temporal de retornos r_t , la cual es secuencialmente correlacionada con media cero, por tanto la varianza condicional de r_t dada la rentabilidad pasada, no es constante.

Denotamos la varianza condicional de r_t como $\sigma_{t|t-1}^2$, el condicionamiento se refiere al periodo t-1. Para obtener un estimador insesgado de $\sigma_{t|t-1}^2$ elevamos r_t al cuadrado.

En el modelo ARCH(1), sabemos que la estructura de la varianza condicionada es un AR(1). De modo que la serie de retornos se genera así:

$$r_t = \sigma_{t|t-1} \varepsilon_t$$

$$\sigma_{t|t-1}^2 = \alpha_0 + \alpha_1 r_{t-1}^2$$

Donde α_0 y α_1 son parámetros desconocidos y ε_t es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza uno, además es independiente de r_{t-j} , $j = 1, 2, \dots$

Como ε_t tiene varianza 1, la varianza condicional de r_t es igual a $\sigma_{t|t-1}^2$, por consiguiente tenemos:

$$E(r_t^2 \setminus r_{t-j}) = E(\sigma_{t \setminus t-1}^2 \epsilon_t^2 \setminus r_{t-j})$$

$$E(r_t^2 \setminus r_{t-j}) = \sigma_{t \setminus t-1}^2 E(\epsilon_t^2) = \sigma_{t \setminus t-1}^2$$

El modelo ARCH es similar a un modelo de regresión, su uso no es igual ya que la varianza condicional unitaria no es directamente observable. Tomemos un valor observado $\epsilon_t = r_t^2 - \sigma_{t \setminus t-1}^2$, donde ϵ_t representa una serie no correlacionada de media cero y también no correlacionada con el retorno pasado. Despejando y sustituyendo $\sigma_{t \setminus t-1}^2 = r_t^2 - \epsilon_t$ tenemos:

$$r_t^2 = \sigma_{t \setminus t-1}^2 + \epsilon_t$$

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \epsilon_t$$

Lo que indica que la serie de retornos al cuadrado satisface un modelo AR(1) bajo el supuesto de una serie de retornos con modelos ARCH(1). Dado que r_t^2 debe ser no negativo, los α_i también deben ser no negativos.

Por otro lado, sabemos que la serie de retornos es estacionaria con varianza $r_{t-1}^2 = r_t^2 = \sigma^2$, sustituyendo en la ecuación anterior tenemos:

$$\sigma^2 = \alpha_0 + \alpha_1 \sigma^2$$

$$\sigma^2 = \frac{\alpha_0}{(1 - \alpha_1)}$$

Con $0 \leq \alpha_1 \leq 1$, condición suficiente y necesaria para la estacionalidad débil del ARCH(1).

En un proceso estacionario la varianza es constante, la condición $0 \leq \alpha_1 \leq 1$ implica que existe una distribución inicial r_0 de tal forma que r_t para $t \geq 1$ sea débilmente estacionario. Por tanto para el modelo ARCH(1) la estacionariedad débil no excluye un proceso de varianza condicional, es decir el modelo ARCH(1) es de ruido blanco y admite

un proceso de varianza condicional no constante que varia con el retardo de uno de los procesos al cuadrado (Cryer y Chan, 2008).

En un modelo ARCH(1), si los errores estandarizados tienen distribución normal, su distribución estacionaria con $1 > \alpha_1 > 0$ es de cola gruesa, es decir su curtosis $\frac{E(r_t^4)}{\sigma^4} - 3$ es mayor que cero (Cryer y Chan, 2008).

Características de un modelo ARCH:

- La esperanza marginal y condicional son iguales a cero.
- La varianza marginal es constante.
- La varianza condicional depende de los valores que haya tomado r_{t-1}^2 , luego no es constante.
- La distribución del proceso ARCH(1) tiene una forma desconocida.

1.1.2. Modelo ARCH(r)

El modelo anterior puede generalizarse permitiendo una dependencia de la varianza condicional con r retardos. De manera que el modelo ARCH(r) para $r_t = \sigma_{t \setminus t-r} \varepsilon_t$ de varianza condicionada sea:

$$\sigma_{t \setminus t-r}^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \cdots + \alpha_r r_{t-r}^2$$

Donde $\alpha_0 > 0$ corresponde a la mínima varianza condicional observada y $0 \leq \alpha_1 \leq 1$ es una condición necesaria y suficiente para la existencia de la varianza incondicional y la condicionada.

Este modelo implica que las posibilidades de rachas de alta volatilidad dependerán de los r últimos valores. Así tenemos:

$$\sigma^2 = E(r_t^2) = E[E(r_{t \setminus t-1}^2)] = \alpha_0 + \alpha_1 E(r_{t-1}^2) + \cdots + \alpha_r E(r_{t-r}^2)$$

De donde obtenemos que:

$$\sigma^2 = \alpha_0 + \alpha_1 \sigma^2 + \dots + \alpha_r \sigma^2$$

$$\sigma^2 = \frac{\alpha_0}{(1 - \alpha_1 - \dots - \alpha_r)}$$

Con la restricción $\sum_{i=1}^r \alpha_i < 1$, por el hecho de varianza positiva.

Ahora tomemos un valor observado $\epsilon_t = r_t^2 - \sigma_{t|t-r}^2$, donde ϵ_t representa una serie no correlacionada de media cero y también no correlacionada con el retorno pasado.

Despejando y sustituyendo $\sigma_{t|t-r}^2 = r_t^2 - \epsilon_t$ tenemos:

$$r_t^2 = \sigma_{t|t-r}^2 + \epsilon_t$$

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_r r_{t-r}^2 + \epsilon_t$$

Donde la positividad de r_t^2 exige que $-(\alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_r r_{t-r}^2) < \epsilon_t$.

Manteniendo las características listadas antes, para un modelo ARCH(r):

- La esperanza marginal y condicional son iguales a cero.
- La varianza marginal es constante.
- La varianza condicional depende de los valores que haya tomado $r_{t-1}^2, \dots, r_{t-r}^2$, luego no es constante.
- Es un proceso de ruido blanco pero no es independiente y no esta idénticamente distribuido.

1.2. Modelos GARCH

En un número importante de casos, los modelos ARCH se caracterizan por: requerir un gran número de parámetros autoregresivos para representar adecuadamente el comportamiento dinámico de la varianza e imponen una estructura fija de retardos. Como

respuesta a estas restricciones Bollerslev (1986) propone los modelos GARCH, estos modelos permiten flexibilizar esta restricciones.

La generalización de los modelos ARCH, mediante modelos GARCH se asemeja mucho a la extensión de los procesos autoregresivos AR y ARMA. Permitiendo una representación más parsimoniosa de la volatilidad.

Bollerslev plantea que la varianza condicionada, σ_{t-r}^2 depende de los valores previos de la variable y de sus valores anteriores, podemos representar esta idea de la siguiente forma:

$$\sigma_{t-r}^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \cdots + \alpha_r r_{t-r}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2$$

Donde $\alpha_0 > 0$, $\alpha_i \geq 0$ y $\beta_i \geq 0$ con $i = 1, \dots, r$ y $j = 1, \dots, p$ respectivamente, para garantizar que la varianza sea positiva.

Nelson (1990) demuestran que si el modelo GARCH admite una representación ARCH(∞), esto es suficiente para exigir que los coeficientes del polinomio de retardos en dicha representación sean todos positivos.

Este nuevo modelo se denomina GARCH(p,r), y se reduce al ya visto ARCH(r) cuando $p = 0$. Bollerslev establece las condiciones de estacionariedad de este modelo, probando que r_t es débilmente estacionaria con:

$$E(r_t) = 0$$

$$Var(r_t) = \frac{\alpha_0}{1 - \sum_{i=1}^r \alpha_i - \sum_{j=1}^p \beta_j}$$

$$Cov(r_t, r_s) = 0; \text{ para todo } t \neq s$$

Lo que implica:

$$\sum_{i=1}^r \alpha_i + \sum_{j=1}^p \beta_j < 1$$

La relación que existe entre los modelos GARCH y ARMA es importante. Definamos $v_t = r_t^2 - \sigma_t^2$, será un ruido blanco formado por variables estacionarias incorreladas de media cero y varianza marginal constante, podemos expresar la dependencia de los modelos GARCH como un proceso ARMA, según la siguiente ecuación:

$$r_t^2 = \alpha_0 + \sum_{i=1}^{\max(p,r)} (\alpha_i + \beta_i) r_{t-i}^2 + v_t - \sum_{j=1}^p \beta_j v_{t-j}$$

1.2.1. Modelo GARCH(1,1)

En el estudio de varias series financieras, se ha visto que el modelo más sencillo GARCH(1,1), es suficiente para modelizar con éxito los cambios temporales en la varianza condicional. El modelo GARCH(1,1) se obtiene cuando $p = r = 1$, de forma que la varianza condicionada se expresa así:

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Donde $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$. Con $\alpha_1 + \beta_1 < 1$, la serie r_t tiene varianza finita, y es ruido blanco con media cero y varianza.

$$Var(r_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

Bollerslev prueba que si $(\beta_1 + \alpha_1)^2 + 2\alpha_1^2 < 1$, el momento de cuarto orden 4 de r_t existe y es finito, la curtosis de r_t es:

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

Cuando $\alpha_1 + \beta_1 < 1$, este valor es mayor que 3 y, por tanto, el proceso GARCH(1,1) estacionario es leptocúrtico.

Dado que $p = r = 1$, la ecuación se escribe como:

$$r_t^2 = \alpha_0 + (\alpha_1 + \beta_1)r_{t-1}^2 + v_{t-1} - \beta_1 v_{t-1}$$

Este modelo puede interpretarse como un proceso ARMA(1,1) para la serie r_t^2 , cuya función de autocorrelación será:

Si $k = 1$:

$$\rho_c(1) = \frac{\alpha_1(1 - \alpha_1\beta_1 - \beta_1^2)}{(1 - 2\alpha_1\beta_1 - \beta_1^2)}$$

Si $k > 1$:

$$\rho_c(k) = (\alpha_1 + \beta_1)^{k-1} \rho_c(1)$$

1.2.2. Modelo IGARCH

En la estimación de modelos GARCH para series de rendimientos financieros es común encontrarse con valores $\alpha_1 + \beta_1$ cercanos a uno. Es te hecho ocasiona que la función de autocorrelación de r_t apenas decrezca, indicando que los cambios en la varianza condicionada son relativamente lentos y por ende los cambios bruscos como shocks persistan en la volatilidad.

Esta importante característica de las series financieras muestra que, aunque la serie original esta incorrelada, existe correlación en la serie de sus cuadrados y estas correlaciones decrecen lentamente, mostrando valores significativos en retardos altos del estudio.

Este hecho despierta el interés en estudiar modelos GARCH(1,1), donde $\alpha_1 + \beta_1 = 1$, a estos modelos les llamaremos IGARCH. Modelos GARCH integrados, mismos que fueron propuestos por Engle y Bollerslev (1986) bajo la siguiente ecuación para la varianza condicionada:

$$\sigma_t^2 = \alpha_0 + \sigma_{t-1}^2 + \alpha_1(r_{t-1}^2 - \sigma_{t-1}^2)$$

Podemos ver que este modelo no es débilmente estacionario, porque su varianza marginal no es finita. Organizando convenientemente la ecuación anterior podemos ver que admite la siguiente representación:

$$\Delta r_t^2 = \alpha_0 + v_t + \beta_1 v_{t-1}$$

Donde Δ es el operador de primeras diferencias, y $v_t = r_t^2 - \sigma_t^2$ es un proceso de ruido blanco, formado por variables estacionarias incorreladas de media cero y varianza marginal constante

Esta nueva ecuación permite interpretar el modelo IGARCH(1,1) como un proceso MA(1) estacionario para las primeras diferencias de r_t^2 , lo que muestra cierta similitud con los proceso ARIMA(0,1,1), pero no son iguales.

1.2.3. Modelo EGARCH

De los estudios realizados por Nelson(1991), sabemos que los modelos GARCH tienen las siguientes limitaciones:

- Las condiciones impuestas sobre los parámetros para asegurar que σ_t^2 no sea negativo son violadas en algunas aplicaciones empíricas.
- El modelo GARCH es incapaz de modelizar una respuesta asimétrica de la volatilidad ante las subidas y bajadas de la serie.

Ante estas observaciones Nelson propone el modelo EGARCH o GARCH exponencial. Este nuevo modelo garantiza la no negatividad de la varianza condicional formulando la ecuación de volatilidad en términos del logaritmo de σ_t^2 mediante la siguiente ecuación lineal:

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{j=1}^p \varphi_j \ln(\sigma_{t-j}^2) + g(\varepsilon_{t-1}) + \sum_{i=1}^r \Psi_i g(\varepsilon_{t-i-1})$$

Donde $g(\varepsilon_t) = \lambda_1 \varepsilon_t + \lambda_2 (|\varepsilon_t| - E|\varepsilon_t|)$. Mediante la función g , que depende del signo y de la magnitud de ε_t , el modelo EGARCH puede capturar una respuesta asimétrica de la volatilidad ante innovaciones de distintos signo, permitiendo así modelizar un efecto contrastado empíricamente en muchas series financieras. Malas noticias (rendimientos negativos) provocan mayor aumento de la volatilidad que las buenas noticias (rendimientos positivos).

Por definición, las perturbaciones $g(\varepsilon_t)$ son variables independientes e idénticamente distribuidas con media cero y varianza constante, por tanto el modelo EGARCH puede considerarse como una representación ARMA para la serie $\ln(\sigma_t^2)$.

1.3. Construcción de Modelos

Para la estimación de estos modelos vamos a seguir tres pasos secuenciales, identificación, estimación y validación del modelo. A continuación la descripción de estos:

1.3.1. Identificación del Modelos

Para saber si una serie requiere ser modelada mediante modelos ARCH o GARCH, necesitamos analizar los residuos estandarizados al cuadrado de un modelo ARIMA previamente ajustado para la serie. De aquí que el primer paso es realizar un ajuste en media a la serie de tiempo, para esto podemos utilizar la metodología de Box-Jenkins.

Del paso anterior obtenemos los residuos estandarizados del modelo ARIMA, con este insumo procederemos a calcular las autocorrelaciones simples y parciales de los residuos estandarizados al cuadrado, si las autocorrelaciones calculadas son significativos

sabremos que la serie requiere ser estimada utilizando procesos ARCH o GARCH. En el caso de series de retornos financieros es habitual ver que los residuos estandarizados están incorrelados pero no son independientes.

Si el modelo para los residuos estandarizados es un ARCH puro, el número de autocorrelaciones parciales significativas de los residuos estandarizados, nos indica el orden del modelo.

Además, podemos utilizar pruebas de hipótesis que permitan ratificar lo visto en los correlogramas. Las pruebas más utilizadas para este fin son: Ljung-Box, Multiplicador de Lagrange para efectos ARCH-GARCH y el McLeod y Li.

1.3.2. Estimación del Modelo

Desde el punto de vista teórico, para la estimación de los modelos ARCH y GARCH es habitual utilizar Máximos de Verosimilitud o métodos de momentos generalizados, partiendo del modelo de regresión ARCH.

En la práctica es importante limpiar la serie de valores atípicos antes de empezar con la estimación del modelo, ya que estos valores pueden crear confusión sobre la necesidad de hacer una estimación de varianza condicional, en especial cuando aparecen en rachas. Por otro lado, pasar por alto la necesidad de una estimación ARCH conllevar a creer que existen muchos valores atípicos en la estimación de la media condicional, modelos ARIMA.

1.3.2.1. Estimación de los parámetros, Modelo ARCH

A continuación vamos a describir la estimación de parámetros del modelo ARCH mediante el método de máxima verosimilitud.

Empezamos construyendo la función de verosimilitud, esta función se construye con el producto de las densidades condicionales, asumiendo que los errores ϵ_t en $e_t = \sigma_t \epsilon_t$ son variables aleatorias independientes e idénticamente distribuidas con media cero y varianza uno. Los modelos ARCH son condicionalmente gaussianos y la distribución condicionada $e_t \setminus e_{t-1}$ es $N(0, \sigma_t^2)$. Entonces tenemos como logaritmo de la función de verosimilitud:

$$\ln L(\theta) = \sum_{t=1}^T \ln f(e_t; \theta \setminus e_{t-1}) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^T \frac{e_t^2}{\sigma_t^2}$$

Donde θ es el vector de parámetros desconocidos del modelo y $f(e_t; \theta \setminus e_{t-1})$ denota la densidad condicionada de e_t dadas las observaciones previas hasta el instante t-1.

Bajo ciertas condiciones de regularidad se demuestra que, si los momentos de primer y segundo orden están correctamente especificados. Los estimadores de máxima verosimilitud son consistentes y asintóticamente normales.

Considerando el modelo ARCH(1), tenemos la siguiente función de verosimilitud:

$$f(e_t, \dots, e_T \setminus \alpha_0, \alpha_1) = f(e_1) * f(e_2 \setminus e_1) \dots f(e_T \setminus e_{T-1})$$

Donde, las funciones de densidad $f(e_t \setminus e_{t-1})$ son normales. Como $e_t = \sigma_t \epsilon_t$, considerando e_{t-1} , el valor de σ_t es una constante y la única variables es ϵ_t que tiene distribución normal.

Entonces tenemos que la media condicionada de la distribución $f(e_t \setminus e_{t-1})$ es cero y la varianza $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2$. La log-función de verosimilitud condicionada a e_1 será:

$$L(e_2, \dots, e_T \setminus \alpha_0, \alpha_1) = -\frac{T-1}{2} \ln 2\pi - \frac{1}{2} \sum_{t=2}^T \ln(\alpha_0 + \alpha_1 e_{t-1}^2) - \frac{1}{2} \sum_{t=2}^T \frac{e_t^2}{\alpha_0 + \alpha_1 e_{t-1}^2}$$

Derivando respecto a los parámetros, llamando $\widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 e_{t-1}^2$ e igualando a cero se obtiene las ecuaciones:

$$\sum \frac{e_t^2}{\widehat{\sigma}_t^4} = \sum \frac{1}{\widehat{\sigma}_t^2}$$

$$\sum \frac{e_{t-1}^2}{\widehat{\sigma}_t^2} = \sum \frac{e_{t-1}^2 e_t^2}{\widehat{\sigma}_t^4}$$

Multiplicamos y dividimos el primer miembro por $\widehat{\alpha}_0 + \widehat{\alpha}_1 e_{t-1}^2$ y obtenemos:

$$\widehat{\alpha}_0 \sum \frac{1}{\widehat{\sigma}_t^4} + \widehat{\alpha}_1 \sum \frac{e_{t-1}^2}{\widehat{\sigma}_t^4} = \sum \frac{e_t^2}{\widehat{\sigma}_t^4}$$

$$\widehat{\alpha}_0 \sum \frac{e_{t-1}^2}{\widehat{\sigma}_t^4} + \widehat{\alpha}_1 \sum \frac{e_{t-1}^2}{\widehat{\sigma}_t^4} \sum \frac{e_{t-1}^2 e_t^2}{\widehat{\sigma}_t^4}$$

Son las ecuaciones de mínimos cuadrados para obtener los parámetros del modelo.

Resolviendo este sistema de ecuaciones, tenemos la estimación de $\widehat{\alpha}_0$ y $\widehat{\alpha}_1$.

1.3.2.2. Estimación de los parámetros, Modelo GARCH

A continuación vamos a describir la estimación de parámetros del modelo GARCH mediante el método de máxima verosimilitud.

Consideremos el modelo GARCH(r,s), donde $r \geq s$, además definamos la función de verosimilitud de un proceso estacionario como $W_t = (w_1, \dots, w_t)$.

$$f(w_1, \dots, w_t | \alpha) = f(w_n | W_{n-1}) f(w_{n-1} | W_{n-2}) \dots f(w_{r+1} | W_r) f(w_1, \dots, w_r | \alpha)$$

$$f(w_1, \dots, w_t | \alpha) = \prod_{t=r+1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left[-\frac{(w_t - E(w_t | W_{t-1}))^2}{2\sigma_t^2} \right] * f(w_1, \dots, w_r | \alpha)$$

Ya que la varianza condicional de las variables es σ_t^2 . Condicionando a las primeras r observaciones, tenemos la función soporte condicionada:

$$L(w_{r+1}, \dots, w_n | \alpha, W_r) = -\frac{1}{2} \sum_{t=r+1}^n \ln \sigma_t^2 - \frac{1}{2} \sum_{t=r+1}^n \frac{(w_t - E(w_t | W_{t-1}))^2}{\sigma_t^2}$$

Que puede optimizarse mediante un algoritmo no lineal, para obtener los estimadores de los parámetros que aparecen en la media condicional y en la varianza condicional. De aquí importante notar que, la estimación puede realizarse en dos etapas ya que la correlación entre los parámetros ARMA y los parámetros GARCH habitualmente es pequeña.

- Etapa 1.- se estiman los parámetros de la media condicionada, es decir, el modelo ARMA, para obtener los errores estándar del modelo.

$$e_t = w_t - E(w_t | W_{t-1})$$

- Etapa 2.- se estima los parámetros de la varianza condicional, para esto maximizamos la verosimilitud de los residuos. O bien, se puede calcular las ecuaciones de media condicional y varianza condicional conjuntamente.

1.3.3. Validación del Modelo

Definamos a e_t como la serie de residuos resultantes de la Etapa 1, estimación de la media condicional y $\hat{\sigma}_t$ a la serie de varianzas resultantes de la Etapa 2.

Entonces los residuos estandarizados $e_t/\hat{\sigma}_t$, deben cumplir con las características de un ruido blanco, para verificar esto podemos utilizar correlogramas y pruebas de hipótesis como: Ljung-Box y Multiplicadores de Lagrange, mismas que al inicio de la Etapa 1 fueron utilizadas para decidir si la serie requiere estimación de variancia condicional. Si el proceso fue correctamente estimado, esperamos que en los correlogramas de los errores y de los errores al cuadrado no tengamos autocorrelaciones significativas. Y que las pruebas de hipótesis no rechacen la hipótesis nula, en favor de la alternativa.

2. R, como herramienta para la estimación de Modelos

Heterocedásticos Condicionales.

Para realizar los cálculos necesarios en el presente trabajo, se ha elegido el Software Estadístico de uso libre R, este software permite manejar la información de una serie de tiempo con facilidad. Además posee una gran cantidad de paquetes y funciones que se pueden utilizar para realizar la estimación de modelo heterocedásticos condicionales ARCH y GARCH.

En este capítulo nos enfocaremos en explicar las funciones necesarias para realizar los cálculos, pruebas de hipótesis y estimaciones, presentadas en el marco teórico. Para este fin vamos a utilizar, la información de cotización diaria de las acciones de la empresa estadounidense Amazon, comprendida en el periodo del 04/Ago/2016 hasta el 01/Ago/2017. Esta información ha sido obtenida de la plataforma: Google Finance y alojada en Github, en la dirección:

https://raw.githubusercontent.com/jorgepaguay86/EstadisticaAplicada_TFM_UGR/master/SerieDeTiempo_Amazon_Ago2016Ago2017.csv

Para facilitar el entendimiento del material expuesto en este capítulo.

2.1. Lectura de datos

En R podemos leer archivos de varios tipos, para nuestro caso particular vamos a enfocarnos en la lectura de un archivo .csv, que se encuentra en una url publica. Para este fin utilizamos la función `read.csv()` esta función nos permite cargar archivos .csv indicando su ubicación. Para ver esta información y de esta manera saber si la información se cargó correctamente utilizamos la función `View()`, esta función abre una ventana con la tabla

que contienen los datos leídos por la función `read.csv()`. A continuación mostramos la estructura y argumentos de estas funciones:

```
read.csv(file, header = TRUE, sep = ",", quote = "\"",  
         dec = ".", fill = TRUE, comment.char = "", ...)
```

Argumentos¹:

- `file`.- El nombre del archivo que contiene los datos, también puede ser una URL completa.
- `header`.- Indicador T/F que define si el archivo contiene el nombre de los campos en la primera línea del archivo o no.
- `sep`.- El carácter separador de campos.
- `quote`.- El conjunto de caracteres que citan.
- `dec`.- El carácter utilizado en el archivo para los puntos decimales.
- `fill`.- Indicador T/F en caso de que las filas tengan una longitud desigual, se añaden implícitamente campos en blanco.
- `comment.char`.- Un vector de caracteres de longitud uno que contiene un solo carácter o una cadena vacía.

```
View(x, title)
```

Argumentos:

- `x`.- Objeto R con valores distintos de cero, que puede ser representado en filas y columnas.
- `title`.- Título para la ventana del visor.

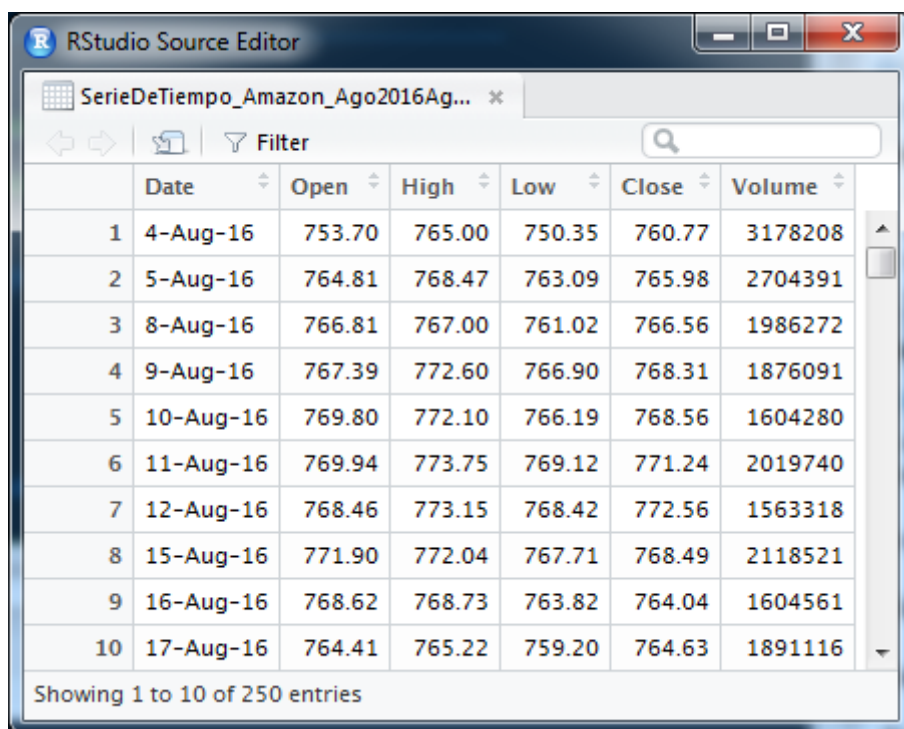
Ahora vamos a mostrar su uso, mediante el siguiente Código:

¹ Cada vez que revisemos los argumentos de una función. Estos corresponderán a la descripción tomada de la ayuda de la función en R e interpretada al español por el autor del presente trabajo.

```
ejemplo_data <-
read.csv('https://raw.githubusercontent.com/jorgepaguay86/EstadisticaAplicada_TFM_UGR/master/SerieDeTiempo_Amazon_Ago2016Ago2017.csv')

View(ejemplo_data, 'SerieDeTiempo_Amazon_Ago2016Ago2017')
```

De esta forma cargamos los datos que vamos a utilizar en este capítulo, podemos ver a continuación que la información se ha cargado correctamente y que contamos con 250 registros en la tabla.



	Date	Open	High	Low	Close	Volume
1	4-Aug-16	753.70	765.00	750.35	760.77	3178208
2	5-Aug-16	764.81	768.47	763.09	765.98	2704391
3	8-Aug-16	766.81	767.00	761.02	766.56	1986272
4	9-Aug-16	767.39	772.60	766.90	768.31	1876091
5	10-Aug-16	769.80	772.10	766.19	768.56	1604280
6	11-Aug-16	769.94	773.75	769.12	771.24	2019740
7	12-Aug-16	768.46	773.15	768.42	772.56	1563318
8	15-Aug-16	771.90	772.04	767.71	768.49	2118521
9	16-Aug-16	768.62	768.73	763.82	764.04	1604561
10	17-Aug-16	764.41	765.22	759.20	764.63	1891116

Figura 2.1. Tabla de datos de ejemplo

2.2. Transformación de los Datos en un Objeto Serie de Tiempo

Una vez que hemos leído la información en R. Es recomendable transformar esta información en un objeto Time-Series. Esto contribuye a un mejor rendimiento del software en los cálculos que se van a realizar y evita errores al utilizar funciones relacionadas con series de tiempo. Para realizar la transformación utilizamos la función

ts(), esta función recibe un vector y lo convierte en un objeto Time-Series. A continuación mostramos la estructura y argumentos de esta función:

```
ts(data = NA, start = 1, end = numeric(), frequency = 1,  
    deltat = 1, ts.eps = getOption("ts.eps"), class = ,  
    names = )
```

Argumentos:

- data.- Un vector o matriz de los valores observados de las series temporales.
- start.- El tiempo de la primera observación.
- end.- El tiempo de la última observación.
- frequency.- El número de observaciones por unidad de tiempo.
- deltat.- La fracción del período de muestreo entre observaciones sucesivas.
- ts.eps.- Tolerancia de comparación de series temporales.
- class.- Clase a ser dada al resultado.
- names.- Un vector de caracteres de nombres para la serie en una serie múltiple.

Ahora vamos a mostrar su uso, mediante el siguiente Código:

```
ejemplo_data_ts <- ts(ejemplo_data$Close)  
View(ejemplo_data_ts)
```

De esta forma hemos tomado la columna que es de nuestro interés y la hemos transformado en un objeto Time-Series.

2.3. Estadística descriptiva de la Serie de Tiempo

Como primer acercamiento a la estructura de la serie de tiempo, se grafican y calculan sus principales estadísticos. Esto permitirá tener una idea del comportamiento de la serie. Para lo cual se debe hacer uso de la función: plot.ts().- esta función nos permite graficar un objeto Time-Series y la función summary().- esta función nos muestra los principales

estadísticos descriptivos de la serie. A continuación mostramos la estructura y argumentos de esta función:

```
plot.ts(x, y = NULL, plot.type = c("multiple", "single"),
        xy.labels, xy.lines, panel = lines axes = TRUE, ...)
```

Argumentos:

- x,y.- Objetos de series temporales, por lo general heredando de la clase "ts".
- plot.type.- Para las series temporales multivariadas.
- xy.labels.- Lógico, indicando si las etiquetas de texto () deberían usarse para un gráfico x-y.
- xy.lines.- Lógico, indicando si se deben dibujar líneas para un gráfico x-y.
- panel.- Una función (x, col, bg, pch, tipo, etc) que da la acción a realizar en cada panel de la pantalla para plot.type = "multiple".
- axes.- Lógico, indicando si deben dibujarse los ejes xy yy.

```
summary(object, ...)
```

Argumentos:

- object.- Un objeto para el que se desea un resumen.

Ahora vamos a mostrar su uso, mediante el siguiente Código:

```
plot.ts(ejemplo_data_ts, main = 'Serie de Tiempo Amazon
Ago2016Ago2017', xlab = 'Tiempo', ylab = 'Precio de Cierre')
summary(ejemplo_data_ts)
```

De esta forma graficamos la serie que vamos a utilizar.



Figura 2.2. Precios Ajustados de Amazon

```
> summary(ejemplo_data_ts)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 719.1  773.0   836.6   856.8   940.2  1053.0
```

2.4. Calculo de Retornos

De la parte teórica sabemos que el retorno de un activo financiero está definido por:

$$r_t = \log(p_t) - \log(p_{t-1})$$

Para realizar este cálculo en R utilizamos la siguiente formula:

```
ret <- diff(log(ejemplo_data_ts))*100
```

Hemos multiplicamos por cien para obtener los rendimientos en porcentaje. Además graficamos la serie de retornos calculados.

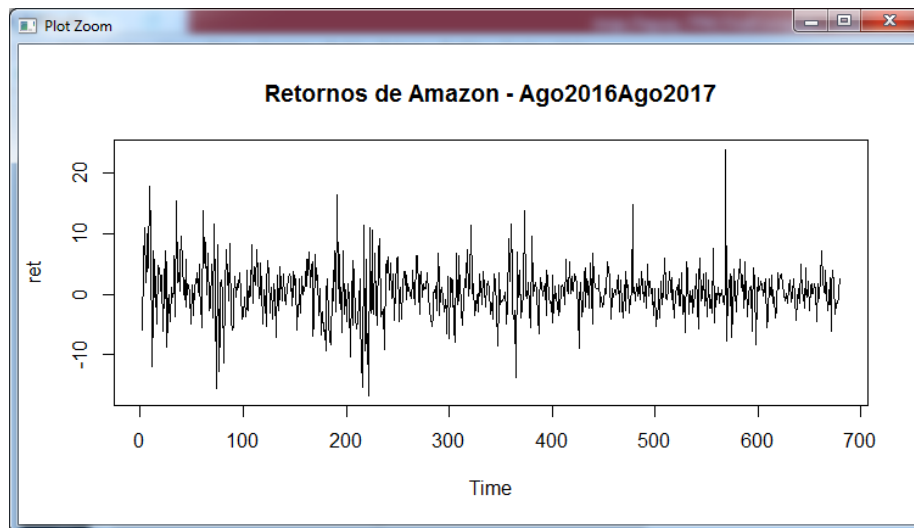


Figura 2.3. Retornos de Amazon

2.5. Pruebas para detectar efectos hetorecedásticos

Para detectar la presencia de efectos hetorecedásticos en la serie de residuos asociada a un Modelo de Media vamos a utilizar las pruebas: Ljung-Box y Multiplicadores de Lagrange.

2.5.1. Prueba Ljung-Box, para independencia

Consiste en probar si existe dependencia lineal entre los errores del modelo de media condicional estimado, ARIMA. Para esto utilizamos la prueba de hipótesis Ljung-Box.

$$H_0: \rho_1 = \rho_2 = \dots = \rho_h = 0$$

$$H_1: \exists \rho_i \neq 0$$

La prueba de hipótesis Ljung-Box está implementada en R con la función: `box.test()`. A continuación mostramos la estructura y argumentos de esta función:

```
Box.test(x, lag = 1, type = c("Box-Pierce", "Ljung-Box"),
fitdf = 0)
```

Argumentos:

- `x`.- Un vector numérico o series temporales univariadas.

- lag.- La estadística se basará en coeficientes de autocorrelación de retardo.
- type.- Prueba a realizar: se utiliza la coincidencia parcial.
- fitdf.- Número de grados de libertad que se restará si x es una serie de residuos.

Ahora vamos a mostrar su uso, mediante el siguiente Código:

```
install.packages('xts')
library('xts')
Box.test(coredata(ret), type = 'Ljung-Box', lag = 12)
```

```
> Box.test(coredata(ret), type = 'Ljung-Box', lag = 12)
      Box-Ljung test
data:  coredata(ret)
X-squared = 15.616, df = 12, p-value = 0.2095
```

Podemos ver que el p-value de la prueba es 0.2095, con este valor no rechazamos H_0 , entonces no existe relación lineal entre los retornos.

2.5.2. Prueba Multiplicador de Lagrange, para efectos ARCH, GARCH

Consiste en probar si es apropiado modelar la varianza de un modelo de media, utilizando un modelo ARCH o GARCH. Para esto debemos utilizar los errores al cuadrado del modelo de media condicional estimado.

Con el fin de ejemplificar su uso, realizamos la prueba con los retornos al cuadrado. Nótese que en esta sección lo importante es entender cuando se rechaza la hipótesis nula de la prueba Multiplicador de Lagrange para efectos ARCH.

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_h = 0$$

$$H_1: \exists \alpha_i \neq 0$$

La prueba de hipótesis LM está implementada en R con la función: Arch.Test(). A continuación los detalles de esta función:

```
ArchTest (x, lags=12, demean = FALSE)
```

Argumentos:

- x.- Vector numérico
- lags.- Número entero positivo, que representa los retrasos.
- demean.- Lógico, elimine la media antes de calcular la estadística de prueba.

Ahora vamos a mostrar su uso, mediante el siguiente Código:

```
install.packages('FinTS')
```

```
library('FinTS')
```

```
ArchTest(ret, lag=1)
```

```
> ArchTest(ret^2, lag=1)
      ARCH LM-test; Null hypothesis: no ARCH effects
data:  ret^2
Chi-squared = 0.0018301, df = 1, p-value = 0.9659
```

Podemos ver que el p-value de la prueba es 0.9659, con este valor no rechazamos H_0 , entonces no existe varianza heterocedástica, cuando analizamos los datos con un retardo ARCH(1).

A continuación realizaremos la misma prueba contemplando cinco retardos, para ver si en este caso existen efectos ARCH.

```
ArchTest(ret, lag=5)
```

```
> ArchTest(ret^2, lag=5)
      ARCH LM-test; Null hypothesis: no ARCH effects
data:  ret^2
Chi-squared = 4.323, df = 5, p-value = 0.5039
```


Podemos ver que el p-value de la prueba es 0.5039, con este valor no rechazamos H_0 , entonces no existe varianza heterocedástica, con cinco retardos. De esta forma confirmamos que la serie no tiene efectos ARCH significativos.

2.6. Función de Auto-correlaciones

Para calcular las auto-correlaciones de una serie de tiempo, utilizaremos la función: `acf()`,

A continuación los detalles de esta función:

```
acf(x, lag.max = NULL,  
    type = c("correlation", "covariance", "partial"),  
    plot = TRUE, na.action = na.fail, demean = TRUE, ...)
```

Argumentos:

- `x`.- Un objeto numérico de serie temporal univariante o multivariado.
- `lag.max`.- Retardo máximo en la que calcular el acf.
- `type`.- Cadena de caracteres que da el tipo de acf a calcular.
- `plot`.- Lógico F/T, dibuja el acf.
- `na.action`.- Función a ser llamada para manejar valores perdidos.
- `demean`.- Lógico. ¿Deben las covarianzas ser sobre los medios de la muestra?

Ahora vamos a mostrar su uso, mediante el siguiente Código:

```
acf(ret)
```

El resultado que obtenemos es el siguiente:

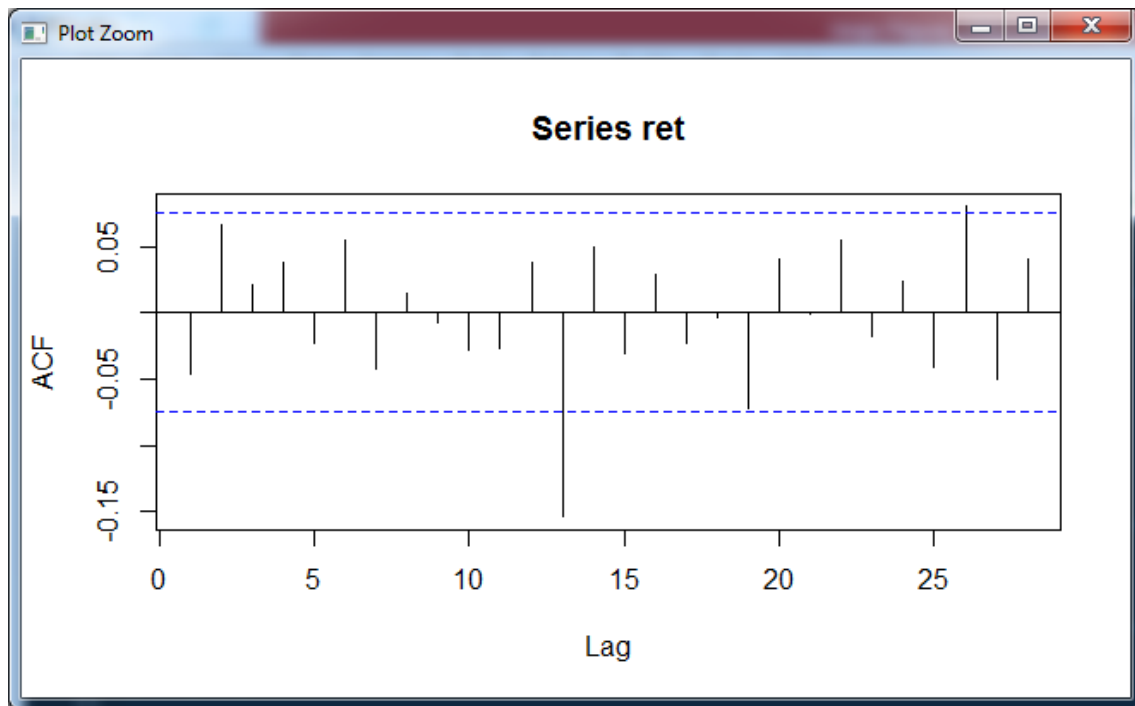


Figura 2.4. Autocorrelaciones Simples de los Retornos de Amazon

En este caso vemos que las autocorrelaciones simples no son significativas a excepción de la autocorrelación 13 misma que podría tratarse de un valor atípico.

Nótese que, se utiliza el mismo comando para el cálculo de auto-correlaciones simples en modelos de media y varianza condicional, el investigador debe elegir la serie según lo necesite.

2.7. Función de Auto-correlaciones Parciales

Para calcular las auto-correlaciones parciales de una serie de tiempo, utilizaremos la función: `pacf()`, A continuación los detalles de esta función:

```
pacf(x, lag.max, plot, na.action, ...)
```

Argumentos:

- `x`.- Un objeto numérico de serie temporal univariante o multivariado.
- `lag.max`.- Retardo máximo en la que calcular el acf.

- `type`.- Cadena de caracteres que da el tipo de acf a calcular.
- `plot`.- Lógico F/T, dibuja el acf.
- `na.action`.- Función a ser llamada para manejar valores perdidos.

Ahora vamos a mostrar su uso, mediante el siguiente Código:

```
pacf (ret)
```

El resultado que obtenemos es el siguiente:

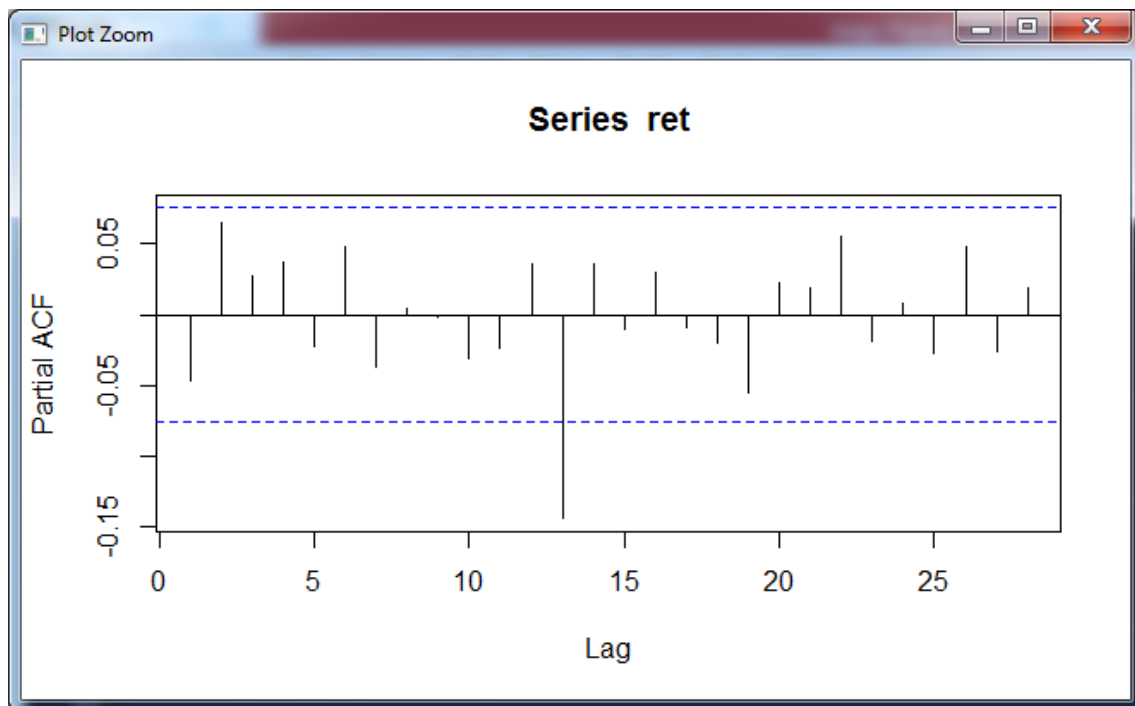


Figura 2.5. Autocorrelaciones Simples de los Retornos de Amazon

Al igual que en las autocorrelaciones simples vemos que la única autocorrelación parcial significativa es la número 13. Nótese que, se utiliza el mismo comando para el cálculo de auto-correlaciones parciales en modelos de media y varianza condicional, el investigador debe elegir la serie según lo necesite.

2.8. Modelos ARIMA(p,d,q)

En esta sección vamos a utilizar los ejemplos de ayuda de R, con el fin de entender el uso de la función `arima()`. Para esto utilizamos la serie `USAccDeaths`, esta serie guarda la información mensual de muertes accidentales en EEUU de 1860 observaciones, iniciando en Enero de 1970.

En el siguiente capítulo realizaremos una explicación más detenida del método para estimar series de tiempo con media condicional y utilizaremos la función que estamos revisando en esta sección. A continuación mostramos la estructura y argumentos de esta función:

```
arima(x, order = c(0L, 0L, 0L),  
      seasonal = list(order = c(0L, 0L, 0L), period = NA),  
      xreg = NULL, include.mean = TRUE,  
      transform.pars = TRUE,  
      fixed = NULL, init = NULL,  
      method = c("CSS-ML", "ML", "CSS"),  
      SSinit = c("Gardner1980", "Rossignol2011"),  
      optim.method = "BFGS",  
      optim.control = list(), kappa = 1e6)
```

Argumentos:

- `x`.- Una serie temporal univariada.
- `order`.- Una especificación de la parte no estacional del modelo ARIMA: los tres componentes enteros (`p`, `d`, `q`) son el orden AR, el grado de diferenciación y el orden MA

- `seasonal`.- Una especificación de la parte estacional del modelo ARIMA, más el período (que por defecto a la frecuencia (x)). Esto debe ser una lista con el orden de los componentes y el período.
- `xreq`.- Opcionalmente, un vector o matriz de regresores externos, que debe tener el mismo número de filas que x.
- `include.mean`.- ¿Debería el modelo ARMA incluir un término medio / intercepto? T/F.
- `transform.pars`.- Los parámetros AR se transforman para asegurar que permanezcan en la región de estacionariedad, T/F
- `fixed`.- Vector numérico opcional de la misma longitud que el número total de parámetros.
- `init`.- Vector numérico opcional de los valores de los parámetros iniciales.
- `method`.- Método de ajuste: máxima verosimilitud o minimizar la suma de cuadrados condicionales.
- `SSinit`.- Una cadena que especifica el algoritmo para calcular la inicialización del estado-espacio de la verosimilitud.
- `optim.method`.- El valor pasado como el argumento del método a `optim`.
- `optim.control`.- Lista de parámetros de control para `optim`.
- `kappa`.- La varianza previa (como un múltiplo de la varianza de las innovaciones) para las observaciones pasadas en un modelo diferenciado. No reduzca esto.

Ahora vamos a mostrar su uso, mediante el siguiente Código:

```
arima(USAccDeaths, order = c(2,1,3))
```

```
> arima(USAccDeaths, order = c(2,1,3))
Call:
arima(x = USAccDeaths, order = c(2, 1, 3))
Coefficients:
```

	ar1	ar2	ma1	ma2	ma3
	1.7290	-0.9994	-2.4603	2.2397	-0.7080
s.e.	0.0084	0.0028	0.1720	0.3432	0.2092

sigma^2 estimated as 302702: log likelihood = -553.18, aic = 1116.36

De esta forma hemos obtenido la estimación de los parámetros para un modelo ARIMA(2,1,3), mediante el uso de la función `arima()`.

2.9. Modelos GARCH(r,s)

En esta sección vamos a utilizar los ejemplos de ayuda de R, con el fin de entender el uso de la función `grach()`. Para esto utilizamos la serie `EuStockMarkets`, esta serie guarda el precio de cierre diario de los principales índices bursátiles europeos: Germany DAX (Ibis), Switzerland SMI, France CAC, y UK FTSE, la serie guarda información desde 1991 hasta 1998, 1860 observaciones.

En el capítulo anterior hemos detallado el proceso de estimación de estos modelos. A continuación mostramos la estructura y argumentos de esta función:

```
garch(x, order = c(1, 1), series = NULL, control =
garch.control(...), ...)
```

Argumentos:

- `x`.- Un vector numérico o una serie de tiempo.
- `order`.- Un vector entero de dos dimensiones que da los órdenes del modelo para que se ajuste
- `seies`.- Nombre de la serie.
- `control`.- Una lista de parámetros de control configurada por `garch.control`.
- `garch.control`.- Otros configuraciones de la estimación del modelo.
- `trace`.- Trace optimizador de salida? T/F.

Ahora vamos a mostrar su uso, mediante el siguiente Código:

```
install.packages('fGarch')
```

```
library('fGarch')

data(EuStockMarkets)

dax <- diff(log(EuStockMarkets))[, "DAX"]

dax.garch <- garch(dax) # Fit a GARCH(1,1) to DAX returns
```

```
> summary(dax.garch) # ARCH effects are filtered. However,

Call:
garch(x = dax)

Model:
GARCH(1,1)

Residuals:
    Min       1Q   Median       3Q      Max
-12.18398  -0.47968   0.04949   0.65746   4.48048

Coefficient(s):
      Estimate Std. Error t value Pr(>|t|)
a0 4.639e-06   7.560e-07   6.137 8.42e-10 ***
a1 6.833e-02   1.125e-02   6.073 1.25e-09 ***
b1 8.891e-01   1.652e-02  53.817 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:
    Jarque Bera Test

data:  Residuals
X-squared = 12947, df = 2, p-value < 2.2e-16

    Box-Ljung test

data:  Squared.Residuals
X-squared = 0.13566, df = 1, p-value = 0.7126
```

De esta forma hemos obtenido la estimación de los parámetros para un modelo GARCH(1,1), mediante el uso de la función `garch()`.

2.10. Criterio de elección entre Modelos

Como resultado de la estimación de modelos en media o varianza condicional, podemos encontrarnos con el caso que más existe más de un modelo que puede explicar la serie en análisis.

En este momento debemos elegir uno de los modelos, para hacer esta elección, vamos a utilizar el AIC² como criterio de elección, entonces el modelo con AIC asociado de menor valor, será elegido como modelo final.

Ahora vamos a mostrar su uso, mediante el siguiente Código:

```
library('tseries')
library('fGarch')

# Para Modelos ARMA(p,d,q)
arima001 <- arima(ret,c(0,0,1)) # -->> MA(1)
arima100 <- arima(ret,c(1,0,0)) # -->> AR(1)
arima101 <- arima(ret,c(1,0,1)) # -->> ARMA(1,1)
arima111 <- arima(ret,c(1,1,1)) # -->> ARIMA(1,1,1)

ModeloMedia <-
c('arima001','arima100','arima101','arima111')

ModeloMedia_ValorAIC <-
c(arima001$aic,arima100$aic,arima101$aic,arima111$aic)

ModeloMedia_TablaComparacion <-
data.frame(ModeloMedia,ModeloMedia_ValorAIC)

subset(ModeloMedia_TablaComparacion,ModeloMedia_ValorAIC ==
min(ModeloMedia_TablaComparacion$ModeloMedia_ValorAIC))
```

Obteniendo el siguiente modelo como final, en la parte de media condicional.

```
> subset(ModeloMedia_TablaComparacion,ModeloMedia_ValorAIC == min(Mode
loMedia_TablaComparacion$ModeloMedia_ValorAIC))
  ModeloMedia ModeloMedia_ValorAIC
2    arima100             800.1129
```

² El criterio de información de Akaike (AIC) es una medida de la calidad relativa de un modelo estadístico. $AIC = 2k - 2 \ln(L)$;

Donde: $k = \text{num. parámetros}$; $L = \text{max. valor de la función de verosimilitud.}$


```

# Para Modelo GARCH(r,s)

garch01 <- garch(arima100$residuals,order=c(0,1),trace=F)
# -->> ARCH(1)

garch11 <- garch(arima100$residuals,order=c(1,1),trace=F)
# -->> GARCH(1,1)

ModeloVarianza <- c('garch01','garch11')

ModeloVarianza_ValorAIC <- c(AIC(garch01),AIC(garch11))

ModeloVarianza_TablaComparacion <-
data.frame(ModeloVarianza,ModeloVarianza_ValorAIC)

subset(ModeloVarianza_TablaComparacion,ModeloVarianza_Valor
AIC ==
min(ModeloVarianza_TablaComparacion$ModeloVarianza_ValorAIC
))

```

Obteniendo el siguiente modelo como final, en la parte de media condicional.

```

> subset(ModeloVarianza_TablaComparacion,ModeloVarianza_ValorAIC == mi
n(ModeloVarianza_TablaComparacion$ModeloVarianza_ValorAIC))
  ModeloVarianza ModeloVarianza_ValorAIC
2      garch11      782.8495

```

3. Estimación de un modelo ARIMA-GARCH, para una serie de rendimientos financieros.

Para realizar la estimación de un modelo ARIMA-GARCH, hemos seleccionado la serie de precios ajustados semanales de Google, comprendida desde el 23/08/2004 hasta el 28/08/2017. Esta información fue consultada en Yahoo! Finance y guarda en GitHub para que el lector pueda replicar los cálculos, a continuación la dirección:

https://raw.githubusercontent.com/jorgepaguay86/EstadisticaAplicada_TFM_UGR/master/SerieDeTiempo_Google_Ago2004Ago2017Semanal.csv

Como inicio de esta estimación vamos a familiarizarnos con la serie de datos, mediante un análisis gráfico-descriptivo de la misma.

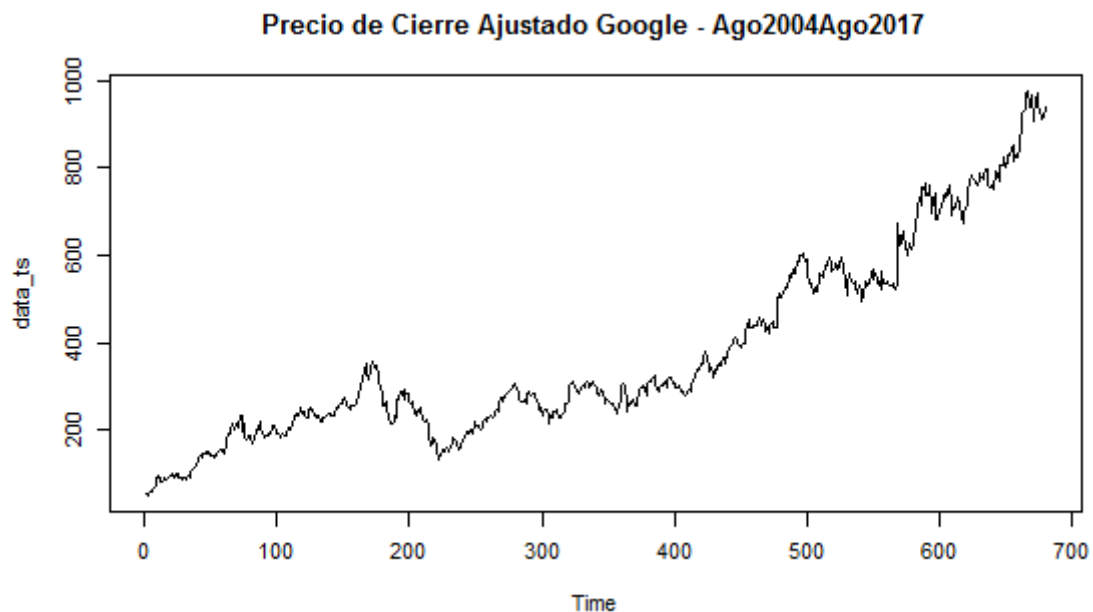


Figura 3.1: Gráfico, Serie de Precios de Cierre Ajustados Google Ago2004 Ago2017

Estadísticos descriptivos:

```
> summary(data_ts)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 49.82  228.50  294.20  376.30  534.20  975.60
```

En la serie de precios ajustados podemos ver una tendencia creciente con picos moderados, en la parte de media. Y en la parte de varianza podemos ver tramos de mucha dispersión y otros de menor dispersión, lo que nos da los primeros indicios de estar tratando con una serie de tiempo heterocedástica.

Sin embargo nuestro interés radica en estudiar la serie de retornos, a continuación realizaremos su cálculo como hemos visto en capítulos anteriores.

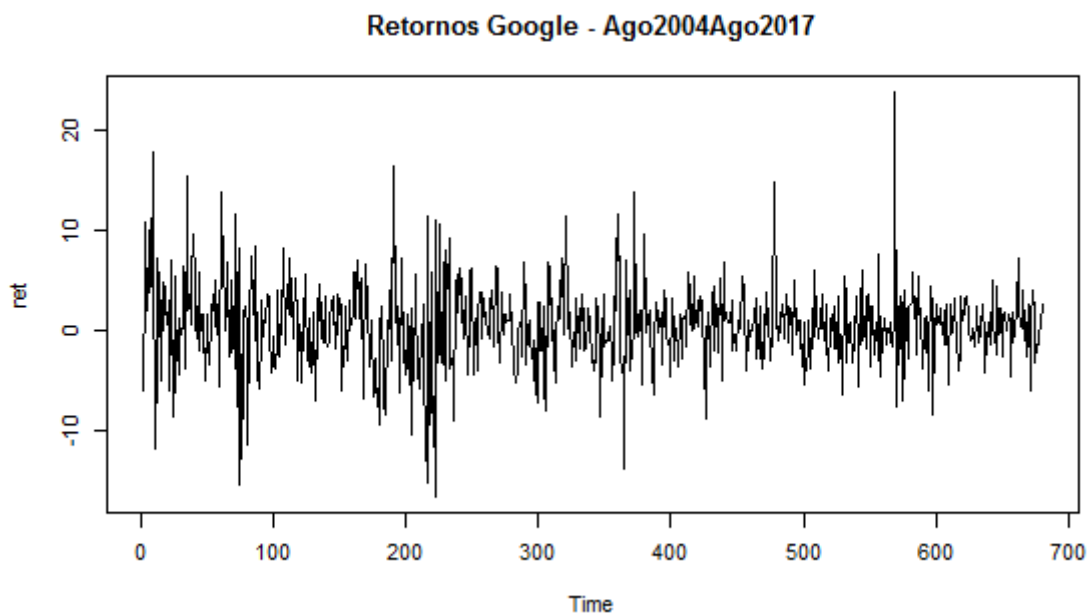


Figura 3.2: Gráfico, Serie de Retornos Google Ago2004 Ago2017

Estadísticos descriptivos:

```
> summary(ret)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-16.6700 -1.9320  0.5336  0.4237  2.6500  23.8500
```

Podemos ver que la serie de Retornos no tiene tendencia, pero si cuenta con periodos de fluctuación diferente. Antes de la observación 400 se ve mayor dispersión, que después de esta observación, esto sugiere heterocedasticidad en la serie de rendimientos.

Además, vemos que esta serie tiene un retorno promedio de 0.42%, y que a alcanzado un valor máximo de 23.85% y mínimo de -16.67%

Para realizar la estimación del ARIMA-GARCH, se seguirá los siguientes pasos:

1. Verificar que la serie es estacionaria mediante la Prueba de Raíces Unitarias.
2. Estimar un modelo de media ARIMA para la serie de retorno, con el fin de eliminar cualquier dependencia lineal en la serie.
3. Utilizar los residuos estandarizados y los residuos estandarizados al cuadrado del modelo de media para probar los efectos de GARCH.
4. Estimar un modelo de volatilidad si los efectos GARCH son estadísticamente significativos.
5. Comprobar el modelo ajustado cuidadosamente y refinarlo si es necesario.

3.1. Análisis de la serie de rendimientos, prueba de raíces unitarias.

Para la estimación de media condicional vamos a seguir la metodología de Box – Jenkins, esta metodología es válida en series estacionarias. Por lo cual necesitamos garantizar que la serie de rendimientos es estacionaria.

En el gráfico de la serie, hemos visto que no existe tendencia, esto nos sugiere que la serie de retornos es estacionaria. Para confirmar esta sospecha realizaremos la prueba de raíces unitarias sobre la serie de retornos.

```
> adf.test(ret) # p-value < 0.05 => no unit-root  
  
Augmented Dickey-Fuller Test  
  
data: ret  
Dickey-Fuller = -9.1746, Lag order = 8, p-value = 0.01  
alternative hypothesis: stationary
```

Como resultado de la prueba obtenemos que el p- value es 0.01, lo cual indica rechazo de H_0 en favor de H_1 . Es decir, la serie de Retornos es estacionaria, permitiéndonos continuar con el análisis.

3.2. Estimación del modelo ARIMA

En esta sección vamos a revisar brevemente la Metodología Box – Jenkins, la idea principal de esta metodología es seguir un conjunto de pasos para encontrar el modelo ARMA, ARIMA o ARIMA estacional que ajuste a la serie de datos analizada. A continuación los pasos que componen esta metodología:

1. Identificación.- en este paso se analiza la serie de retornos y sus correlogramas, para determinar los valores más apropiados para los parámetros p, d, q . De esta etapa pueden resultar varios modelos candidatos que ajustan a la serie, en la parte de validación se elegirá al mejor de ellos, mediante el criterio de información AIC.

En la sección anterior, hemos visto la serie de retornos no presenta tendencia, es estacional y tiene varianza diferente por periodos. De aquí que el valor d es igual a cero y para determinar el valor de los parámetros p y q , utilizaremos los correlogramas de la serie.

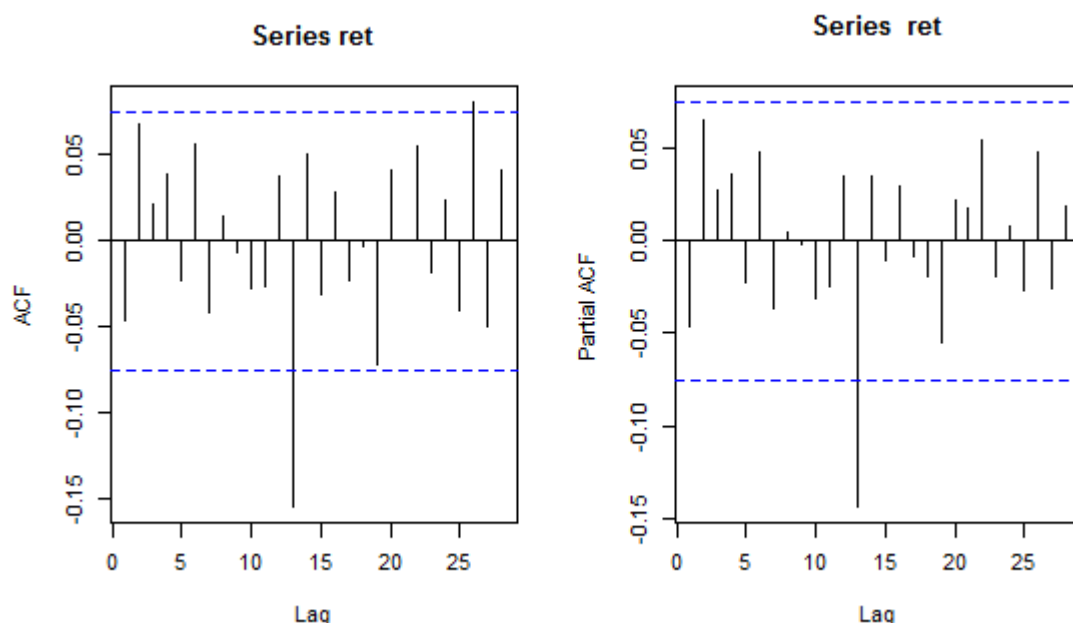


Figura 3.3: Gráfico, Autocorrelaciones Simples y Parciales, Serie de Retornos Google

En las ACF podemos ver que el retardo 2 está muy cercano a la banda de significancia por lo cual podemos asumirlo como válido. Por otro lado, debemos tomar en cuenta que el correlograma está construido al nivel de confianza 5%, esto implica que una de cada veinte autocorrelaciones puede ser significativa fruto del azar, lo cual parece ocurrir con la autocorrelación 13. Además vemos que las autocorrelaciones simples tiene un comportamiento decreciente hasta el retardo 10 y en adelante empiezan a crecer, como formado un cono que se cierra en 10, para la primera mención realizada y otro que se abre desde aquí, para la segunda mención realizada. Concluimos que el valor de q puede ser 2, ya que la autocorrelación 13 parece ser significativa fruto del azar y al ser lejana no influye en explicación del modelo.

En las PACF podemos ver algo similar, en cuanto a la significancia de las autocorrelaciones parciales. Las primeras 2 autocorrelaciones parciales son las más altas de la serie y la autocorrelación 13 parece ser significativa fruto del azar. Por tanto, estableceremos el valor de 2 para p .

2. Estimación.- en este paso determinaremos los posibles modelos que pueden estimar la serie de retornos, verificaremos la significancia de sus coeficientes y elegiremos el mejor de ellos mediante el criterio de información AIC.

Del análisis realizado, obtenemos que la media condicional de la serie de retornos de Google, puede explicarse mediante uno o varios de los siguientes modelos ARIMA(1,0,1), ARIMA(1,0,2), ARIMA(2,0,1) y ARIMA(2,0,2).

A continuación mostraremos los resultados de la estimación para cada uno de los modelos propuestos:

Modelo ARIMA(1,0,1)

```
> coeftest(ModeloArima101)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1    -0.969230   0.023723 -40.856 < 2e-16 ***
ma1     0.940202   0.031987  29.393 < 2e-16 ***
intercept 0.425248   0.160777   2.645 0.00817 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo ARIMA(1,0,1), vemos que todos los coeficientes son significativos, al nivel de confianza 5%. Sin embargo, el coeficiente ar1 es muy cercano a 1, lo que lleva a pensar que la serie no es estacionaria y necesita una diferencia para serlo. Entonces un nuevo candidato es el modelo ARIMA(1,1,1).

```
> coeftest(ModeloArima111)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  -0.0525507  0.0393077   -1.3369   0.1813
ma1  -0.9897745  0.0096084 -103.0117 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo ARIMA(1,1,1), vemos que solo el coeficiente del termino r_{t-1} no es significativo, al nivel de confianza 5%. Por tanto descartamos este modelo, sospechando que un mejor estimador es el modelo ARIMA(0,1,1).

```
> coeftest(ModeloArima011)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1 -0.992703   0.010351 -95.903 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo ARIMA(0,1,1), vemos que el coeficiente ma1 es significativo, Por tanto el modelo ARIMA(0,1,1) es otro candidato para ajustar la serie de retornos Google.

Modelo ARIMA(1,0,2)

```
> coeftest(ModeloArima102)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1    0.521451   0.363071  1.4362  0.15094
ma1   -0.568315   0.363329 -1.5642  0.11777
ma2    0.088951   0.038205  2.3283  0.01990 *
intercept 0.425533   0.178012  2.3905  0.01683 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo ARIMA(1,0,2), vemos que solo el coeficiente del termino ϵ_{t-2} es significativo, al nivel de confianza 5%. Por lo cual, podemos pensar que el modelo ARIMA(0,0,2) es un mejor candidato.

```
> coeftest(ModeloArima002)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1   -0.047452   0.038532 -1.2315  0.21814
ma2    0.066267   0.037488  1.7677  0.07711 .
intercept 0.423923   0.166870  2.5404  0.01107 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


De la estimación realizada vemos que solo el término independiente es significativo, al nivel de confianza 5%.

Por lo cual descartamos ambos modelos.

Modelo ARIMA(2,0,1)

```
> coeftest(ModeloArima201)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1      0.446553   0.328332   1.3601  0.17381
ar2      0.089359   0.038462   2.3233  0.02016 *
ma1     -0.493409   0.329037  -1.4996  0.13373
intercept 0.425093   0.178597   2.3802  0.01730 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo ARIMA(2,0,1), vemos que el coeficiente del termino r_{t-2} y el termino independiente son significativos, al nivel de confianza 5%. Por lo cual, podemos pensar que el modelo ARIMA(2,0,0) es un mejor candidato.

```
> coeftest(ModeloArima200)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1     -0.043685   0.038329  -1.1397  0.25439
ar2      0.065573   0.038335   1.7105  0.08717 .
intercept 0.424087   0.167455   2.5325  0.01132 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

De la estimación realizada vemos que solo el término independiente es significativo, al nivel de confianza 5%.

Por lo cual descartamos ambos modelos.

Modelo ARIMA(2,0,2)

```
> coeftest(ModeloArima202)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	0.029339	0.023563	1.2451	0.213091
ar2	0.968604	0.023446	41.3115	< 2.2e-16 ***
ma1	-0.059522	0.032233	-1.8466	0.064803 .
ma2	-0.940475	0.032226	-29.1833	< 2.2e-16 ***
intercept	0.409382	0.135359	3.0244	0.002491 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

En el modelo ARIMA(2,0,2), vemos que los coeficientes de los términos r_{t-1} y ϵ_{t-1} no son significativos, al nivel de confianza 5%. Por lo cual descartamos el modelo.

Como resultado del proceso de estimación vemos que los modelos que tiene todos sus coeficientes significativos, son: el modelo ARIMA(1,0,1) y el modelo ARIMA(0,1,1).

```
> head(TablaDiagnosticoArima)
  Modelo AllCoef<0.05? AIC
4 ModeloArima101      TRUE 3900.619
9 ModeloArima011      TRUE 3907.670
10 ModeloArima112     FALSE 3902.051
7 ModeloArima202     FALSE 3904.310
2 ModeloArima002     FALSE 3905.727
1 ModeloArima200     FALSE 3905.757
```

Elegiremos el modelo de menor AIC para continuar con nuestro análisis, en la tabla de resumen, vemos que el modelo de menor AIC es el modelo ARMA(1,0,1). Por tal motivo elegiremos este modelo para estimar la media condicional de la serie de retornos de Google.

$$r_t = -0.9692 * r_{t-1} + 0.4252 + 0.9402 * \epsilon_{t-1}; AIC = 3900.619$$

- Validación.- en esta etapa estudiaremos los residuos del modelo elegido para asegurar este cumpla con el supuesto de ruido blanco, mediante el empleo de la serie de residuos, curva de normalidad, autocorrelogramas y pruebas de hipótesis.

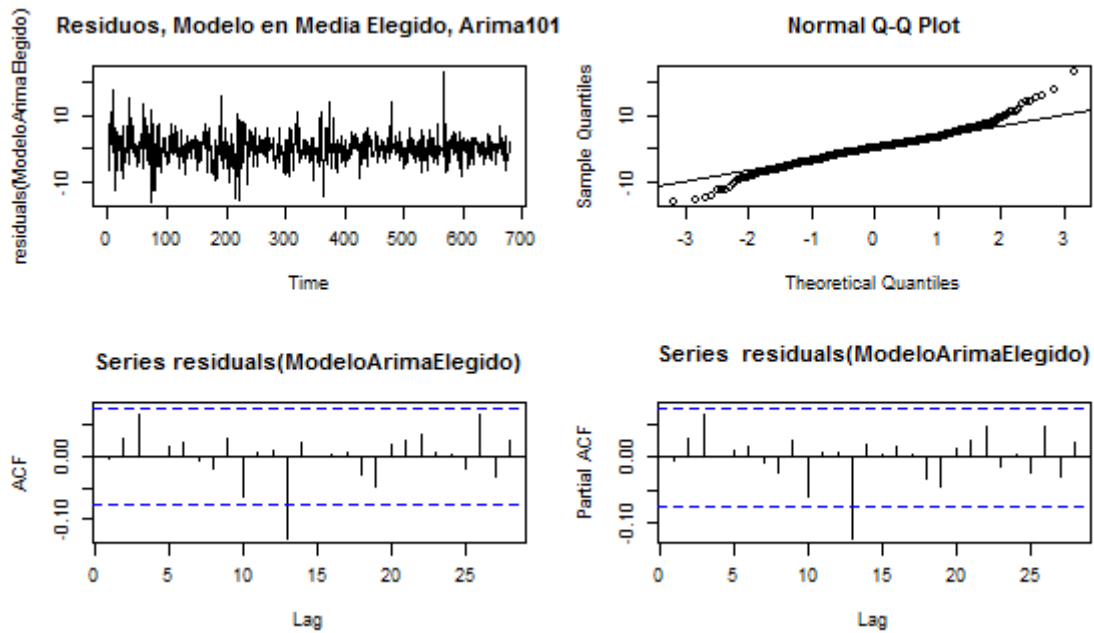


Figura 3.4: Gráfico, Validación de Ajuste Modelo Elegido ARIMA(1,0,1)

En el qqnorm vemos que los valores se aproximan a la línea de normalidad, pero las colas están un poco distantes lo cual nos hace pensar en heterocedasticidad condicional. Y en los correlogramas podemos ver que las autocorrelaciones simples y parciales son nos significativas, entonces podemos concluir que el ModeloArima101 proporciona un buen ajuste de la Serie de Retornos de Google, en el periodo analizado.

3.3. Análisis de la Serie de Residuos del Modelo ARIMA, efectos GARCH

Para determinar si existen efectos GARCH en la serie de residuos del modelo ARIMA, realizamos las pruebas de hipótesis: Ljung-Box y Multiplicador de Lagrange para efectos GARCH. El resultado de estas pruebas, nos indicará si es necesario realizar la estimación de un modelo GARCH.

```
> Box.test(residuals(ModeloArimaElegido)^2,type = 'Ljung-Box',lag = 12)

Box-Ljung test

data: residuals(ModeloArimaElegido)^2
X-squared = 72.792, df = 12, p-value = 9.585e-11
```

```

> library('FinTS')
> ArchTest(residuals(ModeloArimaElegido),lag = 5)

      ARCH LM-test; Null hypothesis: no ARCH effects
data:  residuals(ModeloArimaElegido)
Chi-squared = 37.547, df = 5, p-value = 4.653e-07

```

Como resultado de la prueba de hipótesis Box Ljung, obtenemos un p-value igual a $9.585e-11$. Con este valor rechazamos la hipótesis nula, en favor de la alternativa, es decir los errores al cuadrado están correlacionados entre sí.

Por tanto es necesario realizar una estimación GARCH para esta serie de tiempo, esto nos permite tener un buen estimador de la serie de retornos Google, ya que el modelo ARIMA(1,0,1) por sí solo no lo es.

En la prueba de hipótesis Multiplicador de Lagrange para efectos GARCH, obtenemos un p-value igual a $4.653e-07$. Con este valor rechazamos la hipótesis nula, en favor de la alternativa, es decir los errores tienen efectos GARCH significativos, que necesitan ser estimados mediante un Modelo GARCH.

Esta prueba, confirma lo visto en la prueba anterior. Y confirma la necesidad de estimar un modelo GARCH para serie de retornos Google, ya que el modelo ARIMA(1,0,1) no cumple con la condición requerida de homocedasticidad de los errores, para ser un buen estimador de media condicional.

3.4. Estimación del modelo GARCH

Para estimar un modelo de varianza condicional GARCH, debemos identificar su orden, estimar los coeficientes del modelo y realizar las pruebas de validación pertinentes. Sin embargo, es importante mencionar que en la mayor parte de casos basta un modelo GARCH(1,1) para ajustar el comportamiento de la varianza condicional de una serie de retornos financieros.

1. Identificación.- en este paso vamos a determinar los valores r y s , del modelo, mediante el análisis de los correlogramos de los residuos y de los residuos al cuadrado.

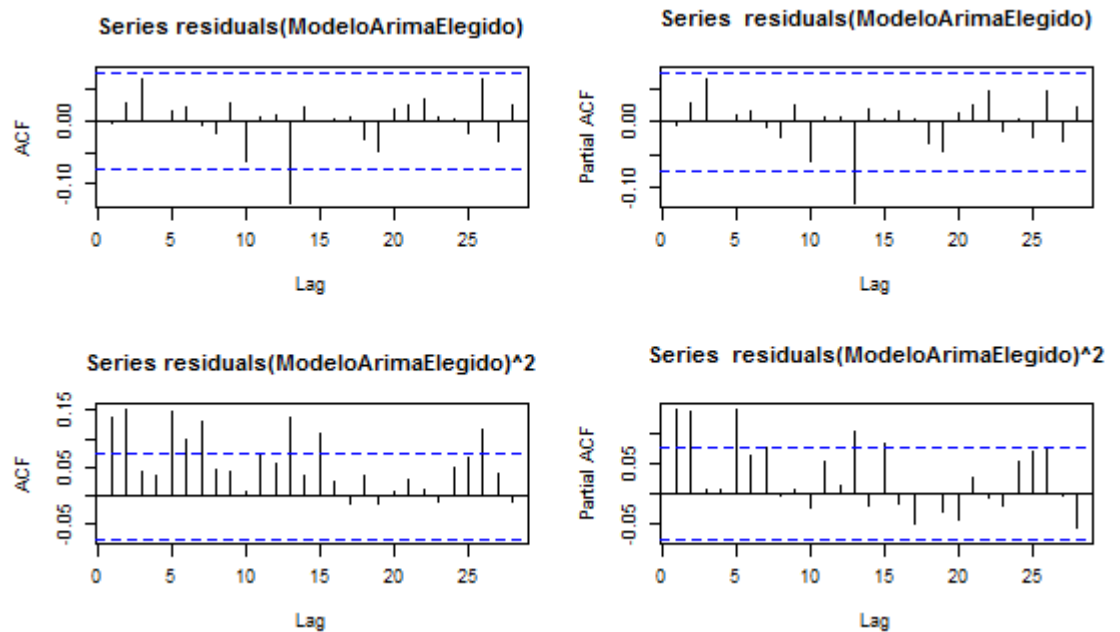


Figura 3.5: Gráfico, Autocorrelaciones Simples y Parciales de la Serie de Residuos.

En los correlogramas de los residuos vemos que las autocorrelaciones simples y parciales no pasan la banda de significancia, por tanto no son significativas. Mientras que los correlogramas de los errores al cuadrado tienen varios valores significativos, tanto en las autocorrelaciones simples como en las parciales.

En las autocorrelaciones simples podemos observar rachas de autocorrelaciones significativas, que van decreciendo conjuntamente de forma lenta. Y en las autocorrelaciones parciales vemos que existe clara significancia se las autocorrelaciones de orden 1, 2 y 5.

Esto nos indica que estamos tratando con un proceso GARCH, ya que tenemos autotcorrelaciones significativas de orden dos en los ACF y PACF, antes de ver una caída de las mismas en el siguiente retardo.

2. Estimación.- en este paso determinaremos los posibles modelos que pueden estimar la varianza condicional de la serie de retornos, verificaremos la significancia de sus coeficientes y elegiremos el mejor de ellos mediante el criterio de información AIC.

En el análisis anterior hemos visto que nos encontramos en presencia de un modelo GARCH, probaremos con el orden (2,2), esperando sea un buen estimador de varianza condicional. Por otro lado, de la experiencia sabemos que las series de retornos financieros, habitualmente son explicadas con un modelo GARCH(1,1). Teniendo así dos posibles candidatos para estimar la varianza condicional de la serie de retornos.

Modelo GARCH(2,2)

```
> coeftest(ModeloGarch22)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
a0 1.4482e+01 2.8945e+00  5.0033 5.635e-07 ***
a1 9.9873e-02 4.1972e-02  2.3795 0.017335 *
a2 1.2711e-01 4.7490e-02  2.6764 0.007441 **
b1 4.4482e-03 2.5733e-01  0.0173 0.986208
b2 5.7912e-14 2.4713e-01  0.0000 1.000000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo GARCH(2,2), vemos que los coeficientes b_1 y b_2 no son significativos, al nivel de confianza 5%. Por lo cual descartamos este modelo, y sospechamos que un mejor candidato es el modelo GARCH(0,2).

```
> coeftest(ModeloGarch02)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
a0 12.642526  0.600719 21.0456 < 2.2e-16 ***
a1 0.133134  0.044099  3.0190 0.0025361 **
a2 0.172055  0.044912  3.8309 0.0001277 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo GARCH(0,2), vemos que todos los coeficientes son significativos, al nivel de confianza 5%. Por tanto es un candidato.

Modelo GARCH(1,1)

```
> coeftest(ModeloGarch11)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
a0 0.944583    0.378338  2.4967 0.0125367 *
a1 0.066551    0.017924  3.7130 0.0002049 ***
b1 0.877930    0.037007 23.7235 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo GARCH(1,1), vemos que todos los coeficientes son significativos, al nivel de confianza 5%. Por tanto es un candidato.

Como resultado del proceso de estimación vemos que los modelos que tiene todos sus coeficientes significativos, son: el modelo GARCH(1,1) y el modelo GARCH (0,2).

```
> head(TablaDiagnosticoGarch)

  Modelo AllCoef<0.05?    AIC
1 ModeloGarch11      TRUE 3828.586
2 ModeloGarch02      TRUE 3842.366
3 ModeloGarch22     FALSE 3849.380
```

Elegiremos el modelo de menor AIC. En la tabla de resumen, vemos que el modelo de menor AIC es el modelo GARCH(1,1). Por tal motivo elegiremos este modelo para estimar la varianza condicional de la serie de retornos de Google.

$$\sigma_t^2 = 0.9445 + 0.0665 * ret_{t-1}^2 + 0.8779 * \sigma_{t-1}^2; AIC = 3828.58$$

3. Validación.- Para comprobar el ajuste del modelo estimado, utilizaremos la serie de residuos, el gráfico de normalidad asociado y correlogramas.

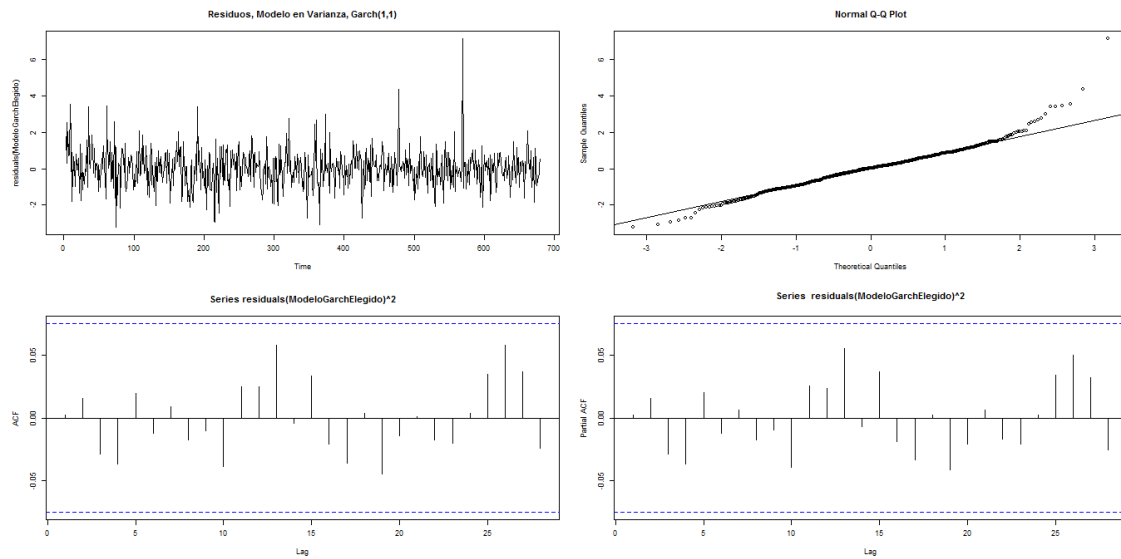


Figura 3.6: Gráfico, Evaluación de los Residuos Estandarizados del ModeloGarch11

En la serie de residuos vemos un comportamiento aleatorio, lo cual habla bien de ellos. En el gráfico de normalidad vemos que los valores se aproximan a la línea de normalidad y existe un valor atípico. En los correlogramas de los errores al cuadrado, podemos ver que las autocorrelaciones simples y parciales son no significativas, entonces podemos concluir que el ModeloGarch11 proporciona un buen ajuste de la Serie de Residuos del Modelo Arima101, estimado anteriormente.

3.5. Presentación Modelo ARIMA-GARCH estimado

Como resultado de la estimación planteada al inicio de este capítulo, obtenemos que el Modelo ARIMA(1,0,1) - GRACH(1,1) es el mejor estimador de la Serie de Retornos de Google en el periodo analizado.

Modelo: ARIMA(1,0,1) - GRACH(1,1)

$$r_t = -0.9692 * r_{t-1} + 0.4252 + 0.9402 * \epsilon_{t-1}$$

Donde:

r_t : Es el rendimiento de una acción de Google en el tiempo t .

$$\epsilon_t = \sqrt{\sigma_t} \varepsilon_t$$

$$\sigma_t^2 = 0.9445 + 0.0665 * ret_{t-1}^2 + 0.8779 * \sigma_{t-1}^2$$

ε_t : Es un Ruido Blanco de variables independientes e idénticamente

Gráficamente tenemos:

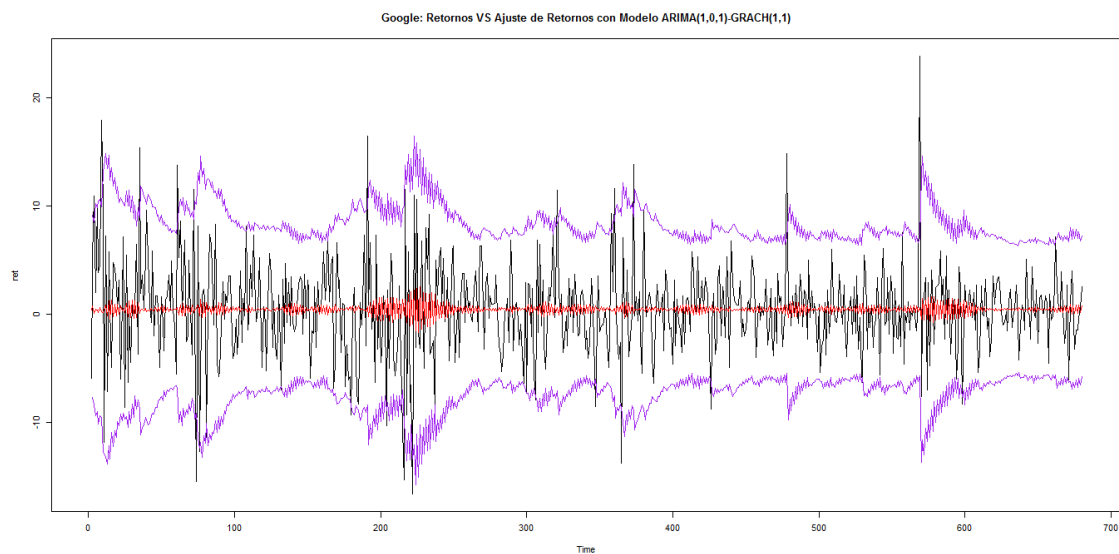


Figura 3.7: Gráfico, Modelo ARIMA(1,01,)-GARCH(1,1) Ajustado

Bibliografía.

1. Cryer. J., Chan. K. (2012). Time series analysis with applications in R, Springer.
2. Gujarati, D., Porter, D.. (2010). Econometría, Mc Graw Hill.
3. Peña, D. (2005). Análisis de series temporales. Alianza Editorial.
4. Tsay, R. S. (2002). Analisis of Financial Time Series. New York: John Wiley & Sons Inc.
5. Eric Zivot (2004). Working with Financial Time Series Data in R, Department of Economics, University of Washington.
6. ENGLE, R.F. (1982): Autorregresive Conditional Heterocedasticity with Estimates of the Variance of the U.K. Inflation *Econométrica*, 50. Pgs: 987-1008.
7. Bollerslev, T. (1986): Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*.
8. Cryer y Chan, J.(2008). Time Series Analysis With Applications in R. New York: Springer Science + Business Media LLC.
9. Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: New Approach” *Econometria*.
10. Nelson, D.B. (1990). ARCH models as diffusion approximations. *J. Econometrics*. 45, 7-39.

Anexo 1: Código R

```
# 3. Estimación de series de tiempo heterocedásticas-----

{

#install.packages('readr')

library('readr')

data <-
read.csv('https://raw.githubusercontent.com/jorgepaguay86/EstadisticaAplicada_TFM_UGR/master/SerieDeTiempo_Google_Ago2004Ago2017Semanal.csv')

data$Date <- as.Date(data$Date)

#data <- subset(data,data$Date <= '2016-01-01')

data_ts <- ts(data$AdjClose) #ojo

ret = diff(log(data_ts))*100

win.graph(width=5.875, height=3.5,pointsize=8)

plot(data_ts,main = 'Precio de Cierre Ajustado Google - Ago2004Ago2017')

win.graph(width=5.875, height=3.5,pointsize=8)

plot(ret,main = 'Retornos Google - Ago2004Ago2017')

summary(data_ts)

summary(ret)

}

# 3.1. Análisis de la serie de rendimientos, prueba de raíces unit -----

install.packages('tseries')

library('tseries')

adf.test(ret) # p-value < 0.05 => no unit-root

# 3.2. Estimación del modelo ARIMA -----
```

```

#1 Correlogramas
{
win.graph(width=5.875, height=3.5,pointsize=8)
par(mfrow=c(1,2))
acf(ret)
pacf(ret)
}

#2 Modelos Plausibles
{
ModeloArima200 <- arima(ret, order = c(2,0,0))
ModeloArima002 <- arima(ret, order = c(0,0,2))
ModeloArima001 <- arima(ret, order = c(0,0,1))
ModeloArima101 <- arima(ret, order = c(1,0,1))
ModeloArima102 <- arima(ret, order = c(1,0,2))
ModeloArima201 <- arima(ret, order = c(2,0,1))
ModeloArima202 <- arima(ret, order = c(2,0,2))
ModeloArima111 <- arima(ret, order = c(1,1,1))
ModeloArima011 <- arima(ret, order = c(0,1,1))
ModeloArima112 <- arima(ret, order = c(1,1,2))
}

{
library('lmtest')

TablaDiagnosticoArima <-
  rbind.data.frame(
    list('ModeloArima200' ,
all(coeftest(ModeloArima200)[,4] <=
0.05),AIC(ModeloArima200)),

```

```

        list('ModeloArima002' ,
all(coeftest(ModeloArima002)[,4] <=
0.05),AIC(ModeloArima002)),

        list('ModeloArima001' ,
all(coeftest(ModeloArima001)[,4] <=
0.05),AIC(ModeloArima001)),

        list('ModeloArima101' ,
all(coeftest(ModeloArima101)[,4] <=
0.05),AIC(ModeloArima101)),

        list('ModeloArima102' ,
all(coeftest(ModeloArima102)[,4] <=
0.05),AIC(ModeloArima102)),

        list('ModeloArima201' ,
all(coeftest(ModeloArima201)[,4] <=
0.05),AIC(ModeloArima201)),

        list('ModeloArima202' ,
all(coeftest(ModeloArima202)[,4] <=
0.05),AIC(ModeloArima202)),

        list('ModeloArima111' ,
all(coeftest(ModeloArima111)[,4] <=
0.05),AIC(ModeloArima111)),

        list('ModeloArima011' ,
all(coeftest(ModeloArima011)[,4] <=
0.05),AIC(ModeloArima011)),

        list('ModeloArima112' ,
all(coeftest(ModeloArima112)[,4] <=
0.05),AIC(ModeloArima112))

    )

    names(TablaDiagnosticoArima) <-
c('Modelo','AllCoef<0.05?','AIC')

```

```

    TablaDiagnosticoArima <- TablaDiagnosticoArima[order(-
TablaDiagnosticoArima$`AllCoef<0.05`,TablaDiagnosticoArima$
AIC),]

    head(TablaDiagnosticoArima)

    coeftest(ModeloArima101)

    ModeloArimaElegido <- ModeloArima101

    #confint(ModeloArima101, level = 0.9)

    }

#3. Validación

{

win.graph(width=5.875, height=3.5,pointsize=8)

par(mfrow=c(2,2))

plot.ts(residuals(ModeloArimaElegido),main = 'Residuos,
Modelo en Media Elegido, Arima101')

qqnorm(residuals(ModeloArimaElegido))

qqline(residuals(ModeloArimaElegido))

acf(residuals(ModeloArimaElegido),na.action = na.omit)

pacf(residuals(ModeloArimaElegido),na.action = na.omit)

}

# 3.3. Análisis de los efectos GARCH -----

Box.test(residuals(ModeloArimaElegido)^2,type = 'Ljung-
Box',lag = 12)

library('FinTS')

ArchTest(residuals(ModeloArimaElegido),lag = 5)

# 3.4. Estimación del modelo GARCH -----

# 1. Correlogramas

{

```

```

win.graph(width=5.875, height=3.5,pointsize=8)

par(mfrow=c(2,2))

acf(residuals(ModeloArimaElegido),na.action = na.omit)
pacf(residuals(ModeloArimaElegido),na.action = na.omit)
acf(residuals(ModeloArimaElegido)^2,na.action = na.omit)
pacf(residuals(ModeloArimaElegido)^2,na.action = na.omit)
}

# 2. Modelos Plausibles
{
  library('fGarch')
  library('forecast')
  library('TSA')

  ModeloGarch11 <-
garch(na.omit(residuals(ModeloArimaElegido)),order =
c(1,1),na.action=na.omit)

  ModeloGarch02 <-
garch(na.omit(residuals(ModeloArimaElegido)),order =
c(0,2),na.action=na.omit)

  ModeloGarch22 <-
garch(na.omit(residuals(ModeloArimaElegido)),order =
c(2,2),na.action=na.omit)
}

#3. Selección del Mejor Modelo
{
  library('lmtest')

  TablaDiagnosticoGarch <-
  rbind.data.frame(

```

```

        list('ModeloGarch11' ,
all(coeftest(ModeloGarch11)[,4] <=
0.05),AIC(ModeloGarch11)),

        list('ModeloGarch02' ,
all(coeftest(ModeloGarch02)[,4] <=
0.05),AIC(ModeloGarch02)),

        list('ModeloGarch22' ,
all(coeftest(ModeloGarch22)[,4] <=
0.05),AIC(ModeloGarch22))

    )

    names(TablaDiagnosticoGarch) <-
c('Modelo','AllCoef<0.05?','AIC')

    TablaDiagnosticoGarch <- TablaDiagnosticoGarch[order(-
TablaDiagnosticoGarch$`AllCoef<0.05`,TablaDiagnosticoGarch$
AIC),]

    head(TablaDiagnosticoGarch)

    coeftest(ModeloGarch11)

    ModeloGarchElegido <- ModeloGarch11
}

# 4. Comprobación del modelo ajustado
{

win.graph(width=4.875, height=2.5,pointsize=8)

par(mfrow=c(2,2))

plot.ts(residuals(ModeloGarchElegido),main = 'Residuos,
Modelo en Varianza, Garch(1,1)')

qqnorm(residuals(ModeloGarchElegido))

qqline(residuals(ModeloGarchElegido))

acf(residuals(ModeloGarchElegido)^2,na.action = na.omit)

pacf(residuals(ModeloGarchElegido)^2,na.action = na.omit)

```



```

}

# 3.5.    Presentación Modelo ARIMA-GARCH estimado -----

{
  library('forecast')

  fit <- fitted.values(ModeloArimaElegido)
  fitgarch <- fitted.values(ModeloGarchElegido)[,1]
  low <- fit - (1.96 * fitgarch)
  high <- fit + (1.96 * fitgarch)
  win.graph(width=4.875, height=2.5, pointsize=8)
  plot(ret,
        main = 'Google: Retornos VS Ajuste de Retornos con
Modelo ARIMA(1,0,1)-GRACH(1,1)',
        type = 'l')
  lines(low,col = 'purple')
  lines(high,col = 'purple')
  lines(fit,col = 'red')
}

```

Anexo 2: Complementos Teóricos

El presente anexo busca dar definiciones puntuales de conceptos relevantes en el estudio de series de tiempo.

1.1. Series de Tiempo y Procesos Estocásticos

Una serie de tiempo es una secuencia de valores ordenados cronológicamente, que representan un concepto: ventas, ingresos, egresos, etc. Además, sabemos que una secuencia de variables aleatorias es un proceso estocástico y sirve como modelo para observar una serie de tiempo, entonces definiremos una serie de tiempo como un proceso estocástico.

$$\{Y_t: t \in Z\}$$

Por otro lado, sabemos que gran parte de la información de este proceso se puede describir en términos de medias, varianzas y covarianzas, por tanto concentraremos nuestra atención en estos momentos.

1.1.1. Función de Medias

Para una serie de tiempo $\{Y_t\}$, la función de medias está definida por:

$$\mu_t = E(Y_t); t \in Z$$

μ_t es el valor esperado del proceso en el tiempo t , por tanto este puede cambiar en cada momento del tiempo.

1.1.2. Función de Auto-covarianzas

Para una serie de tiempo $\{Y_t\}$, la función de auto-covarianzas está definida por:

$$\gamma_{t,s} = Cov(Y_t, Y_s); t, s \in Z$$

Donde:

$$Cov(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)]$$

Es importante, tener clara la siguiente propiedad:

$$\gamma_{t,t} = Var(Y_t)$$

$$\gamma_{t,s} = \gamma_{s,t}$$

$$|\gamma_{t,s}| \leq \sqrt{\gamma_{t,t}\gamma_{s,s}}$$

1.1.3. Función de Auto-correlaciones

Para una serie de tiempo $\{Y_t\}$, la función de auto-correlaciones está definida por:

$$\rho_{t,s} = Corr(Y_t, Y_s); t \in Z$$

Donde:

$$Corr(Y_t, Y_s) = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t)Var(Y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}$$

Es importante, tener clara la siguiente propiedad:

$$\rho_{t,t} = 1$$

$$\rho_{t,s} = \rho_{s,t}$$

$$|\rho_{t,s}| \leq 1$$

1.1.4. Función de Auto-correlaciones Parciales

Para una serie de tiempo normalmente distribuida $\{Y_t\}$, la función de auto-correlaciones parciales está definida por:

$$\phi_{kk} = Corr(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}); t \in Z; k \geq 0$$

ϕ_{kk} : es la correlación en la distribución bivariada de Y_t y Y_{t-k} condicionada en $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$.

Es importante, tener claro lo siguiente:

$$\phi_{kk} = \rho_0; k = 1$$

1.1.5. Componentes de una serie de tiempo

De acuerdo al modelo de descomposición clásico, la serie de tiempo está formada por las siguientes componentes: tendencia, estacionalidad y aleatoriedad, mismas que pueden unirse de forma aditiva o multiplicativa. A continuación una breve descripción de las mismas:

Tendencia

Podemos decir que la tendencia constituye los aumentos o disminuciones que sufre la media de la serie a lo largo del tiempo. Diremos que:

- La serie tiene tendencia creciente, si los valores aumentan a medida que el tiempo aumenta.
- La serie tiene tendencia decreciente, si los valores disminuyen a medida que el tiempo aumenta.
- La serie no tiene tendencia, si los valores no tienen variaciones significativas a medida que el tiempo aumenta.

Estacionalidad

Podemos decir que la estacionalidad constituye las variaciones que la serie sufre con cierta periodicidad, es común que estas periodicidades sean trimestrales, semestrales o anuales.

Aleatoriedad

Esta componente no describe ninguna característica de la serie, corresponde a los factores externos y su ocurrencia se atribuye al azar.

1.2. Procesos Estocásticos Estacionarios

En el presente trabajo, asumiremos que un proceso estocástico es una familia de variables aleatorias, denotada por:

$$\{Y_t: t \in Z\}$$

1.2.1. Procesos Estocásticos Estrictamente Estacionarios

Un proceso estocástico $\{Y_t: t \in Z\}$, es estricta o fuertemente estacionario si:

$$\text{ley}(Y_{t_1+h}, \dots, Y_{t_k+h}) = \text{ley}(Y_{t_1}, \dots, Y_{t_k}); t, k, h \in Z$$

1.2.2. Procesos Estocásticos Débilmente Estacionarios

Un proceso estocástico $\{Y_t: t \in Z\}$, es débilmente estacionario si:

1. $\{Y_t: t \in Z\}$ es real y de segundo orden.
2. $\mu_t = E(Y_t); t \in Z$ es constante.
3. $Cov(Y_t, Y_s) = Cov(Y_{t+h}, Y_{s+h}); t, k, h \in Z$

Al cual se llamará estacionario en adelante.

1.3. Procesos Lineales

En series de tiempo es habitual ver, que la información presente depende de la información pasada y esta idea permite buscar una combinación lineal que explique la relación existente entre el presente y el pasado. Esta idea se puede representar mediante Procesos Lineales, a continuación su definición:

Un proceso estocástico $\{Y_t: t \in Z\}$ Centrado, de segundo orden. Es un Proceso Lineal, si se puede expresar como:

$$Y_t = \sum_{j=0}^{\infty} \Psi_j u_{t-j}; t \in Z$$

Donde:

$$\Psi_0 = 1$$

$$\sum_{j=0}^{\infty} |\Psi_j|^2 < \infty; \Psi_j \in R$$

$\{u_t: t \in Z\}$ Es ruido blanco de varianza σ^2

1.3.1. Proceso Lineal Inversible

Un proceso lineal $\{Y_t: t \in Z\}$ es inversible si:

$$u_t = \sum_{j=0}^{\infty} \pi'_j Y_{t-j}; t \in Z$$

Donde:

$$\pi'_j = 1$$

$$\sum_{j=0}^{\infty} |\pi'_j|^2 < \infty; \pi'_j \in R$$

1.3.2. Procesos Autoregresivos, AR(p)

Llamaremos Proceso Autoregresivo de orden (p), AR(p). A todo proceso lineal estacionario $\{Y_t: t \in Z\}$, que satisface:

$$Y_t = \sum_{j=1}^p \phi_j Y_{t-j} + u_t; t \in Z$$

Donde:

$$\phi_p \neq 0$$

$\{u_t: t \in Z\}$ es un Ruido Blanco³.

$$u_t \perp u_{t-1}; t \in Z$$

1.3.3. Procesos Medias Móviles, MA(q)

Llamaremos Proceso Medias Móviles de orden (q), MA(q). A todo proceso lineal estacionario $\{Y_t: t \in Z\}$, tal que:

$$Y_t = u_t - \sum_{j=1}^p \theta_j u_{t-j}; t \in Z$$

Donde:

$$\theta_p \neq 0$$

$\{u_t: t \in Z\}$ es un Ruido Blanco.

$$u_t \perp u_{t-1}; t \in Z$$

1.3.4. Procesos Medias Móviles - Autorregresivos, ARMA(p,q)

Llamaremos Procesos Medias Móviles - Autoregresivos de orden (p,q), ARMA(p,q) a todo proceso lineal estacionario $\{Y_t: t \in Z\}$, que satisface:

$$Y_t - \sum_{j=1}^p \phi_j Y_{t-j} = u_t - \sum_{j=1}^q \theta_j u_{t-j}$$

Donde:

$$\phi_p \neq 0$$

³ Proceso estacionario de variables aleatorias idénticamente distribuidas con media cero, varianza sigma cuadrado.

$$\theta_p \neq 0$$

$\{u_t: t \in Z\}$ es un Ruido Blanco.

Φ y Θ los polinomios característicos⁴ no tienen raíces de módulo 1.

1.3.5. Procesos Medias Móviles - Autorregresivos Integrados, ARIMA

(p,d,q)

Llamaremos Procesos Medias Móviles - Autorregresivos Integrados de orden (p,d,q),

ARIMA(p,d,q) a todo proceso lineal no estacionario $\{Y_t: t \in Z\}$, que satisface:

$$\Phi(B)(1 - B)^d Y_t = \Theta(B)u_t; t \in Z$$

Donde:

$\{u_t: t \in Z\}$ es un Ruido Blanco.

Φ y Θ los polinomios característicos tienen sus raíces fuera del círculo unidad.