Analisi delle serie storiche 2010/2011

Analisi della serie storica del titolo Italcementi SpA
Con modelli Arch e Garch

Italcementi
Italcementi Group

Docente:
Lea Petrella

Studenti:
Massimo De Giorgi
Galina Moisei

Introduzione

L'obiettivo del presente lavoro è quello di analizzare una serie storica di natura finanziaria e di identificare il processo teorico ipotizzabile come generatore delle osservazioni.
Oggetto della nostra analisi è una serie storica con frequenza giornaliera dell'andamento dei prezzi del titolo Italcementi SpA quotato sul MTA (Mercato Telematico Azionario) della Borsa Italiana. L'indagine abbraccia il periodo che va dal 1-1-2002 fino al 31-8-2011. I prezzi di chiusura aggiustati (adjusted close) sono stati estratti dal sito http://it.finance.yahoo.com/ e sono stati successivamente copiati e quindi salvati su un file di testo.
Per lo studio del suddetto titolo ci siamo serviti del Software statistico *R*.

Caricamento dei dati

Prima di poter iniziare l'analisi della serie storica occorre in primo luogo caricare i dati da una sorgente esterna (nel nostro caso un file di testo). Per far ciò useremo:

ital=read.table("ITALCEMENTI.txt")

Visto che il file testo è formato da 7 vettori-colonna dobbiamo comunicare a R di considerare solo il vettore-colonna 7 dove abbiamo i dati che ci interessano, cioè i prezzi di chiusura aggiustati.

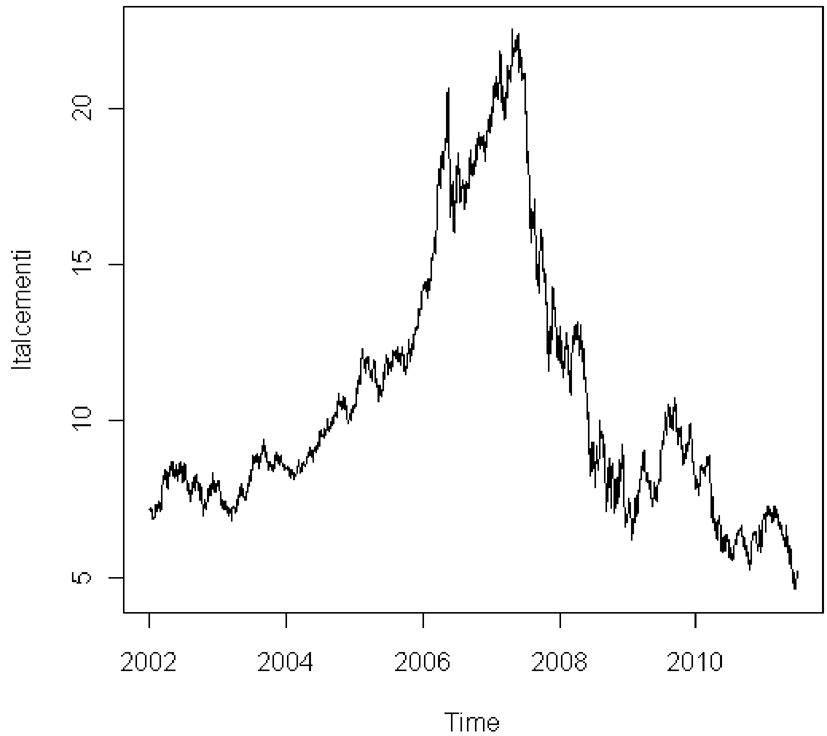
ital=ital[,7]

a questo punto creiamo la nostra serie storica

IC=ts(ital,start=c(2002,01,01),frequency=260)

Come si può notare sono state considerate 260 osservazioni annue, cioè i giorni effettivi di negoziazione escludendo pertanto quelli di chiusura e quelli festivi.
A questo punto ne "plottiamo" l'andamento

plot(IC,ylab="Italcementi")

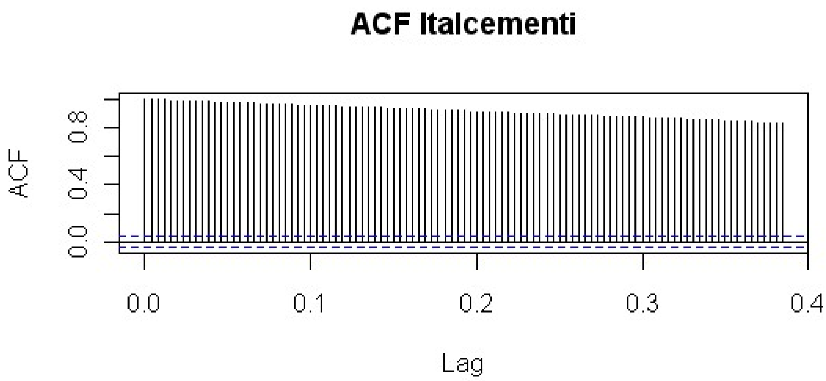


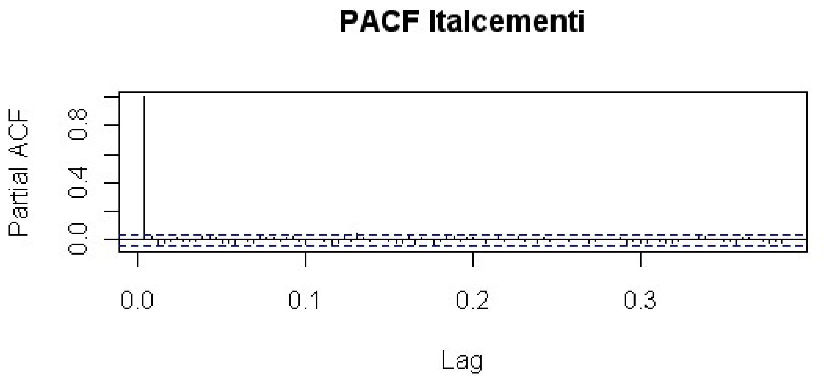
Dalla semplice osservazione grafica, si nota subito che la serie storica è caratterizzata da una *non stazionarietà in media,* evidenziata dal *trend*. Per avere un'ulteriore prova vediamo il correlogramma dell' ACF (funzione di autocorrelazione) e della PACF (autocorrelazione parziale).

par(mfrow=c(2,1))

acf(IC,main="ACF Italcementi",lag.max=100)

pacf(IC,main="PACF Italcementi",lag.max=100)





E' evidente che l'ACF tende a zero molto lentamente.

Un ulteriore verifica può essere fatta empiricamente caricando il pacchetto "Tseries" da "R" e usando lo strumento Augmented Dickey-Fuller Test che serve per individuare la presenza di una radice unitaria:

adf.test(IC)

Augmented Dickey-Fuller Test

data: IC
Dickey-Fuller = -1.0346, Lag order = 13, **p-value = 0.9332**
alternative hypothesis: stationary

Il p-value maggiore di 0,05 ci indica che può essere accettata l'ipotesi nulla H_0, quindi presenta radice unitaria.

Quando si considerano serie finanziarie, più che ai prezzi del titolo si è interessati alle variazioni, perché sono queste a dare informazioni su perdite e guadagni. Generalmente poi, invece di fare riferimento alle variazioni semplici si considerano le variazioni logaritmiche, che in finanza vengono chiamate rendimenti. Le variazioni logaritmiche/rendimenti sono importanti perché rappresentano una approssimazione del tasso di variazione percentuale, che ci da un'idea di come, ad esempio un titolo, sta performando. Inoltre, da un punto di vista statistico fare le differenze prime dei prezzi o i logaritmi dei prezzi, non fa grande differenza, essendo il logaritmo una trasformazione monotona e biunivoca. Quindi i rendimenti corrispondono ad una differenziazione del primo ordine, che ha il compito di eliminare la non stazionarietà in media. Vediamo:

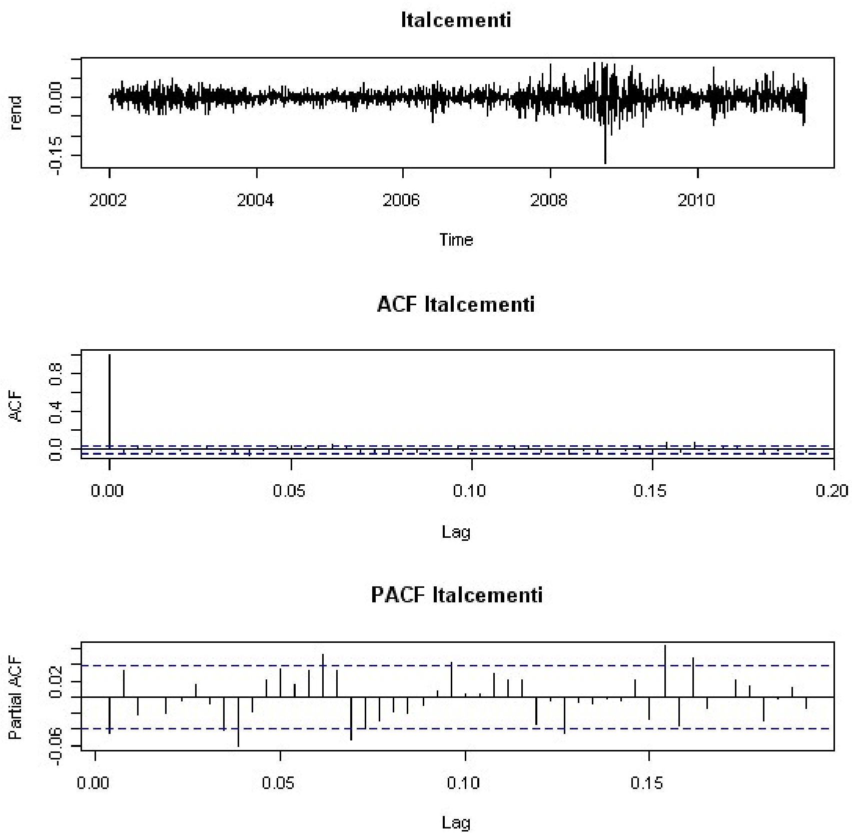
rend=diff(log(IC))

par(mfrow=c(3,1))

plot(rend,main="Italcementi")

acf(rend,main="PACF Italcementi",lag.max=50)

pacf(rend,main="PACF Italcementi",lag.max=50)



Da un'analisi del grafico dei rendimenti si nota che perde il trend, inoltre dal correlogramma dell'ACF i rendimenti risultano incorrelati quindi proprio uno WN (White Noise).

Riproviamo ora a fare il Augmented Dickey-Fuller Test

adf.test(rend)

Augmented Dickey-Fuller Test

data: rend
Dickey-Fuller = -13.6196, Lag order = 13, **p-value = 0.01**
alternative hypothesis: stationary

Warning message:
In adf.test(rend) : **p-value smaller than printed p-value**

Il p-value questa volta è minore di 0,05, questo ci porta a rifiutare l'ipotesi nulla.

Giunti a questo punto si potrebbe pensare che sia impossibile modellare la serie dei rendimenti, vediamo cosa succede se consideriamo la serie dei rendimenti al quadrato. Andiamo a calcolarci il grafico dei rendimenti al quadrato e il relativo correlogramma dell'ACF:

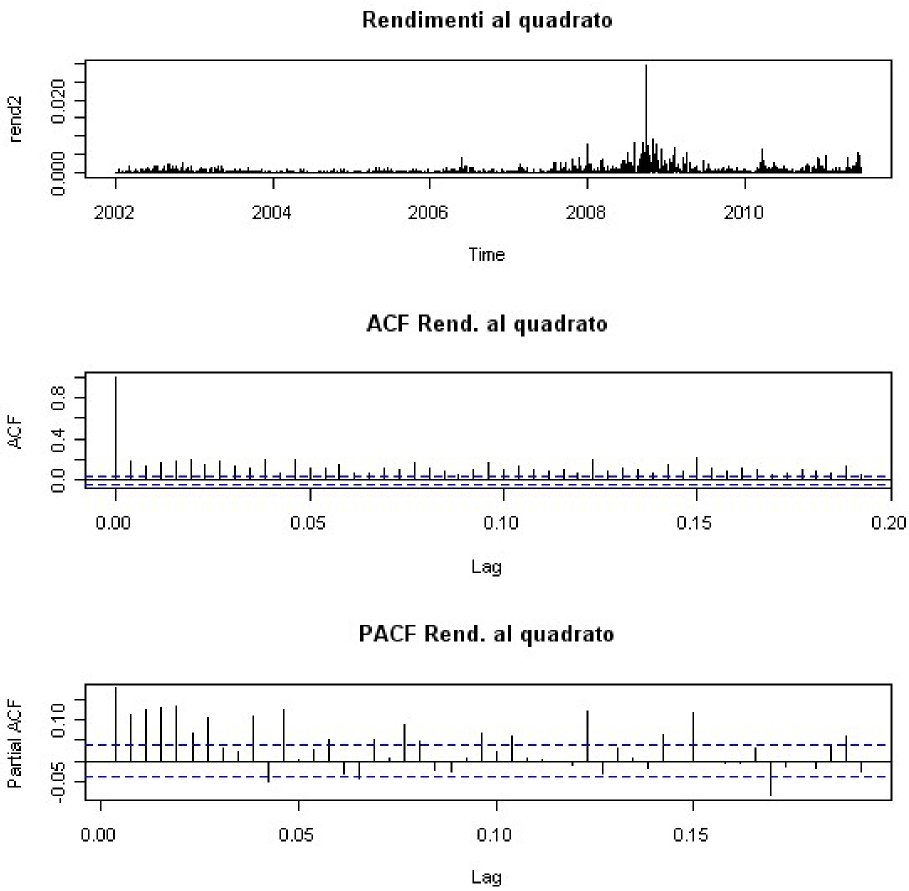
rend2=(rend^2)

par(mfrow=c(3,1))

plot(rend2,main="Rendimenti al quadrato")

acf(rend2,main="ACF Rend. al quadrato",lag.max=50)

pacf(rend2,main="PACF Rend. al quadrato",lag.max=50)



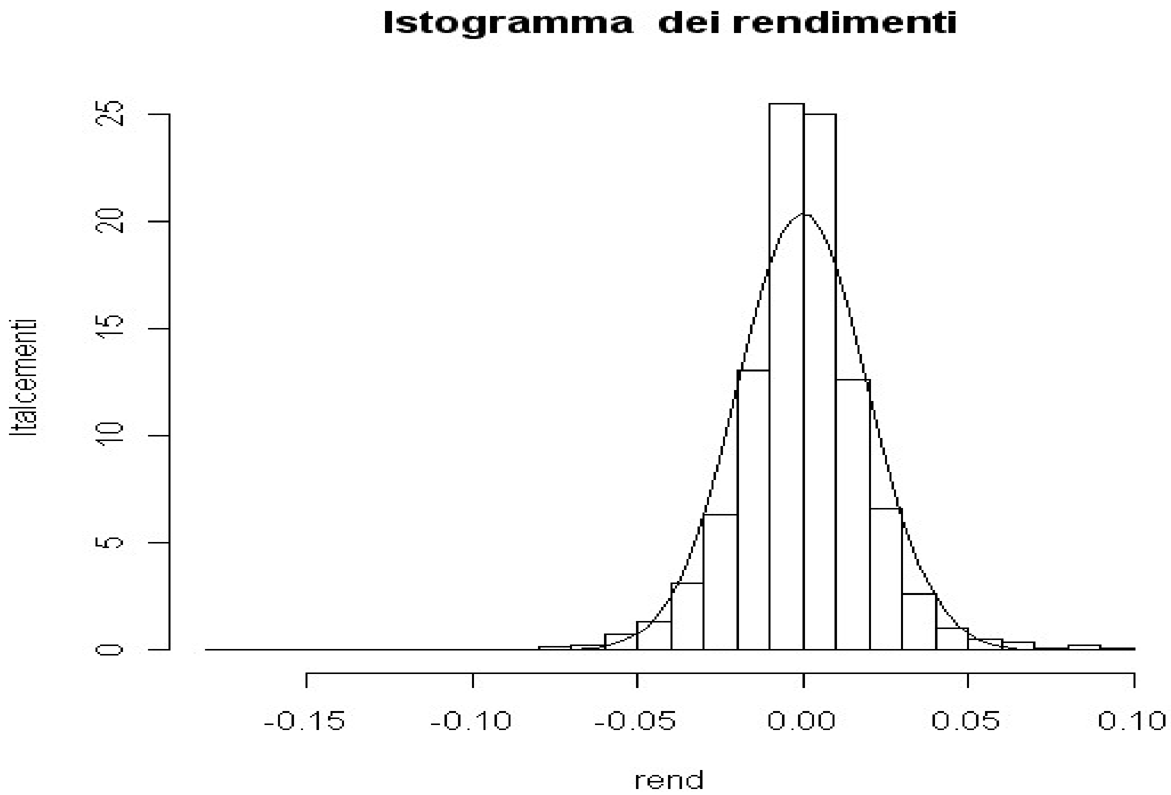
Il grafico e il correlogramma ci dicono che esiste dipendenza lineare sui quadrati dei rendimenti e pertanto è possibile modellare la serie dei rendimenti al quadrato attraverso la modellistica ARMA. Inoltre supponendo che i rendimenti abbiamo valore atteso nullo, potremmo costruire il nostro modello lavorando direttamente sulla volatilità dei rendimenti, cioè potremmo modellare la varianza dei rendimenti.

Leptocurtosi della serie dei rendimenti

La serie dei rendimenti, non presenta una distribuzione normale, ma leptocurtica ovvero più appuntita e con code più pesanti, questo vuol dire che rendimenti grandi e piccoli in valore assoluto si presentano con una frequenza superiore rispetto a quella attesa nel caso normale. Vediamo:

hist(rend,main="Istogramma dei rendimenti",ylab="Italcementi",nclass=20,freq=F)

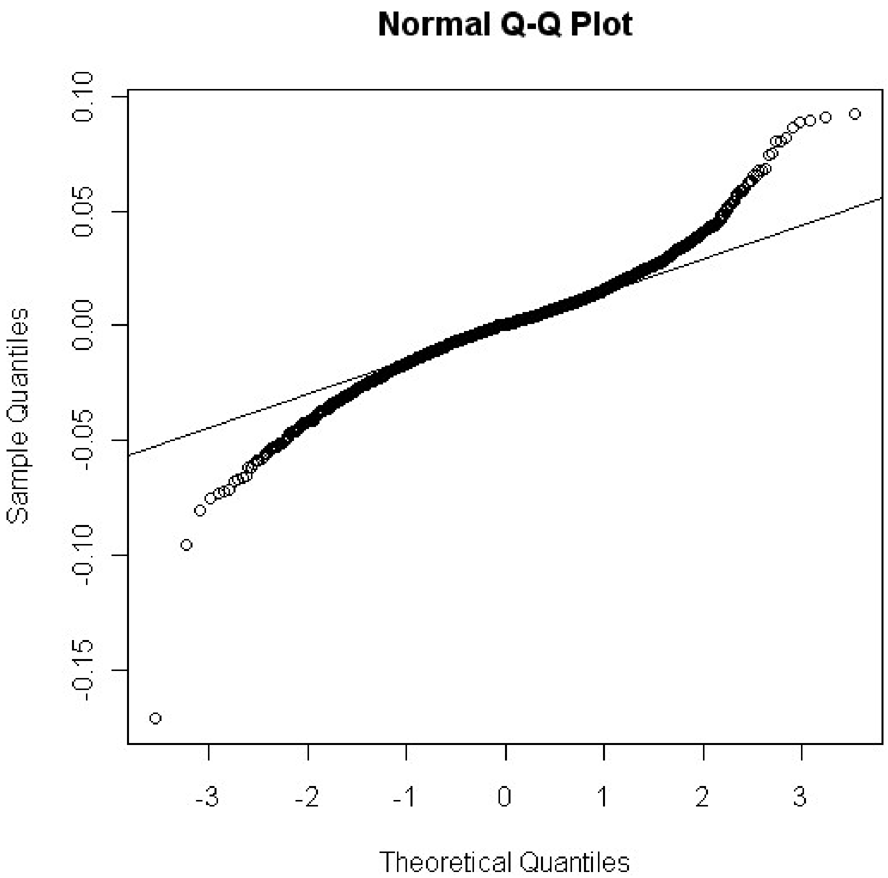
curve(dnorm(x,mean(rend),sd(rend)),add=T)



Un'ulteriore conferma grafica viene dal seguente commando di "R" che permette di confrontare i quantili della distribuzione dei rendimenti con quelli teorici di una normale. Se i punti della serie osservata si distribuiscono come una linea retta, allora la distribuzione è normale, altrimenti, se i punti seguono una linea curva, la distribuzione devia da quella della normale. Vediamo:

qqnorm(rend)

qqline(rend)



Nel nostro caso come è evidente, i quantili della distribuzione dei rendimenti deviano dalla linea retta della distribuzione normale proprio sulle code, evidenziando quindi la presenza di leptocurtosi.

Usiamo ora il pacchetto "*fBasic*" di R e verifichiamo empiricamente le conclusioni a cui siamo giunti dalla semplice osservazione grafica. Useremo gli indici di Curtosi e di Asimmetria. Nel caso di una distribuzione normale dovranno risultare rispettivamente pari a 3 e a 0, vediamo cosa accade.

kurtosis(rend)

[1] 4.66181

skewness(rend)

[1] -0.1603977

Riguardo l'indice di curtosi vediamo che è decisamente maggiore di 3 quindi la distribuzione è più appuntita.

Riguardo l'indice di simmetria (il secondo), essendo diverso da zero e per di più negativo, ci dice che shocks negativi incrementano in misura maggiore la volatilità rispetto a quelli positivi. Inoltre è possibile costruire un test (Jarque-Bera Normality test) che verifica simultaneamente se l'asimmetria e la curtosi sono coerenti con i valori che dovrebbero assumere sotto l'ipotesi nulla di normalità.

jarque.bera.test(rend)

Jarque Bera Test

data: rend
X-squared = 2249.559, df = 2, **p-value < 2.2e-16**

Il *p-value* risulta essere minore di *0,05* e quindi rifiutiamo l'ipotesi nulla di normalità della distribuzione dei rendimenti e tale risultato risulta essere perfettamente coerente con i due precedentemente riscontrati.

Volatility clustering e modelli ARCH e GARCH

Nelle serie storiche finanziarie si nota che periodi di alta volatilità tendono ad essere seguiti da periodi ad alta volatilità e periodi di bassa volatilità tendono ad essere seguiti da periodi a bassa volatilità, fenomeno che prende il nome di volatility clustering.
Tali considerazioni portano a preferire l'uso di modelli che colgono la presenza di eteroschedasticità, come i modelli ARCH e la loro generalizzazione GARCH.

Stima del modello

Per poter individuare il modello che meglio si adatta alla serie in esame, dobbiamo procedere alla stima di differenti parametri e verificare il loro livello di significatività. Inoltre sceglieremo osservando il valore del Akaike, del p-value per i test di Jarque-Bera e Ljung-Box. Dovremo infine tenere presente anche che il modello dovrà essere *parsimonioso*, cioè con il minor numero di parametri stimati.
A questo punto non ci rimane altro che caricare il pacchetto "fGarch" in R e usare il seguente comando:

Tipo Modello=garch(rend,order=c(p,q))

Dove p e q identificheranno il modello considerato. Ad esempio:

Arch1=garch(rend,order=c(0,1))

Per poter ottenere i valori empirici di cui abbian detto sopra, useremo i comandi:

summary(Tipo Modello)

e

AIC(Tipo Modello)

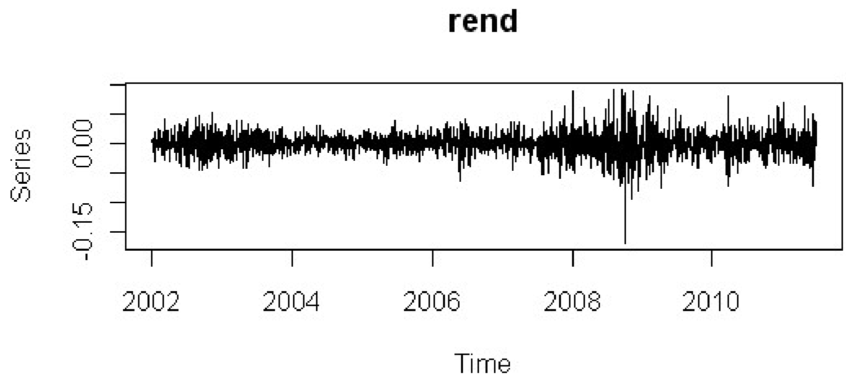
Abbiamo deciso infine di sintetizzare il tutto in una tabella in modo da poter confrontare con più facilità i dati e poi scegliere.

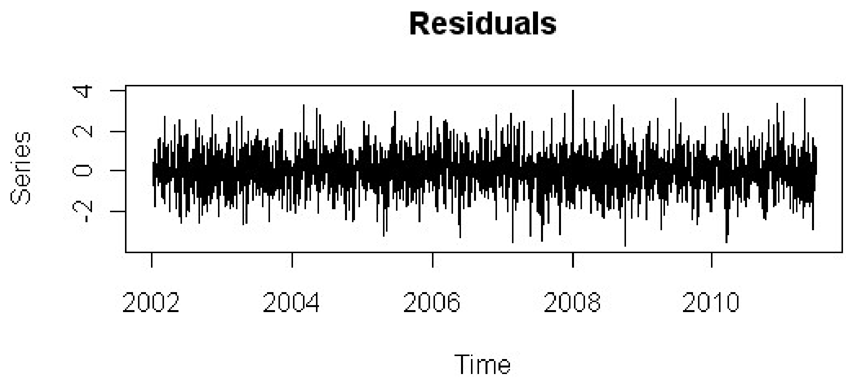
Tipo Modello	Parametri	Jarque-Bera (p-value)	Ljung-Box (p-value)	Akaike
Arch (1)	a0 2.878e-04 a1 2.500e-01	2.2e-16	0.2662	-12572.29

Arch (2)	a0 2.263e-04 a1 2.172e-01 a2 2.119e-01	2.2e-16	0.2398	-12662.49
Arch (3)	a0 1.792e-04 a1 1.775e-01 a2 1.719e-01 a3 2.071e-01	2.2e-16	0.3088	-12760.75
Arch (4)	a0 1.589e-04 a1 1.495e-01 a2 1.597e-01 a3 2.024e-01 a4 9.912e-02	2.2e-16	0.5776	-12789.09
Arch (5)	a0 1.378e-04 a1 1.500e-01 a2 1.226e-01 a3 1.423e-01 a4 8.028e-02 a5 1.638e-01	2.2e-16	0.4783	-12838.55
Garch (1.1)	a0 1.767e-06 a1 5.435e-02 b1 9.421e-01	1.236e-12	0.5934	-13067.08

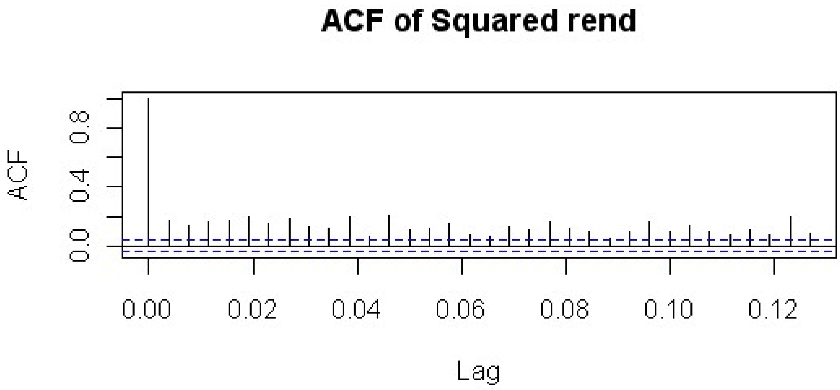
Premesso che l'indice Jarque – Bera viene impiegato per testare la normalità degli errori, leggendo la tabella, verifichiamo come per ogni modello tale condizione non si verifichi. Infatti i valori del p – value sono tutti inferiori a *0,05* portandoci quindi a rifiutare l'ipotesi nulla di normalità.
Il test di Ljiung – Box inoltre consente di verificare la presenza o meno di autocorrelazione dei residui e anche qui ci accorgiamo che ogni modello presenta il p-value maggiore di 0,05 il che conferma l'ipotesi nulla di incorrelazione dei residui.
Rimane il valore del Akaike da valutare. Dovremmo scegliere il modello con l'Akaike minore in valore assoluto e cioè il modello Garch (1.1). Abbiamo a nostro favore anche il fatto che il modello prescelto è anche *parsimonioso*.

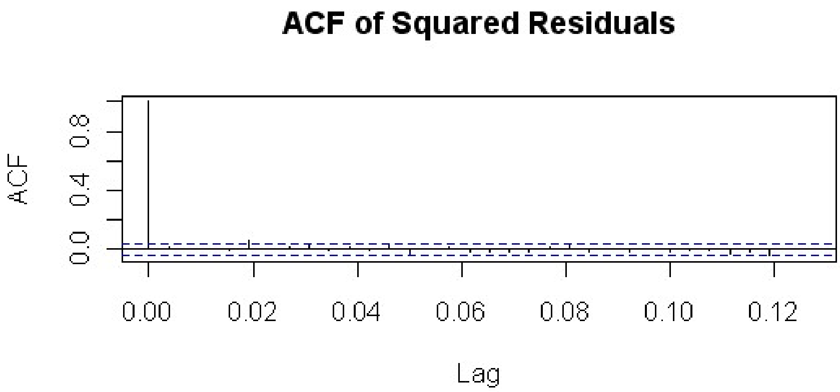
"Plottiamo" il Garch (1.1) e confrontiamo l'andamento dei rendimenti con i residui





Risulta evidente che i rendimenti e i residui non hanno gli stessi andamenti. Il modello quindi riesce a spiegare buona parte della dipendenza nel tempo della serie.
Inoltre il correlogramma dei residui il cui grafico viene riportato sotto, si annulla dopo il lag 1 e questo dimostra come la correlazione risulti nulla, esattamente come accade in un processo White Noise.





L'equazione del modello

Concludiamo il nostro lavoro con l'equazione del modello Garch (1.1)

$$\sigma_t^2 = \mathbf{1.767e\text{-}06 + 5.435e\text{-}02}\ r_{t-1}^2\ \mathbf{+ 9.421e\text{-}01}\ \sigma_{t-1}^2$$