



Università degli Studi di Padova
FACOLTÀ DI SCIENZE STATISTICHE

TESI DI LAUREA IN
SCIENZE STATISTICHE ED ECONOMICHE

Modelli GARCH multivariati
con correlazione condizionata dinamica

Relatore: Prof. SILVANO BORDIGNON

Laureando: Enrico Baggio

Matric. n. 450492/SE

ANNO ACCADEMICO 2002-2003

INDICE

1. Introduzione	1
2. Modelli GARCH univariati	
2.1 Il modello GARCH	7
2.2 Il modello TGARCH	13
2.3 Il modello EGARCH	14
3. Modelli GARCH multivariati	
3.1 Il modello VECH	17
3.2 Il modello BEKK	22
3.3 Il modello CCC	25
4. Il modello DCC	
4.1 Specificazione e caratteristiche del modello DCC	29
4.2 Stima e standard errors	37
4.3 Test per la correlazione costante	43
4.4 Previsione a più passi avanti	46

5. Implementazione del software e applicazioni	
5.1 Il software per il modello DCC	49
5.2 Applicazioni	58
Conclusione	65
Appendice A	69
Appendice B	89
Appendice C	93
Bibliografia	

1 Introduzione

Gli aspetti relativi al comportamento degli oggetti che possono essere studiati con metodi basati sull'analisi delle serie storiche, non possono essere completi senza una valutazione dell'incertezza, cosa che appare naturale vista la dipendenza dei singoli rendimenti da fattori di rischio presenti nel mercato. La volatilità dei rendimenti deve quindi essere analizzata attraverso degli appositi modelli.

Nel 1982 Engle introduce un primo approccio per l'analisi della volatilità con il modello ARCH, il precursore di tutte le ricerche che verranno proposte in seguito. Dopo questo grande successo, Bollerslev nel 1986 generalizzò il modello di Engle impostando definitivamente la strada ad una nuova generazione di modelli capaci di catturare la volatilità dinamica delle serie temporali, il GARCH.

Questa direzione portò all'estensione del GARCH con modelli più particolari come (EGARCH) Exponential GARCH, (TGARCH) Threshold GARCH e (APGARCH) Asymmetric Power GARCH.

Mentre questi modelli riuscivano a catturare eccessi di curtosi e asimmetria, che spesso si riscontrano nelle serie finanziarie, essi non erano in grado di analizzare l'interdipendenza di effetti tra diversi mercati o rendimenti.

Mentre il GARCH univariato aveva incontrato un diffuso successo empirico, il problema associato alla stima del modello GARCH multivariato con correlazione variabile nel tempo, costrinse la ricerca a

stimare modelli sia con limitata possibilità, sia con restrizioni considerevoli.

Infatti lo studio delle relazioni tra variabili finanziarie, spesso necessita di grandi matrici di covarianza dinamiche. Si va dalla possibile interrelazione tra i mercati (contagio), oggi fortemente accelerato dalla globalizzazione economica e dalla comunicazione via Internet, ai titoli appartenenti allo stesso comparto che possono reagire in modo simile alle stesse informazioni, e alla gestione e ottimizzazione di un portafoglio. L'interesse viene posto sulla modellazione di queste dipendenze, soprattutto dinamiche, e sulla possibilità di sfruttarle concretamente.

Descriveremo un modello che può essere usato per stimare matrici di covarianza variabile nel tempo estremamente grandi e le sue proprietà; stiamo parlando del Dynamic Conditional Correlation Multivariate GARCH model.

Questa classe di modelli MV-GARCH differisce dalle altre specificazioni in quanto il modello GARCH univariato è stimato per ogni serie finanziaria, e poi usando i residui standardizzati risultanti dalla prima stima, sarà calcolata la matrice di correlazione dinamica utilizzando una semplice specificazione.

Questa parametrizzazione conserva la semplice interpretazione del modello GARCH univariato, con un facile calcolo degli stimatori della correlazione. Bollerslev, Wooldridge ed Engle nel 1988 originariamente proposero un modello GARCH multivariato nella familiare forma vettoriale, il quale prevede una struttura molto generale per il modello multivariato della volatilità. Il modello più generale richiede $O(k^4)$ parametri da stimare con la massima verosimiglianza, dove k è il numero delle serie inizialmente stimate.

Un più semplice modello, il Diagonal-vech che fu proposto in seguito ammette coefficienti non nulli solo per i propri ritardi e per i ritardi dei loro residui, riducendo il numero dei parametri che devono essere stimati a $O(k^2)$.

Comunque, derivando le restrizioni necessarie ai parametri per assicurare che la matrice di covarianza condizionata sia definita positiva inizia ad essere estremamente difficoltosa con k che aumenta anche a misure moderate.

La formulazione BEKK, proposta da Engle e Kroner, sviluppò una generale forma quadratica per l'equazione della covarianza condizionata, la quale eliminò il problema di assumere definita positiva la stima della covarianza condizionata dell'originale modello vech. Nell'ordine, per il modello BEKK generalmente più completo, il numero di parametri che devono essere stimati è $O(k^4)$, ma la stima di un modello BEKK più standard può assumere $O(k^2)$ parametri.

Altre formulazioni del modello BEKK, più trattabili, sono la forma diagonale e scalare, che pongono restrizioni ad alcuni parametri, sebbene queste restrizioni siano tipicamente rifiutate.

In aggiunta a un largo numero di parametri necessari alla stima di una formulazione generale, l'esatta interpretazione dei singoli coefficienti è difficile da distinguere.

Nel 1990 Bollerslev introdusse uno specifico GARCH multivariato con Correlazione Condizionata Costante (CCC), dove i modelli GARCH univariati sono stimati per ogni serie e poi la matrice di correlazione viene calcolata usando la massima verosimiglianza sugli stimatori di correlazione dei residui che sono stati trasformati usando la stima della loro deviazione standard condizionata. L'assunzione di costanza della correlazione ci permette di stimare un modello anche con molte serie (k può assumere

dimensioni più grandi) e garantisce che la matrice stimata sia definita positiva, richiedendo semplicemente che ogni varianza condizionata (univariata) sia diversa da zero e che la matrice di correlazione sia a rango pieno. Comunque lo stimatore di correlazione costante, non ha fornito un metodo di costruzione degli standard errors consistenti usando il processo di stima a multi-stadio.

Bollerslev trovò plausibile la nozione di correlazione costante, ma già recenti lavori di Tsui e Yu (1999) hanno trovato che la correlazione costante può essere rifiutata per certi portafogli.

Nel 2001, Engle propose una nuova classe di stimatori, totalmente rivoluzionari, che preserva sì la facilità di stima del modello a correlazione costante di Bollerslev, ma che lascia una correlazione dinamica (variabile nel tempo).

Il modello DCC MV-GARCH preserva la parsimonia del modello GARCH univariato per la volatilità dei singoli rendimenti e calcola una semplice stima di verosimiglianza per la correlazione dinamica.

Il numero dei parametri da stimare usando la massima verosimiglianza è $O(k)$, un considerevole miglioramento, al di sopra sia del modello Vech che del modello BEKK. Ancora più importante, il numero dei parametri da stimare simultaneamente è $O(1)$.

Punto centrale di questa tesi è studiare sia la teoria, ma soprattutto le proprietà empiriche del DCC MV-GARCH.

Per questo analizzeremo da vicino sia modelli univariati che gli altri modelli multivariati, soffermandoci sulle modalità di stima e sui vincoli necessari ai parametri. Mostriamo poi, attraverso il software appositamente costruito, il notevole passo avanti avvenuto quando si deve stimare una matrice di covarianza condizionata molto grande.

Questo nuovo stimatore dimostrerà delle grandi prestazioni, specialmente considerando la facilità d'implementazione degli stimatori.

2. Modelli GARCH univariati

2.1 Il modello GARCH

I rendimenti di un'attività finanziaria sono generalmente incorrelati (o caratterizzati da una debole dipendenza seriale a piccoli ritardi) ma certamente non sono indipendenti. In particolare questa dipendenza si manifesta soprattutto nei momenti secondi dei rendimenti che, condizionatamente all'informazione disponibile, variano continuamente con t .

Pertanto il modello omoschedastico (a varianza condizionata costante)

$$r_t = \mu_t + \varepsilon_t$$

dove

$$E(r_t | I_{t-1}) = \mu_t$$
$$Var(r_t | I_{t-1}) = \sigma^2$$

non va bene.

Consideriamo quindi il modello più generale eteroschedastico (a varianza condizionata variabile)

$$r_t = \mu_t + \varepsilon_t$$

con

$$E(r_t | I_{t-1}) = \mu_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

$$\sigma_t^2 = Var(r_t | I_{t-1}) = Var(\varepsilon_t | I_{t-1}) = E[(r_t - \mu_t)^2 | I_{t-1}]$$

- μ_t rappresenta l'equazione per la media (condizionata) dei rendimenti;
- ε_t rappresenta lo shock o il rendimento corretto per la media;
- il modello per σ_t^2 rappresenta l'equazione per la volatilità dei rendimenti.

Specificando opportunamente l'equazione per la varianza condizionata σ_t^2 si ottengono varie classi di modelli eteroschedastici, in particolare i modelli GARCH.

Il termine GARCH, dovuto a Bollerslev (1986), sta per Generalized AutoRegressive Conditional Heteroskedasticity e riguarda una generalizzazione del modello ARCH. La sua idea fu quella di riprodurre la parsimonia del modello ARMA rispetto alla rappresentazione AR o MA in termini del numero di parametri utilizzati. Rispetto al modello di Engle (ARCH), si introducono i valori ritardati della varianza condizionata, in modo da risparmiare parametri da stimare rispetto alla struttura ARCH.

Il GARCH (p,q) è perciò un modello in cui la varianza condizionata al tempo t è una combinazione lineare di p ritardi dei residui al quadrato – ricavati dall'equazione della media condizionata – e da q ritardi della varianza condizionata.

In sintesi un GARCH(p,q) può essere espresso come:

$$E(\varepsilon_t^2 | I_{t-1}) \equiv h_t = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}$$

Nel modello GARCH, quindi, le informazioni passate sono sintetizzate dai ritardi della varianza, mentre le “ novità ”, e la capacità di variazione nel tempo delle stime della varianza condizionata sono racchiuse nel termine ε^2 al tempo $t-i$.

Le condizioni rilevanti nel caso generale GARCH (p,q) sono:

$$\begin{aligned} \omega &> 0 \\ \left| \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j \right| &< 1 \\ \alpha_i &\geq 0 \\ \beta_i &\geq 0 \end{aligned}$$

come condizioni sufficienti per la non negatività della varianza condizionata.

Un importante strumento per stimare i parametri di interesse è la tecnica di stima della massima verosimiglianza.

Applicando la tecnica di massima verosimiglianza nel caso più generale, arriviamo al seguente problema di ottimizzazione:

$$\text{Max}_{\mathcal{g}} \{ \log L_T(\mathcal{g}) \}$$

dove

$$\log L_T(\mathcal{g}) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log h_t(\mathcal{g}) - \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2}{h_t(\mathcal{g})}$$

dove $\mathcal{g} = (\alpha, \beta, \omega)$, soggetta alle condizioni di stazionarietà $\sum_{i=1}^{\max(p,q)} \alpha_i + \beta_i < 1$

per la varianza condizionata e alla non negatività di ω, α, β .

Quindi per il GARCH(p,q), possiamo formulare il seguente problema di ottimo:

$$\begin{aligned} \max & \left\{ -\frac{1}{2} \sum_{i=1}^T \log h_t - \frac{1}{2} \sum_{i=1}^T \frac{\varepsilon_t^2}{h_t} \right\} \\ \text{s.t.} & \quad \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} = h_t \quad \forall t = 1, \dots, T \end{aligned}$$

$$\left\{ \begin{array}{l} \sum_{i=1}^{\max\{p,q\}} \alpha_i + \beta_i \leq 1 \\ \omega \geq 0 \\ \alpha_i \geq 0 \quad \forall i = 1, \dots, p \\ \beta_i \geq 0 \quad \forall i = 1, \dots, q \end{array} \right.$$

Il problema di trovare un valore di ϑ che massimizzi la funzione di log-verosimiglianza può essere visto come quello di trovare un valore di ϑ che sia soluzione dell'equazione che uguaglia a zero la derivata prima. A questo fine sono state sviluppate tecniche dette di *soluzione numerica*, basate su algoritmi che, opportunamente programmati su computer, forniscono soluzioni che rendono le condizioni del primo ordine approssimativamente valide. Si tratta di procedure iterative che a partire da una condizione iniziale suggeriscono soluzioni parziali che diminuiscono via via il grado di approssimazione fino a che esso raggiunga una soglia definita accettabile dall'utilizzatore.

Il punto di partenza che giustifica questa procedura è considerare la derivata prima come una generica funzione $g(\vartheta)$ ed uno sviluppo in serie di Taylor per un valore iniziale $\hat{\vartheta}_0$. Si ha quindi:

$$g(\vartheta) \approx g(\hat{\vartheta}_0) + g'(\hat{\vartheta}_0)(\vartheta - \hat{\vartheta}_0).$$

Dato che $g(\vartheta)$ deve essere uguale a zero nel punto di massimo per $l(\vartheta)$, possiamo scrivere

$$0 \approx g(\hat{\vartheta}_0) + g'(\hat{\vartheta}_0)(\vartheta - \hat{\vartheta}_0) = g(\hat{\vartheta}_0) + g'(\hat{\vartheta}_0)\vartheta - g'(\hat{\vartheta}_0)\hat{\vartheta}_0$$

e quindi,

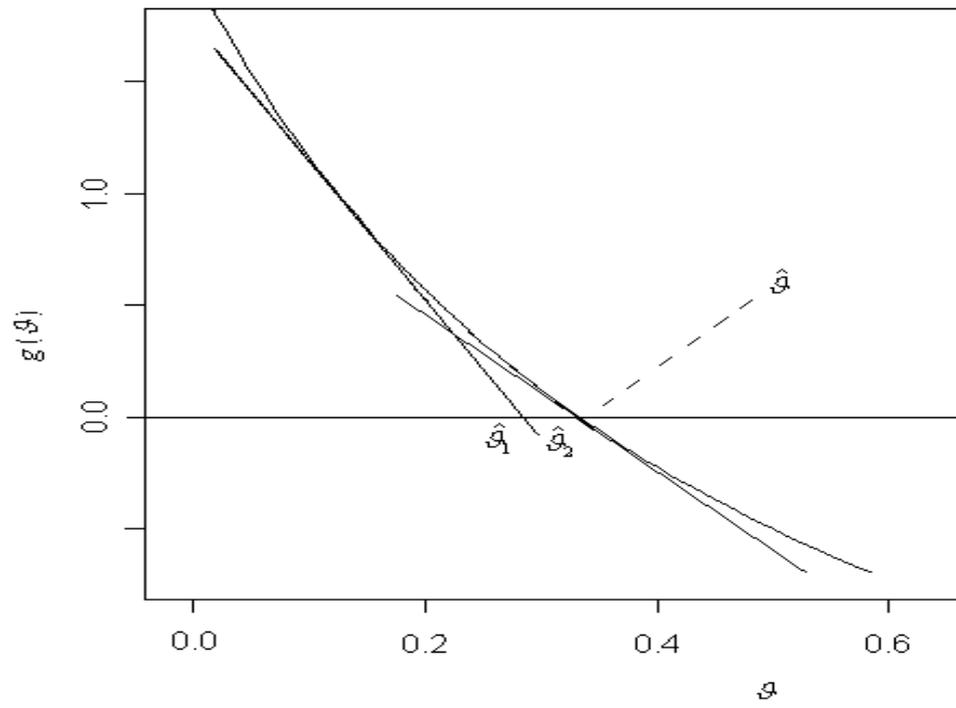
$$\vartheta \approx \hat{\vartheta}_0 - \frac{g(\hat{\vartheta}_0)}{g'(\hat{\vartheta}_0)}.$$

La procedura iterativa, dunque, suggerisce di adottare come valore $\hat{\vartheta}_1$ al passo successivo

$$\hat{\vartheta}_1 = \hat{\vartheta}_0 - \frac{g(\hat{\vartheta}_0)}{g'(\hat{\vartheta}_0)},$$

al fine di avvicinarsi maggiormente alla soluzione. Si noti che, affinché il punto $\hat{\vartheta}$ al quale converge sia un punto di massimo, la funzione $g(\cdot)$ (ricordiamo che per noi questa funzione è la derivata prima) deve essere decrescente (condizione del secondo ordine).

Se $\hat{\vartheta}_0$ è alla sinistra del punto $\hat{\vartheta}$, varrà dunque che $g(\hat{\vartheta}_0) > 0$ e $g'(\hat{\vartheta}_0) < 0$. Il valore $\hat{\vartheta}_0$ sarà dunque incrementato di una quantità positiva e $\hat{\vartheta}_1$ sarà più vicino alla soluzione. Viceversa, se $\hat{\vartheta}_0$ è alla destra del punto $\hat{\vartheta}$, allora $g(\hat{\vartheta}_0) < 0$ e $g'(\hat{\vartheta}_0) > 0$. Il valore $\hat{\vartheta}_1$ sarà più vicino alla soluzione in quanto più piccolo di $\hat{\vartheta}_0$. La procedura può essere ripetuta fino a quando la differenza fra stime successive non sia sufficientemente piccola. In tal caso si dice che si è raggiunta la convergenza (in n passi) e l'ultimo candidato $\hat{\vartheta}_n$ è la stima di massima verosimiglianza.



2.2 Il modello TGARCH

Il modello TGARCH (Glosten, Jagannathan, e Runkle, 1993; Zakoian, 1994) introduce un diverso comportamento in corrispondenza dell'attraversamento da parte dell'innovazione ritardata di una soglia (threshold), posta di solito a zero.

Un modello che riproduca gli effetti di asimmetria riscontrati come regolarità empiriche nelle serie finanziarie può essere formalizzato nel seguente modo:¹

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma S_{t-1}^- \varepsilon_{t-1}^2$$

Se $\varepsilon_{t-1} < 0$ allora

$$h_t = \omega + (\alpha + \gamma) \varepsilon_{t-1}^2 + \beta h_{t-1} \quad \text{con } \gamma > 0$$

Se $\varepsilon_{t-1} \geq 0$ allora

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}$$

Il coefficiente γ misura l'effetto differenziato per *shock* negativi: il segno atteso di γ è positivo, con la conseguenza che la volatilità è più elevata in corrispondenza di innovazioni negative e lo sarà tanto più quanto maggiore è la dimensione dell'innovazione. Quando ci sono shock negativi c'è un aumento della volatilità immediato, misurato da $(\alpha + \gamma)$ moltiplicato per la dimensione dello shock al quadrato.

Se si ipotizza invece che ci sia uno shock positivo, abbiamo semplicemente una naturale diminuzione della volatilità in relazione alle notizie positive.

¹Ci scusiamo per l'abuso di notazione con il quale indichiamo con S_{t-1}^- una variabile dicotomica (dummy) che assume il valore 1 quando $\varepsilon_{t-1} < 0$.

2.3 Il modello EGARCH

Capita spesso che parametri ottenuti con modelli GARCH violino le condizioni di non negatività. In questo caso, dato che il parametro stimato non appartiene all'insieme dei valori ammissibili, dovremmo procedere ad una stima vincolata imponendo che i parametri rispettino le condizioni alla base del modello. La procedura di stima diventa allora più complessa, così nella letteratura sono apparse proposte di specificazione alternative.

E' il caso dell'Exponential GARCH (Nelson 1991) le cui caratteristiche sono:

- l'impossibilità di ottenere una varianza h_t negativa (senza bisogno di imporre alcuna condizione sui parametri);
- la presenza di asimmetria nelle reazioni della volatilità alle innovazioni;
- la possibilità di misurare un effetto asimmetrico proporzionale all'entità delle innovazioni.

A differenza del modello TGARCH, nel quale è inserita una variabile dummy da moltiplicare per gli ε_{t-1}^2 , il modello di Nelson è specificato in termini logaritmici.

Nel caso di un EGARCH(1,1), abbiamo:

$$\ln(h_t) = \omega + \beta \ln(h_{t-1}) + \alpha \left(\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}} \right) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}$$

Qualsiasi valore sia attribuito ai parametri, la trasformazione esponenziale assicura la non negatività della varianza.

Nel modello EGARCH il termine $\beta \ln(h_{t-1})$ cattura l'effetto di persistenza nella volatilità. Dato che l'espressione ha un termine autoregressivo, la

stazionarietà è assicurata dalla condizione $0 < \beta < 1$. La sua dimensione determinerà quanto rapido sia l'assorbimento degli shock passati.

Il termine

$$\left(\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}} \right)$$

è una variabile casuale a media zero nel caso in cui le innovazioni standardizzate siano distribuite normalmente² che consente di tenere conto della possibilità di una reazione asimmetrica proporzionale alle innovazioni.

I termini ω e $\sqrt{\frac{2}{\pi}}$, nella pratica, sono stimati congiuntamente come costante.

L'effetto misurato dal termine espresso in valore assoluto

$$\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}}$$

è evidentemente indipendente dal fatto che gli shock siano positivi o negativi, mentre l'effetto asimmetrico viene evidenziato dall'ultimo elemento dell'espressione,

$$\gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}$$

dove ε_{t-1} può assumere qualunque segno.

Il segno atteso per il parametro γ sarà negativo, in quanto ci aspettiamo un effetto amplificatore sulla volatilità nel caso di innovazioni negative ed un impatto ridotto nel caso di innovazioni positive.

²Nelson (1991) specifica il suo modello in maniera più generale riferendosi ad una v.c. z_t con distribuzione *Generalized Error*, ponendo questo termine uguale a $z_{t-1} - E(z_{t-1})$.

Supponendo $\gamma < 0$, se lo shock ε_{t-1} è positivo, esso avrà un impatto complessivo pari a $\alpha + \gamma < \alpha$; mentre se lo shock è negativo, ε_{t-1} avrà un effetto pari a $\alpha - \gamma > \alpha$, ovvero un effetto amplificativi.

Se γ non risulta significativamente diverso da zero, la specificazione EGARCH ha comunque il vantaggio di non dover imporre restrizioni di positività sui parametri.

Il modello EGARCH ha, quindi, proprietà teoriche migliori rispetto ad altri modelli³, con un costo aggiuntivo in termini di tempo di calcolo, dato che, generalmente, la stima richiede un numero di iterazioni maggiori per raggiungere la convergenza.

³ Vedi appendice

3. Modelli GARCH multivariati

3.1 Il modello VECH

Arrivati a ottime specificazioni per il GARCH univariato, il passo successivo fu quello di cimentarsi nei più complessi modelli multivariati.

Il primo modello GARCH multivariato fu proposto nel 1988 da Bollerslev, Wooldridge ed Engle, il quale prevedeva una rappresentazione vettoriale nella seguente formulazione:

$$vech(\Sigma_t) = vech(\Gamma) + \sum_{i=1}^p A_i \cdot vech(\varepsilon_{t-i} \varepsilon_{t-i}') + \sum_{j=1}^q B_j \cdot vech(\Sigma_{t-j})$$

dove:

$$\varepsilon_t = (\varepsilon_{1,t} \quad \dots \quad \varepsilon_{k,t})^T$$

Γ è una matrice $k \times k$ definita positiva

A_i e B_j sono matrici quadrate di dimensioni $\frac{k(k+1)}{2}$

Questa espressione rappresenta semplicemente la generalizzazione nel multivariato della formula:

$$\sigma_t^2 = \gamma + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

cioè di un GARCH (p,q) univariato.

Vediamo ora un esempio di un GARCH (1,1) calcolato su due serie di rendimenti.

$$\text{vech}(\Sigma_t) = \begin{pmatrix} \sigma_{1,t}^2 \\ \sigma_{12,t} \\ \sigma_{2,t}^2 \end{pmatrix} = \begin{pmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \cdot \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{12,t-1} \\ \sigma_{2,t-1}^2 \end{pmatrix}$$

Risolvendo l'equazione troviamo:

$$\begin{aligned} \sigma_{1,t}^2 &= \gamma_{11} + a_{11} \varepsilon_{1,t-1}^2 + a_{12} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{13} \varepsilon_{2,t-1}^2 + b_{11} \sigma_{1,t-1}^2 + b_{12} \sigma_{12,t-1} + b_{13} \sigma_{2,t-1}^2 \\ \sigma_{12,t} &= \gamma_{12} + a_{21} \varepsilon_{1,t-1}^2 + a_{22} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{23} \varepsilon_{2,t-1}^2 + b_{21} \sigma_{1,t-1}^2 + b_{22} \sigma_{12,t-1} + b_{23} \sigma_{2,t-1}^2 \\ \sigma_{2,t}^2 &= \gamma_{22} + a_{31} \varepsilon_{1,t-1}^2 + a_{32} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{33} \varepsilon_{2,t-1}^2 + b_{31} \sigma_{1,t-1}^2 + b_{32} \sigma_{12,t-1} + b_{33} \sigma_{2,t-1}^2 \end{aligned}$$

Come si può notare la formulazione diventa subito molto pesante e stiamo lavorando con due sole serie di rendimenti.

Nel nostro caso dobbiamo stimare 7 parametri per ogni elemento della nostra matrice di varianza e covarianza, ma più in generale, si hanno:

$$\frac{k(k+1)}{2} \cdot \left[1 + \frac{k(k+1)}{2} \cdot (p+q) \right]$$

dove k è il numero di serie iniziali, p e q sono rispettivamente i ritardi dei residui al quadrato e i ritardi della covarianza condizionata del modello GARCH multivariato.

Una rappresentazione particolare del modello VECH è il **DIAGONAL VECH**, che con una semplice restrizione riduce in maniera considerevole il numero di parametri.

Assumiamo che le matrici A_i e B_j siano diagonali, nel caso $k=2$ abbiamo

$$vech(\Sigma_t) = \begin{pmatrix} \sigma_{1,t}^2 \\ \sigma_{12,t} \\ \sigma_{2,t}^2 \end{pmatrix} = \begin{pmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{pmatrix} \cdot \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{12,t-1} \\ \sigma_{2,t-1}^2 \end{pmatrix}$$

A questo punto avremo:

$$\begin{aligned} \sigma_{1,t}^2 &= \gamma_{11} + a_{11}\varepsilon_{1,t-1}^2 + b_{11}\sigma_{1,t-1}^2 \\ \sigma_{12,t} &= \gamma_{12} + a_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + b_{22}\sigma_{12,t-1} \\ \sigma_{2,t}^2 &= \gamma_{22} + a_{33}\varepsilon_{2,t-1}^2 + b_{33}\sigma_{2,t-1}^2 \end{aligned}$$

Il numero di parametri si riduce a 9, anziché 21 del modello VECH generale.

Ecco due esempi di modelli GARCH con k serie iniziali.

K	VECH	DIAGONAL-VECH
3	78	18
5	465	45
10	6'105	165
50	3'252'525	3'825
100	51'010'050	15'150

Tabella 1 GARCH(1,1)

<i>K</i>	<i>VECH</i>	<i>DIAGONAL-VECH</i>
3	114	24
5	690	60
10	9'130	220
50	4'878'150	5'100
100	76'507'500	20'200

Tabella 2 GARCH(2,1)

Come si può vedere dalle tabelle qui sopra, il numero di parametri da stimare sale vertiginosamente con l'aumentare del numero delle serie, soprattutto con la rappresentazione più generale.

Già usando il DIAGONAL-VECH il numero dei parametri resta, per così dire, moderato rispetto alla forma generale, ma solo aumentando l'ordine del GARCH, i parametri da stimare aumentano e non di poco.

Un altro problema della rappresentazione vettoriale sono i vincoli sui parametri.

Nella rappresentazione più generale le condizioni che assicurano che la matrice Σ_t sia definita positiva sono complesse e difficili da verificare.

Riportiamo l'esempio di un modello ARCH(1) bivariato proposto da Engle, Granger e Kraft (1986). Per tale modello le condizioni necessarie affinché Σ_t sia definita positiva sono:

$$\begin{aligned} &\gamma_{11} > 0, \quad \gamma_{22} > 0, \quad \gamma_{11}\gamma_{22} - \gamma_{12}^2 > 0 \\ &a_{11} \geq 0, \quad a_{13} \geq 0, \quad a_{31} \geq 0, \quad a_{33} \geq 0, \quad a_{11}a_{33} - a_{22}^2 \geq 0, \\ &a_1 - \frac{(1/a_1)^2}{3} \geq 4 \quad -\frac{(1/a_1)^2}{3} \geq 2 \quad a_3 \quad a_3 \quad a_3 \quad a \\ &a_1 - \frac{1}{1}^2 \geq 0 \quad 1, -a_2 \geq 2 \quad 1 \quad a_1 \quad a_3 \quad a_3 \quad a_2 \end{aligned}$$

C o n u n A R C
i r r i i à s o l r 2 i o v , i
G A R C H (p , q)

di 50×50 , oltre all'estremo numero di parametri, avrei altrettanti vincoli, con una difficoltà di gestione a dir poco impossibile.

Nell'esempio di Engle, Granger e Kraft (1986) i vincoli del modello diagonal-vech diventerebbero:

$$\begin{aligned} \gamma_{11} > 0, \quad \gamma_{22} > 0, \quad \gamma_{11}\gamma_{22} - \gamma_{12}^2 > 0 \\ a_{11} \geq 0, \quad a_{33} \geq 0, \quad a_{11}a_{33} - a_{22}^2 \geq 0, \end{aligned}$$

3.2 Il modello BEKK

Per vincere le difficoltà di gestione della matrice di varianza e covarianza, Baba, Engle, Kraft, e Kroner (1995) introdussero la rappresentazione BEKK. In questo modello il numero dei parametri da stimare è diminuito utilizzando una semplice riparametrizzazione.

La rappresentazione BEKK per un generale processo GARCH(p,q) parte da questa formulazione:

$$\Sigma_t = \Gamma + \sum_{i=1}^p A_i^{*'} \cdot \varepsilon_{t-i} \cdot \varepsilon_{t-i}' \cdot A_i^* + \sum_{j=1}^q B_j^{*'} \cdot H_{t-j} \cdot B_j^*$$

dove A_i^* e B_j^* sono due matrici quadrate di dimensione k.

Questa rappresentazione consente di definire quasi sempre la matrice di covarianza definita positiva. Se Γ è definita positiva allora anche Σ_t lo è.

Per un modello GARCH(1,1) su due serie di rendimenti, la specificazione di Σ_t nel modello BEKK è:

$$\begin{aligned} \Sigma_t = & \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} + \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix}' \times \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\ \varepsilon_{2,t-1} \varepsilon_{1,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \times \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix} + \\ & \begin{bmatrix} b_{11}^* & b_{12}^* \\ b_{21}^* & b_{22}^* \end{bmatrix}' \times \begin{bmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{21,t-1} & \sigma_{2,t-1}^2 \end{bmatrix} \times \begin{bmatrix} b_{11}^* & b_{12}^* \\ b_{21}^* & b_{22}^* \end{bmatrix} \end{aligned}$$

Tralasciamo i pedici che si riferiscono al tempo e continuiamo a svolgere il calcolo matriciale.

$$\begin{bmatrix} a_{11}^{*2} \varepsilon_1^2 + a_{11}^* a_{21}^* \varepsilon_2 \varepsilon_1 + a_{11}^* a_{21}^* \varepsilon_1 \varepsilon_2 + a_{21}^{*2} \varepsilon_2^2 & a_{11}^* a_{12}^* \varepsilon_1^2 + a_{12}^* a_{21}^* \varepsilon_2 \varepsilon_1 + a_{11}^* a_{22}^* \varepsilon_1 \varepsilon_2 + a_{21}^* a_{22}^* \varepsilon_2^2 \\ a_{11}^* a_{12}^* \varepsilon_1^2 + a_{22}^* a_{11}^* \varepsilon_2 \varepsilon_1 + a_{12}^* a_{21}^* \varepsilon_1 \varepsilon_2 + a_{22}^* a_{21}^* \varepsilon_2^2 & a_{12}^{*2} \varepsilon_1^2 + a_{22}^* a_{12}^* \varepsilon_2 \varepsilon_1 + a_{12}^* a_{22}^* \varepsilon_1 \varepsilon_2 + a_{22}^{*2} \varepsilon_2^2 \end{bmatrix}$$

La matrice qui sopra è il risultato del prodotto del secondo addendo, qui sotto invece c'è la matrice risultante dal terzo addendo.

$$\begin{bmatrix} b_{11}^{*2}\sigma_1^2 + b_{11}^*b_{21}^*\sigma_{21} + b_{11}^*b_{21}^*\sigma_{12} + b_{21}^{*2}\sigma_2^2 & b_{11}^*b_{12}^*\sigma_1^2 + b_{12}^*b_{21}^*\sigma_{21} + b_{11}^*b_{22}^*\sigma_{12} & b_{21}^*b_{22}^*\sigma_2^2 \\ b_{11}^*b_{12}^*\sigma_1^2 + b_{22}^*b_{11}^*\sigma_{21} + b_{12}^*b_{21}^*\sigma_{12} + b_{22}^*b_{21}^*\sigma_2^2 & b_{12}^{*2}\sigma_1^2 + b_{22}^*b_{12}^*\sigma_{21} + b_{12}^*b_{22}^*\sigma_{12} & b_{22}^{*2}\sigma_2^2 \end{bmatrix}$$

A questo punto non ci resta che sommare queste due matrici con la matrice Γ .

Il risultato che otteniamo è:

$$\sigma_{1,t}^2 = \gamma_{11} + a_{11}^{*2}\varepsilon_{1,t-1}^2 + 2a_{11}^*a_{21}^*\varepsilon_{2,t-1}\varepsilon_{1,t-1} + a_{21}^{*2}\varepsilon_{2,t-1}^2 + b_{11}^{*2}\sigma_{1,t-1}^2 + 2b_{11}^*b_{21}^*\sigma_{12,t-1} + b_{21}^{*2}\sigma_{2,t-1}^2$$

$$\sigma_{12,t} = \gamma_{12} + a_{11}^*a_{12}^*\varepsilon_{1,t-1}^2 + a_{12}^*a_{21}^*\varepsilon_{2,t-1}\varepsilon_{1,t-1} + a_{11}^*a_{22}^*\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{21}^*a_{22}^*\varepsilon_{2,t-1}^2 + b_{11}^*b_{12}^*\sigma_{1,t-1}^2 + b_{12}^*b_{21}^*\sigma_{21,t-1} + b_{11}^*b_{22}^*\sigma_{12,t-1} + b_{21}^*b_{22}^*\sigma_{2,t-1}^2$$

$$\sigma_{2,t}^2 = \gamma_{22} + a_{12}^{*2}\varepsilon_{1,t-1}^2 + 2a_{22}^*a_{12}^*\varepsilon_{2,t-1}\varepsilon_{1,t-1} + a_{22}^{*2}\varepsilon_{2,t-1}^2 + b_{12}^{*2}\sigma_{1,t-1}^2 + 2b_{22}^*b_{12}^*\sigma_{21,t-1} + b_{22}^{*2}\sigma_{2,t-1}^2$$

Ora per risolvere il problema di ottimizzazione dobbiamo riuscire a snellire questa matrice, e per fare ciò utilizziamo dei nuovi parametri.

Per l'espressione di $\sigma_{1,t}^2$ la riparametrizzazione è la seguente:

$$\begin{aligned} \alpha_{11} &= a_{11}^{*2} & \alpha_{12} &= 2a_{11}^*a_{21}^* & \alpha_{13} &= a_{21}^{*2} \\ \beta_{11} &= b_{11}^{*2} & \beta_{12} &= 2b_{11}^*b_{21}^* & \beta_{13} &= b_{21}^{*2} \end{aligned}$$

quindi,

$$\sigma_{1,t}^2 = \gamma_{11} + \alpha_{11}\varepsilon_{1,t-1}^2 + \alpha_{12}\varepsilon_{2,t-1}\varepsilon_{1,t-1} + \alpha_{13}\varepsilon_{2,t-1}^2 + \beta_{11}\sigma_{1,t-1}^2 + \beta_{12}\sigma_{12,t-1} + \beta_{13}\sigma_{2,t-1}^2$$

Lo stesso procedimento deve essere eseguito per gli altri due elementi di Σ_t , per cui abbiamo:

$$\sigma_{12,t}^2 = \gamma_{12} + \alpha_{21}\varepsilon_{1,t-1}^2 + \alpha_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + \alpha_{23}\varepsilon_{2,t-1}^2 + \beta_{21}\sigma_{1,t-1}^2 + \beta_{22}\sigma_{12,t-1} + \beta_{23}\sigma_{2,t-1}^2$$

$$\sigma_{2,t}^2 = \gamma_{22} + \alpha_{31}\varepsilon_{1,t-1}^2 + \alpha_{32}\varepsilon_{2,t-1}\varepsilon_{1,t-1} + \alpha_{33}\varepsilon_{2,t-1}^2 + \beta_{31}\sigma_{1,t-1}^2 + \beta_{32}\sigma_{21,t-1} + \beta_{33}\sigma_{2,t-1}^2$$

In una formulazione generale, il numero dei parametri da stimare per un BEKK(p,q) sono $k(k+1)/2 + (k \cdot k) \cdot (p+q)$ e per un caso bivariato i parametri da stimare sono solamente 11.

3.3 Il modello CCC

Il modello GARCH a correlazione condizionata costante era probabilmente la specificazione più frequentemente applicata per modelli GARCH multivariati.

Questa forma fu introdotta da Bollerslev (1990) che era sicuro di ridurre il numero di parametri della rappresentazione Vech.

Baillie e Bollerslev usarono questa specificazione per fornire un modello per l'interdipendenza dei tassi di cambio e i premi al rischio dipendenti dal tempo per quattro delle maggiori valute europee.

Per semplificare la stime, loro assunsero che tutte le variazioni al di là del tempo nella covarianza condizionata, fossero dovute al cambiamento in ognuna delle corrispondenti due varianze condizionate.

Trovarono che la correlazione condizionata fosse altamente significativa tra i quattro mercati.

Il punto centrale di questa specificazione, quindi, è la matrice di covarianza condizionata.

$$\Sigma_t = \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} & \cdots & \sigma_{1k,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{1k,t} & \cdots & \cdots & \sigma_{k,t}^2 \end{bmatrix}$$

$$\sigma_{1,t}^2 = \omega_i + \sum_{j=1}^p \alpha_{i,j} \varepsilon_{i,t-j}^2 + \sum_{j=1}^q \beta_{i,j} \sigma_{i,t-j}^2 \quad i = 1, \dots, k$$

$$\sigma_{ij,t} = \rho_{ij} \sigma_{i,t} \sigma_{j,t} \quad i, j = 1 \dots k, \quad i \neq j$$

Per semplificare Σ_t quindi si assume $R_t = R$, cioè che la matrice di correlazione condizionata del vettore $\varepsilon_t = (\varepsilon_{1,t} \dots \varepsilon_{k,t})$ sia costante

$$\rho_{ij,t} = \left[\text{corr}(\varepsilon_{i,t}, \varepsilon_{j,t}) \middle| I_{t-1} \right] = \text{corr}(\varepsilon_i, \varepsilon_j) = \rho_{ij}$$

Con questo assunto la matrice di varianza e covarianza è così semplificata:

$$\Sigma_t = \begin{bmatrix} \sigma_{1,t} & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{k,t} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1k} \\ \rho_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \rho_{1k} & \dots & \dots & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t} & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{k,t} \end{bmatrix}$$

Questa formulazione non è altro che la scomposizione della matrice di varianza e covarianza condizionata in 3 matrici⁴:

$$\Sigma_t = D_t R D_t$$

dove R è la matrice di correlazione e D_t è la matrice diagonale della deviazione standard condizionata.

Vediamo ora un caso bivariato:

$$\Sigma_t = \begin{bmatrix} \sigma_{1,t} & 0 \\ 0 & \sigma_{2,t} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t} & 0 \\ 0 & \sigma_{2,t} \end{bmatrix} = \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{1,t} \sigma_{2,t} \rho_{12} \\ \sigma_{1,t} \sigma_{2,t} \rho_{12} & \sigma_{2,t}^2 \end{bmatrix}$$

dove $\rho_{12} = \rho = \text{corr}(\varepsilon_1, \varepsilon_2) = \frac{\text{cov}(\varepsilon_1, \varepsilon_2)}{\sqrt{\text{var}(\varepsilon_1) \cdot \text{var}(\varepsilon_2)}}$.

Un altro assunto che distingue questa specificazione dalle precedenti è che le singole varianze condizionate seguono dei processi GARCH(p,q) univariati standard.

Il modello CCC-MVGARCH, parte, quindi, dallo studio delle singole varianze condizionate, che possono essere modellate con i vari tipi di

⁴ Questa specificazione richiama la formula della correlazione:

$$\text{corr}(x, y) = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}} \Rightarrow \text{cov}(x, y) = \sqrt{\text{var}(x)} \cdot \text{corr}(x, y) \cdot \sqrt{\text{var}(y)}$$

formulazione GARCH, per poi passare al calcolo della correlazione condizionata del modello multivariato.

Alcune differenze tra questa formulazione e le precedenti sono:

- in primo luogo l'assunzione di costanza della correlazione tra le variabili
- il numero totale dei parametri da stimare è ora $(p + q + 1) \cdot k + \frac{n(n+1)}{2}$ che nel caso bivariato si riduce a 7
- la matrice di varianza e covarianza che deve essere definita positiva è ora controllata dalla matrice di correlazione. Inoltre la matrice di correlazione può essere fattorizzata come segue:

$$R = \text{diag} \left(\sqrt{\gamma_{1,1}} \cdots \sqrt{\gamma_{k,k}} \right) \Gamma \Gamma' \left(\sqrt{\gamma_{1,1}} \cdots \sqrt{\gamma_{k,k}} \right)$$

dove le matrici interne assicurano che R sia definita positiva.

4. Il modello DCC

4.1 Specificazione e caratteristiche del modello DCC

Supponiamo che i rendimenti possano essere come:

$$r_{i,t} = \varepsilon_{i,t} = \sigma_{i,t} \cdot v_{i,t}$$

dove $v_{i,t} \sim iid(0,1)$

Mostriamo ora la relazione che esiste tra la varianza condizionata e la correlazione condizionata.

$$\begin{aligned} \rho_{12,t} &= \frac{E(r_{1,t} \cdot r_{2,t} | I_{t-1})}{\sqrt{E(r_{1,t}^2 | I_{t-1}) \cdot E(r_{2,t}^2 | I_{t-1})}} = \\ &= \frac{E(\sigma_{1,t} \cdot v_{1,t} \cdot \sigma_{2,t} \cdot v_{2,t} | I_{t-1})}{\sqrt{E(\sigma_{1,t}^2 \cdot v_{1,t}^2 | I_{t-1}) \cdot E(\sigma_{2,t}^2 \cdot v_{2,t}^2 | I_{t-1})}} = \\ &= \frac{\sigma_{1,t} \cdot \sigma_{2,t} \cdot E(v_{1,t} \cdot v_{2,t} | I_{t-1})}{\sigma_{1,t} \cdot \sigma_{2,t} \cdot \sqrt{E(v_{1,t}^2 | I_{t-1}) \cdot E(v_{2,t}^2 | I_{t-1})}} = \end{aligned}$$

Semplificando numeratore e denominatore arriviamo alla seguente equazione.

$$= \frac{E\left(\nu_{1,t} \cdot \nu_{2,t} \mid \mathbf{I}_{t-1}\right)}{\sqrt{E\left(\nu_{1,t}^2 \mid \mathbf{I}_{t-1}\right) \cdot E\left(\nu_{2,t}^2 \mid \mathbf{I}_{t-1}\right)}}$$

Siamo giunti a dire che la correlazione condizionata tra $r_{1,t}$ e $r_{2,t}$ è la stessa di quella tra $\nu_{1,t}$ e $\nu_{2,t}$.

Analizziamo ora $\nu_{i,t}$. Esso si può anche scrivere $\nu_{i,t} = \frac{r_{i,t}}{\sigma_{i,t}}$ che non sono altro che i rendimenti standardizzati.

$$\begin{aligned} \nu_{i,t} & \text{ iid } (0,1) \\ E\left(\nu_{i,t}^2 \mid \mathbf{I}_{t-1}\right) &= \text{var}\left(\nu_{i,t} \mid \mathbf{I}_{t-1}\right) = \text{var}\left(\nu_{i,t}\right) = 1 \end{aligned}$$

E' del tutto naturale che la loro distribuzione abbia media nulla e varianza 1, visto che sono rendimenti standardizzati.

$$\rho_{12,t} = E\left(\nu_{1,t} \cdot \nu_{2,t} \mid \mathbf{I}_{t-1}\right)$$

Questa è la correlazione condizionata anche se sembra essere la covarianza condizionata. Solo il risultato è uguale visto che la varianza della serie è 1, al denominatore avrei la radice della deviazione standard che sappiamo essere 1.

Indichiamo con $q_{ij,t}$ una stima della covarianza condizionata, $E\left(\nu_{1,t} \cdot \nu_{2,t} \mid \mathbf{I}_{t-1}\right)$.

Una stima di $\rho_{12,t}$ è:

$$\hat{\rho}_{12,t} = \frac{q_{12,t}}{\sqrt{q_{11,t} \cdot q_{22,t}}}$$

N.B. La correlazione condizionata è nuovamente ritornata alla sua forma, questo perché a livello campionario, non necessariamente $q_{ii,t}$ è 1, mentre è vero che $E\left(q_{ii,t}\right) = 1$.

Se noi abbiamo più di due serie, la correlazione diventa una matrice i cui elementi sono le correlazioni tra le coppie di serie.

$$\hat{R}_t = \left[\hat{q}_{ij} \right]_t$$

Il nostro compito è di stimare

$$q_{ij,t}$$

Chiamiamo $z_{i,t}$ i rendimenti standardizzati con una stima di $\sigma_{i,t}$.

$$z_{i,t} = \frac{r_{i,t}}{\hat{\sigma}_{i,t}}$$

Per ogni stima di $\hat{\sigma}_{i,t}$, quindi devo aver già trovato un modello GARCH univariato per ogni singola serie dei dati.

Engle propone di utilizzare la classica struttura di tipo GARCH.

$$\begin{aligned} q_{ij,t} &= \bar{\rho}_{ij} + \alpha \left(z_{i,t-1} z_{j,t-1} - \bar{\rho}_{ij} \right) + \beta \left(q_{ij,t-1} - \bar{\rho}_{ij} \right) \\ &= \bar{\rho}_{ij} (1 - \alpha - \beta) + \alpha \left(z_{i,t-1} z_{j,t-1} \right) + \beta q_{ij,t-1} \\ &= \omega_{ij} + \alpha \left(z_{i,t-1} z_{j,t-1} \right) + \beta q_{ij,t-1} \end{aligned}$$

con $\bar{\rho}_{ij}$ correlazione non condizionata tra r_i e r_j .

Le covarianze vengono modellate individualmente come un GARCH, con parametri α e β comuni ma differenti ω . La correlazione condizionata è di tipo dinamico, ma la dinamica è la stessa per tutte le coppie di serie, ecco l'elemento di rigidità.

Il modello multivariato GARCH che proponiamo assume che la distribuzione dei k rendimenti condizionati siano normali multivariate con media zero e matrice di covarianza H_t .

$$r_t | I_{t-1} \sim N(0, H_t)$$

e

$$H_t \equiv D_t R_t D_t$$

dove D_t è una matrice diagonale $k \times k$ che sulla diagonale ha i valori della deviazione standard calcolati da un modello GARCH univariato e R_t è una matrice di correlazione.

La log-verosimiglianza di questo stimatore può essere scritta:

$$\begin{aligned} L &= -\frac{1}{2} \sum_{t=1}^T \left[k \log(2\pi) + \log(|H_t|) + r_t' H_t^{-1} r_t \right] \\ &= -\frac{1}{2} \sum_{t=1}^T \left[k \log(2\pi) + \log(|D_t R_t D_t|) + r_t' D_t^{-1} R_t^{-1} D_t^{-1} r_t \right] \\ &= -\frac{1}{2} \sum_{t=1}^T \left[k \log(2\pi) + 2 \log(|D_t|) + \log(|R_t|) + \varepsilon_t' R_t^{-1} \varepsilon_t \right] \end{aligned}$$

dove $\varepsilon_t \sim N(0, R_t)$ sono i residui standardizzati dalla loro deviazione standard condizionata.

Proponiamo di scrivere gli elementi di D_t come un modello GARCH univariato, così che:

$$h_{it} = \omega_i + \sum_{p=1}^{P_i} \alpha_{ip} r_{i,t-p}^2 + \sum_{q=1}^{Q_i} \beta_{iq} h_{i,t-q}$$

con le usuali restrizioni del GARCH per la non negatività e la stazionarietà inizialmente imposta.

$$\sum_{p=1}^{P_i} \alpha_{ip} + \sum_{q=1}^{Q_i} \beta_{iq} < 1$$

Gli indici sono presenti in ogni singola P e Q per ogni serie a indicare che la lunghezza dei ritardi scelti non deve necessariamente essere la stessa.

La specificazione di un modello GARCH univariato non è limitata allo standard GARCH(p,q), ma può includere altre specificazioni di modelli GARCH con errori normalmente distribuiti che soddisfino appropriatamente le condizioni di stazionarietà e la costrizione di non

negatività. Per esempio, si può usare un TARARCH per catturare gli effetti asimmetrici nella volatilità o un APARCH per riconoscere la lunga memoria del processo della volatilità.

La proposta della struttura della correlazione dinamica è:

$$Q_t = \left(1 - \sum_{m=1}^M \alpha_m - \sum_{n=1}^N \beta_n\right) \bar{Q} + \sum_{m=1}^M \alpha_m (\varepsilon_{t-m} \varepsilon_{t-m}') + \sum_{n=1}^N \beta_n Q_{t-n}$$

$$R_t = Q_t^{*-1} Q_t^* Q_t^{*-1}$$

dove \bar{Q} è la matrice di covarianza non condizionata dei residui standardizzati risultanti dal primo passo di stima, e

$$Q_t^* = \begin{pmatrix} \sqrt{q_{11,t}} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{q_{kk,t}} \end{pmatrix}$$

cioè Q_t^* è una matrice diagonale composta dalle radici degli elementi diagonali di Q_t .

Il tipico elemento di R_t dovrebbe essere nella forma

$$\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}$$

I seguenti vantaggiosi risultati dell'algebra lineare semplificheranno la ricerca delle condizioni affinché R_t ⁵ sia definita positiva.

⁵ La matrice di correlazione è reale, semi-definita positiva, simmetrica, con 1 nella diagonale principale.

Teorema 1

Sia A una matrice quadrata simmetrica e reale.

A è definita positiva se e solo se $B = A^{*-1}AA^{*-1}$, è definita positiva, dove A^* è una matrice diagonale composta dalle radici degli elementi della matrice A .

Il teorema 1 stabilisce che per definire positiva la matrice di correlazione R_t , dobbiamo solo assicurare che la matrice di covarianza Q_t sia definita positiva. Applicando questo teorema, possiamo descrivere una lista di condizioni sufficienti affinché H_t sia definita positiva uniformemente.

Assumiamo che A sia definita positiva. Ancora, $B = A^{*-1}AA^{*-1}$. Poiché A è reale, simmetrica e definita positiva, noi sappiamo che esiste la fattorizzazione di Cholesky che scompone la matrice attraverso una matrice P

triangolare superiore $A = P'P$.

Riscriviamo allora $B = (A^{*-1})' P'PA^{*-1} = (PA^{*-1})' (PA^{*-1})$

Sappiamo che (PA^{*-1}) ha rango pieno visto che è il prodotto tra una matrice diagonale e una matrice triangolare, entrambe con elementi nella diagonale diversi da zero, e sarà necessariamente una matrice triangolare con elementi nella diagonale diversi da zero.

Usando un risultato familiare dell'algebra lineare, che XX' è definita positiva se e solo se X ha rango pieno, possiamo concludere che B è definita positiva.

Assumendo B reale, simmetrica e definita positiva, possiamo sostituire direttamente A con B , allora abbiamo $A = A^*BA^*$.

Teorema 2

Se le seguenti restrizioni sui parametri del GARCH univariato sono soddisfatte per ogni serie di rendimenti, con $i \in [1 \dots k]$:

- a) $\omega_i > 0$
- b) $\alpha_{ip} \forall p \in [1 \dots P_i]$ e $\beta_{iq} \forall q \in [1 \dots Q_i]$ sono tali che h_{it} sia positiva con probabilità 1
- c) $h_{i0} > 0$
- d) le radici di $1 - \sum_{p=1}^{P_i} \alpha_{ip} Z^p + \sum_{q=1}^{Q_i} \beta_{iq} Z^q$ siano esterne al cerchio unitario

e i parametri del DCC soddisfino:

- e) $\alpha_m \geq 0 \quad \forall m \in [1 \dots M_i]$
- f) $\beta_n \geq 0 \quad \forall n \in [1 \dots N_i]$
- g) $\sum_{m=1}^M \alpha_m + \sum_{n=1}^N \beta_n < 1$
- h) L'autovalore minimo compreso tra $\bar{R} > \delta > 0$

allora H_t può essere definita positiva $\forall t$.

Ogni h_{it} deve essere strettamente positivo, visto che è una somma di ω , un parametro strettamente positivo e di α e β tali che

$$\sum_{p=1}^{P_i} \alpha_{ip} r_{i,t-p}^2 + \sum_{q=1}^{Q_i} \beta_{iq} h_{i,t-q}$$

la quale è non negativa con probabilità 1.

(Q_t) sarà quindi definita positiva per ogni t visto che è una media ponderata di una matrice definita positiva (\bar{Q}) , una matrice semi-definita positiva $(\varepsilon_{t-1} \varepsilon'_{t-1})$ e una matrice definita positiva con parametri non negativi (Q_{t-i}) , e sapendo che Q_0 è definita positiva **dall'assunzione h**.

Dalla proposizione 1, R_t deve essere positiva, e $H_t = D_t R_t D_t$ deve essere positiva definita come prodotto di 3 matrici definite positive.

Essenzialmente i requisiti affinché la covarianza condizionata sia definita positiva sono gli stessi per il modello DCC e per il modello GARCH univariato.

Le restrizioni nei parametri nella proposizione 2 non sono necessari, solo sufficienti per garantire che H_t sia definita positiva.

4.2 Stima e standard errors

Il modello DCC può essere stimato con una procedura a due passi, dove il primo passo è la stima di un modello GARCH univariato per ogni serie di residui, nel secondo passo, i residui, trasformati dalla loro deviazione standard stimata durante il primo passo, sono usati per stimare i parametri della correlazione condizionata.

La verosimiglianza usata nel primo passo, comporta la sostituzione di R_t con I_k , una matrice identità di dimensione k .

Siano i parametri del modello, ϑ , scritti in due gruppi $(\phi_1, \dots, \phi_k, \Psi) = (\phi, \Psi)$ dove gli elementi di ϕ_i corrispondono ai parametri del modello GARCH univariato per i rendimenti delle serie, $\phi_i = (\omega, \alpha_{1i}, \dots, \alpha_{pi}, \beta_{1i}, \dots, \beta_{qi})$.

La funzione di quasi-verosimiglianza risultante dal primo stadio è:

$$\begin{aligned} QL_1(\phi | r_t) &= -\frac{1}{2} \sum_{t=1}^T \left\{ k \log(2\pi) + \log(|I_k|) + 2 \log(|D_t|) + r_t' D_t^{-1} I_k D_t^{-1} r_t \right\} \\ &= -\frac{1}{2} \sum_{t=1}^T \left\{ k \log(2\pi) + 2 \log(|D_t|) + r_t' D_t^{-2} r_t \right\} \\ &= -\frac{1}{2} \sum_{t=1}^T \left\{ k \log(2\pi) + \sum_{n=1}^k \left(\log(h_{it}) + \frac{r_{it}^2}{h_{it}} \right) \right\} \\ &= -\frac{1}{2} \sum_{n=1}^k \left\{ T \log(2\pi) + \sum_{t=1}^T \left(\log(h_{it}) + \frac{r_{it}^2}{h_{it}} \right) \right\} \end{aligned}$$

la quale è semplicemente la somma della log-verosimiglianza di un singolo modello GARCH univariato per ogni serie di rendimenti.

Il secondo passo è una stima usando la verosimiglianza, condizionandola con i parametri stimati nella prima verosimiglianza:

$$\begin{aligned} QL_1 \left(\Psi \mid \hat{\phi}, r_t \right) &= -\frac{1}{2} \sum_{t=1}^T \left\{ k \log(2\pi) + 2 \log(|D_t|) + \log(|R_t|) + r_t' D_t^{-1} R_t^{-1} D_t^{-1} r_t \right\} \\ &= -\frac{1}{2} \sum_{t=1}^T \left\{ k \log(2\pi) + 2 \log(|D_t|) + \log(|R_t|) + \varepsilon_t' R_t^{-1} \varepsilon_t \right\} \end{aligned}$$

Visto che stiamo condizionando a $\hat{\phi}$, la sola parte di log-verosimiglianza che sarà influenzata dal parametro selezionato è $\log(|R_t|) + \varepsilon_t' R_t^{-1} \varepsilon_t$, e nella stima dei parametri del DCC, è spesso più semplice escludere i termini costanti e semplicemente massimizzare:

$$QL_2^* \left(\Psi \mid \hat{\phi}, r_t \right) = -\frac{1}{2} \sum_{t=1}^T \left\{ \log(|R_t|) + \varepsilon_t' R_t^{-1} \varepsilon_t \right\}$$

Nel 1994 White diede una dimostrazione della distribuzione asintotica per lo stimatore QMLE a due stadi.

La dimostrazione per la consistenza e la normalità asintotica dei parametri stimati dal modello DCC comprende i seguenti risultati. Il seguente set di assunzioni sono sufficienti a stabilire la consistenza dei parametri stimati usando questa procedura a due stadi, in aggiunta le assunzioni standard garantiscono la completezza dello spazio di probabilità e la misurabilità della funzione di quasi-verosimiglianza.

Assunzione 1

Per ogni ϕ in Φ , $E(\log f_1(r_t, \phi))$ esiste ed è finito per $t = 1, 2, \dots$,

$\{\log f_1(r_t, \phi)\}$ obbedisce alla legge forte dei grandi numeri.

Per ogni $\vartheta = (\phi, \psi)$ in $\Theta = \Phi \times \Psi$, $E(\log f_2(r_t, \vartheta))$ esiste ed è finito per $t = 1, 2, \dots$,

$\{\log f_2(r_t, \vartheta)\}$ obbedisce alla legge forte dei grandi numeri.

Assunzione 2

$\mathcal{G}_0 = (\phi_0, \psi_0)$ è identificabile unicamente, interna a $\Theta = \Phi \times \Psi$ uniformemente in n , Θ è compatto e \mathcal{G}_0 soddisfa le condizioni della proposizione 2, allora

$\left\{ \bar{L}_{1T}(\phi) = E \left(T^{-1} \sum_{t=1}^T \log f_1(r_t, \phi) \right) \right\}$ è $O(1)$ uniformemente in Φ

$\left\{ \bar{L}_{2T}(\mathcal{G}) = E \left(T^{-1} \sum_{t=1}^T \log f_2(r_t, \mathcal{G}) \right) \right\}$ è $O(1)$ uniformemente in Θ

Assunzione 3

Per ogni ϕ in Φ , $\nabla \bar{L}_{1T}(\phi) = E \left(\nabla L_{1T}(r^T, \phi) \right) < \infty$, dove $r^T = (r_1, r_2, \dots, r_T)$ è un vettore di osservazioni T-dimensionale.

Per ogni \mathcal{G} in Θ , $\nabla \bar{L}_{2T}(\mathcal{G}) = E \left(\nabla L_{2T}(r^T, \mathcal{G}) \right) < \infty$

Assunzione 4

Per ogni ϕ in Φ , $\nabla^2 \bar{L}_{1T}(\phi) = E \left(\nabla^2 L_{1T}(r^T, \phi) \right) < \infty$

$E \left(\nabla^2 L_{1T}(r^T, \cdot) \right)$ è continuo rispetto a Φ uniformemente in $T = 1, 2, \dots$

$\left\{ \nabla^2 \log f_1(r_t, \phi) \right\}$ obbedisce alla legge forte dei grandi numeri.

Per ogni \mathcal{G} in Θ , $\nabla^2 \bar{L}_{2T}(\mathcal{G}) = E \left(\nabla^2 L_{2T}(r^T, \mathcal{G}) \right) < \infty$

$E \left(\nabla^2 L_{2T}(r^T, \cdot) \right)$ è continuo rispetto a Θ uniformemente in $T = 1, 2, \dots$

$\left\{ \nabla^2 \log f_2(r_t, \mathcal{G}) \right\}$ obbedisce alla legge forte dei grandi numeri.

Assunzione 5

$\left\{ A_{11,T} = \nabla_{\phi\phi} \bar{L}_{1T}(\phi_0) \right\}$ è $O(1)$ e uniformemente negativa definita

$\left\{ A_{22,T} = \nabla_{\psi\psi} \bar{L}_{2T}(\mathcal{G}_0) \right\}$ è $O(1)$ e uniformemente negativa definita

Teorema 3

Sotto le assunzioni 1-5, $\hat{\phi}_T \xrightarrow{p} \phi_0$ e $(\hat{\phi}_T, \hat{\psi}_T) = \hat{\mathcal{G}}_T \xrightarrow{p} \mathcal{G}_0$.

Le condizioni per la consistenza sono molto deboli e saranno soddisfatte da numerosi processi generatori di dati.

Aggiungeremo ora una condizione sufficiente per la normalità asintotica dei parametri stimati. Per raggiungere questo risultato, dobbiamo necessariamente inserire un'altra assunzione.

Assunzione 6

$\{T^{-1/2}\nabla'_{\phi} \log f_1(r_t, \phi_0), T^{-1/2}\nabla'_{\psi} \log f_2(r_t, \phi_0, \psi_0)\}$ obbedisce alle condizioni del limite centrale con matrice di covarianza B_{0T} , e B_{0T} è $O(1)$ e uniformemente definita positiva.

Usando queste assunzioni, il seguente teorema stabilisce la distribuzione asintotica dello stimatore a due stadi del modello DCC.

Teorema 4

Sotto le assunzioni 1-6, per f_1 e f_2 ,

$$\sqrt{T} \left(\hat{\mathcal{G}}_T - \hat{\mathcal{G}}_0 \right) \overset{A}{\rightarrow} N \left(0, A_0^{-1} B_0 A_0'^{-1} \right)$$

dove

$$A_0 = \begin{pmatrix} \nabla_{\phi\phi} \log f_1(\phi_0) & 0 \\ \nabla_{\phi\psi} \log f_2(\mathcal{G}_0) & \nabla_{\psi\psi} \log f_2(\mathcal{G}_0) \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ A_{12} & A_{22} \end{pmatrix}$$

e

$$B_0 = \text{var} \left[\sum_{t=1}^T \left\{ T^{-1/2} \nabla'_{\phi} \log f_1(r_t, \phi_0), T^{-1/2} \nabla'_{\psi} \log f_2(r_t, \phi_0, \psi_0) \right\} \right] = \begin{pmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{pmatrix}$$

Dal teorema segue che la varianza asintotica di $\hat{\mathcal{G}}_n$ è data da $A_0^{-1}B_0A_0^{-1}$.

Applicando il teorema per la partizione⁶ di matrici quadrate inverse, la varianze asintotiche dei parametri GARCH per ogni serie, sono stimatori robusti di matrici di covarianza date da $A_{11}^{-1}B_{11}A_{11}^{-1}$.

La varianza asintotica dei parametri del modello DCC del secondo stadio, è comunque una formula molto più complessa, dovendo utilizzare tutti i termini della covarianza.

In aggiunta, avendo modificato gli standard errors, il test rapporto di verosimiglianza con r restrizioni, non sarà più il tipico χ_r^2 .

Sia Foutz e Srivastave (1977) e Liang e Self (1996) hanno discusso il rapporto di verosimiglianza quando le distribuzioni assunte sono incorrette o il modello è stato stimato in più passi.

Entrambi hanno dimostrato che quando non viene presa l'Information Matrix Equality, $A_0 - B_0 \xrightarrow{p} 0$, allora la distribuzione asintotica sarà una somma pesata di r variabili indipendenti χ_1^2 dove i pesi non saranno necessariamente uguali a uno.

In questo caso, dove la distribuzione dei parametri non è interamente conosciuta, Foutz e Srivastave hanno mostrato che :

$$H_0 : \mathcal{G} \in \Theta_r \quad H_1 : \mathcal{G} \in \Theta$$

$$-2\ln \lambda = -2\ln \frac{\max_{\mathcal{G} \in \Theta} \left\{ \prod_{i=1}^T f(r_i, \mathcal{G}) \right\}}{\max_{\mathcal{G} \in \Theta_r} \left\{ \prod_{i=1}^T f(r_i, \mathcal{G}) \right\}} \xrightarrow{d} C$$

dove

$$C \square c(\mathcal{G}_0)_1 \chi_1^2 + c(\mathcal{G}_0)_2 \chi_1^2 + \dots + c(\mathcal{G}_0)_r \chi_1^2$$

⁶ $A_{11}^{-1}B_{11}A_{11}^{-1}$ sono blocchi di matrici diagonali, con la matrice di covarianza per il modello GARCH univariato i^{th} nel i^{th} blocco diagonale.

dove $c(\mathcal{G}_0)$ è l' i^{th} autovalore di $W(\mathcal{G}_0)M(\mathcal{G}_0)$, $W(\mathcal{G}_0) = (-\tilde{A}_1 + \tilde{A}_2\tilde{A}_4^{-1}\tilde{A}_3)$ e $M(\mathcal{G}_0)$ è la matrice superiore $r \times r$ di $\tilde{A}^{-1}\tilde{B}\tilde{A}^{-1}$, dove \tilde{A} è A , come definita nel teorema 4, con le righe e le colonne intercambiate, così che le prime righe e colonne di \tilde{A} corrispondano agli standard errors dei parametri inizialmente testati e \tilde{B} è similmente definita.

4.3 Test per la correlazione costante

Una delle motivazioni primarie di questa tesi è che la correlazione tra serie di rendimenti non è costante nel tempo.

Testare la costanza della correlazione non fu una cosa facile, anzi, si rivelò qualcosa di complesso, visto che a volte veniva rifiutata l'ipotesi di costanza della correlazione anche quando questo era vero e viceversa, sbagliando il modello per la volatilità.

Per primi, Tse (1998) e Bera (1996) iniziarono questi test con ipotesi nulle e alternative che spaziavano molto. Per esempio Tse abbozzò una nulla di correlazione condizionata costante contro l'alternativa di un modello ARCH in correlazione, ma comunque le due ipotesi fondamentali restavano quella nulla di correlazione condizionata costante contro l'alternativa di correlazione condizionata dinamica.

La difficoltà iniziale di questi test fu quella di generalizzarli per dimensioni molto grandi. Proponiamo allora un test che richiede solo la stima consistente della correlazione condizionata costante, e può essere implementato usando un vettore autoregressivo.

Un'altra difficoltà significativa nella conduzione del test, sono i parametri non significativi nello stimatore DCC. Essi non sono identificati sotto l'ipotesi nulla e devono quindi essere trattati come parametri di disturbo.

Ci sono due vie per trattare ciò. Una è quella di testare solamente la nulla contro l'alternativa con uno specifico coefficiente β . Questo test può essere condotto usando il test standard di verosimiglianza con le usuali proprietà del LR, comunque può mancare di potenza se i coefficienti scelti per β sono lontani dal vero valore. Comunque ciò non è restrittivo per il test visto che non è necessario identificare β .

Andrews e Ploberger (1994) hanno stabilito una procedura dalla quale il test può essere condotto senza identificare i parametri e solo recentemente questa struttura può essere estesa ai casi dove, i parametri non identificati sotto la nulla, possono essere sul limite dello spazio dei parametri.

Implementare questo tipo di test è molto difficile, poiché è richiesta una buona ottimizzazione e un valore critico di Monte Carlo.

Il test che proponiamo è:

$$H_0 : R_t = R \quad \forall t \in T$$

contro

$$H_1 : vech^u(R_t) = vech^u(\bar{R}) + \beta_1 vech^u(R_{t-1}) + \dots + \beta_p vech^u(R_{t-p})$$

dove $vech^u$ è una modifica all'operatore $vech$ che seleziona solo gli elementi della diagonale principale.

La procedura per il test è la seguente.

Stimare un processo GARCH univariato e i residui standardizzati per ogni serie. Poi stimare la correlazione dei residui standardizzati e congiuntamente standardizzare il vettore dei residui standardizzati univariati usando la decomposizione in radice quadrata di \bar{R} .

Sotto la nulla cioè con correlazione costante, questi residui devono essere IID con matrice di covarianza data da I_k .

Sia $Y_t = vech^u \left[\left(\bar{R}^{-1/2} D_t^{-1} r_t \right) \left(\bar{R}^{-1/2} D_t^{-1} r_t \right)' - I_k \right]$ dove $\left(\bar{R}^{-1/2} D_t^{-1} r_t \right)$ è un vettore $k \times 1$ dei residui congiuntamente standardizzati sotto la nulla. Il vettore autoregressivo è

$$Y_t = \alpha + \beta_1 Y_{t-1} + \dots + \beta_s Y_{t-s} + \eta_t.$$

Sotto la nulla, l'intercetta e tutti i parametri dei ritardi del modello devono essere zero.

Nell'ordine per ottenere la statistica test, tutto ciò che è necessario fare è produrre il vettore $T \times 1$ di prodotti esterni per ogni regredendo univariato e

una matrice $T \times (s + 1)$ di regressori includendo la costante per ogni insieme di regressori.

Poi i parametri possono essere stimati dall'insieme dei $k(k-1)/2$ vettori di regressori e regressori e compiendo una apparentemente incorrelata regressione.

Il test può essere condotto poiché $\frac{\hat{\delta}' X' X \hat{\delta}}{\hat{\sigma}^2}$ è asintoticamente distribuito come $\chi^2_{(s+1)}$, dove $\hat{\delta}$ sono la stima dei parametri di regressione e X è una matrice di regressori.

In ogni modello considerato, è stata rifiutata l'ipotesi nulla di correlazione costante in favore di una struttura dinamica.

4.4 Previsione a più passi avanti

La previsione della covarianza è una richiesta dei modelli GARCH.

Molti modelli GARCH prevedono un semplice metodo di implementazione per generare previsioni a r-passi avanti.

Per citare un esempio, la previsione a r-passi avanti di uno standard GARCH(1,1) è data da:

$$h_{t+r} = \sum_{i=0}^{r-2} \omega (\alpha + \beta)^i + (\alpha + \beta)^{r-1} h_{t+1}$$

Comunque, il processo di evoluzione DCC è un processo non lineare, dove

$$Q_{t+r} = (1 - \alpha - \beta) \bar{Q} + \alpha (\varepsilon_{t+r-1} \varepsilon'_{t+r-1}) + \beta Q_{t+r-1}$$

dove

$$E[\varepsilon_{t+r-1} \varepsilon'_{t+r-1}] = E[R_{t+r-1}]$$

e

$$R_{t+r} = Q_{t+r}^{*-1} Q_{t+r} Q_{t+r}^{*-1}$$

Così, la previsione a r-passi avanti della correlazione non può essere direttamente risolta.

Nell'esaminare il metodo vengono superate queste difficoltà, due previsioni sembrano essere le più naturali, solo richiedono un differenti insiemi di approssimazioni.

La prima tecnica proposta dovrebbe generare una previsione a r-passi avanti di Q da una approssimazione, cioè

$$E_t[\varepsilon_{t+i} \varepsilon'_{t+i}] \approx Q_{t+i}$$

per $i \in [1, \dots, r]$.

Usando questa approssimazione, abbiamo che la previsione a r-passi di Q_t è:

$$E_t [Q_{t+r}] = \sum_{i=0}^{r-2} (1-\alpha-\beta)\bar{Q}(\alpha+\beta)^i + (\alpha+\beta)^{r-1} Q_{t+1}$$

e $R_{t+r} = Q_{t+r}^{*-1} Q_{t+r} Q_{t+r}^{*-1}$.

Una approssimazione alternativa potrebbe essere $\bar{Q} \approx \bar{R}$ e $E_t [Q_{t+1}] \approx E_t [R_{t+1}]$.

Usando questa approssimazione, possiamo prevedere R_{t+r} direttamente usando la relazione

$$E_t [R_{t+r}] = \sum_{i=0}^{r-2} (1-\alpha-\beta)\bar{R}(\alpha+\beta)^i + (\alpha+\beta)^{r-1} R_{t+1}.$$

E' stato condotto un esperimento (Monte Carlo) per vedere quali di queste approssimazioni mostra la performance migliore.

Sono stati simulati 1000 giorni da una normale bivariata usando $\alpha = 0.01$ e $\beta = 0.98$ e variando la correlazione non condizionata internamente il set $[-0.8, -0.5, 0, 0.35, 0.5, 0.65, 0.8, 0.9]$.

Alla 750-esima osservazione, Q_{750} , r_{750} , R_{750} sono state salvate. La previsione è allora costruita per il giorno 751 fino a 1000 usando i due metodi descritti. Sono state studiate queste due tecniche di previsione trattando $\hat{\alpha}, \hat{\beta}$ e \bar{R} come parametri conosciuti, inserendo solo parametri usati per simulare i dati, per evitare effetti di parametri stimati senza certezza. Vale la pena di notare che questi esperimenti sarebbero stati esattamente gli stessi avendo permesso una volatilità variabile nel tempo, visto che l'errore della volatilità sarebbe stato uguale per entrambe le tecniche.

La prima osservazione che è stato possibile fare è che entrambe le tecniche di previsione producono una previsione con una distorsione veramente

bassa. Questa distorsione è generalmente tendente a 1 (oppure a -1 per serie con correlazione non condizionata negativa).

La previsione prodotta risolvendo per Q_{t+r} (linea tratteggiata) era sempre più vicina ad 1 (o -1) e conseguentemente più distorta.

Ancora di interesse è che entrambe le previsioni appaiono non essere distorte quando la correlazione non condizionata è zero, e che essenzialmente portano allo stessa previsione quando la correlazione non condizionata è zero.

5. Implementazione del software e applicazioni

5.1 Il software per il modello DCC

Piuttosto che definire R come software statistico, esso deve essere definito come un ambiente, ovvero un insieme di macro, librerie, oggetti che possono essere utilizzati per la gestione, l'analisi dei dati e la produzione di grafici.

R è basato sul linguaggio S a cui è strettamente legato un altro 'ambiente' commerciale, S-Plus. R a differenza di S-Plus, è un GNU-Software, ovvero disponibile gratuitamente sotto i vincoli della GPL (General Public Licence).

Il programma che vedremo in seguito, è stato scritto in questo linguaggio, e non mettiamo in dubbio che necessiti di miglioramenti.

Operativamente il programma deve eseguire le seguenti funzioni:

- Stima di un modello GARCH univariato per ogni serie
- Ricavare da ogni serie i residui standardizzati $z_{i,t}$
- A partire dalle serie dei residui standardizzati si stima il modello per le correlazioni
- Per ogni coppia $z_{i,t}$ e $z_{j,t}$ si calcoli la correlazione non condizionata $\bar{\rho}_{ij}$
- A partire dal modello $\bar{\rho}_{ij}(1-\alpha-\beta) + \alpha(z_{i,t-1}z_{j,t-1}) + \beta q_{ij,t-1}$ si stimi $\phi = (\alpha, \beta)$
- Massimizzare $L(\hat{\theta}, \phi) = -\frac{1}{2} \sum_{t=1}^T \{ \log(|R_t|) + \varepsilon_t' R_t^{-1} \varepsilon_t \}$

Prima di iniziare ad analizzare ogni singola parte del programma, definiamo alcune variabili che vengono utilizzate all'interno del suddetto.

n è la lunghezza delle serie che analizziamo (cioè quella che solitamente viene indicata con T), m invece è il numero delle serie di rendimenti che modelliamo.

Siamo sicuri che il modo migliore per analizzare un programma sia quello di seguire passo passo il suo ciclo, e quindi sarà quello che faremo.

Innanzitutto viene richiesto il numero delle serie da utilizzare (m) e la lunghezza di esse (n^7).

```
print ( "Inserisci il numero di serie che vuoi
        analizzare con il modello DCC, poi premi Invio
        due volte" )
m = scan()
```

⁷ Il numero dei dati che dobbiamo analizzare, deve essere necessariamente uguale in tutte le serie, per questo bisogna che n sia un numero sempre minore della lunghezza minima delle serie.

```
print ( "Inserisci il numero delle osservazioni da  
      usare nell'analisi, poi premi Invio due volte" )  
n = scan()
```

Creiamo poi una matrice di dimensione $(n \times m)$ di zeri con il comando 'array'. Il passo successivo è quello di riempire la matrice *serie* con i rendimenti e per fare ciò abbiamo utilizzato un ciclo 'for' in modo da poter inserire i dati in successione.

La nostra matrice, è costruita in modo che in ogni colonna ci sia una serie di rendimenti.

```
serie = array ( 0,c (n,m) )  
print ( "Inserisci il percorso dei file in cui si  
trovano le serie dei dati" )  
  for ( i in 1:m )  
    {  
      print(i)  
      c = readline()  
      s = scan (c)  
      serie[:,i] = s[1:n]  
    }  
}
```

Dobbiamo trovare ora i residui standardizzati di ogni serie. Creiamo quindi una matrice z che naturalmente avrà n righe e m colonne, come la matrice *serie*. Utilizzando un 'for' abbiamo a questo punto tolto la propria media ad ogni serie di rendimenti, in modo tale da ricondurci alla formula

$$r_{i,t} = \varepsilon_{i,t} = \sigma_{i,t} \cdot U_{i,t}.$$

Resi così i rendimenti uguali ai residui, non dobbiamo far altro che modellarli con un GARCH univariato e recuperare i residui standardizzati. Utilizziamo allora una delle due funzioni presenti nel programma.

Una funzione non è altro che un programma all'interno di un altro programma, che deve svolgere una funzione ben determinata, in modo tale da poterla richiamare ogni volta che se ne trovi la necessità, senza appesantire il corpo centrale del programma. Una funzione quindi, essendo un programma, deve lavorare con una o più variabili, e deve restituire un risultato.

La funzione *unigarch* è una funzione nella variabile *y*, che non fa altro che calcolarmi un modello GARCH(1,1) per *y*, e i suoi residui standardizzati. Nel comando 'return' troviamo il valore che la funzione restituisce, che nel nostro caso sono appunto i residui standardizzati.

```
unigarch=function (y)
{
  garch.y=garch( y,order=c( 1,1), trace=F )
  resid.stand=residuals( garch.y, standardize=T )
  return(resid.stand)
}
```

Richiamando all'interno del ciclo for la funzione *unigarch*, dove al posto di *y* inserisco una colonna della matrice *serie* (che non è altro che una singola serie di rendimenti), riesco a riempire la matrice *z* (colonna dopo colonna) con i residui standardizzati. Da notare però che la prima riga della matrice *z* è stata posta uguale a zero.

```
z = array ( 0,c (n,m) )
for (i in 1:m)
{
  serie[,i] = serie[,i] - mean ( serie[,i] )
  z[,i] = unigarch ( serie[,i] )
  z[1,i]=0
}
```

Stimiamo la matrice di varianza e covarianza non condizionata e la matrice di correlazione non condizionata. Entrambe queste matrici sono di dimensione $(m \times m)$.

```
covnc = cov (z)
print (covnc)
corrnc = cor(z)
print (corrnc)
```

La seconda funzione che utilizziamo è la funzione *Rmatrix*, nella variabile *phi* (un vettore di dimensione (1×2)) e che restituisce il valore della log-verosimiglianza del modello DCC. Una nota da evidenziare è che le dimensioni del vettore *Phi* dipendono da che modello GARCH multivariato vado a stimare. Visto che nel nostro caso stimiamo un GARCH(1,1) ho solamente due parametri da stimare simultaneamente, cioè α e β .

Quindi *alpha* è il parametro della parte ARCH e *beta* il parametro della parte GARCH. Essendo però $\alpha \geq 0$ e $\beta \geq 0$ un vincolo del modello, devo necessariamente utilizzarli in valore assoluto.

```
Rmatrix=function(phi)
  alpha=abs(phi [1])
  beta=abs(phi [2])
```

Creiamo poi due array a tre dimensioni, *covc* e *r*, cioè una serie di lunghezza *n*, i cui elementi sono delle matrici di dimensione $(m \times m)$.

Covc sarà per noi la serie delle matrici di covarianza condizionata ed *r* la serie delle matrici di correlazione.

```
covc=array(0, c(m, m, n))
r=array( 0, c(m, m, n))
```

Non potendo calcolare la covarianza condizionata al periodo precedente i nostri dati, assegno alla prima matrice della *covc* la matrice di covarianza non condizionata.

```
covc[, , 1]=covnc
```

Calcoliamo poi la prima matrice di correlazione condizionata utilizzando una semplice formulazione matriciale che riesce a velocizzare⁸ il calcolo.

```
r[, , 1]=covc[, , 1] / (diag(covc[, , 1]) %*%
    (t(diag(covc[, , 1]))))
```

Il comando ‘diag’ estrae gli elementi della diagonale principale e li inserisce in un vettore colonna, anche se apparentemente sembra essere un vettore riga. Al denominatore abbiamo così costruito una matrice ($m \times m$).

In R, dividere una matrice per un’altra delle stesse dimensioni, significa dividere ogni elemento della matrice al numeratore per l’elemento della matrice al denominatore che sta nella stessa posizione.

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} / \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}/b_{11} & a_{12}/b_{12} \\ a_{21}/b_{21} & a_{22}/b_{22} \end{pmatrix}$$

Utilizzando dei cicli ‘for’ concatenati tra loro, calcoliamo tutte le altre matrici di covarianza condizionata, elemento per elemento. Sempre all’interno di uno dei cicli ‘for’, per l’esattezza il ciclo del tempo, calcoliamo le matrici di correlazione condizionata.

Chiusi questi cicli, abbiamo ora tutto quello che ci serve per calcolare la verosimiglianza che poi dovrà essere massimizzata.

```
for ( t in 2:n )
{
  for ( i in 1:m )
  {
    for ( j in 1:m )
```

⁸ Una delle difficoltà dei modelli GARCH multivariati quando si utilizzano un grande numero di serie, è proprio quello di riuscire a semplificare e velocizzare il più possibile i calcoli di stima.

```

        {
            covc[i,j,t]=corrnc[i,j] * ( 1-
            alpha-beta ) + alpha * ( z[t-1,i]
            * z[t-1,j] ) + beta * (
            covc[i,j,t-1] )
        }
    }
    r[:,t]=covc[:,t]/( (diag ( covc[:,t]
    )^(1/2) )   %% ( t ( ( diag (
    covc[:,t] ) )^(1/2) ) ) )
}

```

Prima di iniziare il calcolo della log-verosimiglianza, ricordiamo la sua espressione:

$$QL_2^* \left(\Psi \mid \hat{\phi}, r_t \right) = -\frac{1}{2} \sum_{t=1}^T \left\{ \log \left(|R_t| \right) + \varepsilon_t' R_t^{-1} \varepsilon_t \right\}$$

Per procedere creiamo un contatore, cioè un vettore nullo di dimensioni $(n \times 1)$, c e una serie di lunghezza n di matrici nulle di dimensione $(m \times m)$, $invr$.

La prima cosa da fare è calcolare l'inversa della matrice di correlazione condizionata R_t .

In R il comando 'ginv' calcola la matrice inversa generalizzata (Pseudo inversa di Moore-Penrose), perciò è sufficiente applicare un ciclo for per il tempo e dare il comando 'ginv' che mi inserisce le matrici inverse nella serie di matrici $invr$.

Esiste però un problema.

La pseudo inversa di Moore-Penrose, si può calcolare su tutte le matrici, tranne quelle che contengono forme indeterminate del tipo $0/0$ e ∞/∞ .

Nel nostro caso, può accadere che nel calcolo della matrice di correlazione condizionata, ci si trovi davanti ad una forma indeterminata (naturalmente

del tipo 0/0) e quindi per ovviare questo problema abbiamo dovuto inserire un 'if'. Questo comando, pone la condizione che se il determinante della matrice di correlazione condizionata è in forma indeterminata, al posto dell'inversa generalizzata, metta una matrice contenente tutti zeri, altrimenti proceda con il calcolo dell'inversa.

A questo punto è necessario inserire solo la formula per il calcolo della log-verosimiglianza. Essendo questa una sommatoria di log-verosimiglianze calcolate nella variabile tempo, utilizzando un for, le calcoliamo singolarmente, le inseriamo nel vettore *c* e contemporaneamente iniziamo a sommare gli elementi di questo vettore.

Conclusa questa operazione otteniamo la log-verosimiglianza che la funzione *Rmatrix* restituisce al programma principale.

```

lgv = 0
c = array ( 0,c(n,1) )
invr=array ( 0,c(m,m,n) )
for ( t in 1:n )
  {
    if ( is.nan ( det ( r[, ,t] ) ) )
      {
        invr[, ,t]=0
      }
    else
      {
        invr[, ,t] = ginv ( r[, ,t] )
      }
    c[t] = (-0.5) * ( log ( det ( r[, ,t] ) ) +
      ( z[t,] %*%
        ( invr[, ,t] ) %*% z[t,] ) )
    lgv = lgv + c[t]
  }
return ( lgv)
}

```

Il passo successivo è la massimizzazione per via numerica della verosimiglianza ottenuta dalla funzione precedente e per fare ciò utilizziamo il comando 'optim'.

Affinché questo comando abbia tutto ciò che serve per operare, dobbiamo inserire:

- il valore iniziale dei parametri da stimare
`par = c(0.1,0.1)`
- la funzione da massimizzare o minimizzare
`Rmatrix`
- il metodo di ottimizzazione
`method = "BFGS"`
- se vogliamo massimizzare o minimizzare (nel nostro caso massimizzare)
`control = list (fnscale = -1)`
- se voglio utilizzare la matrice hessiana
`hessian = T`

In poche parole, `optim` calcola la log-verosimiglianza con i parametri iniziali, vede se questa viene massimizzata, altrimenti cambia il valore dei parametri finché non trova quei valori che riescono a massimizzare la funzione. Terminata queste iterazioni, `optim` restituisce l'ultimo valore dato ai parametri, che come si può capire saranno $\hat{\alpha}$ e $\hat{\beta}$.

5.2 Applicazioni

Per testare la validità del programma è necessario fare una prova empirica, e per questo abbiamo utilizzato delle serie storiche che si riferiscono ai maggiori indici borsistici del mondo.

Nel nostro paniere troviamo:

- S&P500: è l'indice più rappresentativo del mercato azionario statunitense nel suo complesso. Esso include 500 Blue Chips quotati nei principali mercati azionari americani.
- Nikkey: è l'indice ufficiale della borsa giapponese, calcolato sulla base di due diversi panieri di titoli quotati prevalentemente nel maggiore mercato di Tokyo.
- FTSE100: è l'indice principale della borsa inglese, il cui mercato azionario principale (LSE) è naturalmente il più importante tra quelli europei. E' basato sulle prime 100 società in termini di capitalizzazione.
- DAX30: è l'indice principale della borsa tedesca (FWB). Esso sintetizza l'andamento dei 30 principali titoli.
- CAC40: è l'indice di riferimento della borsa di Parigi e sintetizza l'andamento dei 40 titoli principali.
- MIB30: sintetizza l'andamento complessivo dei 30 titoli più rappresentativi della borsa italiana.
- BEL20: è l'indice principale della borsa belga
- SWISS: sintetizza l'andamento complessivo della borsa svizzera

Tutte le serie di dati vanno dal 30 novembre 1995 al 5 febbraio del 2001 (fonte www.tesiinborsa.it).

Ogni serie è stata poi trasformata nella serie dei rendimenti utilizzando la seguente espressione:

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

ottenendo 1352 osservazioni per ogni serie.

Sono stati effettuati 4 esperimenti con un modello DCC(1,1), il primo utilizzando tutti gli indici disponibili, il secondo utilizzando solo i titoli europei, il terzo tra MIB30, S&P500 e NIKKEY, e il quarto tra MIB30 e S&P500.

Per ogni esperimento diamo la matrice di covarianza condizionata e la matrice di correlazione condizionata, e naturalmente la stima dei parametri $\hat{\alpha}$ e $\hat{\beta}$ con i relativi test t .

Per semplificare la notazione diamo l'ordine con cui sono stati inseriti gli indici nel programma: MIB30, FTSE100, DAX30, CAC40, BEL20, SWISS, NIKKEY, S&P500.

Matrice di covarianza non condizionata

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	0.9991262	0.5770975	0.5572096	0.6804129	0.4792212	0.5598921	0.1683808	0.3232057
[2,]	0.5770975	0.9906045	0.5912325	0.6944961	0.5284777	0.6155707	0.2423748	0.3888810
[3,]	0.5572096	0.5912325	0.9925209	0.6642497	0.5560543	0.6176683	0.2416616	0.3441782
[4,]	0.6804129	0.6944961	0.6642497	0.9973457	0.5600366	0.6314259	0.2321462	0.3831937
[5,]	0.4792212	0.5284777	0.5560543	0.5600366	0.9947088	0.5820413	0.2004987	0.3512379
[6,]	0.5598921	0.6155707	0.6176683	0.6314259	0.5820413	0.9968820	0.2147978	0.3405536
[7,]	0.1683808	0.2423748	0.2416616	0.2321462	0.2004987	0.2147978	0.9975831	0.0721193
[8,]	0.3232057	0.3888810	0.3441782	0.3831937	0.3512379	0.3405536	0.0721193	0.9955729

Matrice di correlazione non condizionata

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	1.0000000	0.5800813	0.5595496	0.6816156	0.4807042	0.5610122	0.16865837	0.32406509
[2,]	0.5800813	1.0000000	0.5962636	0.6987097	0.5323883	0.6194495	0.24381627	0.39158863
[3,]	0.5595496	0.5962636	1.0000000	0.6676344	0.5596279	0.6209600	0.24286405	0.34623979
[4,]	0.6816156	0.6987097	0.6676344	1.0000000	0.5622708	0.6332536	0.23273637	0.38455543
[5,]	0.4807042	0.5323883	0.5596279	0.5622708	1.0000000	0.5844992	0.20127460	0.35295302
[6,]	0.5610122	0.6194495	0.6209600	0.6332536	0.5844992	1.0000000	0.21539387	0.34184332
[7,]	0.1686584	0.2438163	0.2428641	0.2327364	0.2012746	0.2153939	1.0000000	0.07236698
[8,]	0.3240651	0.3915886	0.3462398	0.3845554	0.3529530	0.3418433	0.07236698	1.0000000

La stima del parametro ARCH è

$$\alpha = 0.018594$$

La stima del parametro GARCH è

$$\beta = 0.9534946$$

TEST t

7.904304 121.473765

Nella seconda prova viene mantenuto l'ordine dei titoli europei.

Matrice di covarianza non condizionata

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.9991262	0.5770975	0.5572096	0.6804129	0.4792212	0.5598921
[2,]	0.5770975	0.9906045	0.5912325	0.6944961	0.5284777	0.6155707
[3,]	0.5572096	0.5912325	0.9925209	0.6642497	0.5560543	0.6176683
[4,]	0.6804129	0.6944961	0.6642497	0.9973457	0.5600366	0.6314259
[5,]	0.4792212	0.5284777	0.5560543	0.5600366	0.9947088	0.5820413
[6,]	0.5598921	0.6155707	0.6176683	0.6314259	0.5820413	0.9968820

Matrice di correlazione non condizionata

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1.0000000	0.5800813	0.5595496	0.6816156	0.4807042	0.5610122
[2,]	0.5800813	1.0000000	0.5962636	0.6987097	0.5323883	0.6194495
[3,]	0.5595496	0.5962636	1.0000000	0.6676344	0.5596279	0.6209600
[4,]	0.6816156	0.6987097	0.6676344	1.0000000	0.5622708	0.6332536
[5,]	0.4807042	0.5323883	0.5596279	0.5622708	1.0000000	0.5844992
[6,]	0.5610122	0.6194495	0.6209600	0.6332536	0.5844992	1.0000000

La stima del parametro ARCH è

$$\alpha = 0.02588099$$

La stima del parametro GARCH è

$$\beta = 0.9481609$$

"TEST T"

8.456769 122.617109

Nella terza prova vengono utilizzate le serie nel seguente ordine: MIB30, S&P500 e NIKKEY.

Matrice di covarianza non condizionata

	[,1]	[,2]	[,3]
[1,]	0.9991262	0.3232057	0.1683808
[2,]	0.3232057	0.9955729	0.0721193
[3,]	0.1683808	0.0721193	0.9975831

Matrice di correlazione non condizionata

	[,1]	[,2]	[,3]
[1,]	1.0000000	0.32406509	0.16865837
[2,]	0.3240651	1.00000000	0.07236698
[3,]	0.1686584	0.07236698	1.00000000

La stima del parametro ARCH è

$$\alpha = 0.0002344196$$

La stima del parametro GARCH è

$$\beta = 0.03512647$$

Nell'ultima prova MIB30 e S&P500

Matrice di covarianza non condizionata

	[,1]	[,2]
[1,]	0.9991262	0.3232057
[2,]	0.3232057	0.9955729

Matrice di correlazione non condizionata

	[,1]	[,2]
[1,]	1.0000000	0.3240651
[2,]	0.3240651	1.0000000

La stima del parametro ARCH è

$$\alpha = 2.103714e-11$$

La stima del parametro GARCH è

$$\beta = 0.3314569$$

TEST T

7.389382e-09	6.479579e+12
--------------	--------------

Dopo aver visto i risultati dei nostri esperimenti dobbiamo solo verificare la loro attendibilità. Per fare ciò, abbiamo guardato delle stime di $\hat{\alpha}$ e $\hat{\beta}$ ottenute dal premio Nobel, Robert Engle.

Le prime stime si basano sullo studio dei titoli che compongono il Dow Jones Average Stocks, mentre le altre sono svolte sui titoli dello S&P.

Prima di trarre le conclusioni, c'è da notare che il professor Engle lavora con titoli appartenenti ad uno stesso mercato invece noi abbiamo lavorato con indici di mercato mondiali.

Non c'è molto da dire, visto che, abbiamo entrambi un parametro ARCH molto basso, $\alpha < 0.01$ per la maggior parte delle prove.

No.of Assets	$\hat{\alpha}$	$\hat{\beta}$
2	0.0104 (1.27)	0.9642 (43.58)
3	0.0038 (0.90)	0.9679 (44.96)
4	0.0048 (1.70)	0.9627 (81.11)
5	0.0042 (2.14)	0.9672 (95.01)
6	0.0043 (2.55)	0.9639 (65.33)
7	0.0061 (3.69)	0.9690 (68.47)
8	0.0061 (4.34)	0.9778 (61.85)
9	0.0061 (5.15)	0.9768 (58.43)
10	0.0066 (5.44)	0.9770 (55.87)
15	0.0048 (2.67)	0.9690 (54.66)
20	0.0044 (4.45)	0.9626 (76.16)
25	0.0047 (5.12)	0.9384 (60.12)
30	0.0043 (4.74)	0.9118 (26.97)
31	0.0059 (8.62)	0.9441 (88.50)

No.of Assets	$\hat{\alpha}$	$\hat{\beta}$
2	0.0335 (2.96)	0.9559 (60.20)
3	0.0266 (2.34)	0.9531 (38.13)
4	0.0163 (4.33)	0.9688 (112.83)
5	0.0123 (6.09)	0.9719 (168.13)
6	0.0103 (6.34)	0.9732 (172.13)
7	0.0101 (7.74)	0.9746 (224.44)
8	0.0097 (6.95)	0.9718 (164.23)
9	0.0090 (5.78)	0.9688 (116.99)
10	0.0255 (3.26)	0.9410 (28.85)
15	0.0176 (3.70)	0.9539 (43.10)
20	0.0133 (3.51)	0.9662 (57.48)
25	0.0097 (3.27)	0.9696 (60.69)
50	0.0072 (5.56)	0.9643 (85.47)
75	0.0052 (9.85)	0.9597 (98.65)
100	0.0049 (14.58)	0.9497 (154.53)

Il parametro GARCH, invece si mantiene sempre vicino a 1, $\beta > 0.97$ per gran parte delle prove effettuate da entrambi.

Possiamo concludere che per quanto possibile le stime ricavate dal nostro programma sembrano essere accettabili.

6 Conclusione

Stilando questa tesi, abbiamo passato in rassegna i principali modelli GARCH univariati, per poi approdare ai più recenti modelli multivariati.

Il passaggio a stime multivariate incontrò subito una prima difficoltà, la grande mole di parametri da stimare e di vincoli da rispettare e interpretare. Attraverso il modello CCC, si è riusciti a superare la difficoltà del grande numero di parametri, sacrificando però la dinamicità del modello.

Un notevole passo avanti, avvenne nel 2001, quando Robert Engle presentò una nuova classe di stimatori GARCH multivariati la quale unisce la semplicità e il successo empirico del modello GARCH univariato con una stima e interpretazione di uno stimatore a correlazione condizionata dinamica. Grazie a questo stimatore è stato possibile costruire un programma in R, capace di stimare grosse matrici di covarianza condizionata dinamica e correlazione condizionata dinamica, nonché i parametri di interesse del modello, α e β .

Utilizzando il software, abbiamo stimato dei modelli DCC(1,1), con un numero di serie che varia da 2 a 8. Ottenuti i risultati, abbiamo verificato se erano ragionevoli osservando delle stime ottenute dal precursore di tutti i modelli GARCH, il prof. Engle.

E' stato anche dimostrato che lo stimatore a due stadi, punto importante del modello, è consistente e asintoticamente normale.

Abbiamo anche visto che assicurare la matrice di covarianza condizionata definita positiva è una cosa facile da raggiungere poiché vengono

semplicemente richieste alcune restrizioni come nel modello GARCH univariato.

E' stato presentato un semplice test per verificare l'ipotesi nulla di correlazione costante contro l'ipotesi alternativa di correlazione condizionata dinamica. Visto il successo di questo modello, il test non poteva che rifiutare l'ipotesi nulla, avvalorando maggiormente le qualità del DCC.

La reale forza di questo processo di stima è la flessibilità fornita nel modellare la dinamicità del processo univariato della volatilità. Si può anche aggiungere che la teoria della specificazione del DCC può essere semplicemente estesa includendo fattori esogeni nel modello della correlazione o parametrizzazioni alternative.

APPENDICE A

In questa tesi trattando tipicamente modelli multivariati è naturale utilizzare strumenti e concetti tipici dell'algebra delle matrici. Questa appendice presenta alcuni elementi di algebra lineare che sono usati in altre parti del testo.

Il primo elemento che viene definito in algebra lineare è il **vettore**.

Definizione 1.1 (Vettore). Un vettore di ordine n , indicato con \underline{a} , è una n -pla ordinata di numeri reali:

$$\underline{a}_{(n \times 1)} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad (0.1)$$

Con la (1.1) si è definito un vettore colonna. Ad un vettore colonna si può applicare l'operazione di trasposizione, indicata con l'apice, mediante cui si può ottenere un vettore riga in cui gli elementi sono posti uno accanto all'altro:

$$\underline{a}_{(1 \times n)} = (a_1 \quad a_2 \quad \dots \quad a_n) \quad (0.2)$$

Si noti infine come le definizioni delle operazioni di somma e di prodotto interno implicano che esse possono essere applicate se e solo se i vettori considerati hanno lo stesso numero di elementi, nel qual caso si dice che i vettori sono conformabili per la somma e per il prodotto scalare.

Le principali operazioni tra vettori sono rappresentate nella tabella A.1.

	Denominazione	Notazione	Definizione
1.	<i>Somma</i>	$\underline{c} = \underline{a} + \underline{b}$	$\underline{a}(n \times 1) \quad \underline{b}(n \times 1)$ $c_i = a_i + b_i$
2.	<i>Prodotto per uno scalare</i>	$\underline{c} = b\underline{a}$	$\underline{a}(n \times 1)$, b reale $c_i = ba_i$
3.	<i>Prodotto scalare o interno</i>	$\underline{c} = \underline{a}'\underline{b}$	$\underline{a}(n \times 1) \quad \underline{b}(n \times 1)$ $c = \sum_{i=1}^n a_i b_i$

Tabella A.1

Si noti infine come le definizioni delle operazioni di somma e di prodotto interno implicano che esse possono essere applicate se e solo se i vettori considerati hanno lo stesso numero di elementi, nel qual caso si dice che i vettori sono conformabili per la somma e per il prodotto scalare.

Il secondo elemento principale dell'algebra lineare è il concetto di **matrice**.

Definizione 1.2 (Matrice). Una matrice di ordine $(n \times k)$, A , è un insieme ordinato rettangolare di numeri reali

$$A_{(n \times k)} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{pmatrix} \quad (0.3)$$

Una matrice ha un numero di elementi pari ad nk ed ognuno di essi è identificato dalla sua posizione nella matrice con il primo indice pari alla

riga di appartenenza ed il secondo dato dalla colonna di appartenenza; cioè l'elemento a_{ij} si trova all'incrocio tra la i -esima riga e j -esima colonna.

Essa può essere vista come un insieme di k vettori colonna, ognuno con n elementi

$$A = \left(a_{(1)} \quad a_{(2)} \quad \dots \quad a_{(k)} \right) \quad a_{(j)} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} \quad (0.4)$$

per $j = 1, 2, \dots, k$; oppure come un insieme di n vettori riga, ognuno con k elementi

$$A = \begin{pmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_n \end{pmatrix} \quad a'_i = (a_{i1} \quad a_{i2} \quad \dots \quad a_{ik}). \quad (0.5)$$

Dalla definizione di matrice risulta che i vettori sono casi particolari di matrici; in particolare un vettore colonna è una matrice con n righe ed una colonna mentre un vettore riga è una matrice con una riga ed n colonne.

Strutture di matrici di particolare interesse sono le seguenti.

- (i) Una matrice è quadrata quando ha lo stesso numero di righe e di colonne

$$A_{(n \times n)} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \quad (0.6)$$

(ii) Una matrice è diagonale se gli unici elementi diversi da zero si trovano sulla diagonale principale

$$A = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{pmatrix} \quad (0.7)$$

(iii) Una matrice è triangolare superiore se gli unici elementi non nulli si trovano al di sopra e sulla diagonale principale (per la triangolare inferiore evidentemente gli elementi devono essere sotto la diagonale principale)

$$U_{(n \times n)} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{pmatrix} \quad (0.8)$$

(iv) La matrice identità è una matrice diagonale i cui elementi sono tutti uguali ad uno

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \quad (0.9)$$

Essa rappresenta l'elemento neutro del prodotto matriciale: $I_n A = A$, se le due matrici sono conformi per il prodotto.

Come per i vettori, anche per le matrici sono definibili le operazioni di somma, di moltiplicazione per uno scalare e di prodotto interno, presentate nella tabella A.2. Come per i vettori la somma ed il prodotto tra matrici sono possibili solo a determinate condizioni: per la somma le matrici

devono avere le stesse dimensioni; per il prodotto, il numero delle colonne della matrice di sinistra deve essere uguale al numero delle righe della matrice di destra. A proposito del prodotti si noti come in generale esso non sia commutativo: $AB \neq BA$.

La trasposta di una generica matrice A di ordine $n \times k$ è definita come la matrice ottenuta scambiando le righe con le colonne di A ; essa, indicata con A' , è data

$$A_{(n \times k)} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{pmatrix} = \begin{pmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_n \end{pmatrix} = (a_{(1)} \quad a_{(2)} \quad \cdots \quad a_{(k)}) \quad (0.10)$$

con k righe ed n colonne.

	Denominazione	Notazione	Definizione
1.	Somma	$C = A + B$	$A(n \times k) \quad B(n \times k)$ $c_{ij} = a_{ij} + b_{ij}$
2.	Prodotto per uno scalare	$C = bA$	$A(n \times k)$, b reale $c_{ij} = ba_{ij}$
3.	Prodotto	$C = AB$	$A(n \times k) \quad B(k \times p)$ $c_{ij} = \sum_{j=1}^n a_{ij} b_{jl}$

Tabella A.2

Dalla definizione di somma, di prodotto matriciale e di trasposta derivano le seguenti relazioni:

$$(A')' = A \quad (A + B)' = A' + B' \quad (AB)' = B'A' \quad (0.11)$$

Una matrice quadrata si dice simmetrica, se è uguale alla sua trasposta: $A = A'$. Si noti che il prodotto di una matrice per la sua trasposta risulta essere una matrice simmetrica, infatti

$$(A'A)' = A'A \quad (0.12)$$

Matrice inversa e determinante

Per matrici quadrate, e soltanto per queste, due concetti fondamentali sono quelli di matrice inversa e determinante.

Nei numeri reali gli elementi invertibili (dotati cioè di un reciproco) sono tutti e solo i numeri diversi da zero e il prodotto di un numero per il suo inverso è pari ad uno. Allo stesso modo possiamo definire l'inversa di una matrice quadrata, indicata con A^{-1} .

Definizione 1.3 (Inversa di una matrice). La matrice quadrata di ordine n si dice invertibile se esiste una matrice A^{-1} per cui

$$AA^{-1} = A^{-1}A = I_n \quad (0.13)$$

Occorre sottolineare che, come vedremo in seguito, non tutte le matrici quadrate hanno inversa, ma, se questa esiste, allora:

$$AA^{-1}A = A \quad (0.14)$$

da cui risulta

$$\left(A^{-1}\right)^{-1} = A \quad (0.15)$$

Sull'inversa di una matrice esiste un teorema di unicità:

Teorema 1.1 (Unicità dell'inversa). Se l'inversa di una matrice quadrata esiste, allora essa è unica.

Definizione 1.4 (Matrice non singolare). Una matrice quadrata invertibile viene detta non singolare. Se la matrice non è invertibile allora essa è singolare.

Sulla base di definizione di matrice inversa possiamo già considerare alcuni esempi in cui è facile esaminare il problema dell'invertibilità.

Una matrice diagonale con $d_{ii} \neq 0$ per ogni i ha come inversa la matrice

$$D^{-1} = \begin{pmatrix} d_{11}^{-1} & 0 & \dots & 0 \\ 0 & d_{22}^{-1} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & d_{mm}^{-1} \end{pmatrix} \quad (0.16)$$

Da quest'ultima espressione risulta evidente come la matrice inversa possa non esistere: se un qualunque d_{ii} è nullo non è possibile calcolare l'inversa di D .

Alcuni risultati notevoli per le matrici inverse sono i seguenti:

$$\begin{aligned} \left(A^{-1}\right)' &= \left(A'\right)^{-1} \\ \left(AB\right)^{-1} &= B^{-1}A^{-1} \end{aligned} \quad (0.17)$$

quest'ultima valida se A e B sono conformabili per il prodotto.

Collegato all'esistenza dell'inversa è il concetto di determinante, che può essere visto come una particolare funzione che associa ad ogni matrice quadrata un numero reale. Per indicare il determinante di una matrice A si userà il simbolo $|A|$ oppure il simbolo $\det(A)$. Per la sua definizione si può

fare ricorso ad un ragionamento ricorsivo cominciando a definire il determinante di una matrice 1×1 , cioè di uno scalare, come uguale allo scalare stesso. Supponendo noto il concetto di determinante di matrici i ordine $n-1$, si definisce il minore complementare dell'elemento a_{ij} il determinante della matrice di ordine $n-1$ ottenuta eliminando la riga di indice i e la colonna di indice j e lo indichiamo con $|M_{ij}|$, ed il complemento algebrico di a_{ij} il numero

$$|A_{ij}| = (-1)^{i+j} |M_{ij}| \quad (0.18)$$

in altre parole il complemento algebrico $|A_{ij}|$ coincide con il minore complementare se la somma degli indici è pari, con il suo opposto se la somma degli indici è dispari.

Definizione 1.5 (Determinante). Si chiama determinante di una matrice quadrata A di ordine $n \geq 1$ la quantità

$$|A| = \begin{cases} a & \text{se } A \text{ è di ordine } 1 \\ \sum_{j=1}^n a_{1j} |A_{1j}| & \text{se } A \text{ è di ordine } n \end{cases} \quad (0.19)$$

Dunque il determinante di una matrice quadrata di ordine superiore al primo è, per definizione, la somma dei prodotti degli elementi della prima riga moltiplicati per i rispettivi complementi algebrici.

Esempio 1.1. per meglio comprendere la definizione si consideri dapprima una matrice (2×2)

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Il determinante viene qui calcolato attraverso una semplice formula:

$$|A| = a_{11}a_{22} - a_{12}a_{21} \quad (0.20)$$

Considerando ora una matrice (3×3)

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

ed applicando la definizione di determinante e quanto ottenuto per la matrice (2×2) si ha:

$$\begin{aligned} |A| &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = & (0.21) \\ &= a_{11} (a_{22}a_{33} - a_{23}a_{32}) - a_{12} (a_{21}a_{33} - a_{23}a_{31}) + a_{13} (a_{21}a_{32} - a_{22}a_{31}) \end{aligned}$$

Su questa base sono calcolabili i determinanti di matrici quadrate di ordine qualunque. Risulta, però, evidente, che all'aumentare di n i calcoli si complicano notevolmente; per semplificarli sono stati elaborati diversi algoritmi che li rendono molto più veloci.

Dagli esempi è immediato verificare che il determinante non varia se invece della prima riga si considera una riga qualunque. Questo è un risultato generale che non dipende dall'ordine della matrice, per cui la definizione di determinante per matrici quadrate di ordine n può scriversi, considerando una generica riga i , come

$$|A| = \sum_{j=1}^n a_{ij} |A_{ij}| \quad (0.22)$$

Il seguente teorema collega il determinante all'inversa di una matrice.

Teorema 1.2 (Esistenza dell'inversa). Condizione necessaria e sufficiente affinché una matrice quadrata A ammette l'inversa è che il suo determinante non sia nullo, $|A| \neq 0$.

In tal caso l'elemento di indici (i,j) dell'inversa, indicato con a^{ij} può essere scritto come il rapporto tra il complemento algebrico di a_{ij} ed il determinante della matrice.

$$a^{ij} = \frac{|A_{ij}|}{|A|} \quad (0.23)$$

Di seguito vengono dati una serie di risultati sui determinanti molto utili per le applicazioni che possiamo incontrare.

1. Il determinante di una matrice cambia di segno se si scambiano due righe o due colonne.
2. La matrice e la sua trasposta hanno lo stesso determinante.

$$|A| = |A'| = |A'| \quad (0.24)$$

3. Se A ha una riga o una colonna di zeri allora il suo determinante è nullo.
4. Per una successione di matrici quadrate A_l , $l=1,\dots,L$, tutte dello stesso ordine, il determinante del prodotto è uguale al prodotto dei determinanti.

$$\left| \prod_{l=1}^L A_l \right| = \prod_{l=1}^L |A_l| \quad (0.25)$$

5. Il determinante del prodotto di uno scalare per una matrice quadrata di ordine n è pari allo scalare elevato alla n per il determinante della matrice.

$$|kA| = k^n |A| \quad (0.26)$$

6. Il determinante di una matrice diagonale o triangolare (inferiore o superiore) è dato dal prodotto degli elementi della diagonale.

$$|D| = \prod_{i=1}^n d_{ii}, \quad |U| = \prod_{i=1}^n u_{ii}, \quad |L| = \prod_{i=1}^n l_{ii} \quad (0.27)$$

7. Il determinante dell'inversa è pari al reciproco del determinante della matrice.

$$|A^{-1}| = |A|^{-1} \quad (0.28)$$

Rango

Un altro elemento dell'algebra collegato al determinante e alle matrici inverse è il rango.

Definizione 1.6 (Rango). Il rango di una matrice $(n \times k)$ A è dato dal numero di colonne (o di righe) linearmente indipendenti.

Qui di seguito viene elencata una serie importante di risultati sul rango delle matrici.

- (i) Il rango di una matrice e della sua trasposta coincidono.

$$\text{rango}(A) = \text{rango}(A') \quad (0.29)$$

- (ii) Il rango di una matrice è minore o tutt'al più uguale al minore tra il numero di righe e il numero di colonne.

$$\text{rango}(A) \leq \min(n, k) \quad (0.30)$$

$(n \times k)$

- (iii) Se A è $(n \times k)$ allora

$$\text{rango}(A) = \text{rango}(AA') = \text{rango}(A'A) \quad (0.31)$$

si noti che le matrici AA' e $A'A$ sono quadrate di ordine, rispettivamente, $(n \times n)$ e $(k \times k)$.

(iv) Il rango del prodotto di due matrici è minore o tutt'al più uguale alla somma dei ranghi.

$$\underset{(n \times k) (k \times n)}{\text{rango}(A B)} \leq \min(\text{rango}(A), \text{rango}(B)) \quad (0.32)$$

Per una matrice quadrata $(n \times n)$ il valore del suo rango è strettamente collegato al problema dell'esistenza dell'inversa. In particolare si dimostra che le seguenti tre proposizioni sono equivalenti

$$A \text{ è non singolare} \quad (0.33)$$

$$\text{rango}(A) = n \quad (0.34)$$

$$|A| \neq 0 \quad (0.35)$$

Di conseguenza una matrice quadrata è invertibile se e solo se tutte le sue colonne (e le righe) sono linearmente indipendenti.

Definizione 1.7 (Pseudo-inversa di Moore-Penrose). Data una matrice $(n \times k)$ A , dicesi *pseudo-inversa* (di Moore-Penrose) della matrice A ogni matrice A^+ soddisfacente alle quattro condizioni:

$$\begin{aligned} AA^+A = A \quad AA^+ &= (AA^+)' \\ A^+AA^+ = A^+ \quad A^+A &= (A^+A)' \end{aligned} \quad (0.36)$$

Evidentemente la matrice A^+ è una matrice $(k \times n)$.

Esaminiamo ora alcune proprietà della pseudo-inversa.

1. A^+ è univocamente individuata dalle condizioni sopra elencate.
2. Se A è quadrata e non singolare, allora

$$A^+ = A^{-1} \quad (0.37)$$

3. Se A ammette inversa sinistra (destra), cioè se $\text{rango}(A) = n$, allora

$$A^+ = (A'A)^{-1} A' \quad \left(A^+ = A' (AA')^{-1} \right) \quad (0.38)$$

e A^+ è un'inversa sinistra di A .

Definizione 1.8 (Autovalori e Autovettori). Gli n autovalori, λ_i , $i = 1, \dots, n$, e gli associati autovettori, \underline{v}_i , della matrice $(n \times n)$ A soddisfano il sistema omogeneo:

$$(A - \lambda_i I_n) \underline{v}_i = \underline{0}_n \quad (0.39)$$

Tale sistema deve avere soluzione diversa dal vettore nullo, per cui gli autovalori devono soddisfare l'equazione caratteristica

$$|A - \lambda I_n| = 0 \quad (0.40)$$

Di seguito vengono dati alcuni risultati utili sugli autovalori e autovettori.

1. Se A è una matrice reale e simmetrica, allora i suoi autovalori ed autovettori sono reali.
2. Il numero di autovalori diversi da zero è pari al rango della matrice.
3. Il determinante di una matrice è pari al prodotto degli autovalori :

$$|A| = \prod_{i=1}^n \lambda_i \quad (0.41)$$

4. Se A è reale e simmetrica, autovettori corrispondenti ad autovalori diversi sono tra loro ortogonali.
5. Se gli n autovalori sono tutti distinti, allora gli autovettori sono linearmente indipendenti.

Gli autovalori permettono una particolare fattorizzazione di una matrice simmetrica.

Forme quadratiche

La funzione scalare

$$Q(x, A) = x'Ax = \sum_{i=1}^n \sum_{j=1}^n x_i x_j a_{ij} \quad (0.42)$$

con A matrice quadrata $(n \times n)$ e x vettore $(n \times 1)$ è denominata forma quadratica. E' facile verificare che il valore della forma quadratica non varia se si considera la matrice $(A + A')/2$, che è simmetrica. Per questo motivo si assume che A sia simmetrica.

Le forme quadratiche e le matrici ad esse associate vengono classificate come nella tabella A.3; una forma quadratica che non rientra nei casi presentati viene detta indefinita.

Classificazione	Definizione
Positiva definita	Per ogni $x \neq 0$ $Q(x, A) > 0$
Positiva semidefinita	Per ogni $x \neq 0$ $Q(x, A) \geq 0$
Negativa definita	Per ogni $x \neq 0$ $Q(x, A) < 0$
Negativa semidefinita	Per ogni $x \neq 0$ $Q(x, A) \leq 0$

Tabella A.3

Le principali proprietà delle forme quadratiche sono le seguenti:

1. Se A è $(n \times k)$, allora $A'A$ e AA' sono entrambe definite positive; se $\text{rango}(A) = n < k$, allora $A'A$ è positiva definita.
2. Se A è simmetrica e positiva definita (semidefinita), allora i suoi autovalori sono tutti positivi (non negativi). Se essa è negativa definita (semidefinita), allora tutti i suoi autovalori sono negativi (non positivi).
3. Ogni matrice simmetrica positiva definita (semidefinita) ha una sua radice quadrata $A^{1/2}$ data da

$$A^{1/2} = V \Lambda^{1/2} V' \tag{0.43}$$

dove V è la matrice degli autovalori e Λ è la matrice diagonale degli autovalori e

$$\Lambda^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{\lambda_n} \end{pmatrix} \quad (0.44)$$

4. Se A è positiva definita allora A^{-1} è positiva definita.

La fattorizzazione LDU e di Cholesky di una matrice

Definizione 1.9 (Fattorizzazione LDU). Una matrice $(n \times n)$ A ha una fattorizzazione LDU se può essere scritta come $A=LDU$, dove D è una matrice diagonale, L è una triangolare inferiore e U è triangolare superiore. Questa fattorizzazione è particolarmente utile per il calcolo dell'inversa, dato che per matrici triangolari questo è molto semplice. Particolarmente utile nella tesi sono i due seguenti risultati sulla fattorizzazione di matrici simmetriche e definite positive.

1. Se A è simmetrica, allora $L = U'$
2. **Fattorizzazione di Cholesky:** Se A è simmetrica e definita positiva, essa può essere scritta come

$$A = TT' \quad (0.45)$$

dove $T = LD^{1/2}$ è triangolare inferiore con gli elementi sulla diagonale tutti positivi.

Traccia e operatore vec

Definizione 1.20(Traccia di una matrice quadrata). La traccia di una matrice quadrata A , $(n \times n)$, è data dalla somma degli elementi della diagonale principale:

$$Traccia(A) = \sum_{i=1}^n a_{ii} \quad (0.46)$$

Le principali proprietà della traccia sono le seguenti.

(i) La traccia è un operatore lineare: per ogni scalare α si ha

$$\text{Traccia}(\alpha A) = \alpha \text{Traccia}(A)$$

e per ogni coppia di α e β vale la relazione

$$\text{Traccia}(\alpha A + \beta B) = \alpha \text{Traccia}(A) + \beta \text{Traccia}(B)$$

con A e B matrici quadrate dello stesso ordine.

(ii) Una matrice quadrata e la sua trasposta hanno la stessa Traccia

$$\text{Traccia}(A) = \text{Traccia}(A')$$

(iii) Sia A una matrice $(p \times q)$ e B una matrice $(q \times p)$, allora

$$\text{Traccia}(AB) = \text{Traccia}(BA) = \sum_{i=1}^p \sum_{j=1}^q a_{ij} b_{ji}$$

(iv) La traccia è uguale alla somma degli autovalori della matrice

$$\text{Traccia}(A) = \sum_{i=1}^n \lambda_i$$

dove λ_i sono gli autovalori della matrice A .

Definizione 1.21 (L'operatore vec). Data una matrice A $(n \times k)$, l'operatore vec incolonna le colonne di A trasformandola in un vettore colonna di dimensione $(nk \times 1)$. Se $A = (a_{(1)} \ a_{(2)} \ \dots \ a_{(k)})$ allora

$$vec(A) = \begin{pmatrix} a_{(1)} \\ a_{(2)} \\ \vdots \\ a_{(k)} \end{pmatrix} \quad (0.47)$$

E' immediato verificare che $vec(A) \neq vec(A')$. Il vec è un operatore lineare, cioè soddisfa le seguenti proprietà: $vec(A+B) = vec(A) + vec(B)$ e $vec(kA) = k \cdot vec(A)$ per ogni scalare k .

Derivazione matriciale

Le derivate di funzioni rispetto agli elementi di un vettore o di una matrice possono essere convenientemente espresse in forma matriciale. Per comodità di notazione indichiamo con $g(A, x)$ una funzione reale, dipendente da una matrice A , di ordine $(n \times m)$, e da un vettore x , di ordine $(n \times 1)$; e con $g(x)$ una funzione vettoriale, cioè

$$g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_p(x) \end{pmatrix}.$$

Per fissare le idee sulla forma matriciale del risultato dell'operazione di derivazione è conveniente il seguente schema:

- (i) Il risultato della derivazione della funzione scalare $g(\cdot)$ rispetto agli elementi del vettore x è un vettore con n elementi così definiti

$$\frac{\partial g(x)}{\partial x_{(n \times 1)}} = \begin{pmatrix} \frac{\partial g(x)}{\partial x_1} \\ \frac{\partial g(x)}{\partial x_2} \\ \vdots \\ \frac{\partial g(x)}{\partial x_n} \end{pmatrix} \quad (0.48)$$

- (ii) Il risultato della derivazione della funzione scalare⁹ $g(\cdot)$ rispetto agli elementi della matrice A è una matrice con le stesse dimensioni di A :

⁹ Nota: stiamo molto attenti che questa è un semplice funzione scalare, che viene derivata per ogni elemento della matrice.

$$\frac{\partial g(A)}{\partial A_{(n \times m)}} = \begin{pmatrix} \frac{\partial g(A)}{\partial a_{11}} & \frac{\partial g(A)}{\partial a_{12}} & \dots & \frac{\partial g(A)}{\partial a_{1m}} \\ \frac{\partial g(A)}{\partial a_{21}} & \frac{\partial g(A)}{\partial a_{22}} & & \frac{\partial g(A)}{\partial a_{2m}} \\ \vdots & & \ddots & \vdots \\ \frac{\partial g(A)}{\partial a_{n1}} & \frac{\partial g(A)}{\partial a_{n2}} & \dots & \frac{\partial g(A)}{\partial a_{nm}} \end{pmatrix} \quad (0.49)$$

(iii) Il risultato della derivazione della funzione vettoriale¹⁰ $g(\cdot)$ con p elementi rispetto alle n componenti del vettore x' è una matrice $(p \times n)$:

$$\frac{\partial g(x)}{\partial x'_{(p \times n)}} = \begin{pmatrix} \frac{\partial g_1(x)}{\partial x_1} & \frac{\partial g_1(x)}{\partial x_2} & \dots & \frac{\partial g_1(x)}{\partial x_n} \\ \frac{\partial g_2(x)}{\partial x_1} & \frac{\partial g_2(x)}{\partial x_2} & & \frac{\partial g_2(x)}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial g_p(x)}{\partial x_1} & \frac{\partial g_p(x)}{\partial x_2} & \dots & \frac{\partial g_p(x)}{\partial x_n} \end{pmatrix} \quad (0.50).$$

Nella tabella A.4 vengono riportate le espressioni delle derivate di alcune funzioni scalari rispetto agli elementi di un vettore o di una matrice. In tali formule si suppone che tutte le operazioni matriciali considerate siano possibili.

Considerando il caso di una matrice $(n \times m)$ A con gli elementi funzione di una variabile, z , $a_{ij}(z)$, la sua derivata rispetto a z è costituita dalla matrice delle derivate degli elementi di A , cioè:

¹⁰ Evidenziamo qui l'importanza della notazione vettoriale, in quanto ora non abbiamo più una semplice equazione lineare, ma bensì un vettore con p elementi.

$$\frac{\partial A}{\partial z} = \begin{pmatrix} \frac{\partial a_{11}}{\partial z} & \frac{\partial a_{12}}{\partial z} & \dots & \frac{\partial a_{1m}}{\partial z} \\ \frac{\partial a_{21}}{\partial z} & \frac{\partial a_{22}}{\partial z} & & \frac{\partial a_{2m}}{\partial z} \\ \vdots & & \ddots & \vdots \\ \frac{\partial a_{n1}}{\partial z} & \frac{\partial a_{n2}}{\partial z} & \dots & \frac{\partial a_{nm}}{\partial z} \end{pmatrix} \quad (0.51)$$

Se A e B sono due matrici conformabili per il prodotto con gli elementi funzione della variabile z , allora, applicando la regola di derivazione di funzione di funzione, si ha:

$$\frac{\partial AB}{\partial z} = \frac{\partial A}{\partial z} B + A \frac{\partial B}{\partial z} \quad (0.52).$$

Adottando tale regola di derivazione, se A è una matrice $(n \times n)$ non singolare, si dimostra che la matrice delle derivate di A^{-1} , rispetto ad un generico elemento di A , a_{ij} , è data da:

$$\frac{\partial A^{-1}}{\partial a_{ij}} = -A^{-1} U_{ij} A^{-1} \quad (0.53)$$

dove U_{ij} è una matrice $(n \times m)$, con l'elemento $u_{ij} = 1$ e tutti gli altri uguali a zero.

Sulla base di tale risultato si dimostra che:

$$\frac{\partial \text{traccia}(A^{-1}F)}{\partial A} = -A^{-1}FA^{-1} \quad (0.54)$$

e che, tenendo conto della formula di derivazione dell'inversa si ha:

$$\frac{\partial \text{vec}(A^{-1})}{\partial \text{vec}(A)'} = -(A^{-1} \otimes A^{-1}) \quad (0.55)$$

$\frac{\partial a'x}{\partial x} = a$	$\frac{\partial A'x}{\partial x'} = A'$
$\frac{\partial x'A'x}{\partial x} = (A + A')x$	$\frac{\partial y'Bx}{\partial B} = yx'$
$\frac{\partial x'A^{-1}y}{\partial A} = -A^{-1}xy'A^{-1}$	$\frac{\partial \text{traccia}(A)}{\partial A} = I_n$
$\frac{\partial \text{traccia}(AF)}{\partial A} = F'$	$\frac{\partial \text{traccia}(AB'B)}{\partial B} = 2B$
$\frac{\partial A }{\partial A} = A (A')^{-1}$ ¹¹	$\frac{\partial \ln A }{\partial A} = (A')^{-1} \quad \text{se } A > 0$

Tabella A.4

¹¹ Sembra evidente che questa derivazione prende questa formulazione solo se il determinante della matrice è diverso da zero, altrimenti il valore della derivazione è zero.

APPENDICE B

```
#-----PROGRAMMA PER IL CALCOLO DEL DCC-----

rm (list=ls())
library(ts)
library(tseries)
library(nls)
library(lattice)
library(nlme)
library(MASS)

#          FUNZIONI
unigarch=function (y)
  {
    garch.y=garch( y,order=c( 1,1), trace=F )
    resid.stand=residuals( garch.y, standardize=T )
    return(resid.stand)
  }

Rmatrix=function(phi)
  {
    #Calcolo della matrice di correlazione
    condizionata
    alpha=abs(phi [1])
    beta=abs(phi [2])
    covc=array(0,c(m,m,n))
    r=array( 0, c(m,m,n))
```

```

covc[, ,1]=covnc
r[, ,1]=covc[, ,1] / (diag(covc[, ,1]) %**% (t
    (diag(covc[, ,1])))
for ( t in 2:n )
    {
    for ( i in 1:m )
        {
        for ( j in 1:m )
            {
            covc[i,j,t]=corrnc[i,j] * ( 1-
                alpha-beta ) + alpha * ( z[t-1,i]
                * z[t-1,j] ) + beta * (
                covc[i,j,t-1] )
            }
        }
    r[, ,t]=covc[, ,t]/( (diag ( covc[, ,t]
        )^(1/2) ) %**% ( t ( ( diag (
        covc[, ,t] ) )^(1/2) ) ) )
    }

```

```
# Calcolo della funzione di log-verosimiglianza
```

```

lgv = 0
c = array ( 0,c(n,1) )
invr=array ( 0,c(m,m,n) )
for ( t in 1:n )
    {
    if ( is.nan ( det ( r[, ,t] ) ) )
        {
        invr[, ,t]=0
        }
    else
        {
        invr[, ,t] = ginv ( r[, ,t] )
        }
    }

```

```

        c[t] = (-0.5) * ( log ( det ( r[, ,t] ) ) +
            ( z[t,] %**%
                ( invr[, ,t] ) %**% z[t,] ) ) )
        lgv = lgv + c[t]
    }
    return ( lgv)
}

# PROGRAMMA
print ( "Inserisci il numero di serie che vuoi
        analizzare con il modello DCC, poi premi Invio
        due volte" )
m = scan()

print ( "Inserisci il numero delle osservazioni da
        usare nell'analisi, poi premi Invio due volte" )
n = scan()

serie = array ( 0,c (n,m) )
print ( "Inserisci il percorso dei file in cui si
        trovano le serie dei dati" )
for ( i in 1:m )
    {
        print(i)
        c = readline()
        s = scan (c)
        serie[,i] = s[1:n]
    }
z = array ( 0,c (n,m) )
for ( i in 1:m)
    {
        serie[,i] = serie[,i] - mean ( serie[,i] )
        z[,i] = unigarch ( serie[,i] )
        z[1,i]=0
    }

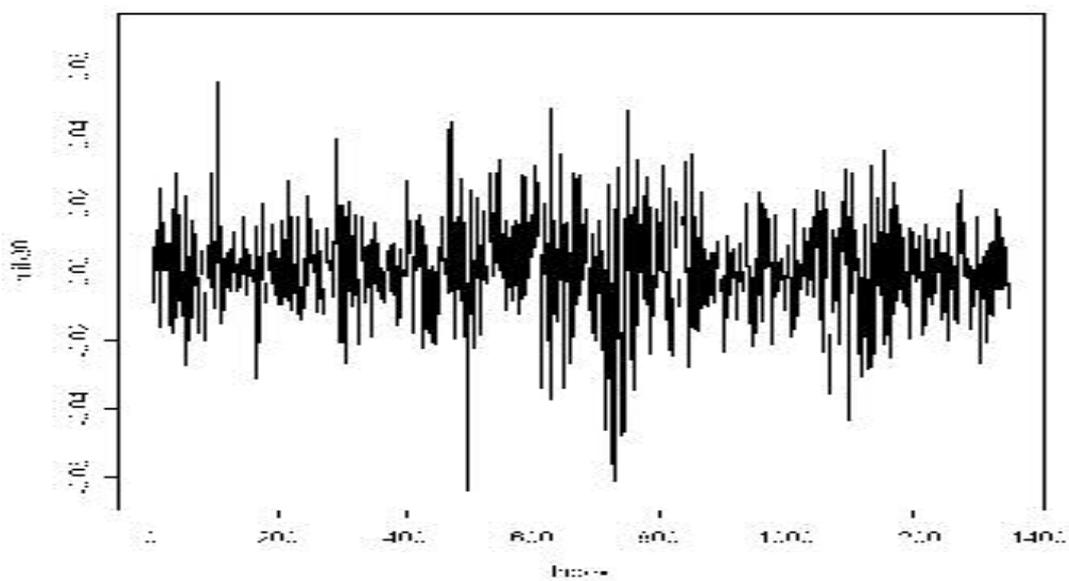
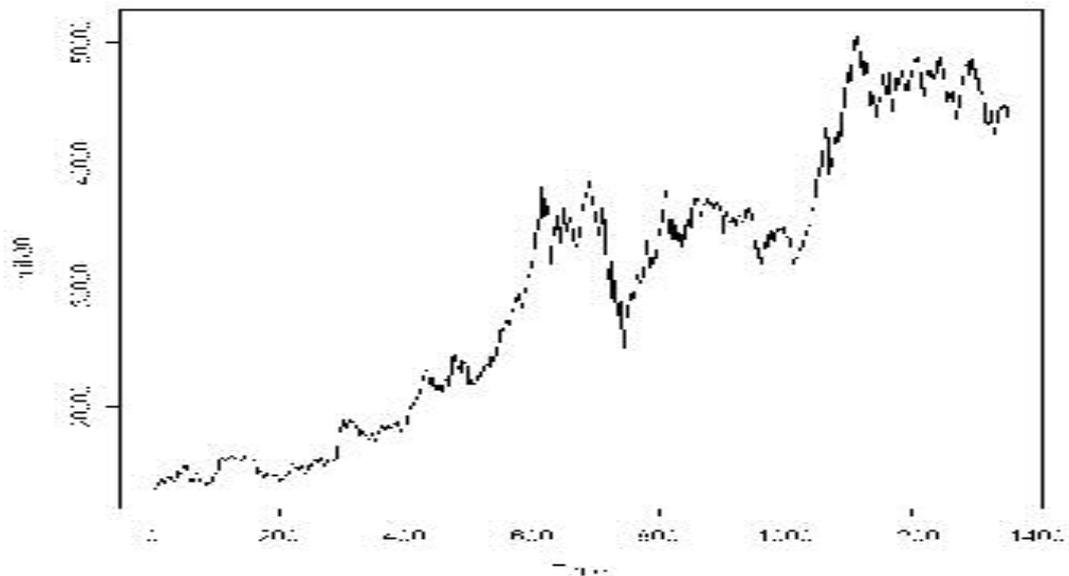
```

```
    }
covnc = cov (z)
print (covnc)
corrnc = cor(z)
print (corrnc)
#Funzione per massimizzare
risu = optim ( par = c(0.1,0.1), Rmatrix, method =
"BFGS", control = list ( fnscale = -1 ), hessian = T
)
parametri = ( abs ( risu$par ) )
print ( "La stima del parametro ARCH è" )
print ( parametri[1] )
print ( "          " )
print ( "La stima del parametro GARCH è" )
print ( parametri[2] )
print ( "          " )

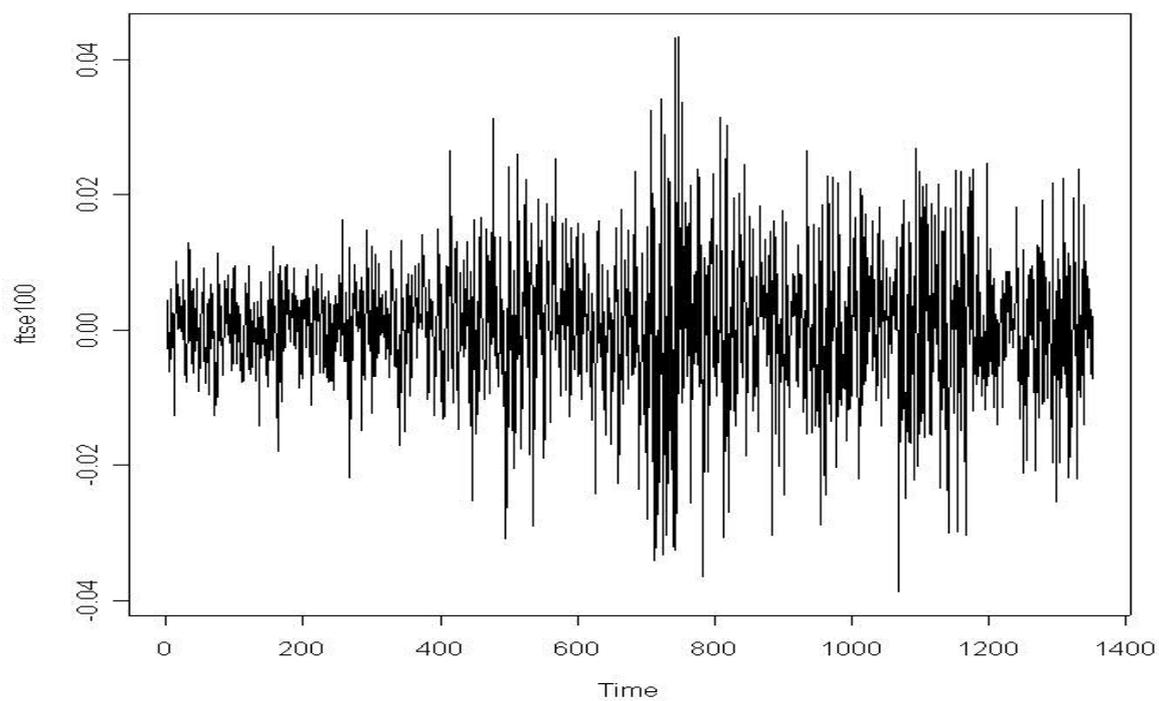
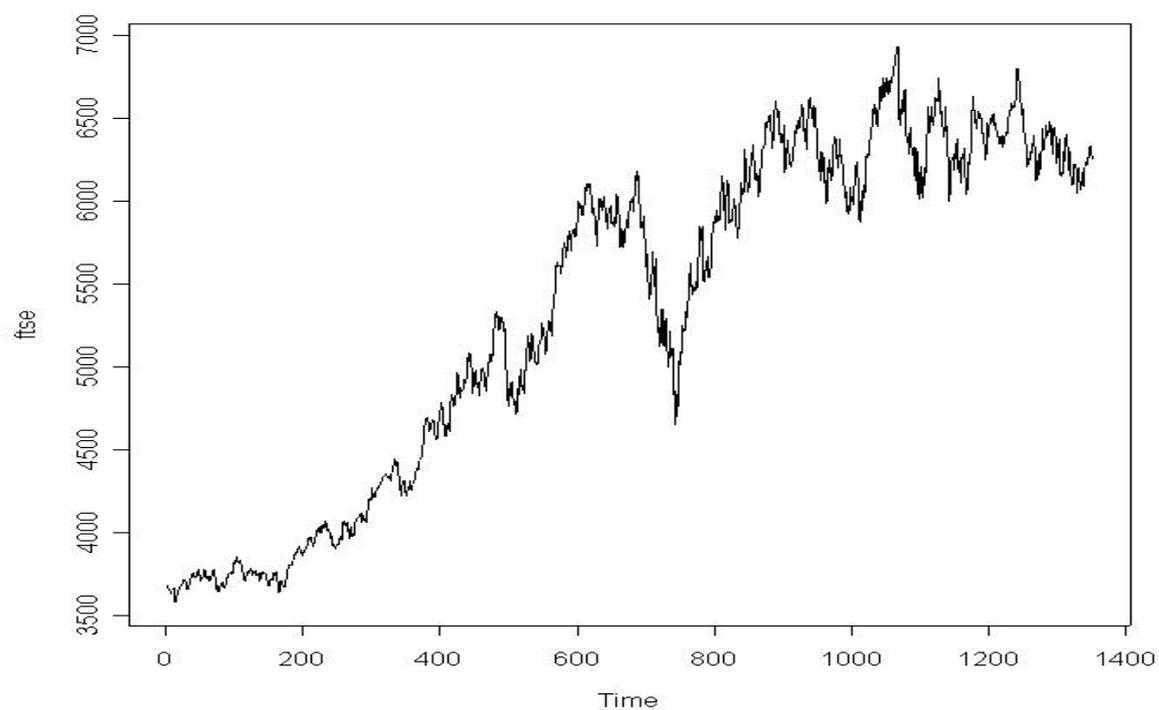
# ----- calcolo di J con metodi numerici -----
jmeno1 = ginv ( - ( risu$hessian / n ) )
varas.qml.num = jmeno1 / n
test = abs ( risu$par ) / sqrt ( diag ( varas.qml.num
) )
print ( "TEST T" )
print ( test )
```

APPENDICE C

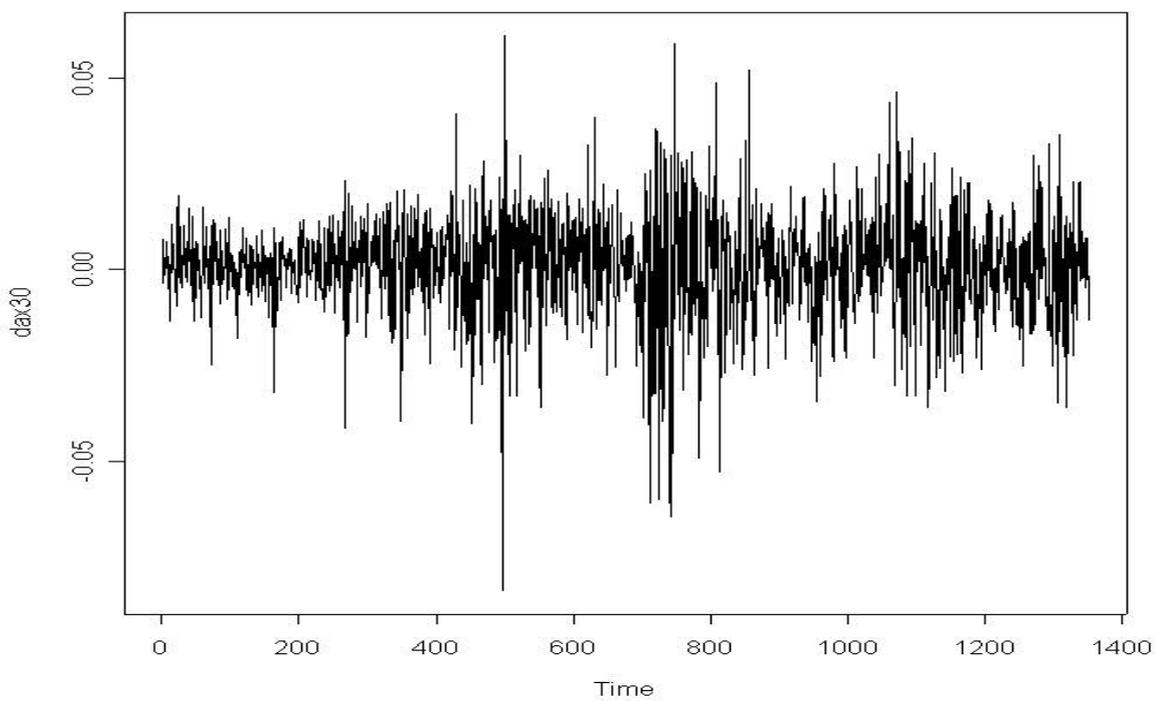
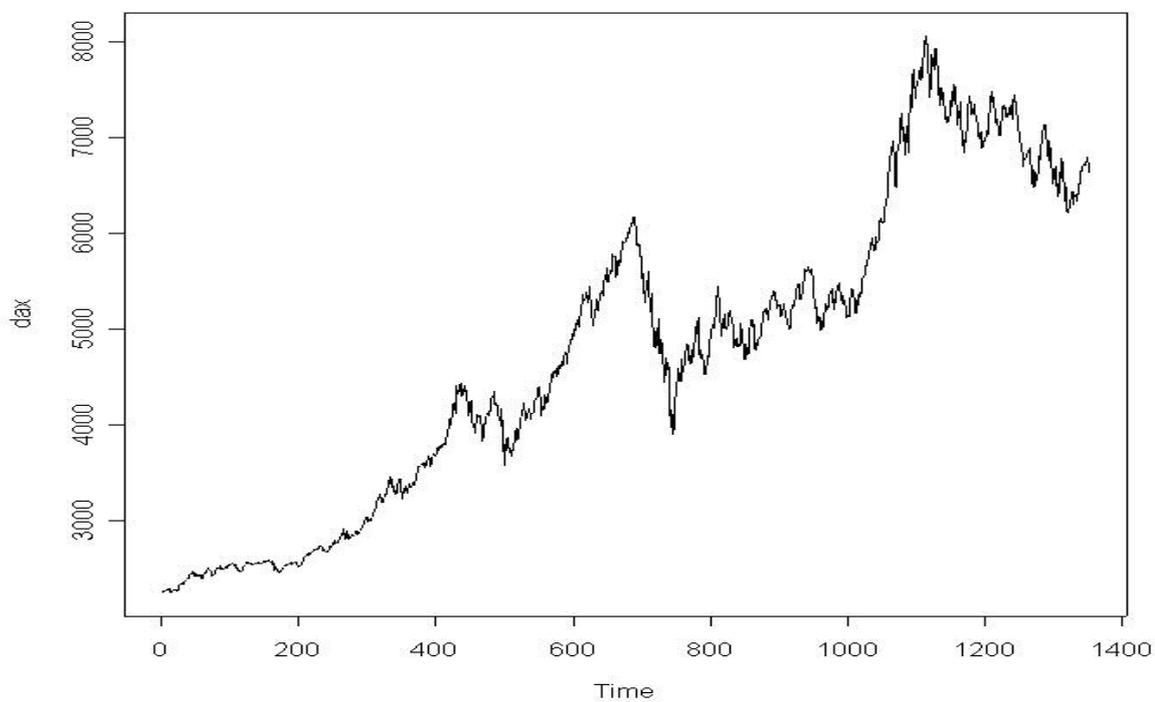
In questa appendice troviamo i grafici delle serie utilizzate nelle prove.



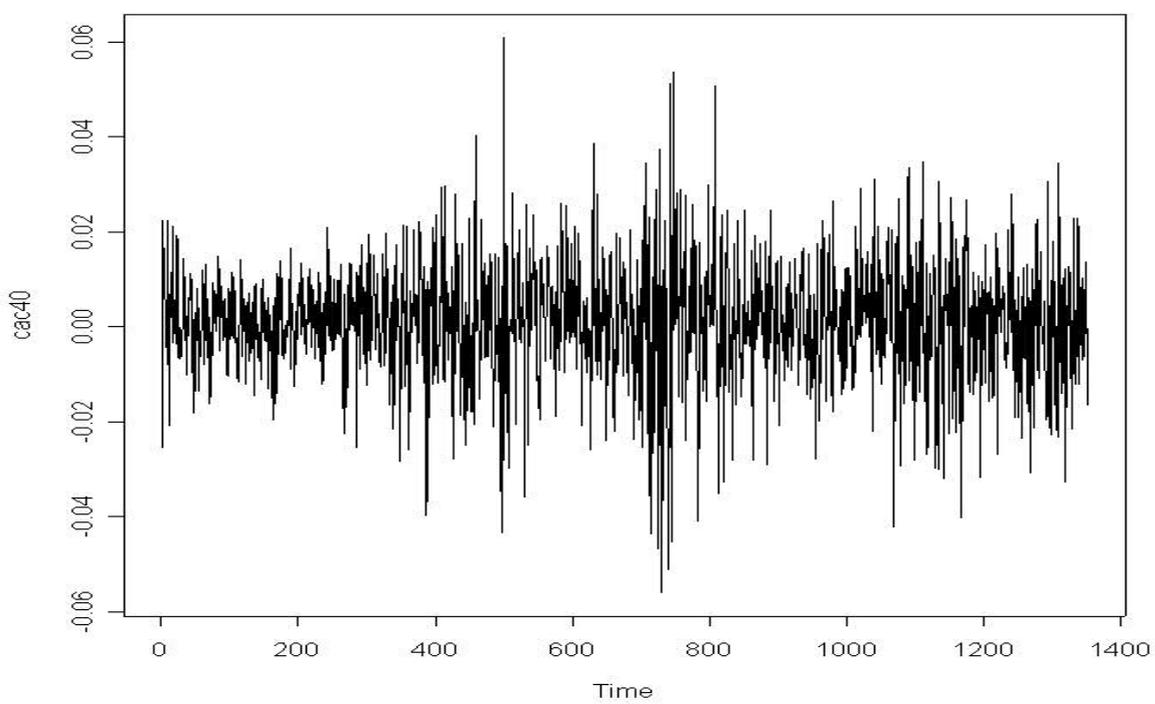
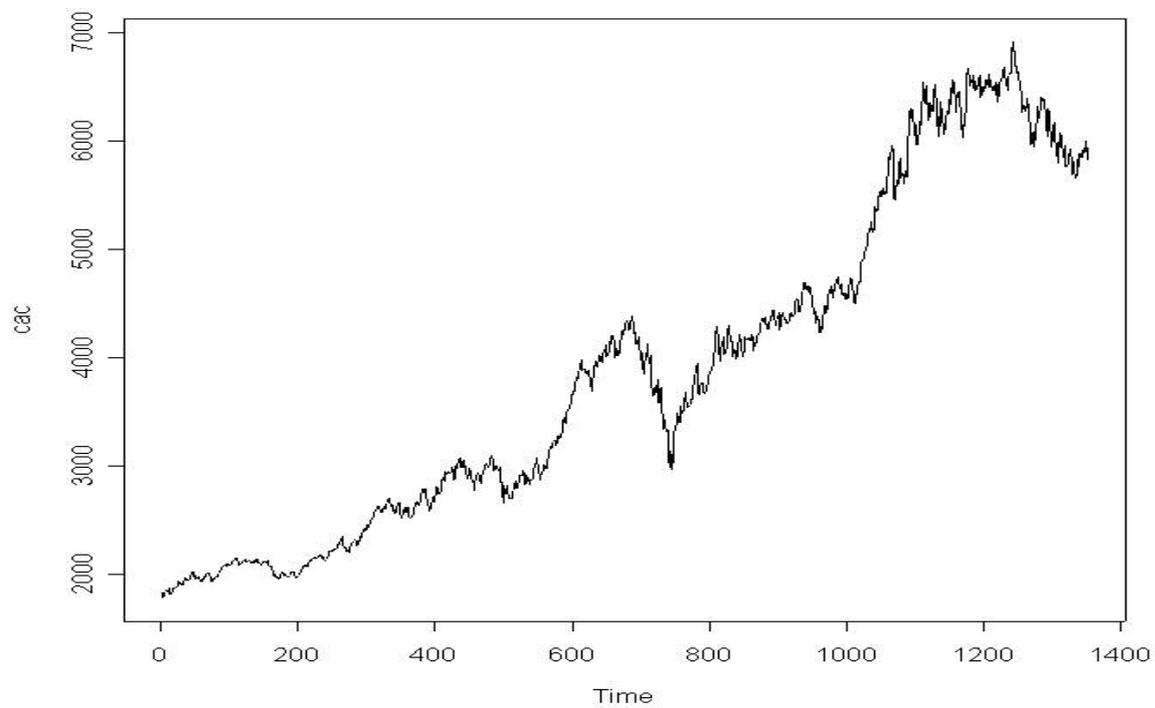
GRAFICI MIB30



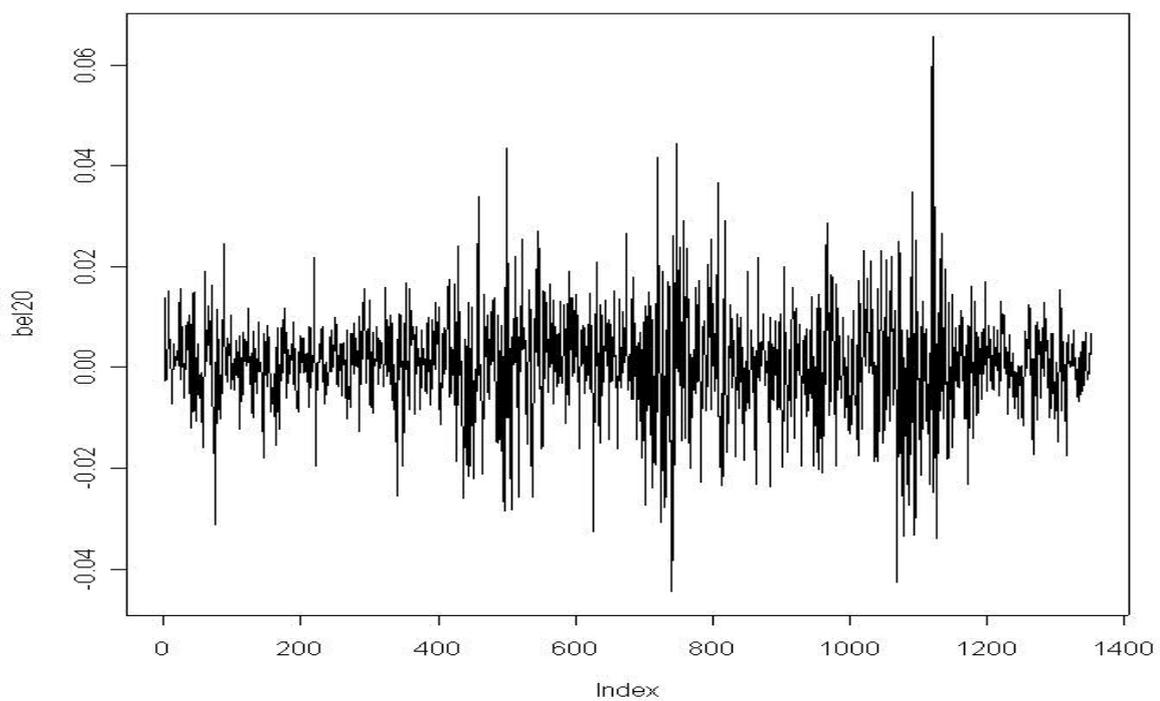
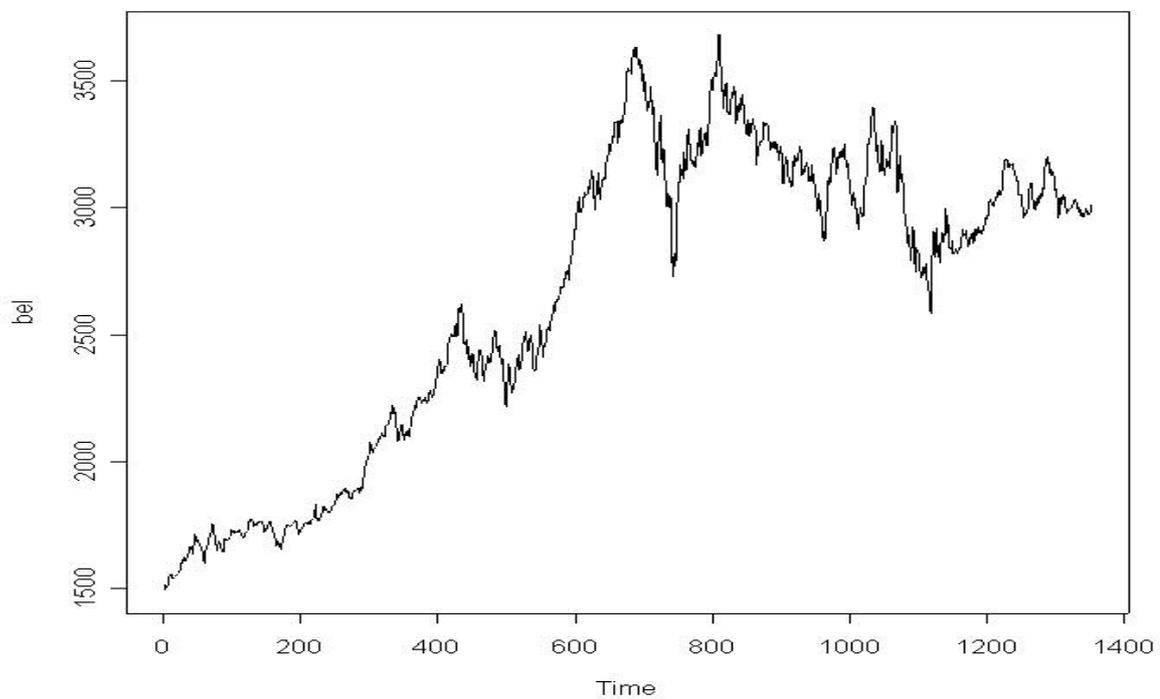
GRAFICI FTSE100



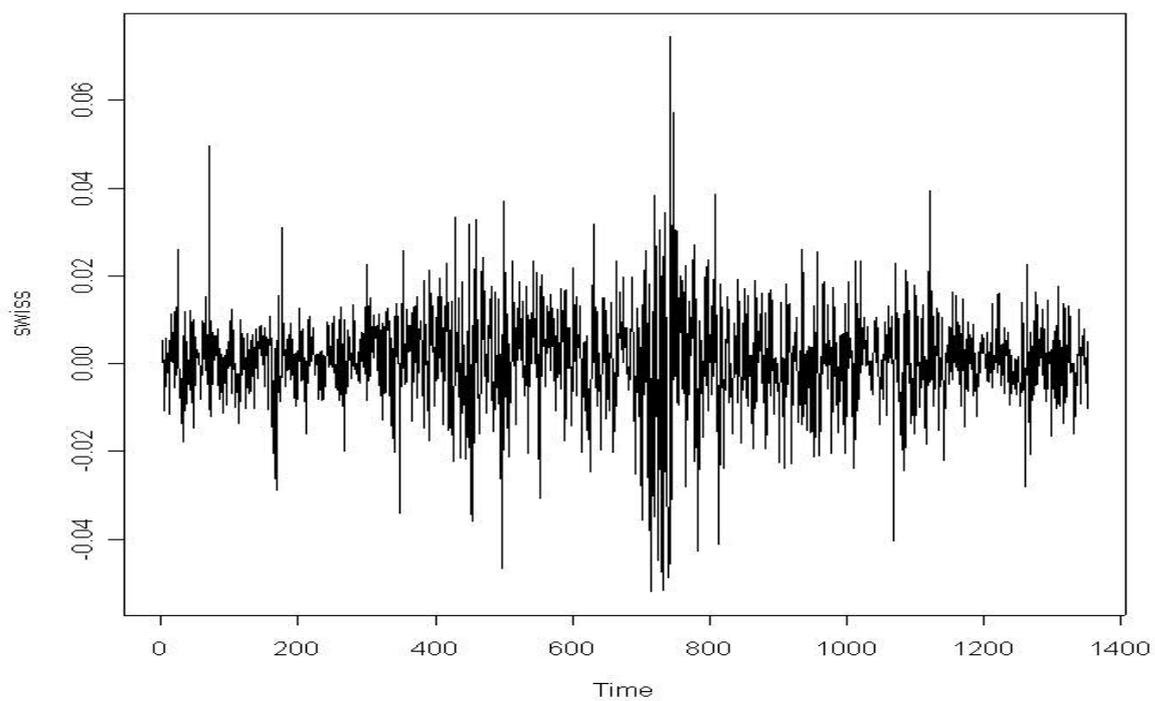
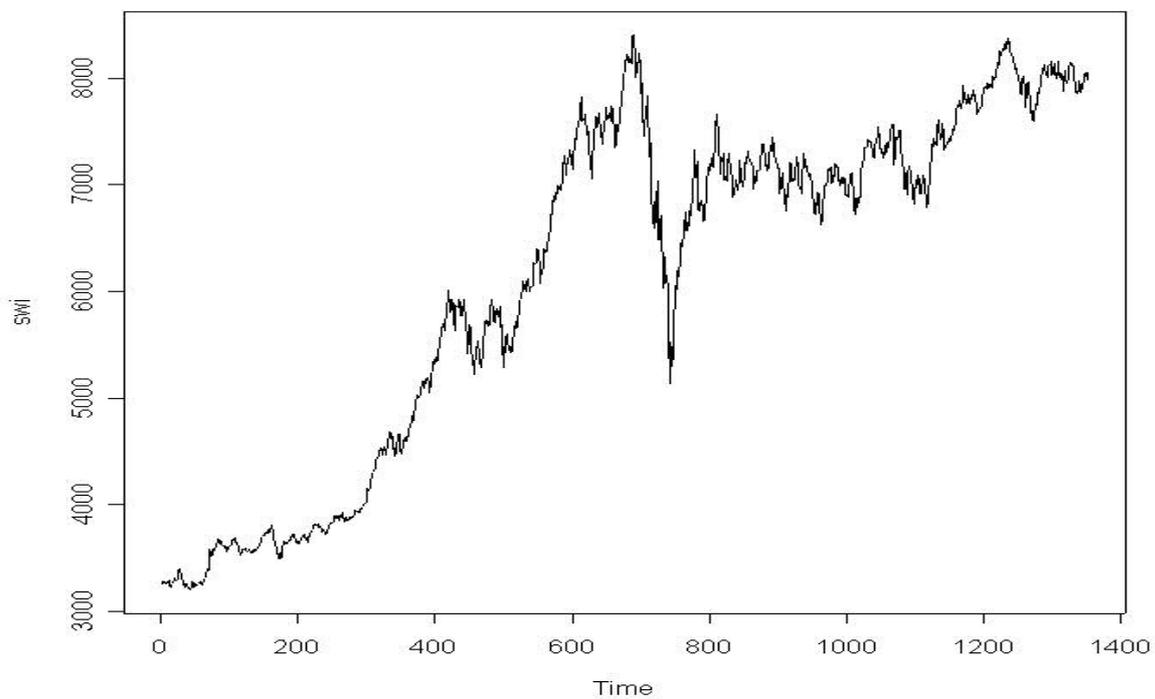
GRAFICI DAX30



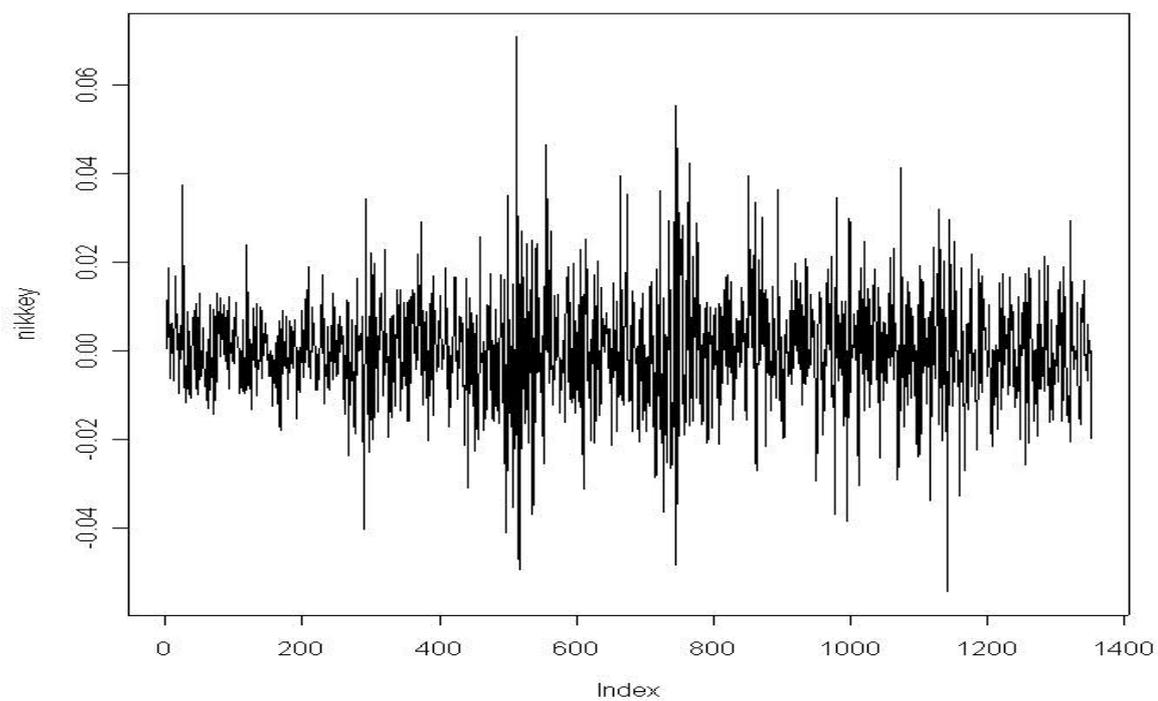
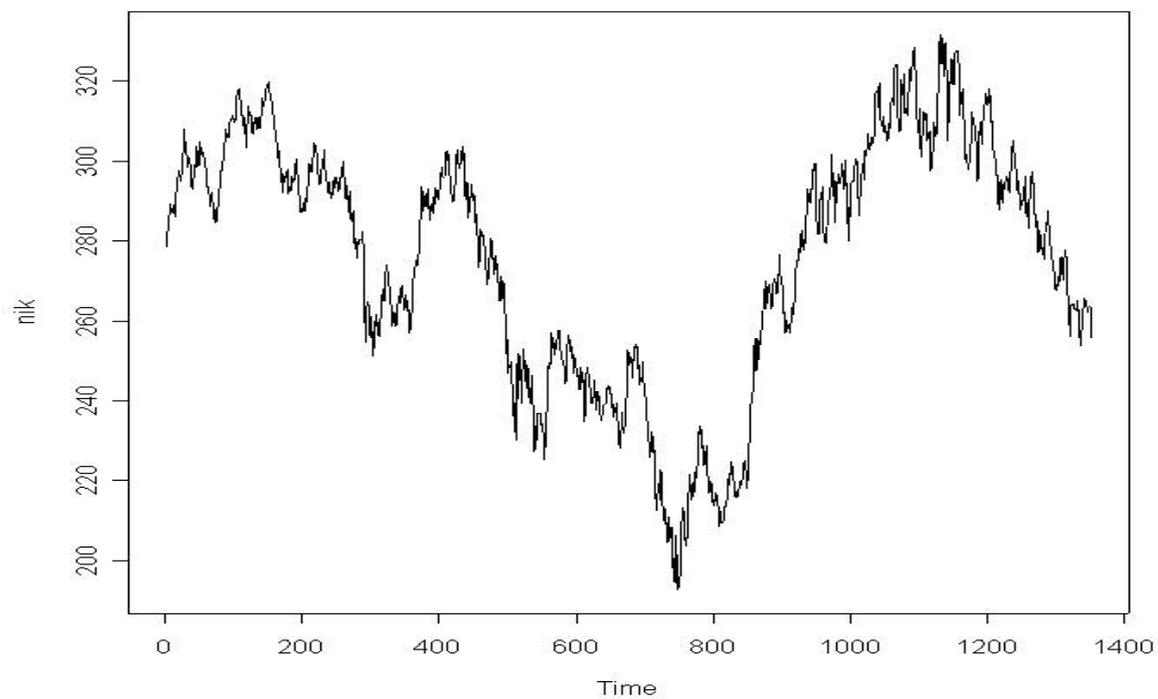
GRAFICI CAC40



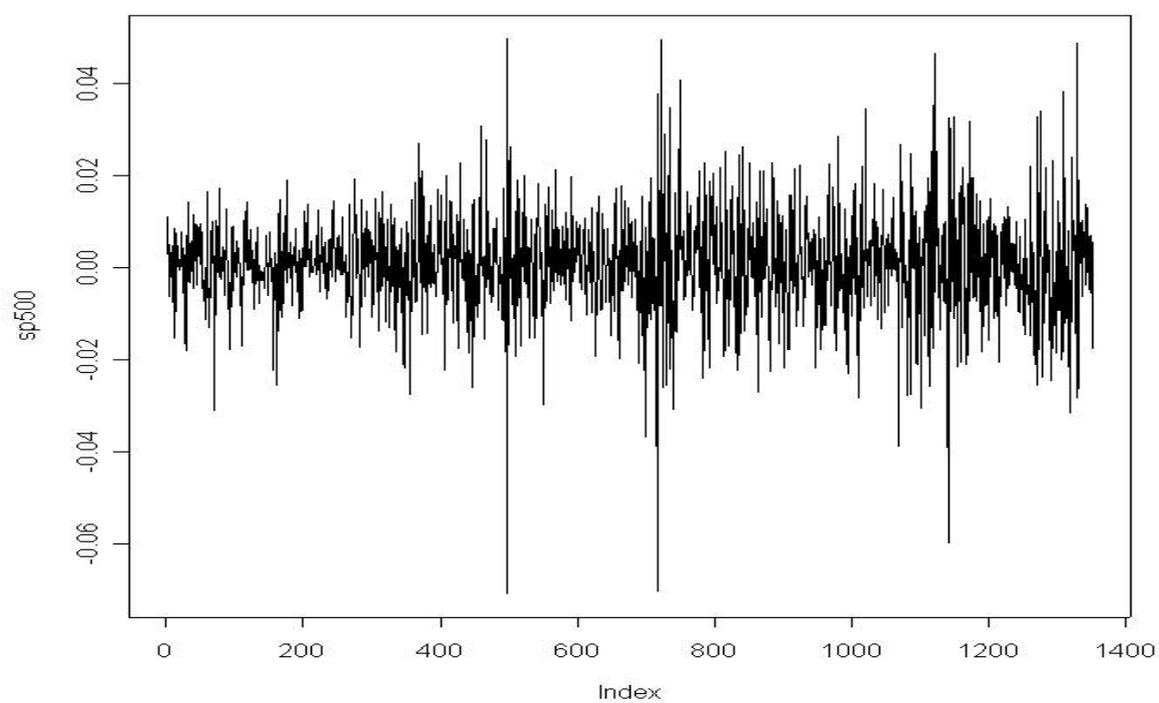
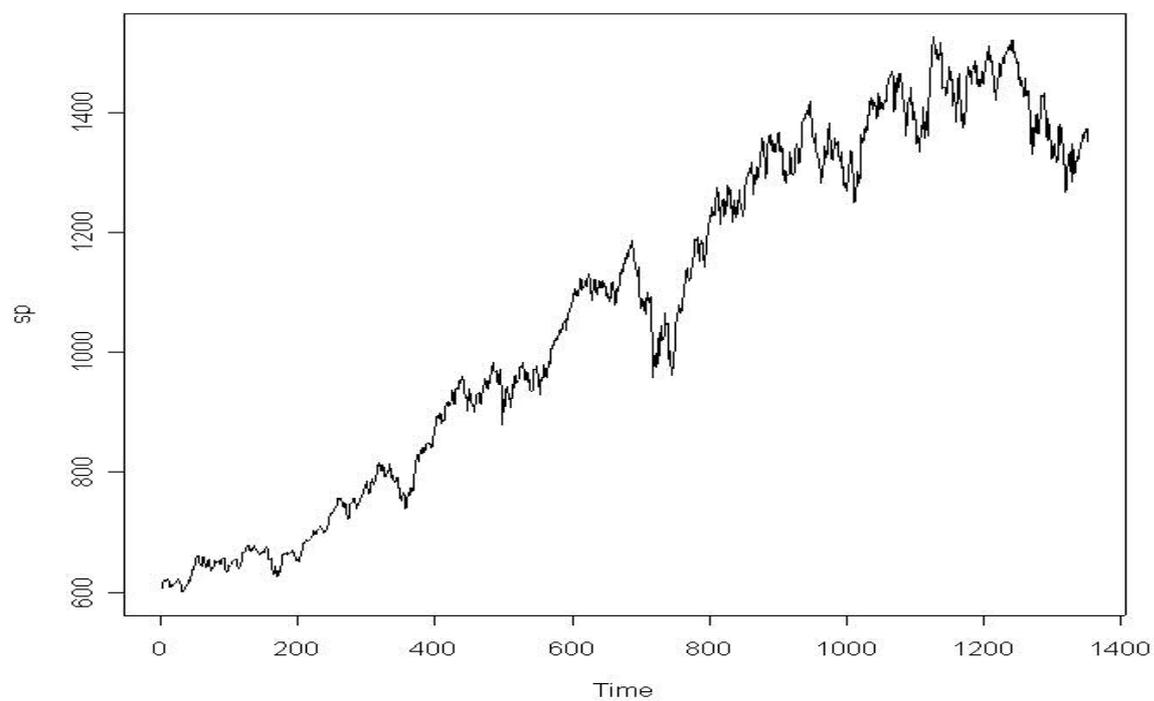
GRAFICI BEL20



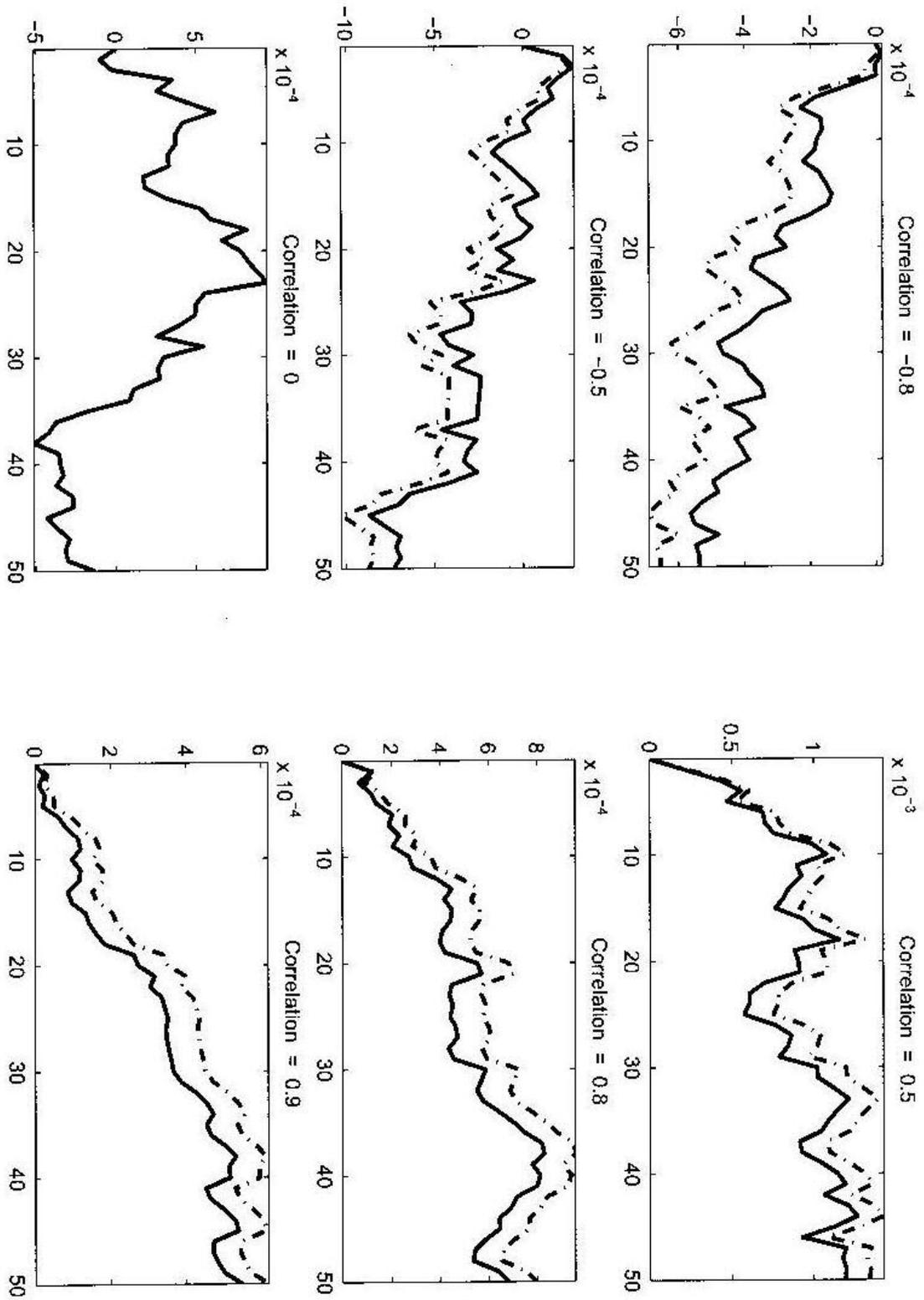
GRAFICI SWISS



GRAFICI NIKKEY



GRAFICI S&P500



Bibliografia

- [1] R. Engle e K. Sheppard, *Theoretical and Empirical properties of Dynamic Conditional correlation Multivariate GARCH*, Dicembre 2001.
- [2] T. Bollerslev e J. M. Wooldridge, *Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances*, *Econometric Reviews*, 11 (1992), pp 143-172.
- [3] J. Conlisk, *Matrix Algebra for Economics Lecture Notes*, 1998
- [4] R. F. Engle, *Dynamic Conditional Correlation – A Simple Class of Multivariate GARCH Models*. UCSD, Maggio 2001. R. Gallant and H. White, *A Unified, Theory for Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, New York, 1998.
- [5] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Application in Statistics and Econometrics*, John Wiley and Sons, New York, 1998.
- [6] Y. K. Tse e A. K. Tsui, *A Multivariate GARCH Model with Time-Varying Correlations*. National University di Singapore, Dicembre 1998.
- [7] R. V. Foutz e R. C. Srivastava, *The Performance of the Likelihood Ratio Test when the Model is Incorrect*, *The annals of Statistics*, 5 (1977), pp 1183-94.
- [8] T. Jeantheau, *Strong Consistency of Estimators for Multivariate GARCH Models*, *Econometric Theory*, 14 (1998), pp 70-86.
- [9] G. M. Gallo e B. Pacini, *Metodi Quantitativi Per i Mercati Finanziari*, Carrocci Editore, 2002.

-
- [10] B. Noble e J. W. Daniel, *Applied Linear Algebra*, Prentice Hall, 1998.

Ringraziamenti

In questi mesi sono state diverse le persone che mi hanno indirizzato – con molta pazienza – verso la comprensione e l’approfondimento della teoria dei modelli GARCH e l’utilizzo del software R, sapendo trovare risposte ai miei dubbi e soluzioni alle difficoltà che di volta in volta si sono presentate. La mia riconoscenza va in particolare al mio relatore, il prof. Bordignon, per avermi sempre dato la sua massima disponibilità.

Devo ringraziare anche il prof. Lisi, che generosamente, mi ha chiarito il problema del mio programma.

Un grande grazie lo devo soprattutto ai miei genitori e a mia moglie che in questi quattro anni di studio mi hanno sempre sostenuto, utilizzando tutta la loro pazienza.

Un grazie deve andare anche a mia sorella Sara, per avermi sempre sopportato prima degli esami.

Infine voglio ringraziare le due comari, tutti gli amici della biblioteca e i compagni di studio della facoltà.