

Università degli Studi di Padova



Facoltà di Scienze Statistiche

Corso di Laurea Specialistica in Scienze Statistiche,
Economiche, Finanziarie e Aziendali

TESI DI LAUREA

**Relazione tra rendimenti e volumi
nei titoli azionari: il caso IBM**

Relatore: Ch.mo Prof. Caporin Massimiliano

Laureanda: Menin Federica

Matricola: 601622

Anno Accademico 2010 / 2011

A chi mi sta sempre accanto

Indice

Introduzione	11
1 Le grandezze finanziarie	13
1.1 I prezzi	13
1.2 I rendimenti	17
1.3 La volatilità	23
1.4 I volumi	25
2 La relazione tra rendimenti e volumi	27
3 Dati ad alta frequenza	31
3.1 Caratteristiche dei dati	32
3.2 Non sincronicità negli scambi	33
4 Modelli GARCH	37
4.1 Il modello ARCH(p)	37
4.2 Stima del modello ARCH(p)	39
4.3 Il modello GARCH(p,q)	42
4.4 Stima del modello GARCH(p,q)	44
4.5 Il modello EGARCH(p,q)	45
4.6 Modelli con variabili esogene	45
5 Applicazioni	47
5.1 Analisi dei dati	48
5.2 I modelli	60

5.3 Conclusioni	74
Ringraziamenti	77
Bibliografia	78

Elenco delle figure

1.1	Grafico dei prezzi del titolo FIAT nel periodo 03/01/2005 - 30/12/2010	15
1.2	Correlogramma dei prezzi del titolo FIAT nel periodo 03/01/2005 - 30/12/2010	16
1.3	Grafico dei rendimenti del titolo FIAT nel periodo 03/01/2005 - 30/12/2010	19
1.4	Istogramma dei rendimenti del titolo FIAT nel periodo 03/01/2005 - 30/12/2010	20
1.5	Q-Q plot dei rendimenti del titolo FIAT nel periodo 03/01/2005 - 30/12/2010	21
1.6	Correlogramma dei rendimenti del titolo FIAT nel periodo 03/01/2005 - 30/12/2010	22
1.7	Grafico dei volumi del titolo FIAT nel periodo 03/01/2005 - 30/12/2010	25
5.1	Grafici dei rendimenti a diverse frequenze per il titolo IBM dallo 02/01/2003 al 30/03/2010	49
5.2	Istogrammi dei rendimenti a diverse frequenze per il titolo IBM dallo 02/01/2003 al 30/03/2010	50
5.3	Grafici dei volumi a diverse frequenze per il titolo IBM dallo 02/01/2003 al 30/03/2010	53
5.4	Medie e varianze giornaliere con frequenza a 1 minuto per rendimenti e volumi del il titolo IBM dallo 02/01/2003 al 30/03/2010	55

5.5	Medie e varianze giornaliere con frequenza a 5 minuti per rendimenti e volumi del il titolo IBM dallo 02/01/2003 al 30/03/2010	56
5.6	Medie e varianze giornaliere con frequenza a 10 minuto per rendimenti e volumi del il titolo IBM dallo 02/01/2003 al 30/03/2010	57
5.7	Medie e varianze giornaliere con frequenza a 15 minuto per rendimenti e volumi del il titolo IBM dallo 02/01/2003 al 30/03/2010	58
5.8	Medie e varianze giornaliere con frequenza a 30 minuto per rendimenti e volumi del il titolo IBM dallo 02/01/2003 al 30/03/2010	59

Elenco delle tabelle

5.1	Statistiche descrittive per le serie dei rendimenti alle frequenze di 1 minuto, 5 minuti, 10 minuti, 15 minuti e 30 minuti	52
5.2	Statistiche descrittive per le serie dei volumi alle frequenze di 1 minuto, 5 minuti, 10 minuti, 15 minuti e 30 minuti	54
5.3	Stima del modello GARCH(1,1) con volume al tempo t	61
5.4	Stima del modello GARCH(1,1) con il logaritmo del volume al tempo t	62
5.5	Stima del modello GARCH(1,1) con la differenza prima del logaritmo del volume al tempo t	63
5.6	Stima del modello GARCH(1,1) con la differenza prima del logaritmo del volume al tempo $t-1$	64
5.7	Stima del modello GARCH(1,1) con volume al tempo $t-1$	65
5.8	Stima del modello EGARCH(1,1) con volume al tempo t	68
5.9	Stima del modello EGARCH(1,1) con logaritmo del volume al tempo t	69
5.10	Stima del modello EGARCH(1,1) con differenza prima del logaritmo del volume al tempo t	70
5.11	Stima del modello EGARCH(1,1) con differenza prima del logaritmo del volume al tempo t	71
5.12	Stima del modello EGARCH(1,1) con volume al tempo $t-1$	72

Introduzione

Lo scopo di questo lavoro è capire la relazione esistente tra i rendimenti e i volumi di un titolo finanziario. Per cercare di spiegare questa relazione sono stati usati dati ad alta frequenza. Questo tipo di dati ha cominciato ad essere molto utilizzato negli ultimi anni perchè permette di catturare molti più movimenti di mercato rispetto ai dati giornalieri usati precedentemente. Al fine di raggiungere lo scopo prefissato sono stati utilizzati i modelli GARCH ed EGARCH ai quali sono state opportunamente aggiunte delle variabili esogene.

Nel primo capitolo verrà fornita una panoramica delle grandezze finanziarie analizzate durante l'elaborazione dei dati. Le grandezze in questione saranno prezzi, rendimenti, volatilità e volumi, dei quali verranno analizzate nel dettaglio le varie caratteristiche.

Nel secondo capitolo cercheremo di fornire una piccola panoramica dei precedenti studi condotti sulla nostra relazione di interesse per cercare di evidenziare, successivamente, se le nostre analisi siano in accordo o meno con quanto ottenuto dagli studi precedenti.

Nel capitolo 3 verranno presentati i dati ad alta frequenza. Per questo tipo di dati verrà anche spiegato il problema della non sincronicità in quanto questi dati registrano tutti i movimenti che avvengono nel mercato, movimenti che non si presenteranno ad intervalli di tempo regolari.

Per capire che tipo di modelli sono stati utilizzati per spiegare la relazione tra rendimenti e volumi, nel capitolo 3 verranno presi in esame i modelli GARCH ed EGARCH. Per prima cosa verranno analizzati i modelli ARCH,

precursori dei GARCH. Sia per i modelli ARCH che per i modelli GARCH sono riportate le funzioni di verosimiglianza utili alla stima dei parametri inseriti nel modello. Successivamente si sono analizzati dei modelli GARCH ed EGARCH in cui sono state inserite delle variabili esogene nell'equazione della media condizionata. Queste modifiche al modello di base si riveleranno utili al fine di capire la relazione tra i rendimenti e i volumi del titolo analizzato.

Nel capitolo 5 verranno forniti i risultati delle analisi. Nella prima parte si analizzeranno le serie storiche dei volumi e dei rendimenti in modo descrittivo così da capire le caratteristiche delle singole serie e confrontarle con quanto riportato nel capitolo 1. Successivamente saranno presentati i modelli che sono stati utilizzati per condurre l'analisi. Verranno inseriti i risultati ottenuti e saranno poi commentati per arrivare a delle conclusioni.

Capitolo 1

Le grandezze finanziarie

In tempi recenti ha attirato molta attenzione l'analisi quantitativa dei mercati finanziari. Questo tipo di analisi ha alla base l'ipotesi di efficienza dei mercati, cioè la proprietà dei prezzi di riflettere completamente tutte le informazioni a disposizione e, quindi, la possibilità di prevederne l'andamento. Un altro importante aspetto dell'analisi quantitativa è il trade-off tra rischio e rendimento. Per trade-off si intende la relazione tra rischio e rendimento che li lega in modo diretto. Infatti, un titolo con maggior rischio rispetto a un altro presenterà anche un maggior rendimento e viceversa. Il punto di partenza dell'analisi quantitativa è la serie storica dei prezzi dove, per *serie storica*, si intende un campione di osservazioni riguardanti lo stesso soggetto ma ad istanti temporali diversi. I prezzi, quindi, sono uno dei soggetti principali necessari ad effettuare questo tipo di analisi. Gli altri soggetti fondamentali sono i rendimenti, la volatilità e i volumi che ora andremo ad analizzare in dettaglio.

1.1 I prezzi

I prezzi rappresentano il valore economico di una determinata attività finanziaria. Possono essere indicizzati dall'istante temporale nel quale sono os-

servati e sono interpretabili come capitalizzazione rispetto al prezzo osservato al tempo precedente effettuata al tasso r_t . Quindi il prezzo P_t è:

$$P_t = P_{t-1}(1 + r_t) \quad (1.1)$$

Nei mercati in cui esiste un intermediario (*market maker*) responsabile della determinazione del prezzo di equilibrio, i prezzi quotati sono diversi a seconda che si riferiscano all'acquisto o alla vendita del titolo stesso. Se ci si riferisce all'acquisto, si parla di prezzo *bid*, se ci si riferisce alla vendita, si parla di prezzo *ask*.

L'evoluzione nel tempo del prezzo di un'azione può essere rappresentato con l'equazione:

$$\ln(p_t) = \mu + \ln(p_{t-1}) + u_t \quad (1.2)$$

con

$$E(u_t | p_{t-1}, p_{t-2}, \dots) = 0 \quad (1.3)$$

Questo processo stocastico prende il nome di *random walk* ed esprime il prezzo p_t come somma di una componente deterministica casuale μ , il prezzo rilevato al tempo precedente p_{t-1} e un termine casuale indipendente dai valori passati dei prezzi u_t . Questo tipo di processo è visibile analizzando il grafico dei prezzi del titolo Fiat in Figura 1.1.

Il processo *random walk* non è stazionario in senso debole perché $E(p_t)$ e $Var(p_t)$ non sono costanti nel tempo ma dipendono, appunto, dall'istante temporale t e $Cov(p_t, p_{t-h})$ non dipende solo dall'istante h ma anche dall'istante t . La non stazionarietà del processo è identificabile attraverso lo studio del correlogramma dei prezzi, presente in Figura 1.2, il quale presenta valori positivi molto elevati e che tendono a zero in modo molto lento.

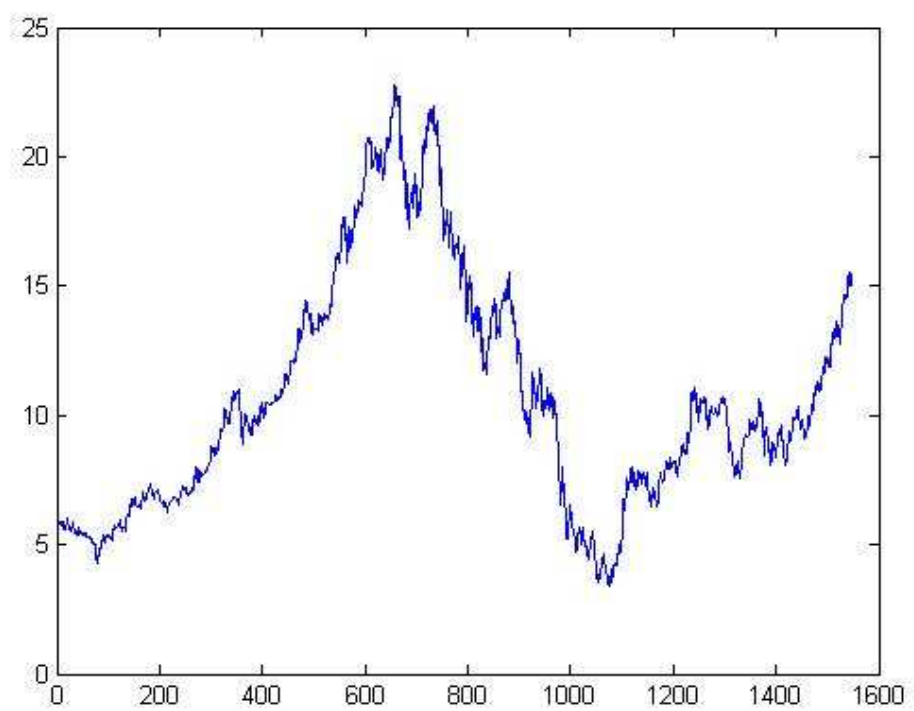


Figura 1.1: Grafico dei prezzi del titolo FIAT nel periodo 03/01/2005 - 30/12/2010

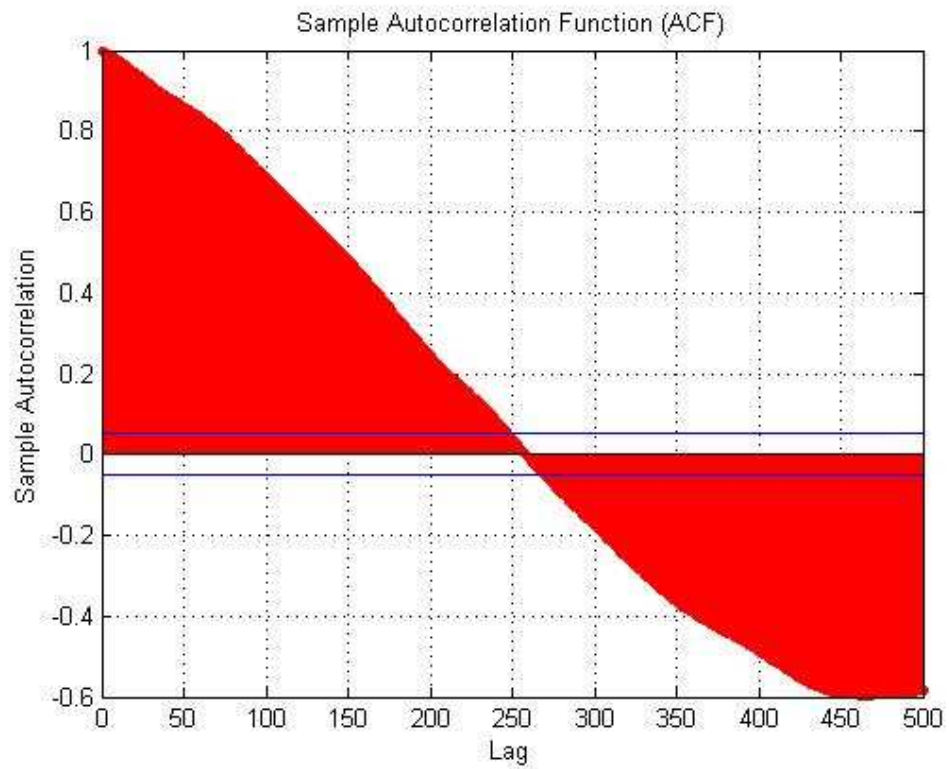


Figura 1.2: Correlogramma dei prezzi del titolo FIAT nel periodo 03/01/2005 - 30/12/2010

1.2 I rendimenti

I rendimenti misurano la variazione percentuale dei prezzi in un certo intervallo di tempo. Molto spesso, anziché analizzare i prezzi, si preferisce analizzare i rendimenti. Infatti, Campbell, Lo e MacKinlay (1997) forniscono due ragioni principali per questa scelta. Il primo motivo è che i rendimenti che guidano l'evoluzione dei prezzi forniscono un'informazione completa sulle opportunità di investimento. Il secondo motivo, invece, riguarda le caratteristiche statistiche dei rendimenti, che permettono una più facile e migliore analisi rispetto a quelle dei prezzi.

Andiamo ora ad analizzare i vari tipi di rendimento.

Rendimento semplice

Il rendimento semplice uniperiodale viene calcolato dal periodo $t-1$ al periodo t nel seguente modo:

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1.4)$$

dove P_t indica il prezzo al periodo t .

Rendimento composto

Il rendimento composto uniperiodale viene anche chiamato *rendimento logaritmico* visto che si calcola come:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (1.5)$$

Rendimento di portafoglio

Se in un portafoglio p sono presenti N asset, in cui ω_i rappresenta il peso dell'asset i , allora il rendimento del portafoglio sarà dato da:

$$R_{p,t} = \sum_{i=1}^N \omega_i r_{it} \quad (1.6)$$

Pagamento dei dividendi

Se un asset presenta periodicamente un pagamento di dividendi si viene a modificare il calcolo dei rendimenti perché si deve tener conto di questo nuovo elemento. Denominato con D_t il valore del dividendo pagato tra il periodo $t-1$ e il periodo t e non incluso nel prezzo P_t , il valore del rendimento sarà dato da:

$$R_t = \frac{P_t + D_t}{P_{t-1}}, \quad r_t = \ln(P_t + D_t) - \ln(P_{t-1}) \quad (1.7)$$

I rendimenti presentano stazionarietà in media ma non presentano stazionarietà in varianza. Dalla Figura 1.3 si vede chiaramente come la media dei rendimenti sia prossima allo zero e tutti i valori si distribuiscano attorno a questo valore. Si può anche vedere come la variabilità dei dati non si mantenga costante nel tempo ma ci siano periodi con maggiore variabilità rispetto ad altri.

Se si analizza la distribuzione dei rendimenti questa risulta centrata sullo zero e abbastanza simmetrica ma non può essere assunta come una distribuzione Normale, in quanto presenta code pesanti. Queste caratteristiche sono visibili anche graficamente, analizzando un istogramma di una serie storica dei rendimenti, come il grafico in Figura 1.4.

Un'ulteriore conferma che la distribuzione dei rendimenti non può essere assunta Normale si ha dal confronto tra i quantili campionari, stimati su una determinata serie storica, con i quantili teorici di una distribuzione normale con l'uso di un q-q plot. Come si nota dalla Figura 1.5, le coppie di punti non si trovano allineate sulla diagonale del grafico, ma si ha uno scostamento rilevante proprio nei valori estremali. Questo conferma quanto detto precedentemente, cioè che i rendimenti non seguono una distribuzione Normale ma presentano code pesanti.

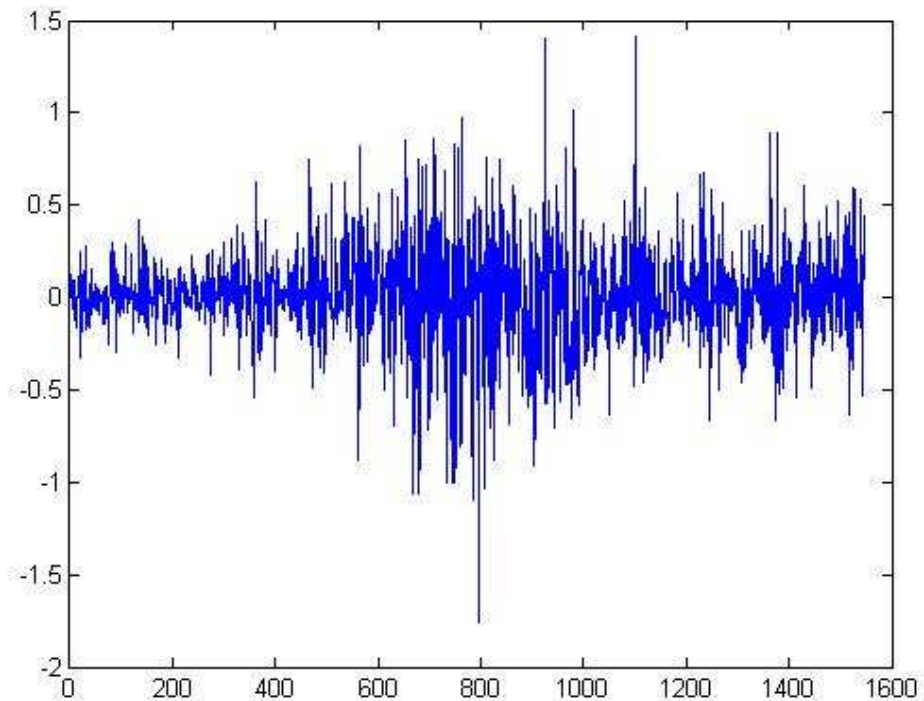


Figura 1.3: Grafico dei rendimenti del titolo FIAT nel periodo 03/01/2005 - 30/12/2010

Altra proprietà dei rendimenti è la loro incorrelazione, o debole correlazione. Questo aspetto può essere facilmente verificabile attraverso il correlogramma di Figura 1.6. Infatti, si vede come i valori del correlogramma non siano molto elevati e rientrino nella bande di confidenza, segnalando la debole correlazione presente in questa serie storica.

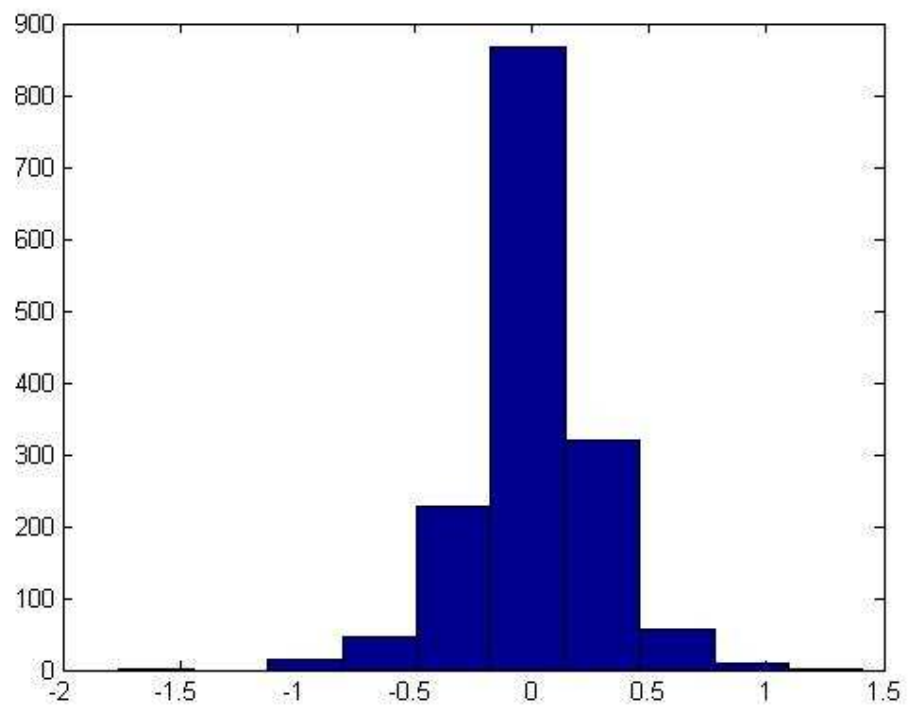


Figura 1.4: Istogramma dei rendimenti del titolo FIAT nel periodo 03/01/2005 - 30/12/2010

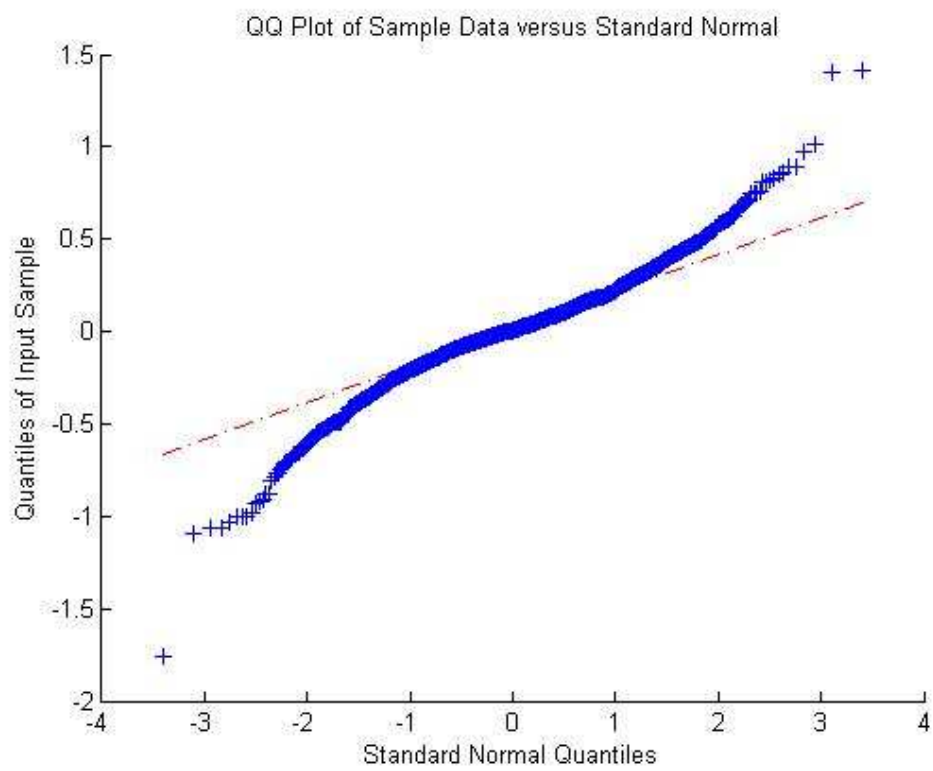


Figura 1.5: Q-Q plot dei rendimenti del titolo FIAT nel periodo 03/01/2005 - 30/12/2010

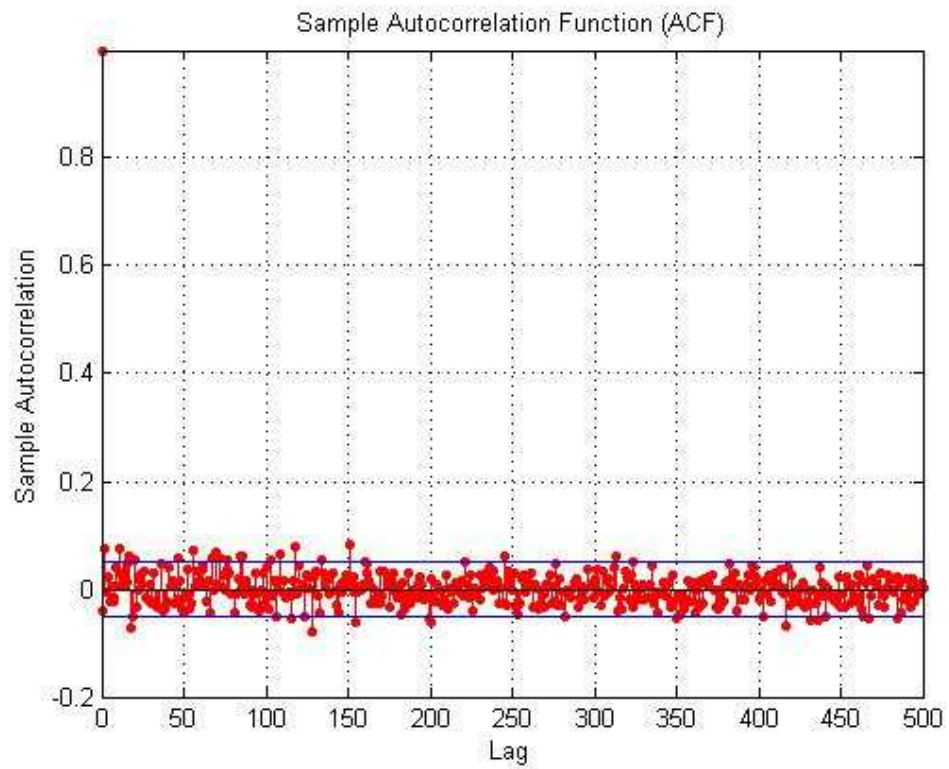


Figura 1.6: Correlogramma dei rendimenti del titolo FIAT nel periodo 03/01/2005 - 30/12/2010

1.3 La volatilità

La volatilità misura le fluttuazioni compiute dai prezzi nel periodo di tempo considerato. Esistono vari modi per misurare questa grandezza finanziaria.

Un indice di variabilità usato come misura di volatilità dei rendimenti su tutto il campione è la *realized volatility*:

$$\hat{\sigma}_r^2 = \frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})^2 \quad (1.8)$$

In questa formulazione, in cui compare la media dei rendimenti, potrebbero esserci delle alterazioni nell'analisi dovute ad eventuali valori anomali. Per questo motivo sarebbe opportuna un'analisi preliminare dei dati indirizzata all'eliminazione di eventuali valori anomali al fine di evitare questo tipo di problema ed ottenere una misura della variabilità in più corretta e veritiera possibile.

Un'altra misura di volatilità è la *varianza mobile*. Questa misura è abbastanza usata e ha un andamento molto smussato, ci potranno essere solamente dei salti nel caso in cui siano presenti dei valori anomali. Per calcolarla si deve dividere il periodo di osservazione T in tanti intervalli di ampiezza τ e ogni volta, aggiungendo un'osservazione più recente ed eliminandole una più vecchia, ricalcolare la varianza. In questo modo il numero di osservazioni rimarrà comunque costante ma ci sarà uno spostamento temporale progressivo. La formula di calcolo sarà quindi:

$$\hat{\sigma}_{r,t}^2 = \frac{1}{\tau} \sum_{i=t-\tau-1}^t (r_i - \bar{r}_t)^2 \quad (1.9)$$

Andiamo ora ad analizzare alcune caratteristiche che può presentare la varianza nelle serie storiche finanziarie.

Code pesanti

Quando si va a confrontare la serie storica dei rendimenti di un titolo con una distribuzione Normale si nota la presenza di code pesanti della serie dei

rendimenti. Questa caratteristica è nota anche come *eccesso di kurtosi*. Il valore del momento di ordine quarto, che indica la kurtosi, per una serie con distribuzione Normale assume valore 3 mentre nelle serie storiche dei rendimenti finanziari supera questo valore segnalando l'impossibilità di una distribuzione di tipo Normale.

Volatility clustering

Per *volatility clustering* si intende la persistenza di uno shock, quindi periodi di grande volatilità seguiranno periodi di grande volatilità. Per vedere questa caratteristica è sufficiente analizzare i correlogrammi della serie storica dei rendimenti del titolo e i corrispondenti valori del test di Box-Ljung che rilevino la presenza di correlazione. Se i valori del correlogramma sono elevati ed escono dalle bande di confidenza allora saremo in presenza di *volatility clustering*

Effetti leva

L'effetto leva si riferisce al fatto che i movimenti del prezzo sono negativamente correlati con la volatilità. Il primo a suggerire questa proprietà fu Black (1976) per i rendimenti azionari. Black, tuttavia, spiegò che l'effetto delle variazioni di prezzo sulla volatilità non potevano essere spiegate solamente dall'effetto leva.

Memoria lunga e co-movimenti della volatilità

Specialmente per i dati finanziari ad alta frequenza, la volatilità è altamente persistente e quindi con memoria lunga. Ciò significa che la volatilità del periodo t sarà presente anche nel periodo $t+1$ sia nel caso in cui questa sia in aumento sia nel caso in cui sia in diminuzione.

Se si guardano le serie storiche di mercati diversi, ad esempio le serie storiche dei tassi di cambio di valute diverse, si può notare la presenza degli stessi movimenti tra le varie serie analizzate. Questo suggerisce l'importanza di modelli multivariati per modellare la cross-correlazione nei diversi mercati.

1.4 I volumi

Il volume rappresenta la quantità di titoli finanziari scambiati nel periodo di tempo considerato. E', quindi, un indicatore della liquidità di una determinata attività finanziaria. Solitamente, un'importante variazione dei volumi è seguita da un'importante variazione dei prezzi del titolo stesso. Questa misura è molto usata nell'analisi tecnica per esaminare l'andamento e la consistenza delle contrattazioni di un titolo, al fine di prevederne l'andamento.

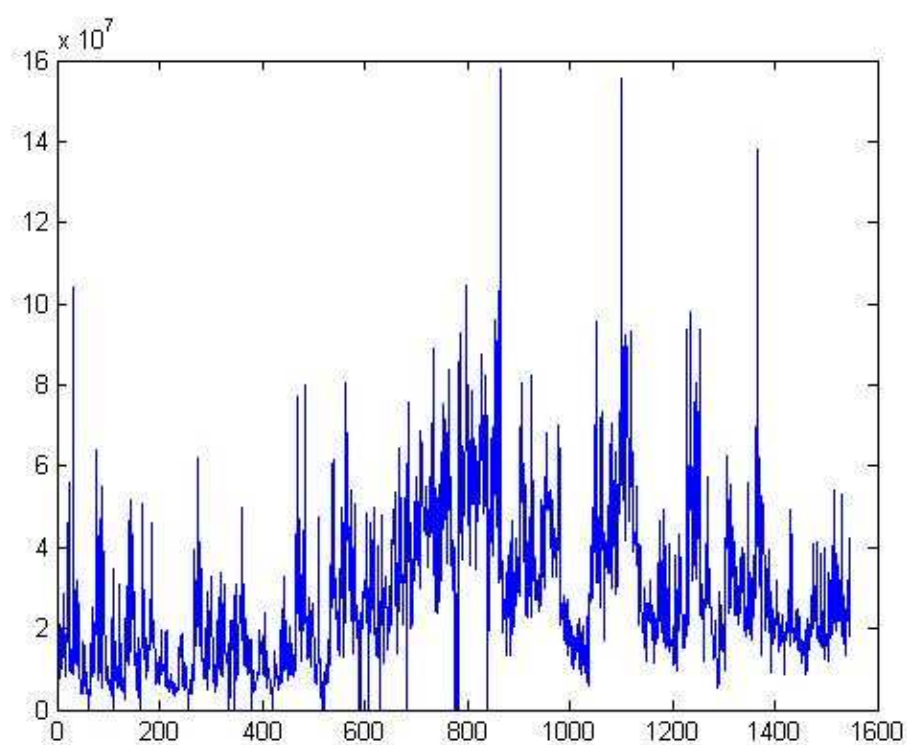


Figura 1.7: Grafico dei volumi del titolo FIAT nel periodo 03/01/2005 - 30/12/2010

Una proprietà importante è la liquidità del mercato da cui proviene la serie storica. Se la transazione è associata a un volume di scambio elevato, questa sarà sicuramente più significativa rispetto a una transazione con volume di scambio non molto alto perché nel primo caso il prezzo rappresenta realmente

una situazione di equilibrio tra domanda e offerta. Se il volume di scambio associato a un determinato prezzo è di livello basso si potrebbe presentare il problema del quantitativo di titoli da comprare o vendere per non alterare in modo sensibile il prezzo del titolo stesso, altrimenti si rischierebbe di effettuare l'operazione a un prezzo di gran lunga diverso da quello atteso, in base all'ultima quotazione disponibile.

Capitolo 2

La relazione tra rendimenti e volumi

Molti studi sono stati effettuati sulla relazione esistente tra i rendimenti di un titolo finanziario e i suoi volumi di scambio. Questa relazione risulta molto importante per vari motivi. Per prima cosa può essere utile a capire in che modo le informazioni sono disseminate nei mercati finanziari. Inoltre, questa relazione è importante nella valutazione della distribuzione dei rendimenti stessi. Importante da capire è anche se è il rendimento a determinare il volume o viceversa e se la relazione è contemporanea o ritardata.

Saatcioglu e Starks (1998) hanno realizzato uno studio sul mercato latino-americano per cercare di capire la relazione tra rendimenti e volumi in questo mercato emergente. Hanno usato dati mensili e sono riusciti ad individuare una correlazione positiva tra rendimenti e volumi contemporanei per quasi tutti i mercati dell'area di analisi. Sono riusciti, inoltre, a concludere che i rendimenti sono determinati dai volumi ma non il viceversa.

Un'analisi sul mercato greco è stata effettuata da Floros e Vougas (2007). Per questo caso gli autori hanno usato il modello GARCH(1,1) e hanno inserito il logaritmo del volume, prima nell'equazione della media, e poi nell'equazione della varianza. Con questa metodologia sono riusciti a concludere che tra i rendimenti e i volumi contemporanei esiste una relazione positiva e che il

volume spiega anche, in maniera significativa, la varianza condizionata.

Un'analisi sul mercato turco è stata realizzata da Baklaci e Kasman, usando i dati giornalieri di 25 titoli dell'Istanbul Stock Exchange. Per lo studio della possibile relazione tra rendimento e volume è stato usato anche in questo caso il modello GARCH(1,1) ed è stata evidenziata una relazione contemporanea tra rendimenti e volumi inserendo questi ultimi nell'equazione della varianza condizionata. Inoltre, è stato verificato che la persistenza della volatilità non diminuisce con l'inserimento del volume nell'equazione della varianza condizionata.

Mubarik e Javid (2009) hanno realizzato uno studio sul mercato pakistano. Anche per l'analisi di questo mercato è stato usato un modello GARCH(1,1) inserendo nell'equazione della varianza il volume. I coefficienti ARCH e GARCH del modello sono risultati significativi alla soglia dell'1% ed è stata evidenziata una relazione positiva contemporanea tra i rendimenti e i volumi. In questo lavoro è stata anche studiata la relazione tra i rendimenti e i volumi al periodo precedente. Anche per questa analisi la relazione è risultata essere positiva e significativa.

Alcuni studi hanno rilevato che inserendo il volume nell'equazione della varianza condizionata in un modello GARCH(1,1) la persistenza della varianza condizionata non diminuiva per i mercati del centro Europa e dell'Europa dell'est. Per questo motivo Bohl e Henke (2003) hanno cercato di studiare il mercato polacco per riuscire a smentire quanto detto precedentemente. Per questo studio sono stati analizzati i dati giornalieri di 20 titoli del Warsaw Stock Exchange ed è stato inserito il volume nell'equazione della varianza condizionata del modello. E' stata rilevata una riduzione della persistenza della varianza soltanto per alcuni dei titoli presi in esame.

De Medeiros e Van Doornik hanno effettuato uno studio sui dati giornalieri del mercato brasiliano. Compiendo un'analisi di cross-correlazione tra rendimenti e volumi sono riusciti a stabilire che tra le due grandezze esiste una relazione di tipo contemporaneo. Un'analisi attraverso le equazioni simultanee è riuscita a catturare una dipendenza dei rendimenti dai volumi ma

non è invece stata riscontrata la relazione inversa. Da questo studio è risultato anche che livelli elevati dei volumi sono associati ad un aumento della volatilità dei rendimenti. Questa relazione non è simmetrica ma è risultata più evidente nei casi in cui i rendimenti erano positivi.

Uno studio molto interessante è stato effettuato da Lee e Rui (2002) in cui sono stati usati i dati giornalieri dei mercati di New York, Londra e Tokio. L'analisi è stata effettuata sia sul singolo mercato sia tra i tre mercati presi in esame. Dall'analisi è risultato che esiste una certa relazione positiva tra i rendimenti e i volumi nei tre mercati ma non sembra esserci una relazione di causalità tra le due grandezze. Dall'analisi incrociata dei tre mercati è risultato che i volumi del mercato americano contengono informazioni sulle variabili finanziarie sia del mercato giapponese che di quello inglese e questo suggerisce la presenza di un flusso informativo che dal mercato americano si espande negli altri paesi.

Nello studio effettuato da Chen, Firth e Rui (2002) sono stati presi in esame i mercati di 9 paesi: Stati Uniti d'America, Giappone, Gran Bretagna, Francia, Canada, Italia, Svizzera, Olanda e Hong Kong. Per tutti i mercati studiati è stata rilevata la presenza di correlazione positiva tra il volume e il valore assoluto dei rendimenti. Il modello migliore per rappresentare i rendimenti dei mercati è risultato il modello EGARCH e nell'equazione della varianza condizionata è stato inserito il volume, sia contemporaneo che ritardato. Con l'inserimento di queste variabili non è diminuita, però, la persistenza della varianza. Questo risultato è stato riscontrato in tutti i 9 mercati presi in esame.

Anche Gallant, Rossi e Tauchen hanno cercato di capire questa importante relazione usando i dati giornalieri del mercato americano. Come per gli altri casi, i loro studi hanno dimostrato l'esistenza di una relazione positiva e non lineare tra rendimenti e volumi.

Analizzando questi precedenti lavori risulta quindi evidente che tra rendimenti e volumi esiste una relazione contemporanea positiva almeno per i dati giornalieri. Molti sono stati i casi in cui è stato usato come riferimento

il modello GARCH(1,1) inserendo il termine del volume nell'equazione della varianza condizionata ma sembra che questo non diminuisca la persistenza della varianza.

L'obiettivo di questa tesi è valutare la stabilità della relazione tra i rendimenti e i volumi usando i dati ad alta frequenza. La nostra analisi sarà basata sia sui modelli GARCH che sui modelli EGARCH. Faremo quindi riferimento principalmente alle analisi compiute da Floros e Vougas (2007) e Chen, Firth e Rui (2002) che, per realizzare i loro lavori, si sono basati sui modelli GARCH ed EGARCH inserendo il volume in questi modelli.

A questo punto non ci resta che studiare la relazione tra rendimenti e volumi nel nostro caso, e questo verrà trattato nel capitolo 5.

Capitolo 3

Dati ad alta frequenza

Fino a poco tempo fa i dati finanziari che venivano utilizzati per lo studio dei mercati erano, principalmente, di tipo giornaliero. Per questo tipo di dati si consideravano i volumi, i prezzi di apertura, massimo, minimo e di chiusura e quest'ultimo era considerato come il prezzo di giornata. In questo modo, però, venivano persi tutti i movimenti intragiornalieri dei prezzi e dei volumi, rischiando di perdere informazioni utili per le stime successive e non riuscendo a percepire effettivamente tutti i movimenti che avevano portato al prezzo di chiusura della giornata di contrattazioni.

Negli ultimi anni, grazie anche ai miglioramenti tecnologici, si è sviluppato lo studio dei *dati ad alta frequenza*. Questo tipo di dati registra ogni singola transazione che avviene sul mercato in modo da minimizzare la perdita di informazioni dovuta alla rilevazione dei dati ad intervalli equispaziati e non molto stretti. In letteratura, questo tipo di dati viene definito *tick-by-tick*, dove per *tick* si intende la singola transazione. La distanza che intercorre tra un tick e l'altro viene definita *duration*.

Questi dati sono raccolti e divulgati, nel caso italiano, dalla Borsa Italiana S.p.A. e vengono forniti in forma di database in formato ASCII che riporta, nell'ordine:

- data della transazione nel formato aaaammgg;
- orario della transazione nel formato hhmmsscc;

- prezzo in euro al quale avviene lo scambio;
- volume scambiato;
- controvalore in euro.

3.1 Caratteristiche dei dati

Con l'uso dei dati ad alta frequenza si devono analizzare grandi quantità di dati rilevati a frequenze altissime. Per analizzare questo tipo di dati si devono tenere in considerazione alcuni aspetti fondamentali che non sono presenti nei dati a frequenze minori.

Dati non equispaziati

I dati ad alta frequenza non sono equispaziati perché registrano ogni singola transazione. Molti modelli econometrici si basano su dati equispaziati e sarà quindi necessario modificare i dati di partenza al fine di poter utilizzare questo tipo di modelli.

Per rendere i dati equispaziati si possono usare vari metodi:

- media aritmetica delle transazioni nell'intervallo di tempo desiderato;
- media ponderata con gli intervalli di tempo che intercorrono tra le transazioni e il termine temporale desiderato;
- interpolazione lineare.

La media ponderata è il metodo che usa il maggior numero di fattori rispetto agli altri metodi ma questo non permette di dire che sia il metodo migliore in termini di risultato. Bisogna comunque ricordare che l'accorpamento dei dati comporta, per quanto alta possa essere la frequenza di campionamento, una perdita di informazione.

Periodicità giornaliera

I dati finanziari intragiornalieri contengono delle periodicità. Solitamente la volatilità è maggiore poco dopo l'apertura del mercato e poco prima della sua chiusura. Anche le transazioni tendono ad avere una certa periodicità, infatti sono maggiori negli istanti vicini all'apertura e alla chiusura del mercato.

Discrezionalità dei prezzi

Le variazioni di prezzo tra una transazione e l'altra sono espresse in multipli di *tick*. Quindi nei dati ad alta frequenza il prezzo è una variabile discreta. Fino al 24 giugno 1997, nel New York Stock Exchange, un tick era un ottavo di dollaro per poi diventare un sedicesimo fino al 29 gennaio 2001. Da quel momento un tick è stato indicato in decimali.

Transazioni multiple in un singolo secondo

Si possono presentare dei casi in cui in un secondo ci siano più transazioni a prezzi diversi. Questo può essere dovuto al fatto che l'intervallo di un secondo potrebbe essere troppo ampio per questo tipo di dati e si dovrebbero considerare intervalli di tempo inferiori. ovviamente rilevare i dati a frequenze inferiori al secondo risulta essere molto difficile, ma bisogna comunque tenere in considerazione questo aspetto.

3.2 Non sincronicità negli scambi

Il mercato presenta molteplici titoli finanziari i quali presentano delle frequenze di trading diverse e che possono cambiare, per ogni singolo titolo, di ora in ora. Solitamente, però, le analisi delle serie dei prezzi e dei rendimenti avvengono a frequenze fisse. Per le analisi di dati giornalieri si considera come prezzo di riferimento del titolo il prezzo di chiusura, che è l'ultimo prezzo della giornata. In questo modo si assume, erroneamente, che la serie dei rendimenti sia una serie equispaziata con un intervallo di 24 ore. Per i dati giornalieri la non sincronicità può introdurre correlazione tra i rendimenti al

primo ritardo, correlazione tra i rendimenti del portafoglio al primo ritardo e, in alcuni casi, correlazione seriale negativa della serie dei rendimenti di ogni singolo titolo finanziario. Molti sono stati gli studi su questo fenomeno. Andiamo ad analizzare la versione del modello proposta da Lo e MacKinlay (1990).

Sia r_t una sequenza di rendimenti indipendenti e identicamente distribuiti con media μ e varianza σ_t^2 . La probabilità, costate nel tempo, che il titolo non venga scambiato sia π . Sia r_t^0 il rendimento osservato, che assume valore $r_t^0 = 0$ se al tempo t non ci dovessero essere scambi. Se, invece, al tempo t dovessero esserci degli scambi r_t^0 sarà la somma di tutti i rendimenti precedenti. La relazione tra r_t e r_t^0 sarà quindi:

$$r_t^0 = \begin{cases} 0 & \text{con } p = \pi \\ r_t & \text{con } p = (1 - \pi)^2 \\ r_t + r_{t-1} & \text{con } p = (1 - \pi)^2 \pi \\ \vdots & \\ \sum_{i=0}^k r_{t-i} & \text{con } p = (1 - \pi)^2 \pi^{k-1} \\ \vdots & \end{cases} \quad (3.1)$$

Le probabilità sono di facile intuizione. Infatti, per ottenere $r_t^0 = r_t$ dovranno esserci degli scambi sia al tempo t sia al tempo $t-1$. Lo stesso ragionamento si deve attuare in tutti gli altri casi.

Andiamo ora ad analizzare i momenti della serie dei rendimenti osservati r_t^0 . Il valore atteso della serie sarà dato da:

$$\begin{aligned} E(r_t^0) &= (1 - \pi)^2 E(r_t) + (1 - \pi)^2 \pi E(r_t + r_{t-1}) + \dots \\ &= (1 - \pi)^2 \mu + (1 - \pi)^2 \pi 2\mu + (1 - \pi)^2 \pi^3 \mu + \dots \\ &= (1 - \pi)^2 \mu [1 + 2\pi + 3\pi^2 + 4\pi^3 + \dots] \\ &= (1 - \pi)^2 \mu \frac{1}{(1 - \pi)^2} = \mu \end{aligned} \quad (3.2)$$

Per il calcolo della varianza della serie dei rendimenti osservati useremo $Var(r_t^0) = E[(r_t^0)^2] - [E(r_t^0)]^2$. Quindi:

$$\begin{aligned}
E(r_t^0)^2 &= (1 - \pi)^2 E[(r_t)^2] + (1 - \pi)^2 \pi E[(r_t + r_{t-1})^2] + \dots \\
&= (1 - \pi)^2 [(\sigma^2 + \mu^2) + \pi(2\sigma^2 + 4\mu^2) + \mu^2(3\sigma^2 + p\mu^2) + \dots] \\
&= (1 - \pi)^2 \sigma^2 [1 + 2\pi + 3\pi^2 + \dots] + \mu^2 [1 + 4\pi + p\pi^2 + \dots] \\
&= \sigma^2 + \mu^2 \left[\frac{2}{1 - \pi} - 1 \right]
\end{aligned} \tag{3.3}$$

La varianza sarà data da:

$$\begin{aligned}
Var(r_t^0) &= \sigma^2 + \mu^2 \left[\frac{2}{1 - \pi} - 1 \right] - \mu^2 \\
&= \sigma^2 + \frac{2\pi\mu^2}{1 - \pi}
\end{aligned} \tag{3.4}$$

Andiamo ora a considerare l'autocovarianza al primo ritardo della serie dei rendimenti osservati r_t^0 . Useremo $Cov(r_t, r_t^0) = E(r_t^0 r_{t-1}^0) - E(r_t^0)E(r_{t-1}^0)$. In questo modo sarà necessario calcolare solamente $E(r_t^0 r_{t-1}^0)$ nel modo seguente:

$$r_t^0 r_{t-1}^0 = \begin{cases} 0 & \text{con } p = 2\pi - \pi^2 \\ r_t r_{t-1} & \text{con } p = (1 - \pi)^3 \\ r_t(r_{t-1} + r_{t-2}) & \text{con } p = (1 - \pi)^3 \pi \\ \vdots & \\ r_t(\sum_{i=1}^k r_{t-i}) & \text{con } p = (1 - \pi)^3 \pi^{k-1} \\ \vdots & \end{cases} \tag{3.5}$$

Sfruttando $E(r_t r_{t-j}) = E(r_t)E(r_{t-j}) = \mu^2$ andiamo a calcolare:

$$\begin{aligned}
E(r_t^0 r_{t-1}^0) &= (1 - \pi)^3 E(r_t r_{t-1}) + \pi E[r_t(r_{t-1} + r_{t-2})] + \pi^2 E\left[r_t \left(\sum_{i=1}^3 r_{t-i}\right)\right] + \dots \\
&= (1 - \pi)^3 \mu^2 [1 + 2\pi + 3\pi^2 + \dots] \\
&= (1 - \pi) \mu^2
\end{aligned} \tag{3.6}$$

Quindi, l'autocovarianza al primo ritardo della serie dei rendimenti osservati è:

$$Cov(r_t^0 r_{t-1}^0) = -\pi\mu^2 \quad (3.7)$$

Se $\mu \neq 0$ si avrà una autocorrelazione negativa al primo ritardo nella serie r_t^0 data da:

$$\rho(r_t^0) = \frac{-(1-\pi)\pi\mu^2}{(1-\pi)\sigma^2 + 2\pi\mu^2} \quad (3.8)$$

L'entità dell'autocorrelazione al primo ritardo dipende dalla scelta di μ , π e σ . Comunque se $\mu \neq 0$ la non simultaneità degli scambi provoca un'autocorrelazione negativa. I risultati ottenuti possono essere generalizzati per un portafoglio di N titoli.

Capitolo 4

Modelli GARCH

Introdotti nel 1986 da Bollerslev, i modelli GARCH(p,q) (*Generalized Autoregressive Conditional Heteroskedasticity*) rappresentano una generalizzazione dei modelli ARCH(p) (*AutoRegressive Conditional Heteroskedasticity*) introdotti da Engle nel 1982. Già nei modelli ARCH(p) veniva introdotta una differenza tra varianza non condizionata e condizionata, permettendo a quest'ultima di cambiare nel tempo in funzione degli errori passati. Bollerslev decise di introdurre nell'equazione della varianza condizionata dei termini ritardati della stessa al fine di ridurre il numero di parametri da stimare rispetto al modello introdotto precedentemente da Engle.

4.1 Il modello ARCH(p)

Nello studio compiuto da Engle nel 1982 sull'inflazione del Regno Unito, viene sfruttata la presenza di autocorrelazione in trasformazioni positive dei rendimenti e viene suggerito che l'andamento della varianza del processo generatore dei dati sia autoregressivo condizionato.

Secondo Engle le innovazioni seguono il seguente processo:

$$\varepsilon_t = \eta_t \sigma_t \tag{4.1}$$

dove σ_t indica la radice quadrata varianza condizionata e

$$\eta_t | I_{t-1} \sim N(0, 1) \quad (4.2)$$

quindi η_t si distribuisce come una Normale standardizzata condizionata all'insieme informativo disponibile al tempo $t-1$. Questo implica che la distribuzione di ε_t condizionatamente all'insieme informativo I_{t-1} sia:

$$\varepsilon_t | I_{t-1} \sim N(0, \sigma_t^2) \quad (4.3)$$

Formalmente, la varianza condizionata dei rendimenti sarà:

$$\sigma_t = \text{Var}(r_t | I_{t-1}) = E(\varepsilon_t^2 | I_{t-1}) \quad (4.4)$$

Quindi, le equazioni di media e varianza condizionata per il modello ARCH(p) sono:

$$r_t = \mu + \varepsilon_t = \mu + \eta_t \sigma_t \quad (4.5)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 \quad (4.6)$$

Come si vede dall'equazione (3.6) la varianza condizionata di questo modello è espressa come funzione lineare dei quadrati dei valori passati delle innovazioni e non dipende in alcun modo dai valori passati della varianza condizionata stessa. Questo modello presenta delle condizioni che devono essere rispettate affinché la varianza condizionata risulti non negativa. Le tre restrizioni al modello sono:

$$\omega > 0 \quad (4.7)$$

$$\alpha_i \geq 0 \quad i = 1, \dots, p \quad (4.8)$$

$$\sum_{i=1}^p \alpha_i < 1 \quad (4.9)$$

Per capire la relazione tra varianza condizionata e non condizionata consideriamo:

$$\sigma^2 = E[\varepsilon_t^2] = E[\text{Var}(\varepsilon_t | I_{t-1})] = E[\sigma_t^2] \quad (4.10)$$

dove σ^2 indica la varianza non condizionata.

Tenendo conto dell'espressione (3.10) si ha che

$$E[\varepsilon_t^2] = E[\sigma_t^2] = E[\omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2] = \omega + \sum_{i=1}^p \alpha_i E[\varepsilon_{t-1}^2] \quad (4.11)$$

Data la stazionarietà delle innovazioni ε_t per la quale

$$E[\varepsilon_t^2] = E[\varepsilon_{t-i}^2] = \sigma^2 \quad (4.12)$$

otteniamo

$$\sigma^2 = \omega + \sigma^2 \sum_{i=1}^p \alpha_i \Leftrightarrow \sigma^2 = \frac{\omega}{1 - \sum_{i=1}^p \alpha_i} \quad (4.13)$$

che rappresenta l'equazione della varianza non condizionata.

Osservando la struttura del modello si vede come valori alti degli shock al quadrato ε_{t-1}^2 implicino valori alti della varianza condizionata σ_t^2 . Quindi per questo modello la probabilità che si realizzi uno shock importante, dopo che ce ne sia già stato uno, è maggiore rispetto a quella che si verifichi uno shock di minore entità.

4.2 Stima del modello ARCH(p)

Il modello ARCH(p) può essere stimato attraverso la funzione di verosimiglianza.

Questa funzione, per una campione y_1, y_2, \dots, y_n , coincide con la funzione di probabilità congiunta

$$L(\theta) = f(y_1, y_2, \dots, y_n | \theta). \quad (4.14)$$

dove θ rappresenta il vettore dei parametri da stimare. Se le y_i sono tra loro indipendenti la funzione di verosimiglianza è il prodotto di tutte le funzioni di densità di probabilità:

$$\begin{aligned}
L(\theta) &= f(y_1, y_2, \dots, y_n | \theta) \\
&= f(y_1 | \theta) f(y_2 | \theta) \dots f(y_n | \theta) \\
&= \prod_{t=1}^n f(y_t | \theta)
\end{aligned} \tag{4.15}$$

Se le y_i non sono tra loro indipendenti, la densità congiunta dell'intero campione è data dal prodotto di una funzione di densità di probabilità di una variabile casuale in un certo istante condizionata ai valori assunti delle variabili casuali precedenti e della funzione di probabilità congiunta di un certo numero p di variabili casuali riferite ai periodi iniziali. Quindi

$$\begin{aligned}
f(y_1, y_2, \dots, y_n | \theta) &= f(y_n | y_{n-1} \dots y_1 | \theta) \cdot \\
&\quad \cdot f(y_{n-1} | y_{n-2}, \dots, y_1 | \theta) \cdot \dots \cdot f(y_{p+1} | y_p, \dots, y_1 | \theta) \cdot \\
&\quad \cdot f(y_p, y_{p-1}, \dots, y_1; \theta)
\end{aligned} \tag{4.16}$$

La funzione di verosimiglianza può essere allora espressa come

$$\begin{aligned}
L(\theta) &= f(y_1, y_2, \dots, y_n | \theta) \\
&= f(y_p, y_{p-1}, \dots, y_1; \theta) \prod_{t=p+1}^n f(y_t | I_{t-1}; \theta)
\end{aligned} \tag{4.17}$$

e, di conseguenza, la funzione di log-verosimiglianza è

$$l(\theta) = \log f(y_p, y_{p-1}, \dots, y_1; \theta) + \sum_{t=p+1}^n \log f(y_t | I_{t-1}; \theta) \tag{4.18}$$

Nel nostro modello, avendo assunto che $\eta_t | I_{t-1}$ segua una distribuzione Normale, la funzione di verosimiglianza per un generico modello ARCH(p) è

$$\begin{aligned}
f(\varepsilon_1, \dots, \varepsilon_n | \omega, \alpha_1, \dots, \alpha_p) &= \\
&= f(\varepsilon_n | I_{t-1}) f(\varepsilon_{n-1} | I_{n-2}) \dots f(\varepsilon_{p+1} | I_p) f(\varepsilon_1, \dots, \varepsilon_n | \omega, \alpha_1, \dots, \alpha_p) \\
&= \prod_{t=p+1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{\varepsilon_t^2}{2\sigma_t^2}\right] f(\varepsilon_1, \dots, \varepsilon_n | \omega, \alpha_1, \dots, \alpha_p) \quad (4.19)
\end{aligned}$$

dove $f(\varepsilon_1, \dots, \varepsilon_n | \omega, \alpha_1, \dots, \alpha_p)$ è la funzione di probabilità congiunta di $\varepsilon_1, \dots, \varepsilon_p$. Quando la numerosità del campione è sufficientemente elevata si può tralasciare la funzione di probabilità congiunta nel calcolo della funzione di verosimiglianza, vista anche la sua complessità. Quindi la funzione di verosimiglianza risultante è:

$$L(\theta) = \prod_{t=p+1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{\varepsilon_t^2}{2\sigma_t^2}\right] \quad (4.20)$$

dove $\theta = (\omega, \alpha_1, \dots, \alpha_p)$ che rappresenta il vettore dei parametri da stimare. Per ottenere il valore dei parametri sarà sufficiente massimizzare la funzione appena specificata. La massimizzazione della funzione di verosimiglianza coincide con la massimizzazione della funzione di log-verosimiglianza, la cui espressione è:

$$l(\theta) = - \sum_{t=p+1}^n \frac{1}{2} \ln(2\pi) - \sum_{t=p+1}^n \frac{1}{2} \ln(\sigma_t^2) - \sum_{t=p+1}^n \frac{1}{2} \left(\frac{\varepsilon_t^2}{\sigma_t^2}\right) \quad (4.21)$$

Per massimizzare questa funzione sarà necessario porre a zero tutte le derivate prime della funzione di log-verosimiglianza calcolate rispetto ai parametri del vettore θ .

In alcuni casi potrebbe essere più appropriato assumere che $\eta_t | I_{t-1}$ abbia una distribuzione con code più pesanti rispetto alla Normale e, quindi, potrebbe avere distribuzione t di Student con v gradi di libertà. In questo caso la funzione di densità di probabilità di η_t è

$$f(\eta_t | v) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2) \sqrt{(v-2)\pi}} \left(1 + \frac{\eta_t^2}{v-2}\right)^{-[(v-2)/2]} \quad v > 2 \quad (4.22)$$

dove $\Gamma(x)$ rappresenta la funzione Gamma definita come

$$\Gamma(x) = \int_0^{+\infty} y^{x-1} e^{-y} dy \quad (4.23)$$

La funzione di verosimiglianza in questo caso è

$$L(\theta) = \prod_{t=p+1}^n \frac{\Gamma((v+1)/2)}{\Gamma(v/2) \sqrt{(v-2)\pi} \sigma_t} \frac{1}{\sigma_t} \left[1 + \frac{\varepsilon_t^2}{(v-2)\sigma_t^2} \right] \quad (4.24)$$

e quella di log-verosimiglianza

$$l(\theta) = - \sum_{t=p+1}^n \left[\frac{v+1}{2} \ln \left(1 + \frac{\varepsilon_t^2}{(v-2)\sigma_t^2} \right) + \frac{1}{2} \ln(\sigma_t^2) \right] \quad (4.25)$$

Queste funzioni possono essere usate quando i gradi di libertà v sono determinati a priori e non devono essere stimati. In caso contrario i gradi di libertà dovranno essere stimati congiuntamente agli altri parametri e la funzione di log-verosimiglianza da massimizzare diventa:

$$\begin{aligned} l(\theta, v) = & (n-p+1)[\ln(\Gamma((v+1)/2)) - \\ & - \ln(\Gamma(v/2)) - 0.5 \ln((v-2)\pi)] - \\ & - \sum_{t=p+1}^n \left[\frac{v+1}{2} \ln \left(1 + \frac{\varepsilon_t^2}{(v-2)\sigma_t^2} \right) + \frac{1}{2} \ln(\sigma_t^2) \right] \quad (4.26) \end{aligned}$$

4.3 Il modello GARCH(p,q)

Il modello GARCH(p,q) rappresenta un'evoluzione del modello ARCH(p) precedentemente analizzato. I due modelli mantengono le stesse assunzioni per quanto riguarda le distribuzioni di $\varepsilon_t | I_{t-1}$ e di $\eta_t | I_{t-1}$ e rimane invariata anche l'equazione della media condizionata. Questo sta a significare che le innovazioni ε_t rappresentano una variabile casuale con varianza dipendente dal tempo. Quindi, per ogni istante di tempo t , si realizzeranno delle distribuzioni Normali che differiscono per la varianza, cioè per la concentrazione dei dati rispetto alla media. L'equazione che differisce è quella della varianza

condizionata. Infatti, nel modello GARCH(p,q) l'equazione della varianza condizionata al tempo t è una combinazione lineare di p ritardi dei residui al quadrato e di q ritardi della varianza condizionata stessa. Quindi, il modello esprime la varianza condizionata come

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad \varepsilon_t | I_{t-1} \sim N(0, \sigma_t^2) \quad (4.27)$$

L'equazione della varianza condizionata deve fornire dei valori positivi e le condizioni affinché questo sia verificato sono:

$$\omega > 0 \quad (4.28)$$

$$\alpha_i \geq 0 \quad i = 1, \dots, p \quad (4.29)$$

$$\beta_j \geq 0 \quad j = 1, \dots, q \quad (4.30)$$

La condizione sufficiente affinché il processo GARCH(p,q) sia stazionario è:

$$\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_j) < 1 \quad (4.31)$$

In questo modello gli α_i costituiscono i termini ARCH, mentre i β_j costituiscono i termini GARCH.

Nel modello GARCH(p,q), vista la struttura della varianza condizionata, assume che valori elevati degli shock passati ε_{t-1} e delle varianze condizionate passate σ_{t-1}^2 producano valori elevati della varianza condizionata σ_t^2 . Il modello implica, inoltre, che un periodo di alta volatilità sia seguito da un altro periodo di alta volatilità con una probabilità maggiore rispetto ad essere seguito da un periodo di volatilità limitata. In molti casi è preferibile usare un modello GARCH(p,q) rispetto a un modello ARCH(p) perchè quest'ultimo spesso richiede un numero di parametri abbastanza elevato per riuscire a stimare delle serie storiche finanziarie, mentre il modello GARCH(p,q) riesce a modellare la stessa serie storica finanziaria con un minor numero di parametri.

4.4 Stima del modello GARCH(p,q)

Come per il modello ARCH(p), anche per il modello GARCH(p,q) viene utilizzata la funzione di verosimiglianza per ottenere la stima dei parametri. Quindi, sotto l'assunzione di normalità della distribuzione condizionata, la funzione di verosimiglianza è

$$L(\theta|\varepsilon) = f(\varepsilon_1, \dots, \varepsilon_p) \prod_{t=p+1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \quad (4.32)$$

Per ottenere la stima dei parametri possiamo tralasciare i termini che non dipendono dai parametri da stimare contenuti nel vettore θ , quindi la funzione di log-verosimiglianza da considerare è

$$l(\theta) = -\frac{1}{1} \sum_{t=p+1}^n \log\sigma_t^2 - \frac{1}{2} \sum_{t=p+1}^n \frac{\varepsilon_t^2}{\sigma_t^2} \quad (4.33)$$

Questa funzione deve essere massimizzata per ottenere la stima dei parametri e, quindi, devono essere poste a zero tutte le derivate prime calcolate rispetto ad ogni singolo parametro da stimare. In generale, quindi, dovrà essere posta a zero:

$$\frac{\partial l(\theta)}{\partial \theta} = -\frac{1}{1} \sum_{t=p+1}^n \log\sigma_t^2 \frac{\partial \sigma_t^2}{\partial \theta} - \frac{1}{2} \sum_{t=p+1}^n \frac{\varepsilon_t^2}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \quad (4.34)$$

Anche per il modello GARCH(p,q), come nel modello ARCH(p), è possibile assumere per η_t una distribuzione t di Student. In questo modello la funzione di log-verosimiglianza da massimizzare, dati a priori i gradi di libertà v , è

$$l(\theta) = - \sum_{t=p+1}^n \left[\frac{v+1}{2} \ln\left(1 + \frac{\varepsilon_t^2}{(v-2)\sigma_t^2}\right) + \frac{1}{2} \ln\sigma_t^2 \right] \quad (4.35)$$

Se, invece, i gradi di libertà vengono stimati congiuntamente agli altri parametri del modello la funzione di log-verosimiglianza diventa

$$l(\theta, v) = (n - p)[\ln(\Gamma[(v + 1)/2]) - \ln(\Gamma[v/2]) - 0.5\ln((v - 2)\pi)] - \\ - \sum_{t=p+1}^n \left[\frac{v + 1}{2} \ln\left(1 + \frac{\varepsilon_t^2}{(v - 2)\sigma_t^2}\right) + \frac{1}{2} \ln\sigma_t^2 \right] \quad (4.36)$$

4.5 Il modello EGARCH(p,q)

Nel modello GARCH(p,q) precedentemente analizzato si è visto come ci siano delle restrizioni sui parametri affinché la varianza non assuma valori negativi. Un modello che non prevede alcuna restrizione sui parametri è il modello EGARCH(p,q). La formulazione della varianza per questo modello è:

$$\log\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \sum_{j=1}^q \beta_j \log(\sigma_{t-1}^2) \quad (4.37)$$

Come si vede, non sono necessarie delle restrizioni sui parametri perchè per ottenere il valore di σ_t^2 è necessario applicare la funzione esponenziale alla precedente formule, funzione che non assume mai valori negativi. Oltre a questo aspetto il modello EGARCH(p,q) viene usato anche per stimare l'effetto asimmetrico delle innovazioni positive o negative sulla volatilità attraverso il parametro γ che, se diverso da zero, evidenzia un impatto asimmetrico delle innovazioni tenendo in considerazione l'effetto leva.

4.6 Modelli con variabili esogene

I modelli analizzati finora possono essere modificati inserendo delle variabili esogene. Aggiungendo queste variabili alla specificazione del modello è possibile analizzare il diretto impatto che queste hanno sui rendimenti e sulla varianza.

Nel nostro caso la variabile esogena di interesse sarà il volume. Infatti, nel capitolo 5, verranno analizzati dei modelli in cui verrà inserito il volume, e alcune sue trasformazioni, nell'equazione della media condizionata.

Per quanto riguarda il modello GARCH(p,q) l'equazione della media condizionata che si ottiene inserendo la variabile esogena è:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \delta x_t + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (4.38)$$

Se la variabile esogena viene inserita, invece, nel modello EGARCH(p,q) si ottiene:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \delta x_t + \sum_{j=1}^q \beta_j \log(\sigma_{t-1}^2) \quad (4.39)$$

Le variabili esogene che si possono inserire possono essere molteplici e riferite a diversi istanti temporali, questi parametri varieranno a seconda dell'analisi che si deve effettuare.

Capitolo 5

Applicazioni

Lo scopo di questo lavoro è capire la relazione esistente tra rendimenti e volumi dei titoli finanziari. L'analisi svolta si basa sulle serie storiche dei prezzi di chiusura e dei volumi di scambio del titolo IBM con frequenza di un minuto in un arco di tempo compreso tra il 02/01/2003 e il 30/03/2010. Quest'analisi, quindi, verrà condotta con dati ad alta frequenza ma equispaziati, al fine di evitare il problema della non sincronicità dei dati evidenziato nel capitolo 2.

Una prima analisi delle serie storiche ha portato ad eliminare i giorni che contenevano dei valori zero. Sono stati eliminati interamente i giorni e non solamente il dato a zero perchè era necessario mantenere la stessa quantità di dati per ogni giorno al fine di poter anche effettuare delle analisi a livello giornaliero. L'orario delle contrattazioni va dalle 9.30 alle 16.00, quindi, per ogni giorno la serie presenta 390 dati relativi a prezzi di chiusura e rendimenti.

Andremo ad effettuare un'analisi delle statistiche descrittive dei dati per poi passare ad analizzare i vari modelli utilizzati per capire la relazione esistente tra i rendimenti e i volumi.

5.1 Analisi dei dati

I dati che verranno utilizzati per l'analisi sono a frequenza di un minuto. Questi dati sono stati successivamente elaborati per formare delle serie storiche a frequenze di 5 minuti, 10 minuti, 15 minuti e 30 minuti. Per ottenere le serie così modificate è stata effettuata la media dei prezzi nell'arco di tempo indicato dalla frequenza della nuova serie, per poi andare a calcolare i rendimenti, mentre per i volumi è stata effettuata la somma. Lo scopo della creazione di serie storiche a frequenze diverse è capire se anche la frequenza dei dati influenza la relazione tra rendimenti e volumi. Andiamo ora ad analizzare le statistiche descrittive e i grafici delle serie storiche alle varie frequenze.

Dai grafici della serie dei rendimenti, riportati in Figura 5.1 si vede come questi risultino stazionari in media attorno al valore zero per le varie frequenze considerate. Questo risultato era sicuramente prevedibile e in accordo con quanto riportato in letteratura, in quanto i rendimenti positivi compensano i rendimenti negativi fornendo, quindi, un valore medio prossimo allo zero. Se si analizzano gli istogrammi dei rendimenti di Figura 5.2 si ha ulteriore conferma del fatto che abbiano una distribuzione centrata sul valore zero della loro media e abbastanza simmetrica rispetto a quest'ultima. Questo, però, non deve portare necessariamente alla conclusione che la distribuzione dei rendimenti sia di tipo Normale.

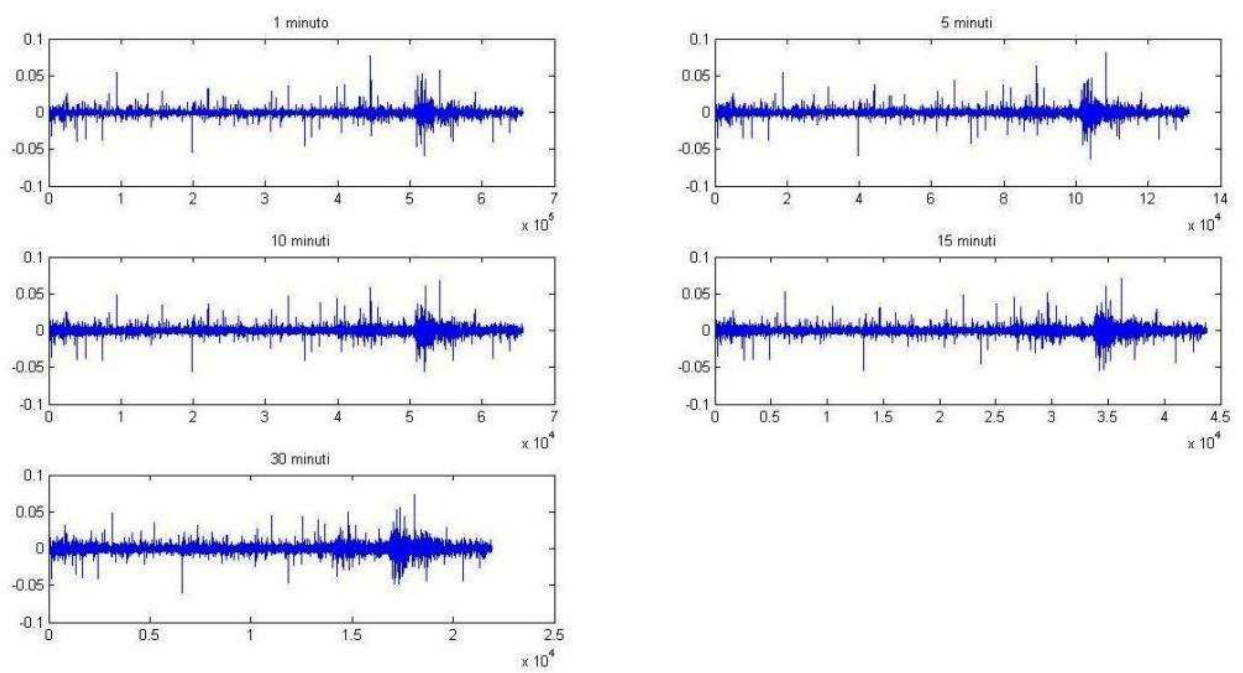


Figura 5.1: Grafici dei rendimenti a diverse frequenze per il titolo IBM dallo 02/01/2003 al 30/03/2010

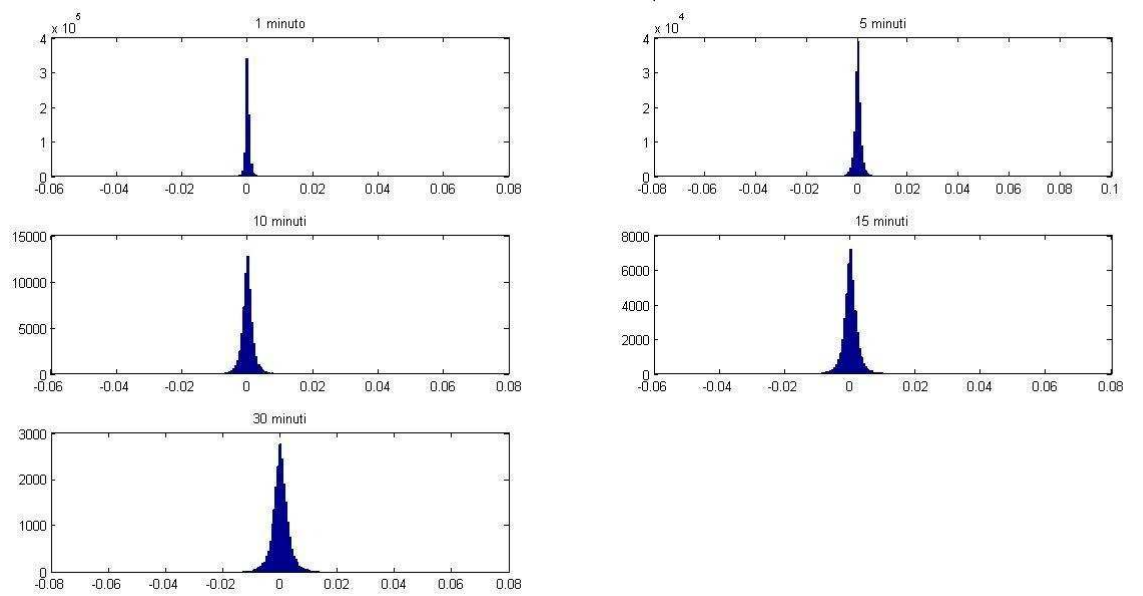


Figura 5.2: Istogrammi dei rendimenti a diverse frequenze per il titolo IBM dallo 02/01/2003 al 30/03/2010

Andiamo ora ad analizzare in dettaglio alcune statistiche descrittive per la serie storica dei rendimenti alle varie frequenze riportate in Tabella 5.1.

Analizzando i valori della media si vede come questa sia praticamente zero per tutte le frequenze considerate. Questo dato è pienamente in accordo con quanto visto precedentemente in entrambi i grafici. Anche il valore della deviazione standard non si discosta molto nelle varie frequenze e risulta essere non molto elevato. Analizzando anche gli istogrammi si vede che effettivamente i dati non si discostano molto dal valore medio in cui la distribuzione dei rendimenti è centrata. Per capire se i rendimenti hanno una distribuzione simmetrica si deve porre l'attenzione sul valore dell'indice di simmetria la cui formula di calcolo è:

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^2 \right)^{\frac{3}{2}}} \quad (5.1)$$

Se il valore di questo indice è zero significa che la distribuzione dei dati risulta simmetrica. Nel nostro caso l'indice di simmetria tende ad avvicinarsi al valore zero con l'aumento della frequenza dei dati. Questi primi risultati potrebbero far pensare che la distribuzione dei rendimenti segua una distribuzione Normale con media zero anche se il valore dell'indice di simmetria potrebbe dare qualche altra indicazione. Quindi, per capire se la distribuzione dei dati segue una distribuzione di tipo Normale, si deve andare ad analizzare il valore dell'indice di kurtosi così calcolato:

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^2 \right)^2} \quad (5.2)$$

Questo particolare indice segnala la presenza di dati con distribuzione Normale se assume un valore prossimo a 3. Nelle varie frequenze dei dati si vede come l'indice di kurtosi assuma valori molto distanti da 3. Il valore dell'indice diminuisce con l'aumentare della frequenza ma non si avvicina mai al valore soglia per la distribuzione Normale. Questo ci porta quindi ad escludere che i rendimenti seguano una distribuzione Normale. Infatti, se si analizzano

congiuntamente anche gli istogrammi si può notare come la distribuzione dei rendimenti presenti delle code più pesanti rispetto alla distribuzione Normale. Al fine di confermare quanto detto finora passiamo ad analizzare il test di Jarque-Bera. Questo test permette di valutare la normalità dei dati usando, congiuntamente, gli indici di asimmetria e kurtosi nel modo seguente:

$$JB = \frac{n}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right) \quad (5.3)$$

Questo test presenta una distribuzione chi-quadro con due gradi di libertà ed è usato per testare l'ipotesi nulla che il campione in analisi sia distribuito come una Normale. L'ipotesi nulla del test di Jarque-Bera è un'ipotesi congiunta che sia l'indice di asimmetria che l'indice di kurtosi siano zero e viene rigettata per valori grandi del test. Nella nostra analisi si vede come il test di Jarque-Bera assuma valori molto grandi e, quindi, viene rigettata sicuramente l'ipotesi nulla di normalità dei dati. Con questa prima analisi descrittiva abbiamo confermato, come riportato in letteratura, che i rendimenti presentano una distribuzione simmetrica e centrata sulla media, che è zero, ma questa distribuzione non è di tipo Normale.

freq	n° oss	media	std. dev.	simmetria	curtosi	JB test	prob
1 min	656760	7.53e-07	0.000864	1.082723	435.2712	5.11e+09	0.000000
5 min	131352	3.76e-06	0.001833	0.754269	127.4137	84727731	0.000000
10 min	65676	7.53e-06	0.002512	0.421017	62.14781	9575485	0.000000
15 min	43784	1.13e-05	0.003049	0.228207	48.54504	3784688	0.000000
30 min	21892	2.26e-05	0.004265	0.186531	28.66687	601051.4	0.000000

Tabella 5.1: Statistiche descrittive per le serie dei rendimenti alle frequenze di 1 minuto, 5 minuti, 10 minuti, 15 minuti e 30 minuti

Passiamo ora ad analizzare i volumi alle varie frequenze prese in esame con la Figura 5.3 e la Tabella 5.2.

Analizzando le serie storiche dei volumi si vede molto bene come questo tipo di dati presenti una grande variabilità mantenuta per tutte le frequenze considerate, variabilità confermata anche dai valori della deviazione standard.

Analizzando anche le statistiche descrittive si vede come i volumi non presentino caratteristiche di normalità e simmetria e quindi il test di Jarque-Bera rifiuta l'ipotesi nulla di normalità degli stessi.

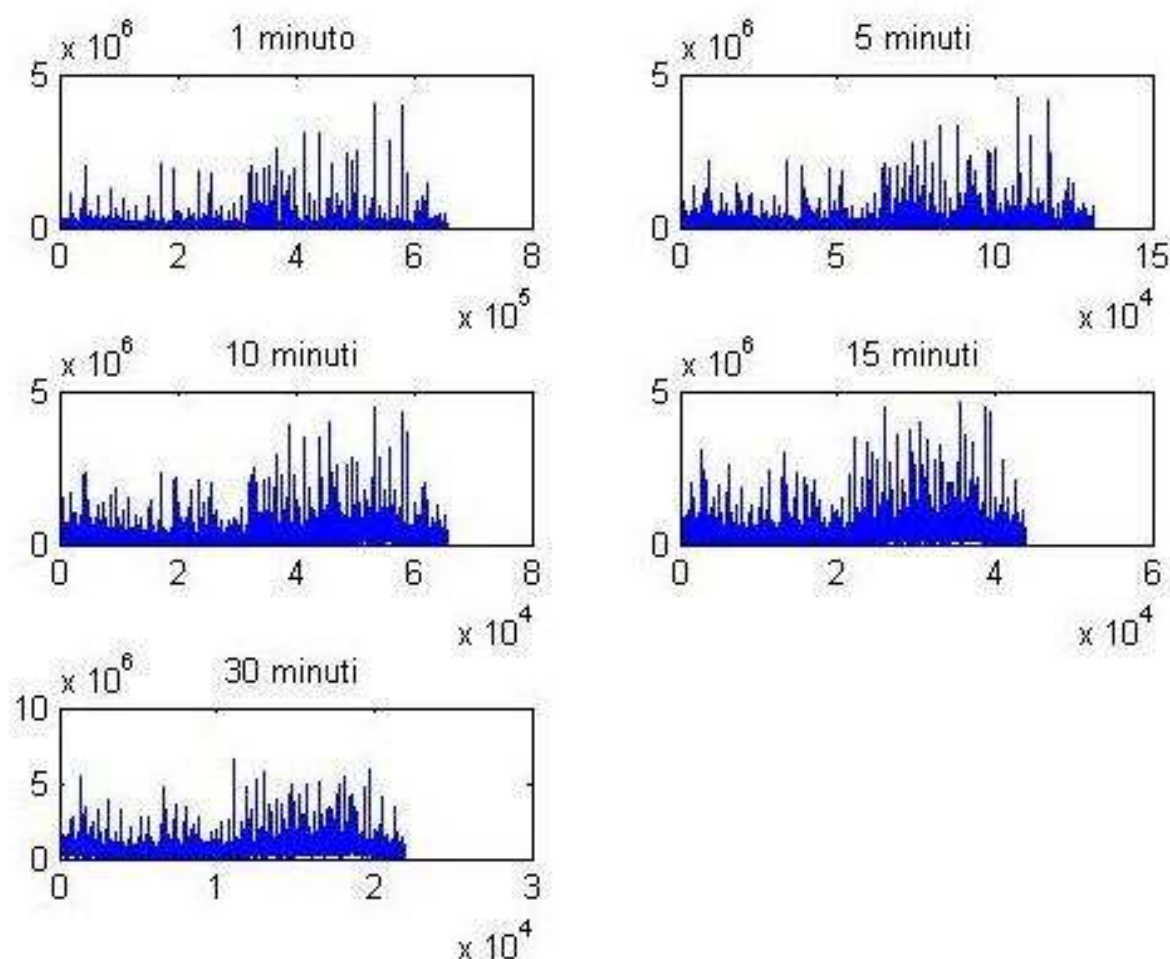


Figura 5.3: Grafici dei volumi a diverse frequenze per il titolo IBM dallo 02/01/2003 al 30/03/2010

Vista l'alta frequenza dei dati cerchiamo ora di capire se ci sono delle periodicità intragiornaliere. Per fare ciò sono state calcolate le medie e varianze

freq	n° oss	media	std. dev.	simmetria	curtosi	JB test	prob
1 min	656760	16970.42	24986.047	42.27661	4388.181	5.26e+11	0.000000
5 min	131352	84852.10	81245.43	9.647996	261.9331	3.69e+08	0.000000
10 min	65676	169704.2	144649.1	6.053980	91.82305	21990889	0.000000
15 min	43784	244556.3	204314.6	4.849400	54.39277	4990069	0.000000
30 min	21892	509112.6	373040.7	3.801805	30.00072	819907.2	0.000000

Tabella 5.2: Statistiche descrittive per le serie dei volumi alle frequenze di 1 minuto, 5 minuti, 10 minuti, 15 minuti e 30 minuti

giornaliere per ogni istante t della serie. In questo modo riusciremo a capire come si comportano le due grandezze considerate nell'arco della giornata di contrattazioni anche attraverso i grafici in Figura 5.4, Figura 5.5, Figura 5.6, Figura 5.7 e Figura 5.8.

I grafici delle medie evidenziano come anche, mediamente, in un giorno, i rendimenti siano distribuiti intorno allo zero. Molto interessante è analizzare i grafici delle varianze. Questi, infatti, mettono in luce una maggiore variabilità dei rendimenti all'apertura e alla chiusura della giornata di contrattazioni. Questo non deve sorprendere in quanto il mercato è molto attivo in questi due momenti della giornata, mentre nella parte centrale della giornata di contrattazioni i movimenti risultano essere minori.

Da questa analisi grafica si nota come rendimenti e volumi presentino le caratteristiche simili per quanto riguarda la periodicità intragiornaliera. Questo fatto sembrerebbe evidenziare la presenza di una certa relazione tra i rendimenti del titolo e i suoi volumi.

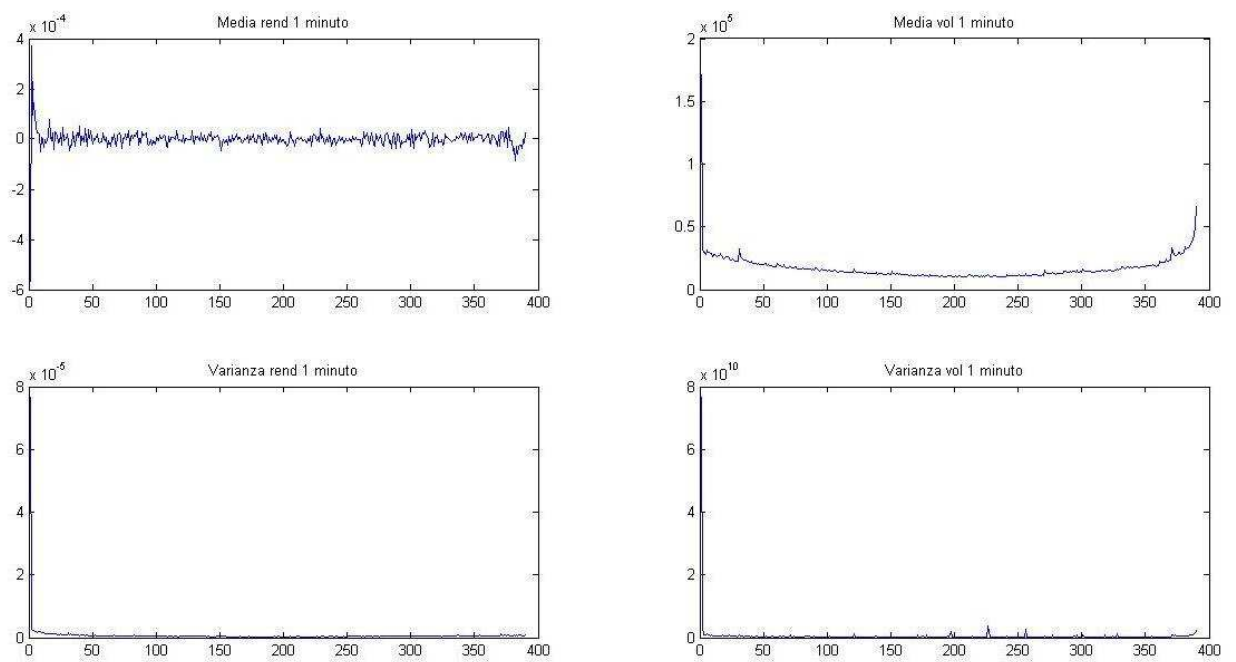


Figura 5.4: Medie e varianze giornaliere con frequenza a 1 minuto per rendimenti e volumi del il titolo IBM dallo 02/01/2003 al 30/03/2010

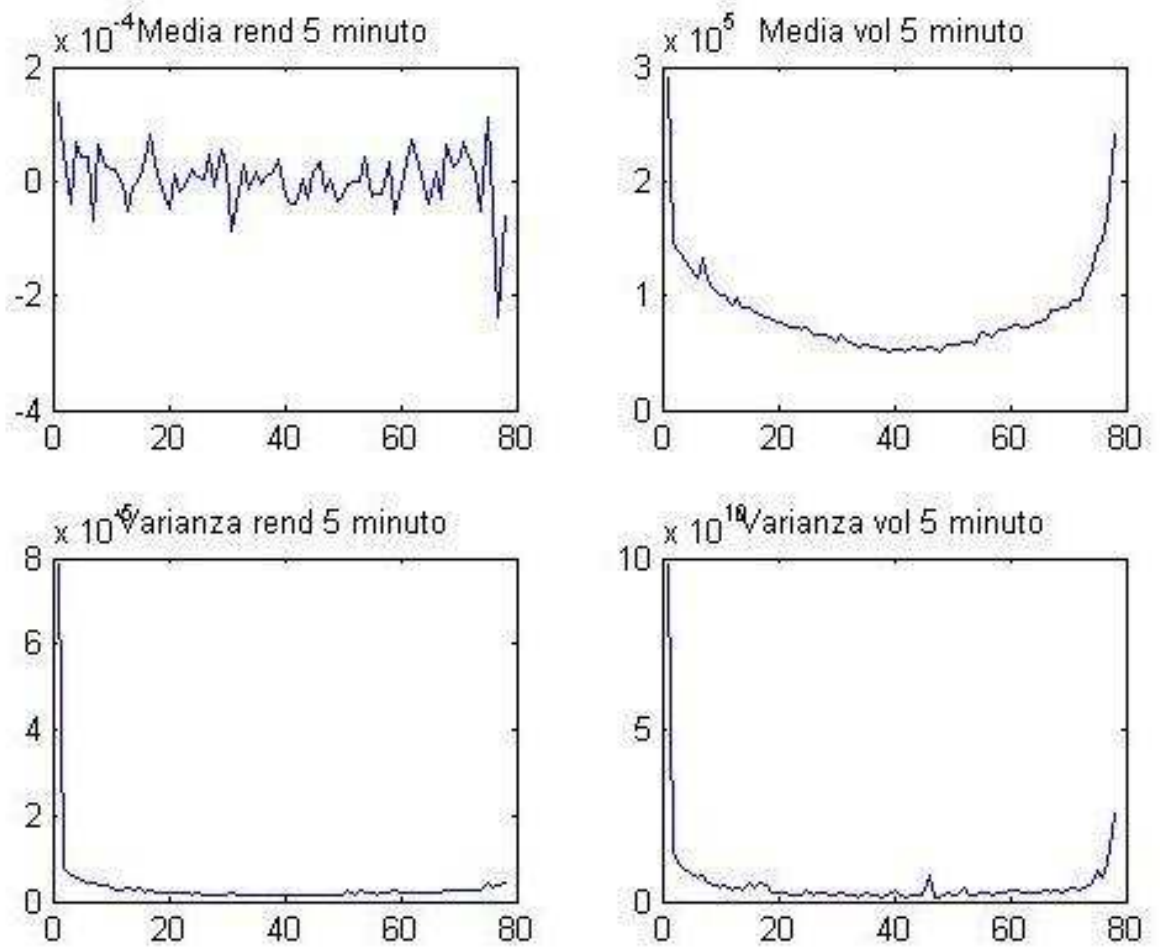


Figura 5.5: Medie e varianze giornaliere con frequenza a 5 minuti per rendimenti e volumi del titolo IBM dallo 02/01/2003 al 30/03/2010

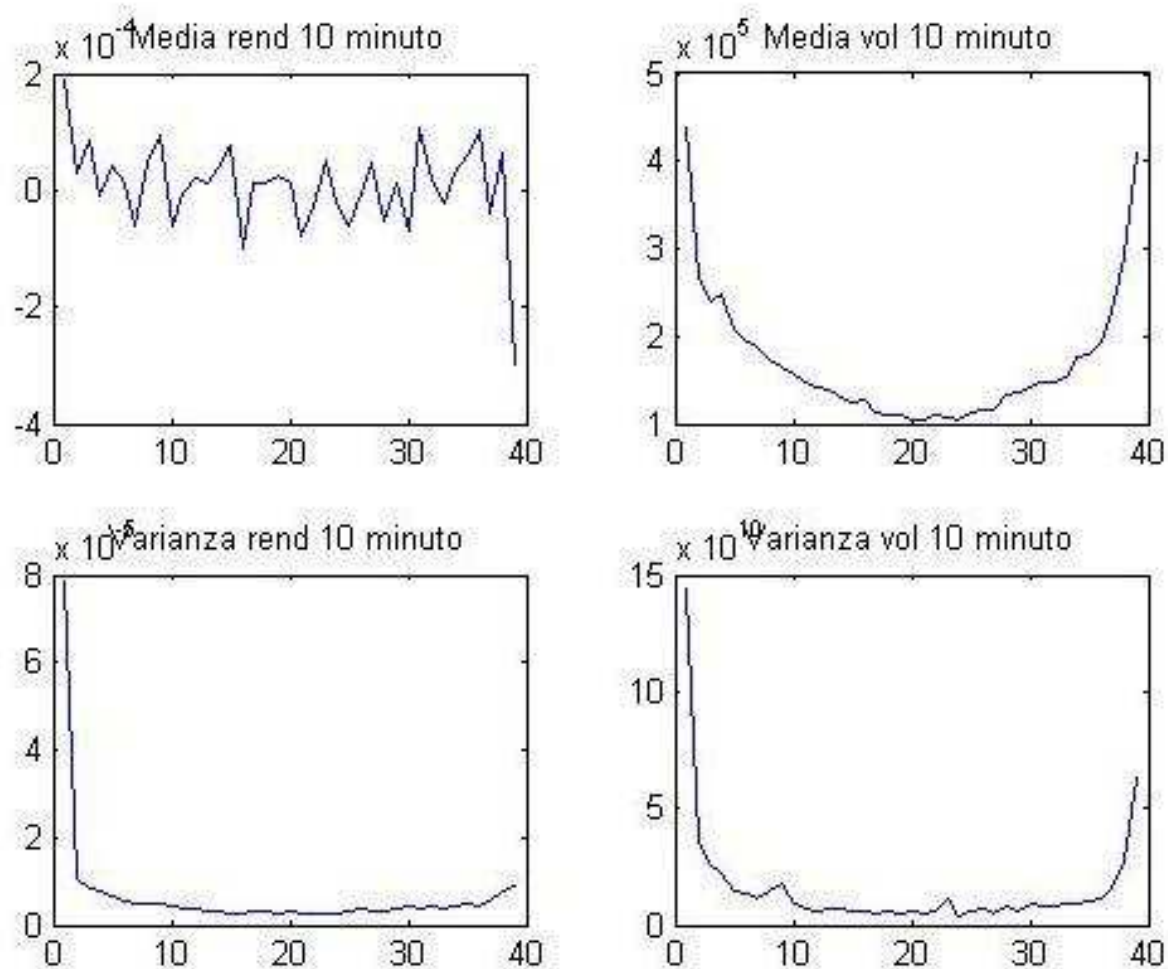


Figura 5.6: Medie e varianze giornaliere con frequenza a 10 minuto per rendimenti e volumi del il titolo IBM dallo 02/01/2003 al 30/03/2010

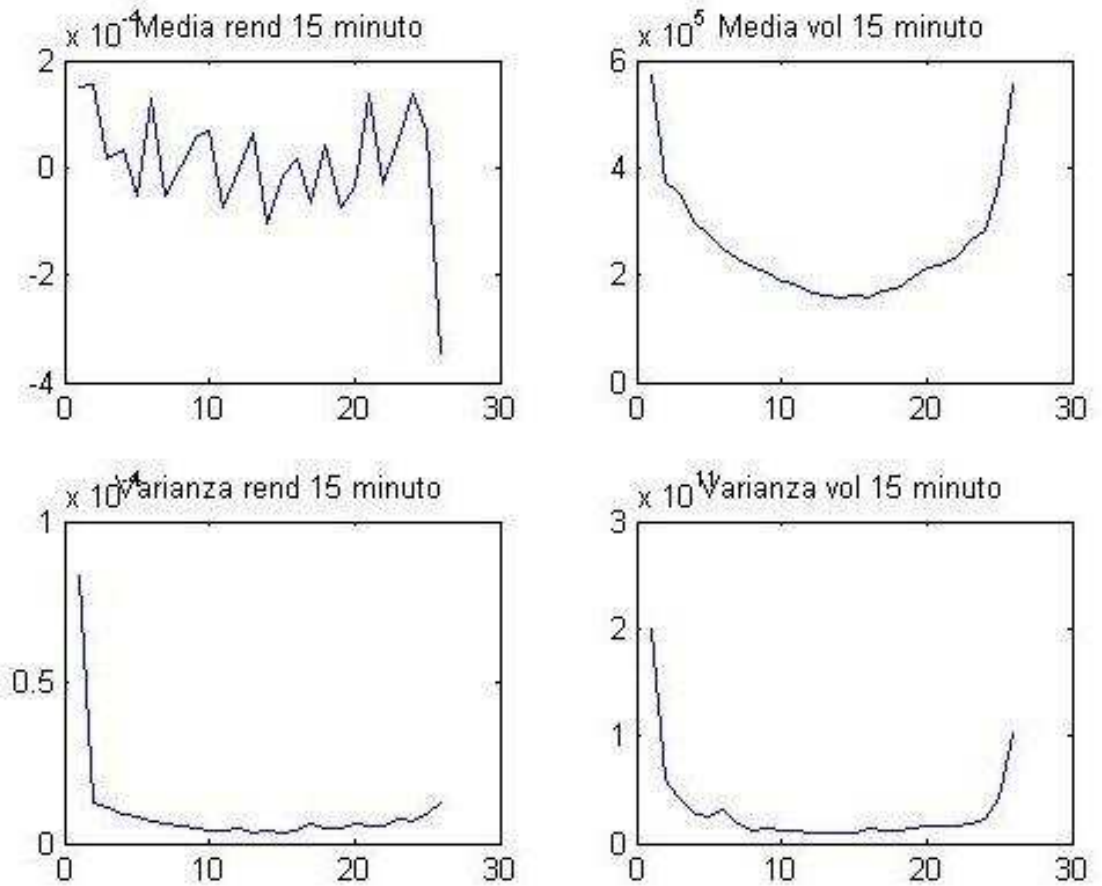


Figura 5.7: Medie e varianze giornaliere con frequenza a 15 minuto per rendimenti e volumi del il titolo IBM dallo 02/01/2003 al 30/03/2010

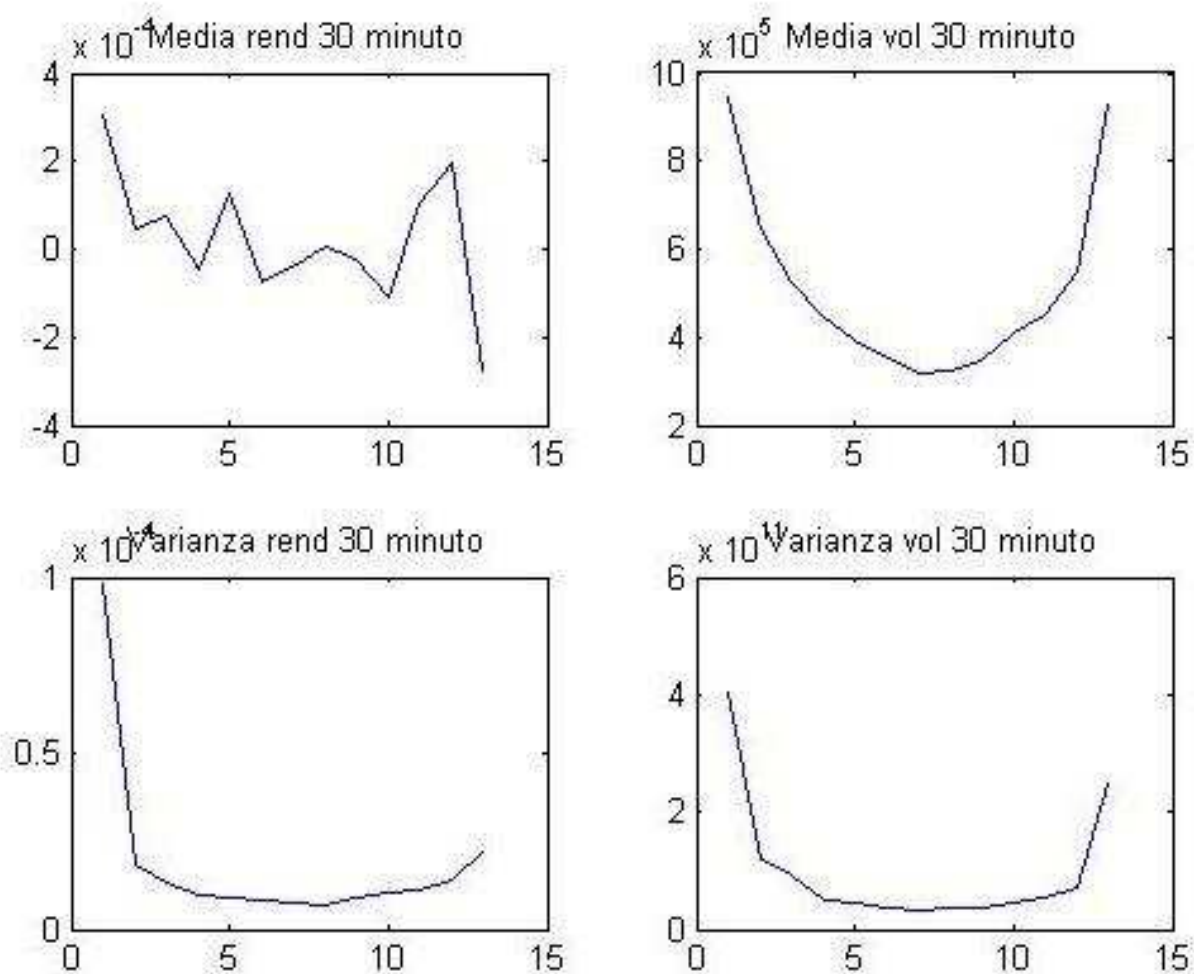


Figura 5.8: Medie e varianze giornaliere con frequenza a 30 minuto per rendimenti e volumi del il titolo IBM dallo 02/01/2003 al 30/03/2010

5.2 I modelli

Nella sezione precedente è stata evidenziata una possibile relazione tra i rendimenti del titolo IBM e i suoi volumi di scambio. In questa sezione cercheremo di approfondire questo aspetto. I modelli alla base dell'analisi sono i modelli GARCH(p,q) ed EGARCH(p,q) ai quali saranno aggiunte delle variabili esogene, in particolare saranno inseriti i volumi o delle funzioni dei volumi. Tutti i modelli usati sono stati applicati alle frequenze di 1 minuto, 5 minuti, 10 minuti, 15 minuti e 30 minuti. Nella sezione precedente oltre ad evidenziare una possibile relazione tra i rendimenti e i volumi, dai grafici si vede bene come ci sia la presenza di una certa periodicità dei dati. Infatti, analizzando i grafici delle varianze, si nota come le serie dei rendimenti e dei volumi presentino una maggiore variabilità all'apertura e alla chiusura del mercato. Nei modelli che verranno presentati di seguito sono state inserite anche delle dummy per ogni istante temporale della giornata. Questo è stato fatto per cercare di catturare questa periodicità. Per chiarezza di notazione non verranno però inserite nelle tabelle delle stime del modello in quanto le dummy inserite sono un numero abbastanza elevato, ma verranno semplicemente riportati i coefficienti dei parametri che forniscono informazioni sulla relazione tra rendimenti e volumi. Andiamo a vedere nel dettaglio i singoli modelli e i risultati ottenuti.

Il primo modello GARCH(1,1) analizzato contiene come esogena il volume:

$$r_t = \mu + \delta vol_t + \varepsilon_t \quad (5.4)$$

$$\sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha\varepsilon_{t-i}^2 + \beta\sigma_{t-j}^2 \quad (5.5)$$

dove $k=1, \dots, \text{numero di transazioni in un giorno}$

Il volume inserito in questo modello è contemporaneo ai rendimenti, quindi, con questo modello l'obiettivo è quello di capire se la relazione tra rendimenti e volumi è di tipo contemporaneo.

Le stime ottenute sono riportate in Tabella 5.3.

freq	par	value	std. error	t stat
1 min	δ	0.0007	1.885e-007	3591.044
	α	0.298	1.971e-005	15120.245
	β	0.699	5.221e-005	13383.830
5 min	δ	0.002	1.906e-006	779.091
	α	0.301	0.0007	432.863
	β	0.698	0.0004	1588.086
10 min	δ	0.002	5.621e-006	390.844
	α	0.299	0.001	210.919
	β	0.645	0.001	514.511
15 min	δ	0.003	60..1e-006	458.549
	α	0.411	0.003	149.718
	β	0.589	0.002	354.516
30 min	δ	0.004	3.681e-005	115.786
	α	0.028	0.0005	52.013
	β	0.954	0.0007	1311.306

Tabella 5.3: Stima del modello GARCH(1,1) con volume al tempo t

Il secondo modello GARCH(1,1) analizzato contiene come esogena il logaritmo del volume:

$$r_t = \mu + \delta \log(vol_t) + \varepsilon_t \quad (5.6)$$

$$\sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha \varepsilon_{t-i}^2 + \beta \sigma_{t-j}^2 \quad (5.7)$$

Anche in questo modello in cui viene considerato il logaritmo del volume di scambio viene comunque mantenuta la contemporaneità con i rendimenti.

Le stime ottenute sono riportate in Tabella 5.4.

freq	par	value	std. error	t stat
1 min	δ	0.0007	1.885e-007	3591.044
	α	0.298	1.971e-005	15120.245
	β	0.699	5.221e-005	13383.230
5 min	δ	0.002	1.906e-006	779.091
	α	0.301	0.0007	432.863
	β	0.698	0.0004	1588.086
10 min	δ	0.002	5.621e-006	390.844
	α	0.298	0.001	210.919
	β	0.645	0.001	514.511
15 min	δ	0.003	6.009e-006	458.549
	α	0.411	0.003	149.718
	β	0.589	0.002	354.546
30 min	δ	0.004	3.682e-005	115.786
	α	0.028	0.0005	52.013
	β	0.954	0.0007	1311.309

Tabella 5.4: Stima del modello GARCH(1,1) con il logaritmo del volume al tempo t

Il modello GARCH(1,1), in questo caso, contiene come esogena la differenza prima del logaritmo del volume:

$$r_t = \mu + \delta[\log(vol_t) - \log(vol_{t-1})] + \varepsilon_t \quad (5.8)$$

$$\sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha\varepsilon_{t-i}^2 + \beta\sigma_{t-j}^2 \quad (5.9)$$

Le stime ottenute sono riportate in Tabella 5.5.

freq	par	value	std. error	t stat
1 min	δ	-4.602e-005	3.67e-007	-125.367
	α	0.326	0.0002	1577.459
	β	0.653	0.0002	4407.472
5 min	δ	-0.0001	2.916e-006	-38.989
	α	0.276	0.0007	390.510
	β	0.724	0.0006	1164.184
10 min	δ	-8.8e-005	5.925e-006	-14.851
	α	0.338	0.002	170.167
	β	0.661	0.001	442.467
15 min	δ	-0.0001	1.281e-005	-9.390
	α	0.329	0.003	133.321
	β	0.616	0.002	285.002
30 min	δ	-0.001	4.428e-005	-2.426
	α	0.027	0.0006	47.3739
	β	0.931	0.0008	1258.016

Tabella 5.5: Stima del modello GARCH(1,1) con la differenza prima del logaritmo del volume al tempo t

Andiamo ora a valutare la differenza prima del volume ma per il periodo precedente rispetto ai rendimenti, quindi il modello diventa:

$$r_t = \mu + \delta[\log(vol_t - 1) - \log(vol_{t-2})] + \varepsilon_t \quad (5.10)$$

$$\sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha\varepsilon_{t-i}^2 + \beta\sigma_{t-j}^2 \quad (5.11)$$

Le stime ottenute sono riportate in Tabella 5.6.

freq	par	value	std. error	t stat
1 min	δ	-1.872e-005	3.803e-007	-49.229
	α	0.328	0.0002	1557.461
	β	0.652	0.0001	4895.988
5 min	δ	-1.2e-005	1.8544e-006	-6.471
	α	0.275	0.0007	369.154
	β	0.725	0.0006	1152.149
10 min	δ	-3.524e-006	6.047e-006	-0.583
	α	0.337	0.002	186.903
	β	0.663	0.001	464.151
15 min	δ	-5.371e-005	1.291e-005	-4.159
	α	0.329	0.003	132.991
	β	0.619	0.002	285.936
30 min	δ	-0.0002	4.626e-005	-3.685
	α	0.027	0.0006	47.315
	β	0.961	0.0008	1260.411

Tabella 5.6: Stima del modello GARCH(1,1) con la differenza prima del logaritmo del volume al tempo t-1

Finora l'esogena inserita nel modello era contemporanea con i rendimenti. In questo modello, invece, il volume inserito nell'equazione della media condizionata sarà ritardato di un periodo per cercare di capire se la relazione esistente tra rendimenti e volumi sia non solo contemporanea ma anche ritardata. Il modello sarà, quindi:

$$r_t = \mu + \delta vol_{t-1} + \varepsilon_t \quad (5.12)$$

$$\sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha \varepsilon_{t-i}^2 + \beta \sigma_{t-j}^2 \quad (5.13)$$

Le stime ottenute sono riportate in Tabella 5.7.

freq	par	value	std. error	t stat
1 min	δ	-4.602e-005	3.67e-007	-125.367
	α	0.356	0.0002	1577.459
	β	0.653	0.0002	4407.472
5 min	δ	-0.0001	2.916e-006	-38.989
	α	0.276	0.0007	390.510
	β	0.724	0.0006	1164.184
10 min	δ	-8.8e-005	5.926e-006	-14.8509
	α	0.339	0.002	170.167
	β	0.661	0.002	442.467
15 min	δ	-0.0001	1.281e-005	-9.390
	α	0.329	0.003	133.321
	β	0.619	0.002	285.002
30 min	δ	-0.0001	4.428e-005	-2.426
	α	0.027	0.0005	47.374
	β	0.961	0.0008	1258.017

Tabella 5.7: Stima del modello GARCH(1,1) con volume al tempo t-1

I modelli analizzati in questa prima parte sono tutti modelli GARCH(1,1) con, inserite nell'equazione della media condizionata, diverse esogene trasformazioni del volume di scambio o il volume di scambio stesso e delle dummy per ogni istante temporale per riuscire a catturare la periodicità dei dati. Affinchè il modello GARCH(1,1) sia stazionario la somma dei coefficienti α e β deve essere inferiore a 1. Se si osservano le stime dei modelli ottenuti si vede come in tutti i modelli e a tutte le frequenze la condizione sufficiente per la stazionarietà sia rispettata. Il parametro β corrisponde a una sintesi delle

informazioni passate. Nei modelli analizzati questo parametro assume un valore piuttosto elevato. Questo aspetto sta a significare una certa persistenza della varianza condizionata, quindi, periodi di grande variabilità saranno preceduti da periodi di grande variabilità e viceversa. Di conseguenza, visto che deve essere rispettata la condizione sufficiente di stazionarietà del modello, il parametro α non assume valori molto elevati. Ciò significa che i modelli non sono troppo sensibili alle news di mercato e la capacità di variazione nel tempo di σ_t^2 è limitata. Queste considerazioni sono valide per tutti i modelli in cui l'esogena, che sia il volume o una sua trasformazione, è contemporanea ai rendimenti del titolo.

Per capire la relazione tra rendimenti e volumi il parametro di interesse è δ . In tutti i modelli in cui il volume è contemporaneo ai rendimenti questo parametro assume valori positivi, evidenziando, quindi, una relazione positiva tra le due grandezze. Il valore del parametro non risulta molto elevato ma comunque significativamente diverso da zero, quindi la relazione esiste anche se risulta essere un po' debole. Quindi, un aumento del volume, in questi modelli, comporta un aumento della e del rendimento.

In tutti i modelli in cui compare il volume non contemporaneo ai rendimenti, sia nei casi in cui si calcola la differenza prima sia nel caso in cui è inserito semplicemente il volume ritardato di un periodo, il coefficiente δ risulta sempre negativo. Quindi, in questi casi un aumento di volume comporta una diminuzione della volatilità.

Come nello studio di Floros e Vougas (2007), che hanno inserito il logaritmo del volume nell'equazione della media condizionata, anche nel nostro caso si è riusciti a stabilire una relazione contemporanea positiva tra i rendimenti e i volumi. Anche l'analisi di Backlaci e Kasman sul mercato turco ha rilevato questa relazione ma, come nel nostro caso, è riuscita a dimostrare che la persistenza della varianza non diminuisce con l'inserimento del volume nel modello GARCH(1,1). Per quanto riguarda invece la relazione non contemporanea tra rendimenti e volumi i dati ad alta frequenza usati nel nostro studio non presentano le stesse caratteristiche dei dati del mercato pakistano

usati nell'analisi di Mubarick e Javid (2009). infatti, questi autori avevano riscontrato che anche nel caso di non contemporaneità di volumi e rendimenti la relazioni trovata risulta positiva, mentre nel nostro caso risulta negativa. Passiamo ora ad analizzare il modello EGARCH(1,1). In questo modello non ci sono particolari restrizioni sui parametri in quanto la varianza condizionata risulta sempre positiva vista l'applicazione della funzione esponenziale. Con questi modelli andremo a valutare non solo la relazione tra rendimenti e volumi ma, anche, l'impatto che il segno del rendimento ha sulla varianza condizionata attraverso il parametro γ

Il primo modello EGARCH(1,1) analizzato contiene come esogena il volume:

$$r_t = \mu + \delta vol_t + \varepsilon_t \quad (5.14)$$

$$\log \sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta \log(\sigma_{t-1}^2) \quad (5.15)$$

Le stime ottenute sono riportate in Tabella 5.8.

freq	par	value	std. error	t stat
1 min	δ	0.0006	1.3e-007	4805.7
	α	0.159	5.517e-005	2885.991
	β	0.982	1.279e-005	76813.947
	γ	-0.001	5.319e-005	-25.601
5 min	δ	0.001	1.181e-006	1171.727
	α	0.359	0.0006	574.964
	β	0.917	0.0002	4527.432
	γ	-0.065	0.0004	-157.987
10 min	δ	0.002	3.982e-006	526.806
	α	0.350	0.001	305.426
	β	0.884	0.0007	1359.304
	γ	-0.040	0.0008	-47.707
15 min	δ	0.003	7.678e-006	368.7026
	α	0.036	0.0003	136.629
	β	0.998	4.252e-005	23477.044
	γ	-0.012	0.0002	-60.416
30 min	δ	0.004	1.826e-005	224.419
	α	0.048	0.0007	68.343
	β	0.998	0.0001	9474.609
	γ	-0.015	0.0005	-32.651

Tabella 5.8: Stima del modello EGARCH(1,1) con volume al tempo t

Il secondo modello EGARCH(1,1) analizzato contiene come esogena il logaritmo del volume:

$$r_t = \mu + \delta \log(vol_t) + \varepsilon_t \quad (5.16)$$

$$\log \sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta \log(\sigma_{t-1}^2) \quad (5.17)$$

Le stime ottenute sono riportate in Tabella 5.9.

freq	par	value	std. error	t stat
1 min	δ	0.0006	1.30e-007	4805.70
	α	0.159	5.519e-005	2885.991
	β	0.982	1.279e-005	76813.947
	γ	-0.001	5.391e-005	-25.601
5 min	δ	0.001	1.181e-006	1171.727
	α	0.359	0.0006	574.964
	β	0.917	0.0002	4527.432
	γ	-0.065	0.0004	-157.987
10 min	δ	0.002	3.982e-006	526.806
	α	0.350	0.001	305.426
	β	0.884	0.0007	1359.3136
	γ	-0.040	0.0008	-47.7071
15 min	δ	0.003	7.678e-006	368.703
	α	0.036	0.0003	136.629
	β	0.998	4.252e-005	23477.044
	γ	-0.012	0.0002	-60.416
30 min	δ	0.004	1.826e-005	224.419
	α	0.048	0.0007	68.343
	β	0.998	0.0001	9474.609
	γ	-0.015	0.0004	-32.651

Tabella 5.9: Stima del modello EGARCH(1,1) con logaritmo del volume al tempo t

Il modello EGARCH(1,1) analizzato ora contiene come esogena la differenza prima del logaritmo dei volumi:

$$r_t = \mu + \delta[\log(vol_t) - \log(vol_{t-1})] + \varepsilon_t \quad (5.18)$$

$$\log\sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta \log(\sigma_{t-1}^2) \quad (5.19)$$

Le stime ottenute sono riportate in Tabella 5.10.

freq	par	value	std. error	t stat
1 min	δ	-5.111e-005	2.223e-007	-229.553
	α	0.175	8.190e-005	2139.476
	β	0.982	1.329e-005	73915.170
	γ	-0.010	7.193e-005	-139.573
5 min	δ	-9.438e-005	2.223e-006	-42.459
	α	0.344	0.0007	455.977
	β	0.936	0.0002	4222.06
	γ	-0.042	0.0005	-86.50
10 min	δ	-8.618e-005	5.291e-006	-16.287
	α	0.425	0.002	225.638
	β	0.890	0.0006	1421.579
	γ	-0.037	0.001	-29.24
15 min	δ	-0.0002	9.173e-006	-18.134
	α	0.363	0.002	160.513
	β	0.878	0.001	808.199
	γ	-0.039	0.0015	-25.510
30 min	δ	-0.0002	3.536e-005	-5.325
	α	0.055	0.001	61.268
	β	0.995	0.002	5844.856
	γ	-0.02	0.001	-35.178

Tabella 5.10: Stima del modello EGARCH(1,1) con differenza prima del logaritmo del volume al tempo t

Il modello EGARCH(1,1) analizzato ora contiene come esogena la differenza prima del logaritmo dei volumi al tempo precedente rispetto ai rendimenti:

$$r_t = \mu + \delta[\log(vol_t - 1) - \log(vol_{t-2})] + \varepsilon_t \quad (5.20)$$

$$\log\sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta \log(\sigma_{t-1}^2) \quad (5.21)$$

Le stime ottenute sono riportate in Tabella 5.11.

freq	par	value	std. error	t stat
1 min	δ	-1.665e-005	3.349e-007	-49.708
	α	0.177	8.430e-005	2093.544
	β	0.981	1.296e-005	75712.369
	γ	-0.010	6.469e-005	-159.995
5 min	δ	-9.405e-006	1.531e-006	-6.143
	α	0.340	0.001	490.617
	β	0.936	0.0002	4299.586
	γ	-0.043	0.0004	-108.333
10 min	δ	-9.255e-006	5.607e-006	-1.651
	α	0.424	0.002	233.859
	β	0.890	0.001	1425.949
	γ	-0.038	0.001	-29.369
15 min	δ	-6.693e-005	1.122e-005	-5.968
	α	0.356	0.002	158.409
	β	0.879	0.001	836.732
	γ	-0.039	0.002	-26.206
30 min	δ	-0.0001	3.561e-005	-4.647
	α	0.054	0.0009	60.898
	β	0.995	0.0002	6045.584
	γ	-0.021	0.0005	-36.284

Tabella 5.11: Stima del modello EGARCH(1,1) con differenza prima del logaritmo del volume al tempo t

In questo caso il modello EGARCH(1,1) analizzato contiene come esogena il volume ritardato:

$$r_t = \mu + \delta vol_{t-1} + \varepsilon_t \quad (5.22)$$

$$\log \sigma_t^2 = \omega(d_1 + \dots + d_k) + \alpha \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta \log(\sigma_{t-1}^2) \quad (5.23)$$

Le stime ottenute sono riportate in Tabella 5.12.

freq	par	value	std. error	t stat
1 min	δ	-5.111e-005	2.226e-007	-229.553
	α	0.175	8.190e-005	2139.476
	β	0.982	1.329e-005	13915.170
	γ	-0.010	7.193e-005	-139.573
5 min	δ	-9.438e-005	2.223e-006	-42.459
	α	0.344	0.001	455.977
	β	0.936	0.0002	4222.060
	γ	-0.041	0.0005	-86.500
10 min	δ	-8.618e-005	5.291e-006	-16.287
	α	0.425	0.002	225.638
	β	0.890	0.001	1421.579
	γ	-0.037	0.001	-29.240
15 min	δ	-0.0002	9.173e-006	-18.194
	α	0.363	0.002	160.513
	β	0.878	0.001	808.199
	γ	-0.039	0.002	-25.510
30 min	δ	-0.0002	3.536e-005	-5.325
	α	0.055	0.001	61.268
	β	0.995	0.0002	5844.859
	γ	-0.021	0.0006	-35.177

Tabella 5.12: Stima del modello EGARCH(1,1) con volume al tempo t-1

In tutti i modelli EGARCH(1,1) analizzati, sia con variabile esogena contemporanea ai rendimenti sia ritardata, il valore del parametro β risulta sempre molto elevato e vicino a 1. Quindi, la persistenza della varianza condizionata risulta essere molto forte in tutti i casi e a tutte le frequenze. Per quanto riguarda l'impatto delle news di mercato e la capacità di variazione nel tempo di σ_t^2 , quindi il parametro α , questo presenta valori molto simili in tutti i modelli e che segnalano un certo impatto non trascurabile.

Il parametro che valuta la presenza di relazione tra rendimenti e volumi, δ , risulta in tutti i casi significativo e, quindi, si può affermare che, anche modellando la serie storica con i modelli EGARCH(1,1) la relazione tra rendimenti e volumi è presente. Come per i modelli GARCH(1,1) anche in questo caso il valore del parametro δ risulta negativo per i modelli che presentano il volume ritardato rispetto ai rendimenti.

Come accennato precedentemente, in questi modelli possiamo andare a valutare l'effetto leva con il parametro γ . Questo parametro assume valori significativamente diversi da zero per tutti i modelli presi in esame, quindi l'effetto leva è presente in tutti i casi considerati e a tutte le frequenze. I valori assunti sono sempre negativi ed in effetti questo è quanto ci si attende da questo modello, in quanto ci si aspetta che in caso di variazioni negative dei prezzi l'impatto sulla volatilità sia amplificativo rispetto a una variazione negativa. Nel modello che presenta come esogena i volumi di scambio ritardati di un periodo rispetto ai rendimenti, il valore del parametro γ è sempre positivo tranne per la frequenza di 1 minuto. Quindi, considerando il volume ritardato di un periodo l'impatto di variazioni positive dei prezzi è amplificativo sulla varianza condizionata.

Nelle precedenti analisi non viene riportato il caso di inserimento del volume, o di sue trasformazioni, nell'equazione della varianza condizionata non perchè non sia stata effettuata un'analisi di questo tipo ma perchè i risultati ottenuti portavano ad avere dei coefficienti del parametro relativo al volume statisticamente pari a zero. Lo studio è stato compiuto sia per i modelli GARCH sia per i modelli EGARCH e con tutte le varie trasformazioni del

volume viste nei precedenti modelli. Questo porta a concludere che non sembra esserci una diretta influenza dei volumi nella varianza condizionata dei modelli considerati per la nostra analisi.

5.3 Conclusioni

In questo lavoro sono stati analizzati dati ad alta frequenza con dei modelli GARCH(1,1) ed EGARCH(1,1). Questi modelli, in origine, erano stati pensati per dati giornalieri, quindi dati equispaziati e a frequenza massima di un giorno. Nella nostra analisi si è riusciti a superare il problema dei dati equispaziati considerando delle serie di partenza a frequenza di un minuto e quindi si sono usati i modelli presenti in letteratura con dati che negli ultimi anni sono molto usati.

Lo scopo di queste analisi era di capire l'eventuale presenza di una relazione tra rendimenti e volumi del titolo IBM. Molti sono stati gli studi precedentemente effettuati da vari autori e tutti hanno portato all'evidenza una relazione positiva contemporanea tra queste due grandezze finanziarie.

Dopo aver effettuato un'analisi di tipo descrittivo delle serie storiche di rendimenti e volumi si è notato come, effettivamente, potesse essere presente anche nei dati ad alta frequenza una certa relazione tra le grandezze in esame. Si è inoltre riscontrata una certa periodicità giornaliera sia dei rendimenti sia dei volumi. Infatti, analizzando la varianza giornaliera delle serie storiche, è risultato evidente come questa fosse maggiore in corrispondenza dell'apertura e della chiusura della giornata di contrattazioni.

I modelli alla base dell'analisi sono stati i GARCH(1,1) ed EGARCH(1,1) ma, per poter raggiungere il nostro scopo, sono state inserite delle variabili esogene e delle dummy. Le variabili in questione sono state il volume e alcune sue trasformazioni, in particolare sono il logaritmo del volume e la sua differenza prima. Si è voluta anche valutare la presenza di una relazione ritardata di un periodo tra i rendimenti e i volumi e non solo la relazione contemporanea, e questo è stato possibile inserendo come variabile esogena

dei modelli il volume ritardato di un periodo. Per tener conto della periodicità giornaliera dei volumi e dei rendimenti si sono inserite nel modello delle dummy per ogni istante temporale.

L'analisi è cominciata con i modelli GARCH(1,1) e si è riscontrato come per i vari modelli con le diverse variabili esogene, e per tutte le frequenze, che la relazione tra rendimenti e volumi è presente anche se i coefficienti non sono molto elevati. La relazione si manifesta sia a livello contemporaneo sia a livello ritardato ma in modo diverso. Nel caso in cui è stato inserito nel modello il volume, o il logaritmo del volume, la relazione è risultata positiva. Quindi, un aumento del volume in un certo istante temporale t comporta un aumento dei rendimenti dello stesso istante temporale. Questo risultato risulta in accordo con molti degli studi precedenti già analizzati nel capitolo 2. Quando nel modello il volume è stato inserito come variabile ritardata o come differenza prima la relazione è risultata negativa. Questo fatto quindi fa emergere che un aumento del volume in periodi precedenti porta ad avere all'istante t una diminuzione della varianza e del rendimento. Il risultato ottenuto in questo caso si scontra con lo studio effettuato da Mubarik e Javid (2009), i quali avevano riscontrato una relazione positiva anche per il caso di non contemporaneità tra rendimenti e volumi.

Con i modelli EGARCH(1,1) si è inserito nell'analisi anche un parametro per valutare l'effetto leva presente. Con questi modelli non solo è stata rilevata la presenza di una relazione contemporanea positiva tra i rendimenti e i volumi ma, anche, un effetto leva presente nei dati. Per tutti i modelli il parametro γ che identifica l'effetto leva risulta negativo segnalando, quindi, la relazione inversa tra la volatilità e le variazioni delle innovazioni.

Anche nei modelli EGARCH(1,1), come nei modelli GARCH(1,1), la relazione tra rendimenti e volumi risulta positiva se il volume è contemporaneo al rendimento, ma risulta negativa se il volume è ritardato rispetto al rendimento.

Ringraziamenti

Per la realizzazione di questo lavoro devo sicuramente ringraziare il prof. Caporin Massimiliano per il suo supporto e la sua disponibilità. Per la realizzazione di questa tesi devo ringraziare anche il dott. Costola Michele per il supporto informatico.

Non posso non ringraziare i miei genitori. Grazie perchè in tutti questi anni non hanno mai dubitato del mio percorso e mi hanno sempre sostenuta. Mi hanno sempre appoggiata e mai scoraggiata, aiutandomi in tutti i modi per raggiungere questo mio obiettivo.

Questi anni di studio mi hanno fatto conoscere delle persone speciali che sono entrate nella mia vita e ci rimarranno. Con loro ho condiviso intere giornate di studio ma anche di risate e divertimento! Abbiamo gioito e ci siamo disperate insieme ad ogni appello d'esame e queste cose non le dimenticherò. Un grazie a chi ha condiviso con me solo qualche breve periodo o chi è stato con me per gran parte del percorso. Non le ringrazio ad una ad una solamente perchè rischerei di dimenticare qualcuno ma loro sanno benissimo a chi mi riferisco. Una di loro, però, deve essere particolarmente ringraziata e questa persona è Cristiana. Con lei ho vissuto tutto il percorso universitario dal primo giorno di lezione. La devo ringraziare di essersi seduta vicino a me quel primo giorno in cui non sapevamo bene dove eravamo! Da quel momento non ci siamo più allontanate e di questo sono molto felice perchè ho avuto la possibilità di conoscere una persona davvero speciale.

Ai miei amici un grazie grande. Mi hanno veramente sopportata anche quando avevo la testa sempre sui libri e magari non li vedevo molto. Loro mi

conoscono letteralmente da una vita e hanno sempre saputo come stare vicino a una persona non troppo facile come me! So che ci saranno per me come io per loro.

Per ultima, ma non ultima, devo ringraziare una persona speciale che ha saputo essere al fianco sempre nel modo giusto e sempre con le parole giuste ad una 'studentessa universitaria' come me, a volte completamente fuori di testa!!!

Bibliografia

- [1] Baklaci H., Kasman A., *An Empirical Analysis of Trading Volume and Return Volatility Relationship in the Turkish Stock Market*, Working paper
- [2] Bohl M., Henke H. (2003), *Trading Volume and Stock Market Volatility: the Polish case*, *International Review of Financial Analysis*, n. 12, pp. 513-525
- [3] Bollerslev T. (1986), *Generalized Autoregressive Conditional Heteroskedasticity*, *Journal of Economics* 31, pp. 307-327, North Holland
- [4] Butterworth, Heinemann Finance, *Forecasting Volatility in the Financial Markets*, second edition, John Knight & Stephen Satchell
- [5] Cecconi M., Lombardi M. J. (2001), *Dati finanziari ad alta frequenza*, *Scienza e Business*, Anno III, n. 9-10, pp. 17-24
- [6] Chen G., Firth M., Rui O. M. (2001), *The Dynamic Relation between Stock Returns, Trading Volume and Volatility*, *The Financial Review*, n. 38, pp. 153-174
- [7] De Medeiros O. R., Van Doornik B. F. N., *The Empirical Relationship between Stock Returns, Return Volatility and Trading Volume in the Brazilian Stock Market*, Working paper

- [8] Engle R. F., Russel R. R. (2004), *Analysis of High Frequency Financial Data*, Working paper
- [9] Floros C., Vougas V. D. (2007), *Trading Volume and Returns Relationship in Greek Stock Index Futures Markets: GARCH vs GMM*, International Research Journal of Finance and Economics, n. 12
- [10] Gallant A. R., Rossi P. E., Tauchen G., *Stock Price and Volume*, Working paper
- [11] Gallo G. M., Pacini B., *Metodi quantitativi per i mercati finanziari*, Carocci editore
- [12] lee B., Rui O. M (2002), *The Dinamic Relationship between Stock Returns anjd Trading Volume: Domestic and cross-country Evidence*, Journal of Banking & Finance, n. 26, pp. 51-78
- [13] Mubarik F., Javid A. Y. (2009), *Relationship Between Stock Return, Tradin Volume and Volatility: Evidence from Pakistani Stock Market*, Asia Pacific Journal of Finance and Banking Research, Volume 3, n. 3
- [14] Saatcioglu K., Starks T. L. (1998), *The Stock Price-Volume Relationship in Emerging Stock Markets: The Case of Latin America*, International Journal of Forecasting, Volume 14, n. 2, pp. 215-225
- [15] Tsay Ruet S., *Analysis of Financial Time Series*, Wiley
- [16] Pastorello S., *Rischio e Rendimento*, Il Mulino